Aplicación de Redes Complejas a la Descripción de la Dinámica de Contaminantes Atmosféricos

Application of Complex Networks to the Description of Atmospheric Pollutants Dynamics

PROGRAMA DE DOCTORADO

Computación avanzada, energía y plasmas

AUTOR

Rafael Carmona Cabezas

DIRECTORES

Prof. Dr. Francisco José Jiménez Hornero

Prof. Dr. Eduardo Gutiérrez de Ravé Agüera

2021

TITULO: *Application of Complex Networks to the Description of Atmospheric Pollutants Dynamics*

AUTOR: *Rafael Carmona Cabezas*

TÍTULO DE LA TESIS:
Aplicación de redes complejas a la descripción de la dinámica de contaminantes atmosféricos.

Application of complex networks to the description of atmospheric pollutans dynamics.

DOCTORANDO:
Rafael Carmona Cabezas

INFORME RAZONADO DE LOS DIRECTORES DE LA TESIS
(se hará mención a la evolución y desarrollo de la tesis, así como a trabajos y publicaciones derivados de la misma).

El desarrollo de la tesis doctoral ha permitido alcanzar los objetivos inicialmente planteados en el plan de investigación en los plazos previstos. En el marco de ejecución de esta tesis doctoral que, opta a la mención internacional, el doctorando ha realizado una provechosa estancia la cual a dado lugar a dos publicaciones en revistas JCR fruto de la colaboración establecida.

Respecto a las tres publicaciones que constituyen el compendio presentado en esta tesis doctoral, hay que destacar que todas ellas están indexadas en el primer cuartil de su categoría de JCR. Como consecuencia, la difusión internacional de los resultados presentados en esas publicaciones y, por tanto, en esta tesis está asegurada. Dos de estas tres publicaciones describen con éxito la dinámica temporal del ozono troposférico tanto en el ámbito urbano como rural usando los gráficos de visibilidad. La tercera publicación presenta una novedad metodológica que permite aplicar los gráficos de visibilidad en el análisis de una serie temporal de forma más eficiente. Este es uno de los aspectos más innovadores de esta tesis doctoral.

Por último, hay que destacar por su relevancia las 5 publicaciones que acompañan a las tres que forman el compendio y que contienen resultados derivados de la tesis doctoral. Estos resultados provienen del análisis multifractal y de la aplicación de los gráficos de visibilidad a la descripción de la evolución temporal de variables hidroclimáticas y de la concentración de contaminantes atmosféricos como el dióxido de nitrógeno el ozono y las partículas en suspensión PM_{10}.

Por todo ello, se autoriza la presentación de la tesis doctoral.

Córdoba, 21 de mayo de 2021

Firma de los directores

Fdo.: Francisco José Jiménez Hornero Fdo.: Eduardo Gutiérrez de Ravé Agüera

Acknowledgments

"Agradece a la llama su luz, pero no olvides el pie del candil que paciente la sostiene."

(Original: "Thank the flame for its light, but do not forget the lamp holder standing in the shade with the constancy of patience")

Rabindranath Tagore

Tras escribir los artículos e investigaciones que conforman esta tesis doctoral, me resulta curioso la poca práctica que tenemos al expresar de corazón y mediante palabra escrita, algo tan humano y sincero como es la gratitud. Todos hemos experimentado alguna vez ese sentimiento, al recibir el apoyo desinteresado de los que nos rodean. Sin embargo, contadas son las ocasiones en que nos detenemos a expresarlo, aún a sabiendas del regalo tan valioso que supone ese acto gratuito. Es por ello que no puedo dejar pasar esta oportunidad para pararme a respirar, mirar al camino andado y expresar de corazón lo que tantas veces he pensado, pero no dicho.

La primera creo que está clara: a mis padres, Dolores y Rafael, sin vosotros es más que obvio que hoy no sería quien soy. Aún no puedo imaginar los retos y duras decisiones que conlleva ser padre, pero seguro que equivaldría a infinitas tesis doctorales. No sé si estaréis orgullosos de lo que hemos conseguido, pero yo os aseguro que estoy muy orgulloso de vosotros.

A mis directores de tesis, Francisco y Eduardo, no sólo porque esta tesis no sería posible sin vosotros, sino por todo vuestro apoyo durante estos años, tanto en lo profesional como en lo personal. Si algo he aprendido con vosotros, es que un grupo de investigación se basa tanto en la parte investigadora, como en la humana. Gracias por hacerme sentir como uno más de la familia.

Aunque previo a esta tesis, no puedo saltarme a todos los y las docentes que, sin duda alguna, me habéis moldeado como persona, desde primaria a la universidad. Agradezco (entre otros) a José María, Paco, Carmen, Jerónimo, Juanma y José Ignacio. Cada vez tengo más claro que mi futuro está en la educación, y no creo que sea casualidad. Aunque no sea mucho, no quería perder la ocasión de mencionaros.

A mis amigos/as, me habéis sostenido constantemente con vuestro insufrible humor, probablemente sin saberlo. Gracias Kike, Pablo, Carlos, Rafa, Adri, Sergio, José, Alex, Arturo, Máximo y los y las que no puedo incluir, pero agradeceré personalmente.

A toda mi familia: abuelos/as, tíos/as y primos/as por vuestro inquebrantable interés y cariño, que hacen muy difícil, si no imposible, el perder la motivación.

A João y todos los compañeros de la Universidad de Évora que me acogieron y enseñaron con tanta amabilidad formas de hacer ciencia distintas a las que estaba

acostumbrado. Also, thanks to Thomas, who has been another fellow colleague to me this last year, although we have never meet in person.

A mis compañeros/as del departamento: Javi, Jesús, Fernando, Valle, Alberto, Ana y Elena. Por esos desayunos (más que necesarios) en los que uno se olvidaba un rato de los fractales, las redes complejas y Matlab. También a los trabajadores de cierta cafetería que, aunque no lean esto, alegraban la mañana con su simpatía y esas tostadas estrella.

Por último, no me había olvidado de vosotros, os estaba reservando este párrafo final con las mejores vistas al resto del documento. A los bichos de la casa: Toby y Bella por vuestra infinita lealtad e inocencia, que me cargan las pilas día tras día. Y por supuesto a ti, Paqui, por estar siempre ahí, rellenando perfectamente mis defectos (que no son pocos) y soportando mis buenos y malos días. Aunque no tuvieras energía ni para ti, siempre has sacado para los dos; y es por ello que esto es también tuyo.

<div align="right">Gracias</div>

Abstract

"Sé breve en tus razonamientos, que ninguno hay
gustoso si es largo."

Miguel de Cervantes

Abstract

Air pollution has been a major concern among environmental scientists due to its importance to public health. Among the different pollutants that can be found in the air, one can point out tropospheric ozone as one of the most studied ones in the recent years, due to the risk derived for living beings. As a result of the many factors involved in the creation and destruction of this gas, the analysis of its dynamics is quite complicated. Traditionally, conventional statistical methods have been employed, while in the last decades multifractal approaches have gained importance. This is due to their suitability describing systems with a great degree of variability.

This thesis focuses on the evaluation and implementation of complex networks for the analysis of tropospheric O_3 dynamics. The studies carried out are based on the Visibility Graph (VG) technique, which transforms time series into complex networks that inherit the properties of the first ones. In the first part, a combination of the VG and the multifractal Sand-Box (SB) algorithms is performed. By doing this, authors analyze the generalized fractal dimensions and the singularity spectra. Then, a comparison was made between these multifractal parameters and the quantities obtainable from the degree distribution of the resulting graphs.

Regarding the second part of this thesis, the VG methodology was used on O_3 time series from rural and urban stations, in order to retrieve the centrality parameters from the obtained networks. This way, degree, shortest path and betweenness are studied to support the use of this technique and find new information. Results show that this methodology can indeed differentiate between ozone measurements in urban and countryside environments, providing new insights about the dynamics.

In the third and last part of this document, authors propose an alternative approach to the VG, called Sliding Visibility Graph (SVG). This new technique takes advantage of the fact that visibility adjacency matrices are mostly empty, since the bast majority of the

nodes are not connected to each other. Thanks to this, it is possible to apply effectively a sliding window approach to lessen considerably the computation time, reducing one order the time efficiency (from $O(N^2)$ to $O(N)$). This is especially convenient when dealing with very large time series. As the resulting network approximates the original VG, it has been evaluated how it converges to the VG case for different types of series, as there lies the actual interest of this tool. As expected, the SVG results converge quite rapidly to the exact values, especially for random and O_3 concentration series.

Resumen

La contaminación atmosférica es uno de los principales problemas estudiados dentro de la ciencia ambiental, debido a su gran impacto en la salud pública. Entre los diferentes contaminantes que podemos encontrar en el aire, merece la pena destacar el ozono troposférico (O_3) como uno de los más estudiados en los últimos años, debido al alto riesgo para los seres vivos. Como resultado de los numerosos factores implicados en la creación y destrucción de este gas, el análisis de sus dinámicas es bastante complejo. Tradicionalmente, se han usado métodos estadísticos convencionales, mientras que en las últimas décadas han ganado importancia las técnicas multifractales. Esto se debe a su adecuación para describir sistemas con un grado elevado de variabilidad.

Esta tesis se centra en la evaluación e implementación de las redes complejas para el análisis de la dinámica del O_3. Los estudios llevados a cabo se basan en el uso de la técnica del Grafo de Visibilidad (GV), que transforma series temporales en redes complejas que heredan propiedades de las primeras. En la primera parte, se utiliza una combinación del GV y del algoritmo multifractal Sand-Box (SB). Gracias a esto, es posible obtener las dimensiones fractales generalizadas y el espectro de singularidades. Por último, se ha realizado una comparación entre los parámetros multifractales y las cantidades obtenibles directamente a partir de la distribución del grado de los grafos resultantes.

En cuanto a la segunda parte de esta tesis, el método del GV es usado en series temporales de O_3 de estaciones rurales y urbanas, con la finalidad de obtener los parámetros de centralidad de las redes conseguidas. De este modo, el grado, el camino mínimo y la intermediación se estudian para refutar la aplicabilidad del GV y buscar nueva información. Los resultados muestran que en efecto esta metodología puede permitir diferenciar entre medidas de ozono en medios rurales y urbanos.

En la tercera y última parte de este documento, los autores proponen un método alternativo al GV, llamado Grafo de Visibilidad Deslizante (GVD). Esta nueva técnica se aprovecha del hecho de que las matrices de adyacencia de los GV son prácticamente vacías, puesto que la mayoría de los vértices no están conectados entre sí. Gracias a ello, es posible aplicar de forma efectiva un algoritmo de ventana deslizante para reducir considerablemente el tiempo de cálculo, bajando en uno el orden de magnitud de la eficiencia (de $O(N^2)$ a $O(N)$). Esto es especialmente provechoso cuando se trata con series temporales muy grandes. Debido a que la red resultante aproxima al GV original, se ha evaluado cómo converge al segundo para diferentes tipos de series temporales, que es donde reside en interés real de esta herramienta. Como era de esperar, los resultado del GVD convergen rápidamente a los valores exactos, especialmente para series aleatorias y concentración de O_3.

Table of Contents

1. Introduction

"Si quieres construir un barco, no empieces por buscar madera, cortar tablas o distribuir el trabajo. Evoca primero en los hombres y mujeres el anhelo del mar libre y ancho."

(Original : "Si tu veux construire un bateau, ne rassemble pas tes hommes et femmes pour leur donner des ordres, pour expliquer chaque détail, pour leur dire où trouver chaque chose... Si tu veux construire un bateau, fais naître dans le cœur de tes hommes et femmes le désir de la mer. ")

Antoine de Saint-Exupéry

Introduction

In the midst of all the challenges that humanity needs to face in order to reach progress, pollution emerges as one the most severe ones. The word *pollution* is defined as "*introduction into the environment (esp. as a result of human activity) of harmful or poisonous substances, or excessive levels of light, noise, organic waste, etc.*"[1]. As stated in its own definition, it has its origin mostly in the human activity and due to its great danger to both society and nature, it gained the attention of the scientific community rather rapidly. One example of its severity can be regarded by looking at the estimations of the number of deaths derived from pollution, that raised to 9 million per year [2].

Figure 1: Photo of Madrid skyline, where photochemical smog is clearly seen [3].

Pollution can be seen in a wide range of different forms, from energy (heat, noise or light) to harmful chemical substances. It also can be classified depending on the medium where it occurs: atmospheric, water or soil pollution, for instance. This work is focused on the analysis of the very first one of them.

Air pollution has been a common problem in big metropolises for centuries. There exist records of measures related to this that go back in time several centuries. For instance, Edward I of England banned in 1272 the burning of coal in London, as the air

Introduction

pollution in the city was becoming a major issue to public health [4]. It was not until the Industrial Revolution when this started to generalize to almost every major city in the developed world, since the burning of coal and consequent gas emissions increased exponentially. An event in the recent past that is worth remarking was the *Great Smog of 1952,* occurred also in London, that killed nearly 4000 people just in that city and gave rise to the first major environmental law of modern times: The Clean Air Act (1956). Since then, air pollution has been a constant concern on population and considered for environmental policies.

Among the different substances that are considered air pollutants, one can highlight several that are extensively studied due to their danger: particulate matter (PM), carbon oxides (NO_x), lead, sulfur dioxide SO_2 and ground-level or tropospheric ozone (O_3), along with others. More precisely, this thesis focuses its work on the pollution by the last one mentioned: the ground-level ozone. Ozone in general can be considered "good" or "bad", depending on which height in the atmosphere it can be found at. Stratospheric O_3 protects earth surface from sun's ultraviolet light. Indeed, its excessive destruction is not desirable (the ozone layer hole). Conversely, the tropospheric one (and object of this thesis) has many disadvantages.

It is a secondary pollutant belonging to the photochemical smog. Classic smog is usually called to large amounts of gas and particles that result from combustion activities and accumulates in a given region. On the other hand, the photochemical one corresponds to the agglomeration of different gases in the atmosphere, that react with each other by the action of ultraviolet light to form secondary pollutants, such as the mentioned O_3. This gas becomes especially dangerous in warm and densely inhabited metropolises.

Recently, ozone has become one of the most widely studied gases in both urban and rural environments, due to its abundance and risks that are derived, making it the main photochemical oxidant. Among the cited perils, one can highlight the effects that ozone has on:

4

❖ <u>Human health:</u> Ozone in high concentrations is known to cause many harms to humans, from minor respiratory irritation to chronic respiratory and heart disease, lung cancer and chronic bronchitis (among others), leading to thousands of deaths every year [5–8].

❖ <u>Harvests and economy:</u> High concentrations of O_3 have been proven to damage severely crops yield, due to its oxidative stress on plants, which reduces photosynthesis and growth. This leads to high economic losses that are estimated to be about 3 billion USD just in China [9,10].

❖ <u>Climate change:</u> It is estimated to have been responsible for 5%–16% of the Earth temperature raise since preindustrial times [11]. Also, due to the reduction of stomatal conductance of plants, their consumption of CO_2 is decreased and therefore the concentration of that greenhouse effect gas increases [12,13].

When it comes to the creation of ozone (it is a secondary pollutant, as referred before), it is formed by photochemical reactions that involve primary pollutants such as nitrogen oxides ($NO_x = NO + NO_2$) and volatile organic compounds (VOCs) [14–16]. One of the main reactions behind the creation and destruction mechanisms for ozone is the following one [17]:

$$NO_2 + O_2 \leftrightarrow O_3 + NO \qquad (1)$$
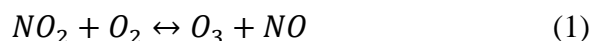
The direction of this photochemical process is rightwards (ozone creation) when there is enough energy in form of light available. The other case (leftwards and ozone destruction) happens when this energy is not present to maintain the balance. For that reason, highest and lowest concentrations of this secondary pollutant are always found during day and night conditions, respectively. It must be pointed out that the primary pollutants that are known to be involved in ground-level ozone creation (usually known as precursors of O_3) also react between them. One example is NO_2 that interacts with VOCs to produce aldehydes, which results in a lower rate of NO_2 - NO reaction to produce O_3.

Introduction

Furthermore, it does not only affect big metropolitan areas with great anthropogenic production as commented before, but rural ones as well. While the precursors of ozone are produced mainly from factories and traffic, sources that are much more present in cities, it can affect other areas due to transport of those precursors [18–21].

In addition to ultraviolet light and the chemical precursors of ozone, there are as well other parameters that play important roles in the dynamics of O_3. Those are meteorological conditions, such as temperature, humidity, and wind, which have been found to be closely related to it [15,22–24]. Another consideration to take into account is the atmospheric mixing layer height (MLH), which exists due to discontinuities in the stratification of the atmosphere layers that give rise to air turbulences. This is an important meteorological factor that influences the vertical diffusion of most atmospheric pollutants, and hence the dynamics of photochemical reactions such as the ones affecting tropospheric O_3 [25,26].

Due to all these factors, the analysis of the dynamics of the ozone becomes quite an intricate task. Thus, traditional statistical methods (which often offer a limited interpretation of series with a high level of variability[27]) may not be sufficient for its description. Recently, multifractal approaches were employed to overcome such shortcomings [28–30].

In short, a fractal is a subset of the Euclidean space with its Hausdorff dimension higher than the topological dimension [31,32]. That means that the way how they scale is different from typical geometric figures. For instance, doubling the size of a square by 2, would enlarge its area by a factor of 4 (which is 2^2, an integer power of the initial factor). Nevertheless, doing so with a fractal set would not necessarily lead to a scaling by an integer power of that number. That non-integer power is known as the fractal dimension of the said set. Another interesting peculiarity lies on the fact that fractals are nowhere differentiable[31,33], and that is a very interesting feature when describing natural phenomena. That is because we usually try to describe and approximate everything in the world surrounding us by employing linear or differentiable functions. Nonetheless, many

events in the universe do not follow such predictable trends and have rather chaotic and nonlinear dynamics [32,34–36].

Multifractality, on the other hand is a generalization of the concept of fractal that is needed when one single fractal dimension is not enough to describe the scaling properties of the system [37]. Indeed, most fractals in nature need more than one scaling exponent to be reproduced, being the examples numerous: meteorological variables [38–42], brain activity [43,44] or earthquake episodes [45], among others. There exist many methods to calculate the multifractal parameters from a dataset and among them it can be highlighted the Box-Counting (BC) fixed-size techniques, which are relatively simple and widely used [46], but also have some limitations, such as the working scales. To overcome these shortcomings, the Sand-Box algorithm was developed [47,48] and has proven to have many advantages when compared to the first ones [49–51].

In the last years, a new family of methods to analyze time series have been introduced. These approaches rely on the transformation of time series into complex networks, in order to analyze the underlying properties. A complex network is a type of graph (or network), which is a mathematical object formed by two finite sets. The first one is called the set of nodes, while the second one is the set of edges, which is made of ordered pairs from the first set. The particular features of complex networks imply that they have topological features that are non-regular nor random, and therefore occur in the description of real complex systems. In the last years, the scientific research around these kind of graphs has been active and fruitful, being applied to many fields ranging from social [52,53], to computer [54,55], to brain networks [56,57].

Although time series and complex networks do not seem to have much in common (indeed they are completely different mathematical objects) there is much to be gained with the mentioned transformation. The vast arsenal of network science techniques can be used to obtain new insights into nonlinear time series dynamics. The different mapping approaches are based on concepts such as visibility [58], correlation [59], phase space reconstruction [60] and recurrence analysis or transition probabilities [61]. All of them result in different networks that capture the main features of the underlying temporal signals.

Introduction

This thesis is focused on the Visibility Graph (VG)[58], which has attracted the interest of many scientists since it was introduced in 2008 by Lacasa et al [58]. Initially, the concept of such complex networks was used for the analysis of mutual visibility between pairs of points in landscape or a given spatial distribution. Its applications are numerous, from collision-free motion for robots to architectural design [62–64]. When applied to time series, each point of the signal is taken as an obstacle that may have (or not) visibility with the rest of them, depending on the values of the intermediate ones. It has been shown that results from VG can describe some properties of the time series related with fractal and multifractal properties of the time series [65,66]. Furthermore, the VG has already been found to be useful in many fields: neurology, image processing, thermodynamics, finance and geosciences, for instance [67–71].

Throughout the works that form this document, a thematic unity has been stablished, which gave entity and motivated the research carried out. The results of this research were materialized in the form of three publications that are the core of this PhD dissertation. To illustrate and clarify its contents, the outline and relations of such reports is shown next.

In the first paper [72], the Visibility Graph approach is applied to ground-level ozone time series from the city of Córdoba (southern Spain) for different times of the year. Then, in order to test whether the multifractal properties of the signal are conserved after the transformation of the time series into a complex network, the Sand-Box algorithm is applied to the resulting graph. The generalized fractal dimensions and the singularity spectrum are closely studied to achieve such goal.
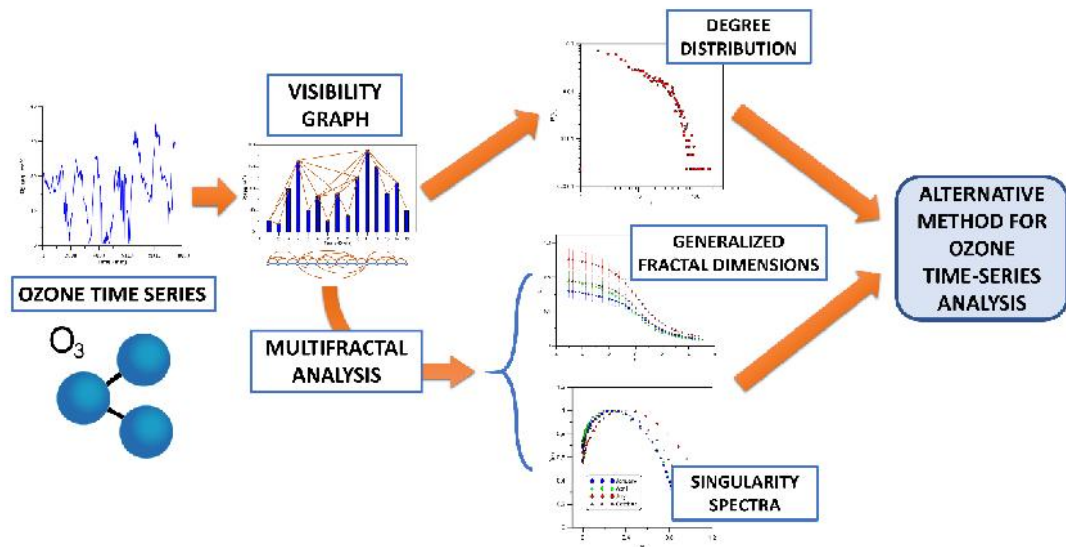
Figure 2: First paper graphical abstract. O_3 concentration time series are transformed into complex network for a later multifractal analysis, to provide an alternative method.

On the other hand, once the applicability of complex networks to reproduce the nature of ozone dynamics has been tested in the first one, the second one is devoted to the application of the measures that can be obtained directly from the network [73]. These are the so-called centrality parameters and have the advantage that are univocal for a given time series and their computation is rather simple compared to the multifractal approaches. In this case it has been applied to the study of the difference between rural and urban environments, using annual data from four different stations near the natural reservoir "Parque Natural de Los Alcornocales" in the province of Cádiz (Spain).
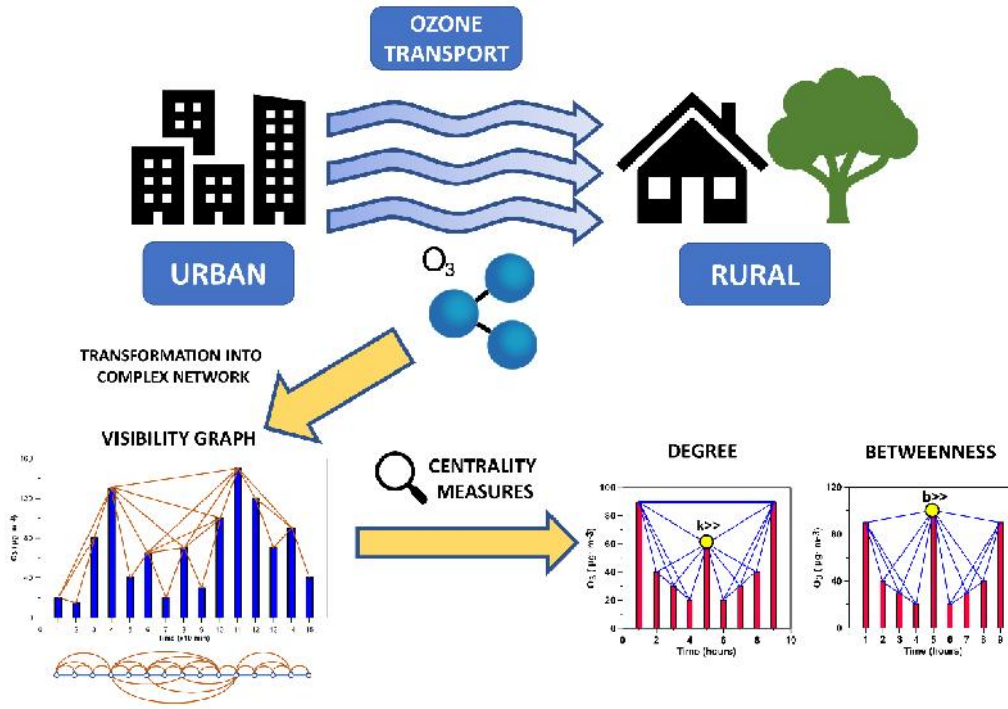
Introduction

Figure 3: Second paper graphical abstract. The degree and betweenness centrality parameters are evaluated with the aim of gaining information about the transport of O_3 from urban to rural areas.

Finally, in the third paper [74], an improvement over the original Visibility Graph is presented. This new method is termed Sliding Visibility Graph (SVG) and it is based on the fact that most of the nodes in the resulting network are usually not connected, due to the nature of the visibility criterion. Therefore, by applying a sliding window approach, it is possible to considerably reduce the required computation time, obtaining a remarkably good approximation of the actual resulting network. In that work, the proposed method is compared to both the original Visibility Graph and to the Horizontal Visibility Graph, which is an alternative that reduces the time consumption as well.
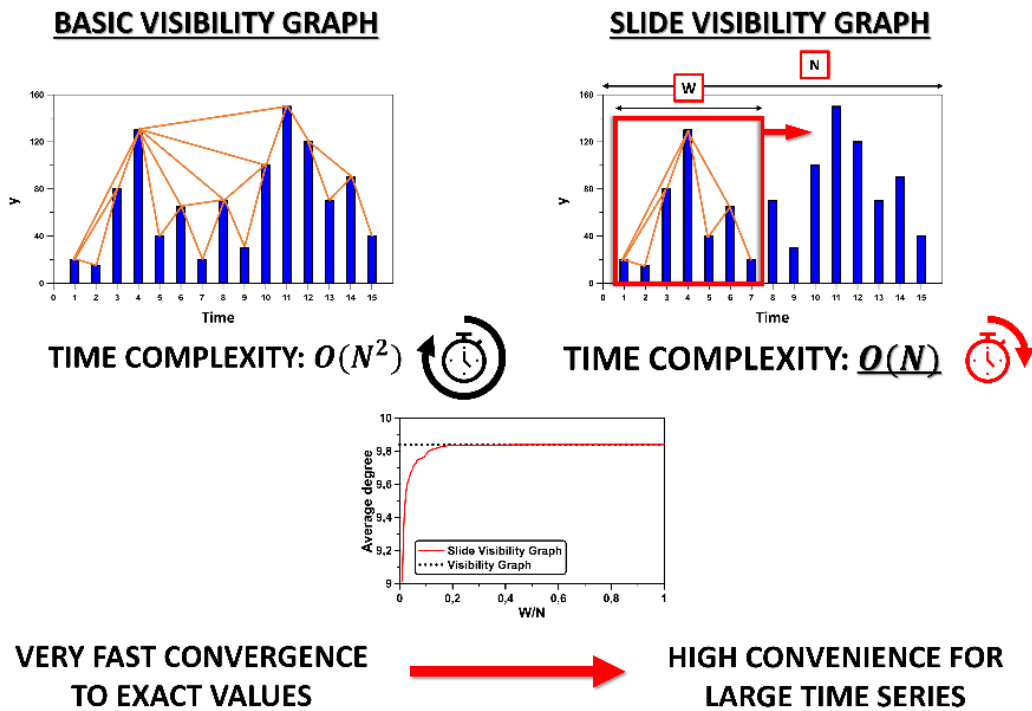
Figure 4: Third paper graphical abstract. An alternative method to obtain the VG from a time series is evaluated. The time efficiency is greatly improved, and it shows a fairly fast convergence to the exact values.

2. Hypothesis and objectives

"No hay viento favorable para el que no sabe a dónde va."

Séneca

General motivation

The number of works that have been devoted to the study of atmospheric pollutants is large. Among them, many used traditional descriptive statistical techniques to tackle this analysis [75–79] both in rural and urban environments. The obtained results, although clarifying in many aspects, have been always influenced by some limitations as the necessity of assuming a certain distribution. Moreover, they base their results on a given (time) scale, although the physical quantity in question might have several characteristic scales due to the number of variables in play [80]. The motivation behind the application complex networks (VG more precisely) was based on earlier ones that employ multifractal approaches to take advantage of the scale-free features of the environmental time series to analyze them [27,29,30,81,82].

Since it was found that the Visibility Graphs derived from a time series can be used to describe fractal series [58,65,66], it motivated their implementation for air pollutant dynamics, which are known to be multifractal in nature, as commented before. Thus, the line of work of this thesis was initiated with the corroboration of that possibility, before tackling more detailed analyses using complex networks. Once that is done, one is able to apply the methodology to larger datasets by employing the powerful tools that are available from graph theory.

Furthermore, the development of this PhD dissertation has been oriented by several goals. The followed aims can be identified and divided as follows, are achieved in the different papers that conform the document, as it will be specified for each one of them.

Specific objectives

In the **<u>first paper</u>**[72], the overall objective was to prove the usability of the complex networks to analyze O_3 time series, based on the multifractal properties that were previously employed for such matter. More precisely:

1) Implementing the VG algorithm for air pollutant time series and checking that its results are consistent with bibliography (i.e. that ozone series map scale-free networks) before tackling more profound and complex research.

2) Applying the SB algorithm to the complex networks that are obtained from the time series through the VG, in order to analyze its multifractal properties and extract relevant information from it.

3) Testing the method proposed by Chhabra and Jensen [83] for the computation of the singularity spectrum by using an alternative that does not rely on the Legendre transform.

After these were achieved, the line of work followed by the **<u>second paper</u>**[73] was focused on characterizing the differences that are encountered between rural and urban environments from the point of view of VG, with several individual goals:

4) Using the degree distribution, previously employed, in order to obtain information about the difference between the dynamics of tropospheric ozone in urban and rural areas.

5) Exploring a new centrality parameter, betweenness, with the same goal as the formerly used degree. This quantity identifies important nodes with a different criterion than the degree, leading to another type of distribution and interpretation of the results.

6) Comparing the information that can be drawn from hubs and the skyline-hubs, which are introduced in the paper itself as a tool that might be used for future works.

Hypothesis and objectives

Finally, after the methodology was employed for the description of ground-level ozone dynamics, for several scenarios and by using different approaches, both multifractal and graph-based, a new method is proposed in the **third paper**[74]. The main motivation was reducing the computation time of the original when the size of the series is sufficiently enlarged, being the following ones, the more detailed objectives that have been followed:

7) Benchmarking the new proposed SVG algorithm by comparing its computation time with the needed to use the original VG method.

8) Studying how rapid is the convergence of the proposed methodology and evaluate its convenience for different types of time series. Each one of those series has a different nature and graph topology, being ozone time series among the studied ones.

Stablishing a criterion to identify an optimal size of window for which the error made is low enough and the computation time is minimum, by searching the point at which the approximation saturates around the real value of the desired parameter.

3. References

"Después de todo, la ciencia es en esencia internacional y es únicamente a través de la falta de sentido histórico que las cualidades de nacionalidad se le han atribuido a ésta."

(Original: "After all, science is essentially international, and it is only through lack of the historical sense that national qualities have been attributed to it.")

Marie Curie

References

1. Pearsall J. *The New Oxford Dictionary of English*. Judy Pearsall. Oxford University Press; 2001.

2. Landrigan PJ, Fuller R, Acosta NJR, et al. The Lancet Commission on pollution and health. *The Lancet*. 2018;391(10119):462-512. doi:10.1016/S0140-6736(17)32345-0

3. Sánchez E. La pandemia hace caer la contaminación por ozono a los niveles más bajos desde hace una década. *El País*. https://elpais.com/clima-y-medio-ambiente/2020-11-10/la-pandemia-hace-caer-la-contaminacion-por-ozono-a-los-niveles-mas-bajos-desde-hace-una-decada.html. Published November 10, 2020. Accessed November 10, 2020.

4. Urbinato D. London's Historic "Pea-Soupers." U.S. Environmental Protection Agency History. Published 1994. Accessed September 11, 2019. https://web.archive.org/web/20061002080012/http://www.epa.gov/history/topics/perspect/london.htm

5. Doherty RM, Heal MR, Wilkinson P, et al. Current and future climate- and air pollution-mediated impacts on human health. *Environ Health*. 2009;8(1):S8. doi:10.1186/1476-069X-8-S1-S8

6. Kampa M, Castanas E. Human health effects of air pollution. *Environmental Pollution*. 2008;151(2):362-367. doi:10.1016/j.envpol.2007.06.012

7. Liu H, Liu S, Xue B, et al. Ground-level ozone pollution and its health impacts in China. *Atmospheric Environment*. 2018;173:223-230. doi:10.1016/j.atmosenv.2017.11.014

8. Soriano JB, Abajobir AA, Abate KH, et al. Global, regional, and national deaths, prevalence, disability-adjusted life years, and years lived with disability for chronic obstructive pulmonary disease and asthma, 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015. *The Lancet Respiratory Medicine*. 2017;5(9):691-706. doi:10.1016/S2213-2600(17)30293-X

9. Miao W, Huang X, Song Y. An economic assessment of the health effects and crop yield losses caused by air pollution in mainland China. *J Environ Sci*. 2017;56:102-113. doi:10.1016/j.jes.2016.08.024

10. Yi F, McCarl BA, Zhou X, Jiang F. Damages of surface ozone: evidence from agricultural sector in China. *Environ Res Lett*. 2018;13(3):034019. doi:10.1088/1748-9326/aaa6d9

References

11. Forster P, Ramaswamy V, Artaxo P, Berntsen T, Betts R, Fahey DW. Changes in atmospheric constituents and in radiative forcing. In: *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. S Solomon, D Qin, M Manning, Z Chen, M Marquis, et al. Cambridge Univ. Press; 2007.

12. Ainsworth EA, Yendrek CR, Sitch S, Collins WJ, Emberson LD. The Effects of Tropospheric Ozone on Net Primary Productivity and Implications for Climate Change. *Annu Rev Plant Biol*. 2012;63(1):637-661. doi:10.1146/annurev-arplant-042110-103829

13. IPCC. *Climate Change 2013 – The Physical Science Basis: Working Group I Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press; 2014. doi:10.1017/CBO9781107415324

14. Sillman S. The relation between ozone, NOx and hydrocarbons in urban and polluted rural environments. *Atmos Environ*. 1999;33(12):1821-1845. doi:10.1016/S1352-2310(98)00345-8

15. Trainer M, Parrish DD, Goldan PD, Roberts J, Fehsenfeld FC. Review of observation-based analysis of the regional factors influencing ozone concentrations. *Atmos Environ*. 2000;34:2045-2061. doi:doi.org/10.1016/S1352-2310(99)00459-8

16. Vingarzan R. A review of surface ozone background levels and trends. *Atmospheric Environment*. 2004;38(21):3431-3442. doi:10.1016/j.atmosenv.2004.03.030

17. Graedel TE, Crutzen PJ. Atmospheric change: an earth system perspective. *J Chem Educ*. 1993;70(9):A252. doi:10.1021/ed070pA252.2

18. Chakraborty T, Beig G, Dentener FJ, Wild O. Atmospheric transport of ozone between Southern and Eastern Asia. *Science of The Total Environment*. 2015;523:28-39. doi:10.1016/j.scitotenv.2015.03.066

19. Domínguez-López D, Adame JA, Hernández-Ceballos MA, Vaca F, De la Morena BA, Bolívar JP. Spatial and temporal variation of surface ozone, NO and NO2 at urban, suburban, rural and industrial sites in the southwest of the Iberian Peninsula. *Environ Monit Assess*. 2014;186(9):5337-5351. doi:10.1007/s10661-014-3783-9

20. Saavedra S, Rodríguez A, Taboada JJ, Souto JA, Casares JJ. Synoptic patterns and air mass transport during ozone episodes in northwestern Iberia. *Science of The Total Environment*. 2012;441:97-110. doi:10.1016/j.scitotenv.2012.09.014

21. Valverde V, Pay MT, Baldasano JM. Ozone attributed to Madrid and Barcelona on-road transport emissions: Characterization of plume dynamics over the Iberian Peninsula. *Science of The Total Environment*. 2016;543:670-682. doi:10.1016/j.scitotenv.2015.11.070

22. Elminir HK. Dependence of urban air pollutants on meteorology. *Science of The Total Environment*. 2005;350(1-3):225-237. doi:10.1016/j.scitotenv.2005.01.043

23. Li K, Chen L, Ying F, et al. Meteorological and chemical impacts on ozone formation: A case study in Hangzhou, China. *Atmospheric Research*. 2017;196:40-52. doi:10.1016/j.atmosres.2017.06.003

24. Wang Y, Du H, Xu Y, Lu D, Wang X, Guo Z. Temporal and spatial variation relationship and influence factors on surface urban heat island and ozone pollution in the Yangtze River Delta, China. *Science of The Total Environment*. 2018;631-632:921-933. doi:10.1016/j.scitotenv.2018.03.050

25. Stull RB, ed. *An Introduction to Boundary Layer Meteorology*. Springer Netherlands; 1988. doi:10.1007/978-94-009-3027-8

26. Aron R. Mixing height—an inconsistent indicator of potential air pollution concentrations. *Atmospheric Environment (1967)*. 1983;17(11):2193-2197. doi:10.1016/0004-6981(83)90215-9

27. Pavón-Domínguez P, Jiménez-Hornero FJ, Gutiérrez de Ravé E. Joint multifractal analysis of the influence of temperature and nitrogen dioxide on tropospheric ozone. *Stochastic Environ Res Risk Assess*. 2015;29(7):1881-1889. doi:10.1007/s00477-014-0973-5

28. He H. Multifractal analysis of interactive patterns between meteorological factors and pollutants in urban and rural areas. *Atmos Environ*. 2017;149:47-54. doi:10.1016/j.atmosenv.2016.11.004

29. Lee C-K, Juang L-C, Wang C-C, et al. Scaling characteristics in ozone concentration time series (OCTS). *Chemosphere*. 2006;62(6):934-946. doi:10.1016/j.chemosphere.2005.05.046

30. Pavon-Dominguez P, Jimenez-Hornero FJ, Gutierrez de Rave E. Multifractal analysis of ground–level ozone concentrations at urban, suburban and rural background monitoring sites in Southwestern Iberian Peninsula. *Atmos Pollut Res*. 2013;4(2):229-237. doi:10.5094/APR.2013.024

31. Mandelbrot BB. *The Fractal Geometry of Nature*. W.H. Freeman; 1982.

32. Briggs J. *Fractals: The Patterns of Chaos*. Thames and Hudson; 1992.

33. Mandelbrot B, van Ness JW. Fractional Brownian Motions, Fractional Noises and Applications. *SIAM Rev*. 1968;10:422-437. doi:http://dx.doi.org/10.1137/1010093

34. Kantz H, Schreiber T. *Nonlinear Time Series Analysis*. 2nd ed. Cambridge University Press; 2004.

35. Ding M, Grebogi C, Ott E, Sauer T, Yorke JA. Estimating correlation dimension from a chaotic time series: when does plateau onset occur? *Phys D*. 1993;69(3-4):404-424. doi:10.1016/0167-2789(93)90103-8

36. Safonov LA, Tomer E, Strygin VV, Ashkenazy Y, Havlin S. Delay-induced chaos with multifractal attractor in a traffic flow model. *EPL*. 2002;57:8. doi:https://doi.org/10.1209/epl/i2002-00555-0

References

37. Harte D. *Multifractals*. Chapman & Hall; 2001.

38. Ariza-Villaverde AB, Pavón-Domínguez P, Carmona-Cabezas R, de Ravé EG, Jiménez-Hornero FJ. Joint multifractal analysis of air temperature, relative humidity and reference evapotranspiration in the middle zone of the Guadalquivir river valley. *Agricultural and Forest Meteorology*. 2019;278:107657. doi:10.1016/j.agrformet.2019.107657

39. García-Marín AP, Estévez J, Alcalá-Miras JA, Morbidelli R, Flammini A, Ayuso-Muñoz JL. Multifractal analysis to study break points in temperature data sets. *Chaos*. 2019;29(9):093116. doi:10.1063/1.5096938

40. Calif R, Schmitt FG. Modeling of atmospheric wind speed sequence using a lognormal continuous stochastic equation. *Journal of Wind Engineering and Industrial Aerodynamics*. 2012;109:1-8. doi:10.1016/j.jweia.2012.06.002

41. Calif R, Schmitt FG, Huang Y, Soubdhan T. Intermittency study of high frequency global solar radiation sequences under a tropical climate. *Solar Energy*. 2013;98:349-365. doi:10.1016/j.solener.2013.09.018

42. Plocoste T, Calif R, Jacoby-Koaly S. Temporal multiscaling characteristics of particulate matter P M 10 and ground-level ozone O 3 concentrations in Caribbean region. *Atmospheric Environment*. 2017;169:22-35. doi:10.1016/j.atmosenv.2017.08.068

43. França LGS, Miranda JGV, Leite M, et al. Fractal and Multifractal Properties of Electrographic Recordings of Human Brain Activity: Toward Its Use as a Signal Feature for Machine Learning in Clinical Applications. *Front Physiol*. 2018;9:1767. doi:10.3389/fphys.2018.01767

44. Papo D, Goñi J, Buldú JM. Editorial: On the relation of dynamics and structure in brain networks. *Chaos*. 2017;27(4):047201. doi:10.1063/1.4981391

45. Yang D, Zhang C, Liu Y. Multifractal characteristic analysis of near-fault earthquake ground motions. *Soil Dynamics and Earthquake Engineering*. 2015;72:12-23. doi:10.1016/j.soildyn.2015.01.020

46. Meisel LV, Johnson M, Cote PJ. Box-counting multifractal analysis. *Phys Rev A*. 1992;45(10):6989-6996. doi:10.1103/PhysRevA.45.6989

47. Tél T, Fülöp Á, Vicsek T. Determination of fractal dimensions for geometrical multifractals. *Physica A*. 1989;159(2):155-166. doi:10.1016/0378-4371(89)90563-3

48. Vicsek T, Family F, Meakin P. Multifractal Geometry of Diffusion-Limited Aggregates. *EPL*. 1990;12(3):217-222. doi:10.1209/0295-5075/12/3/005

49. De Bartolo SG, Gaudio R, Gabriele S. Multifractal analysis of river networks: Sandbox approach. *Water Resour Res*. 2004;40(2). doi:10.1029/2003WR002760

50. Mali P, Manna SK, Mukhopadhyay A, Haldar PK, Singh G. Multifractal analysis of multiparticle emission data in the framework of visibility graph and sandbox algorithm. *Physica A*. 2018;493:253-266. doi:10.1016/j.physa.2017.10.015

51. Jiménez-Hornero FJ, Ariza-Villaverde AB, De Ravé EG. Multifractal description of simulated flow velocity in idealised porous media by using the sandbox method. *Fractals*. 2013;21(01):1350006. doi:10.1142/S0218348X13500060

52. Puga-Gonzalez I, Sueur C. Emergence of complex social networks from spatial structure and rules of thumb: a modelling approach. *Ecological Complexity*. 2017;31:189-200. doi:10.1016/j.ecocom.2017.07.004

53. Carchiolo V, Malgeri M, Mangioni G, Nicosia V. An adaptive overlay network inspired by social behaviour. *Journal of Parallel and Distributed Computing*. 2010;70(3):282-295. doi:10.1016/j.jpdc.2009.05.004

54. Gan C, Yang X, Liu W, Zhu Q, Jin J, He L. Propagation of computer virus both across the Internet and external computers: A complex-network approach. *Communications in Nonlinear Science and Numerical Simulation*. 2014;19(8):2785-2792. doi:10.1016/j.cnsns.2013.12.026

55. Bigdeli N, Haeri M, Choobkar S, Jannesari F. Characterization of complex behaviors of TCP/RED computer networks based on nonlinear time series analysis methods. *Physica D: Nonlinear Phenomena*. 2007;233(2):138-150. doi:10.1016/j.physd.2007.06.017

56. Lee B, Kang U, Chang H, Cho K-H. The Hidden Control Architecture of Complex Brain Networks. *iScience*. 2019;13:154-162. doi:10.1016/j.isci.2019.02.017

57. Stam CJ. Characterization of anatomical and functional connectivity in the brain: A complex networks perspective. *International Journal of Psychophysiology*. 2010;77(3):186-194. doi:10.1016/j.ijpsycho.2010.06.024

58. Lacasa L, Luque B, Ballesteros F, Luque J, Nuño JC. From time series to complex networks: The visibility graph. *Proc Natl Acad Sci*. 2008;105(13):4972-4975. doi:10.1073/pnas.0709247105

59. Zhang J, Small M. Complex Network from Pseudoperiodic Time Series: Topology versus Dynamics. *Phys Rev Lett*. 2006;96(23):238701. doi:10.1103/PhysRevLett.96.238701

60. Gao Z, Jin N. Complex network from time series based on phase space reconstruction. *Chaos*. 2009;19(3):033137. doi:10.1063/1.3227736

61. Campanharo ASLO, Sirer MI, Malmgren RD, Ramos FM, Amaral LANunes. Duality between Time Series and Networks. Perc M, ed. *PLoS ONE*. 2011;6(8):e23378. doi:10.1371/journal.pone.0023378

62. De Floriani L, Marzano P, Puppo E. Line-of-sight communication on terrain models. *International journal of geographical information systems*. 1994;8(4):329-342. doi:10.1080/02693799408902004

References

63. Lozano-Pérez T, Wesley MA. An algorithm for planning collision-free paths among polyhedral obstacles. *Communications of the ACM*. 1979;22(10):560-570. doi:10.1145/359156.359164

64. Turner A, Doxa M, O'Sullivan D, Penn A. From Isovists to Visibility Graphs: A Methodology for the Analysis of Architectural Space. *Environ Plann B Plann Des*. 2001;28(1):103-121. doi:10.1068/b2684

65. Lacasa L, Luque B, Luque J, Nuño JC. The visibility graph: A new method for estimating the Hurst exponent of fractional Brownian motion. *EPL*. 2009;86(3):30001. doi:10.1209/0295-5075/86/30001

66. Ni X-H, Jiang Z-Q, Zhou W-X. Degree distributions of the visibility graphs mapped from fractional Brownian motions and multifractal random walks. *Physics Letters A*. 2009;373(42):3822-3826. doi:10.1016/j.physleta.2009.08.041

67. Iacovacci J, Lacasa L. Visibility graphs for image processing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. Published online 2019:1-1. doi:10.1109/TPAMI.2019.2891742

68. Lacasa L, Nuñez A, Roldán É, Parrondo JMR, Luque B. Time series irreversibility: a visibility graph approach. *The European Physical Journal B*. 2012;85(6). doi:10.1140/epjb/e2012-20809-8

69. Pierini JO, Lovallo M, Telesca L. Visibility graph analysis of wind speed records measured in central Argentina. *Physica A: Statistical Mechanics and its Applications*. 2012;391(20):5041-5048. doi:10.1016/j.physa.2012.05.049

70. Qian M-C, Jiang Z-Q, Zhou W-X. Universal and nonuniversal allometric scaling behaviors in the visibility graphs of world stock market indices. *J Phys A: Math Theor*. 2010;43(33):335002. doi:10.1088/1751-8113/43/33/335002

71. Sannino S, Stramaglia S, Lacasa L, Marinazzo D. Visibility graphs for fMRI data: multiplex temporal graphs and their modulations across resting state networks. *Network Neurosci*. Published online April 3, 2017:208-221. doi:10.1101/106443

72. Carmona-Cabezas R, Ariza-Villaverde AB, Gutiérrez de Ravé E, Jiménez-Hornero FJ. Visibility graphs of ground-level ozone time series: A multifractal analysis. *Sci Total Environ*. 2019;661:138-147. doi:10.1016/j.scitotenv.2019.01.147

73. Carmona-Cabezas R, Gómez-Gómez J, Ariza-Villaverde AB, Gutiérrez de Ravé E, Jiménez-Hornero FJ. Can complex networks describe the urban and rural tropospheric O3 dynamics? *Chemosphere*. 2019;230:59-66. doi:10.1016/j.chemosphere.2019.05.057

74. Carmona-Cabezas R, Gómez-Gómez J, Gutiérrez de Ravé E, Jiménez-Hornero FJ. A sliding window-based algorithm for faster transformation of time series into complex networks. *Chaos*. 2019;29(10):103121. doi:10.1063/1.5112782

75. Adame JA, Lozano A, Bolívar JP, De la Morena BA, Contreras J, Godoy F. Behavior, distribution and variability of surface ozone at an arid region in the south

of Iberian Peninsula (Seville, Spain). *Chemosphere*. 2008;70(5):841-849. doi:10.1016/j.chemosphere.2007.07.009

76. Dueñas C, Fernández MC, Cañete S, Carretero J, Liger E. Analyses of ozone in urban and rural sites in Málaga (Spain). *Chemosphere*. 2004;56(6):631-639. doi:10.1016/j.chemosphere.2004.04.013

77. Dueñas C, Fernández MC, Cañete S, Carretero J, Liger E. Assessment of ozone variations and meteorological effects in an urban area in the Mediterranean Coast. *Sci Total Environ*. 2002;299(1-3):97-113. doi:10.1016/S0048-9697(02)00251-6

78. Cheng N, Li R, Xu C, et al. Ground ozone variations at an urban and a rural station in Beijing from 2006 to 2017: Trend, meteorological influences and formation regimes. *Journal of Cleaner Production*. 2019;235:11-20. doi:10.1016/j.jclepro.2019.06.204

79. Blanchard D, Aherne J. Spatiotemporal variation in summer ground-level ozone in the Sandbanks Provincial Park, Ontario. *Atmospheric Pollution Research*. 2019;10(3):931-940. doi:10.1016/j.apr.2019.01.001

80. Zeleke TB, Si BC. Characterizing scale-dependent spatial relationships between soil properties using multifractal techniques. *Geoderma*. 2006;134(3-4):440-452. doi:10.1016/j.geoderma.2006.03.013

81. Jiménez-Hornero FJ, Gutiérrez de Ravé E, Giráldez JV, Giráldez JV. Description of the seasonal pattern in ozone concentration time series by using the strange attractor multifractal formalism. *Environ Monit Assess*. 2010;160(1-4):229-236. doi:10.1007/s10661-008-0690-y

82. Jiménez-Hornero FJ, Jimenez-Hornero JE, Gutiérrez de Ravé E, Pavón-Domínguez P. Exploring the relationship between nitrogen dioxide and ground-level ozone by applying the joint multifractal analysis. *Environ Monit Assess*. 2010;167(1-4):675-684. doi:10.1007/s10661-009-1083-6

83. Chhabra A, Jensen RV. Direct determination of the f(α) singularity spectrum. *Phys Rev Lett*. 1989;62(12):1327-1330. doi:10.1103/PhysRevLett.62.1327

4. Publications

"La ciencia más útil es aquella cuyo fruto es el más comunicable."

Leonardo da Vinci

Note about the publications

In this section, complete copies of the published papers that form this PhD thesis are displayed next. All the metrics and information related to them that is shown are extracted from the JCR (Journal Citation Reports) annual publications by Clarivate Analytics. The reference year correspond to the manuscript's publication date or, failing that, the closest to it and last available one. This is applicable to the additional publications derived from this thesis as well.

Publication 1
Visibility Graphs of Ground-Level Ozone Time Series: A Multifractal Analysis

Journal

Science of the Total Environment (Ed. Elsevier)

Position / Quartile

(22/265) / Q1-D1 (Environmental Sciences)

Impact factor

6.551 (2019)

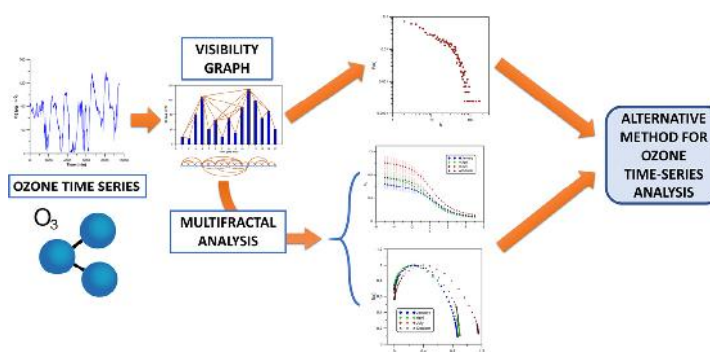# Visibility graphs of ground-level ozone time series: A multifractal analysis

Rafael Carmona-Cabezas *, Ana B. Ariza-Villaverde, Eduardo Gutiérrez de Ravé, Francisco J. Jiménez-Hornero

University of Córdoba, Spain

HIGHLIGHTS

- Ground-level ozone concentration time series have a multifractal behavior.
- Visibility graphs can be used to analyze multifractality of ozone time series.
- Many aspects of ozone dynamics can be observed through the degree distribution.
- This technique gives supplementary information within the study of complex signals.

GRAPHICAL ABSTRACT

ABSTRACT

A recent method based on the concurrence of complex networks and multifractal analyses is applied for the first time to explore ground-level ozone behavior. Ozone time series are converted into complex networks for their posterior analysis. The searched purpose is to check the suitability of this transformation and to see whether some features of these complex networks could constitute a preliminary analysis before the more thorough multifractal formalism.

Results show effectively that the exposed transformation stores the original information about the ozone dynamics and gives meaningful knowledge about the time series. Based on these results, the multifractal analysis of the complex networks is performed. Looking at the physical meaning of the multifractal properties (such as fractal dimensions and singularity spectrum), a relationship between those and the degree distribution of the complex networks is found.

In addition to all the promising results, this novel connection between time series and complex networks can deal with both stationary and non-stationary time series, overcoming one of the main limitations of multifractal analysis. Therefore, this technique can be regarded as an alternative to give supplementary information within the study of complex signals.

© 2019 Elsevier B.V. All rights reserved.

1. Introduction

Many studies have been performed about ground-level ozone over the last decades. The importance of ozone characterization and analysis lies on the fact that it is one of the main photochemical oxidants (due to its abundance). This irritant gas has serious repercussions for human

---

health and harvests when its concentration is high (Doherty et al., 2009). Those exposed damages have an impact from the economical point of view, and according to Miao et al. (2017), they lead every year to losses of several billions of dollars.

Ozone is a secondary pollutant whose chemical formation and destruction mechanisms are known to be photochemical and nonlinear processes (Graedel and Crutzen, 1993; Trainer et al., 2000). These mechanisms highly depend on meteorological variables such as temperature, wind direction and mainly solar radiation (Graedel and Crutzen, 1993; Guicherit and van Dop, 1977), as it has been studied for the Mediteranean area in several works (Güsten et al., 1994; Kouvarakis et al., 2000; Ribas and Peñuelas, 2004). In addition to that, ozone concentration depends as well on chemical precursors, such as nitrogen oxides and volatile organic compounds derived from the urban and industrial activity (Sillman, 1999). All these factors make the analysis of the temporal evolution of ozone a very complex task indeed.

Due to the facts exposed above, ozone studies based on traditional statistical analysis may provide a limited description of more complex dynamics of time series where the variability is high. The reason for this limitation is that these methods approximate and smooth the signal by means of fitting to functions, with the derived loss of information. Besides, they base their results on one (time) scale, while the physical phenomenon can appear at several scales due to the number of variables in play (Zeleke and Si, 2006). On the contrary, multifractal methods can be used to save this drawback, since they work directly with the raw data extracting the information from their singularities. Furthermore, fractals (and multifractals) are characterized for being self-similar when divided into smaller parts (i.e. they are scale-independent) or at least their statistical properties (Mandelbrot, 1982). That way, if a natural phenomenon can be characterized by means of multifractal parameters, these will be able to describe it for a range of scales.

In the presented work, a link between the multifractal analysis and complex networks has been tested for the description of ground-level ozone dynamics. To that purpose, ozone concentration time series have been transformed into visibility graphs (VG) (whose topology inherits the features of the associated time series) and then evaluated using two methods for multifractal analysis: the sandbox method in order to compute the generalized fractal dimensions (Rényi spectrum) and the approach introduced by Chhabra and Jensen (1989) for the calculation of the α-spectrum as an independent value from the other.

With this study, the main purpose is to check the suitability of the multifractal analysis performed over the VGs by connecting their degree distribution with Rényi and α-spectra. It could be expected, since the resulting graph stores indeed much of the original information and properties of the original time series (Lacasa et al., 2008).

2. Materials and methods

2.1. Data

The information that has been used in the analysis of this work corresponds to a 10-min ozone concentration data collected for the months of January, April, July and October in 2007. These time series of ozone concentration can be seen in Fig. 1, for the four months analyzed in this study.

The chosen region is the western part of Andalusia (Spain) since as exposed by Domínguez-López et al. (2014), this area meets the weather conditions (high temperatures and solar radiation), orographic
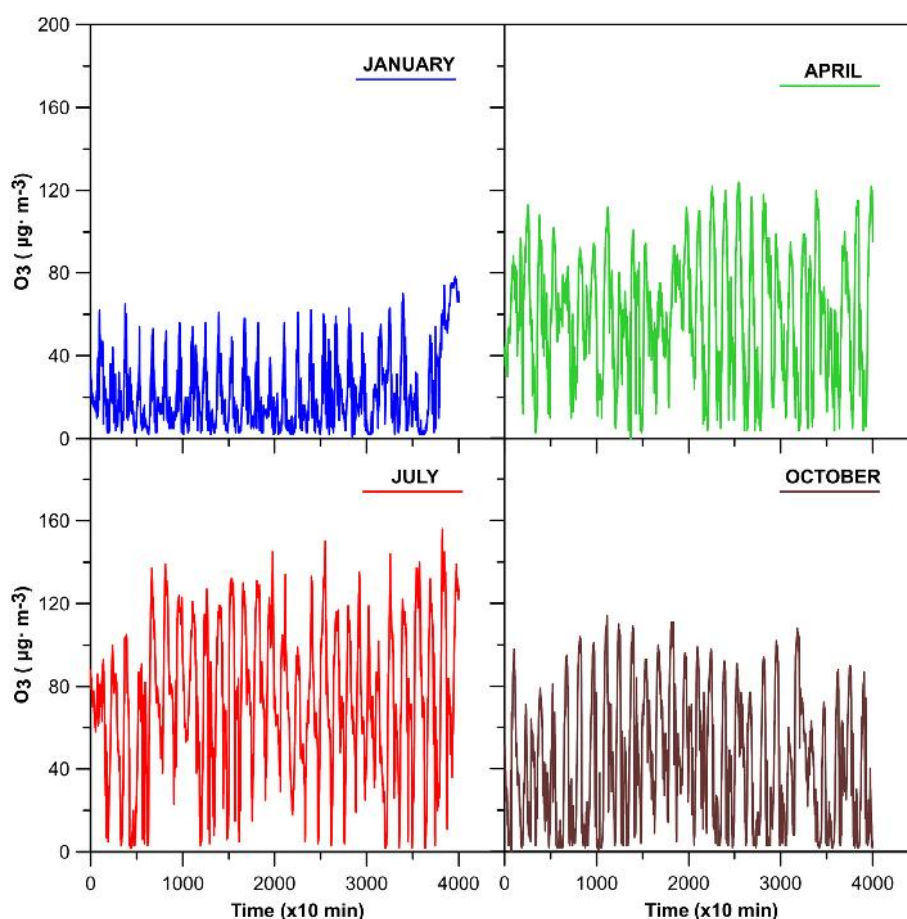

Fig. 1. Ozone time series for the different analyzed months.

(the valley of the Guadalquivir river) and anthropic (four capitals and two important industrial centers such as the chemical pole of Huelva and the Bay of Algeciras) to be potentially vulnerable to pollution by surface-level ozone. The measurements were performed at the urban station located in Lepanto, Córdoba (37.53° N, 4.47° W). The cited station belongs to the regional network in charge of controlling the air quality in Andalusia, co-financed by the Consejería de Medioambiente (Regional Environmental Department) and the European Union. This station is located at 117 m of altitude and the average temperature for each moth is 8, 15, 28 and 18 °C (January, April, July and October, respectively). The average direct solar radiation is 310.2, 577.8, 991.4 and 419.3 MJ/m^2 (again for the four months ordered). The climate of the zone of study, according to the Köppen-Geiger classification, is defined as Csa with warm average temperatures and hot and dry summer.

As it can be seen in Fig. 1, the ozone concentrations are especially high in summer (July) and low in Winter (January). As it has been previously discussed for this region by Adame et al. (2008) and Jiménez-Hornero et al. (2010a), that is due the fact that the conditions for the ozone creation are more suitable from the end of January. The progressive raise of temperature and solar radiation that reaches its peak in July allows those higher creation rates and thus its concentration. One of the reactions that governs the ozone production can be found below (Graedel and Crutzen, 1993).

$$NO_2 + O_2 \leftrightarrow O_3 + NO \tag{1}$$

It is a reversible photochemical reaction which tends to the ozone production when there is energy available in form of light (right sense of the arrow) and in the other way when there is not. For that reason, the higher values of ozone are always found during the day and vice versa, happening the same with summer and winter respectively, as discussed before.

2.2. Visibility graph

A graph can be defined as a set of vertices, points or nodes connected to each other by lines that are usually called *edges*. A tool to transform time series into a graph was presented by Lacasa et al. (2008). This new complex network receives the name of *visibility graph* and has been proven to inherit many of the properties of the original series; being some of them related with fractality, such as the Hurst exponent and fractal dimension of the Brownian motion (Lacasa et al., 2009).

In order construct the visibility matrix which contains the information of all the nodes of the new system, it is necessary to stablish a criterion to discern whether two points would be connected or not. This criterion reads as follows: two arbitrary data from the time series (t_a, y_a) and (t_b, y_b) have visibility (and would become two connected nodes in the graph) if any other data point (t_c, y_c) between them $(t_a < t_c < t_b)$ fulfills the following condition:

$$y_c < y_a + (y_b - y_a) \frac{t_c - t_a}{t_b - t_a} \tag{2}$$

In Fig. 2, an application of this method for a simple time series is given as illustration. As it can be seen, the original time series has been transformed into a complex network. The complexity of the original series is inherited by the new graph, as it has been found by Lacasa et al. (2008) and Lacasa and Toral (2010), meaning that for instance a periodic time series, would lead to a regular graph.

The result of applying this visibility method is a NxN adjacency binary matrix, being N the number of points in the set. Each row of the matrix contains the information of a different node, such as $a_{ij} = 1$ means that the node i and j have visibility; whereas $a_{ij} = 0$ means that there is no edge between them. The resulting matrix has several properties that can be used to simplify the algorithm and thus reduce the computational required time.
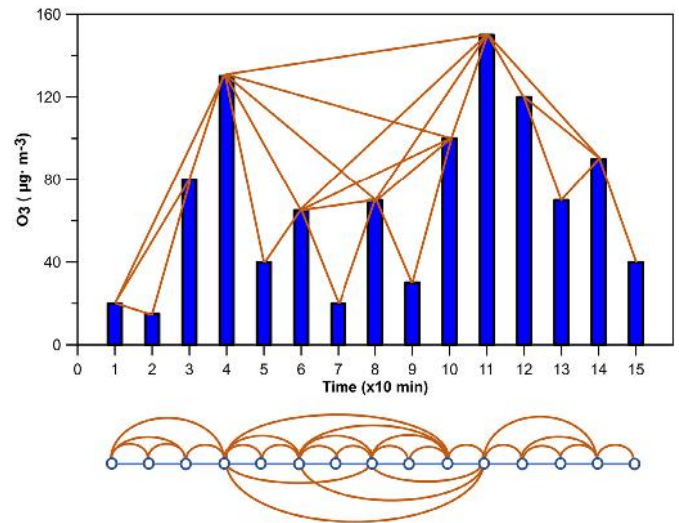

Fig. 2. Sample time series transformed into a complex network through the visibility algorithm. Below, all the connections are shown in a more visual way.

- Hollow matrix: All the elements in the diagonal are zero ($a_{ii} = 0$), because there is no visibility of an element with itself, since there are no intermediate points to fulfills the criterion.
- Symmetric matrix: The elements satisfy $a_{ij} = a_{ji}$, due to the reciprocity of the visibility between two nodes.
- Nearest neighbors: Because each point always sees the closest previous and next node, the elements surrounding the diagonal are always 1 ($a_{ij} = 1$ for $j = i \pm 1$).

Taking all of that into account, the visibility matrix A has a general form as shown below:

$$A = \begin{pmatrix} 0 & 1 & \dots & a_{1,N} \\ 1 & 0 & 1 & \vdots \\ \vdots & 1 & \ddots & 1 \\ a_{N,1} & \dots & 1 & 0 \end{pmatrix}$$

The degree of a node (k_i) can be defined as the number nodes that have reciprocal visibility with the first one ($k_i = \sum_j a_{ij}$). In Fig. 2, the degree of the first node is $k = 3$, for the second one $k = 2$, for the third one $k = 3$, and so on. From the degree of each one of the nodes present in the VG, it is possible to obtain the degree distribution of the sample ($P(k)$), which is nothing but the probability that every degree has.

Previous works have shown how the analysis of this degree distribution built from the VG can effectively describe the nature of the time series (Lacasa et al., 2008; Mali et al., 2018), distinguishing between periodic, random or fractal series for instance. Therefore, by studying the degree distribution, one can get a first insight on the behavior of the ozone time series before stepping into a more complex multifractal analysis. As stated by Lacasa et al. (2009) and Lacasa and Toral (2010), time-series that have VGs with degree distributions that follow power laws such as $P(k) \propto k^{-\gamma}$, can be considered as scale free.

2.3. Multifractal measurements

While fractal analysis is based on the complexity of a fractal set, the multifractal approach can describe the distribution of a given measure over a fractal object (Mandelbrot, 1974; Halsey et al., 1986). It implies the possibility of having different densities depending on the region of application.

There are two ways of representing multifractals: the generalized fractal dimensions D_q and the singularity or multifractal spectrum

$(f(\alpha))$. Both of them are discussed below separately. Typically, the multifractal analysis has been widely performed by means of the fixed-size algorithms (FSA), that rely on the subdivision of the system into smaller parts with equal size and then that size is changed iteratively. The method used in this work is the "sand-box algorithm" and will be discussed later on.

As stated before, the first of the measurements when it comes to multifractal analysis are the generalized fractal or Rényi dimensions D_q, which describe the scaling exponents of the qth moments of the system and can be defined (Feder, 1988) as:

$$D_q = \frac{1}{q-1} \lim_{\delta \to 0} \frac{\ln Z_q(\delta)}{\ln \delta} \ \forall q \neq 1 \tag{3}$$

$$D_1 = \lim_{\delta \to 0} \frac{\sum_{i=1}^{N_c(\delta)} \mu_i(\delta) \ \ln \mu_i(\delta)}{\ln \delta} \tag{4}$$

With q the moment order, δ the size of the used cells to cover the system, $Z_q(\delta)$ the partition function, $N_c(\delta)$ the number of cells with length δ and $\mu_i(\delta)$ the probability measurement of each cell. The expression for D_1 is obtained taking the limit of D_q when $q \to 1$. The cited partition function can be defined as follows:

$$Z_q(\delta) = \sum_{i=1}^{N_c(\delta)} [\mu_i(\delta)]^q \tag{5}$$

From these generalized fractal dimensions, it must be pointed out that $D_{q=0}$ corresponds to the fractal dimension of the given system or *box-counting dimension*, $D_{q=1}$ is the so-called *information entropy* and $D_{q=2}$ the *correlation dimension*. The limits of D_q when q goes to $-\infty$ and $+\infty$ describe the scaling properties of the regions where the measure is more rarified and concentrated, respectively.

The other set of multifractal parameters is the so-called singularity spectrum $(f(\alpha))$, as commented previously. A frequent method to determine is based on the use of a Legendre transform from mass exponents $\tau(q)$. However, some authors as Chhabra and Jensen (1989) and Veneziano et al. (1995), state that it can lead to some errors due to the inclusion of spurious points and error amplification from the derivative. Furthermore, it does not yield independent measurements from the Rényi spectrum, as $\tau(q) = (1-q)D_q$. As an alternative to this and more focused on experimental data, Chhabra and Jensen (1989) proposed a direct method to determine the α-spectrum, overcoming the drawbacks referred before. This technique relies on the normalized measure $\beta_i(q)$, μ_i in the original work, for computing the probabilities of the boxes of radius r:

$$\beta_i(q,r) = [P_i(r)]^q / \sum_j [P_j(r)]^q \tag{6}$$

With $P_i(r)$ the different fractal measurements for each box of radius r (number of nodes in this case). From this, $f(\alpha)$ and α can be retrieved using the next expressions:

$$f(q) = \lim_{r \to 0} \frac{\sum_i \beta_i(q,r) \ \log[\beta_i(q,r)]}{\log r} \tag{7}$$

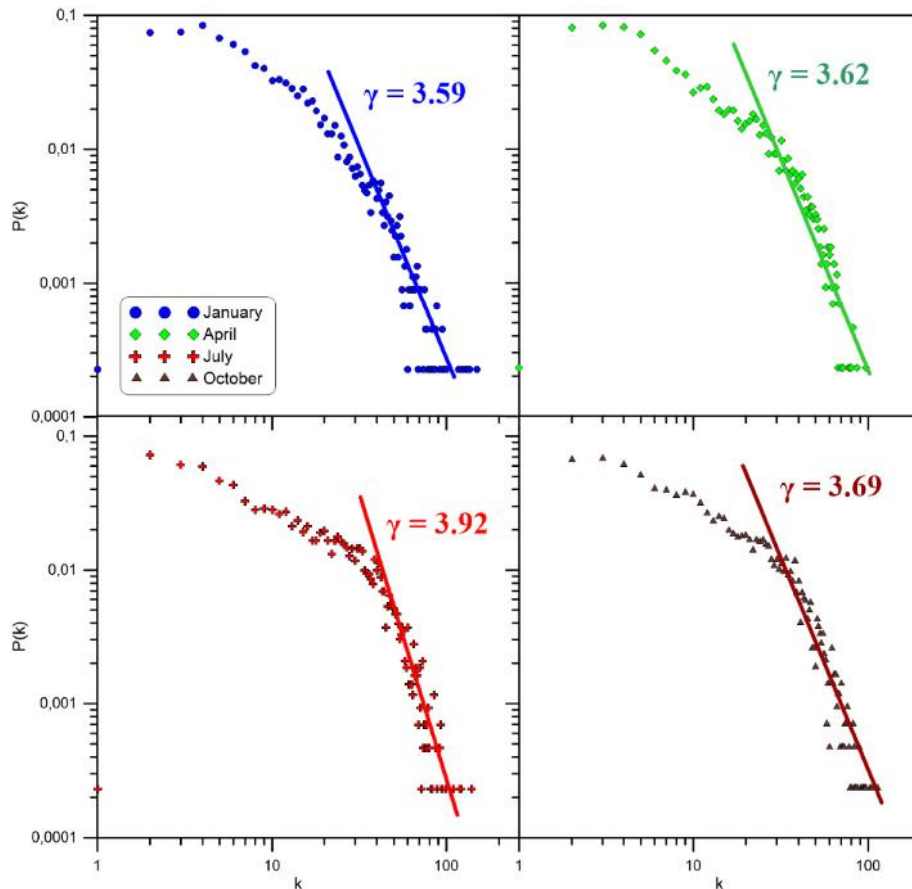$$\alpha(q) = \lim_{r \to 0} \frac{\sum_i \beta_i(q,r) \ \log[P_i(r)]}{\log r} \tag{8}$$

Fig. 3. Degree distribution of the visibility graph from each month in logarithmic scale. As it can be observed, the tail of the distribution shows a scale-free behavior because it can be fitted by a power law.
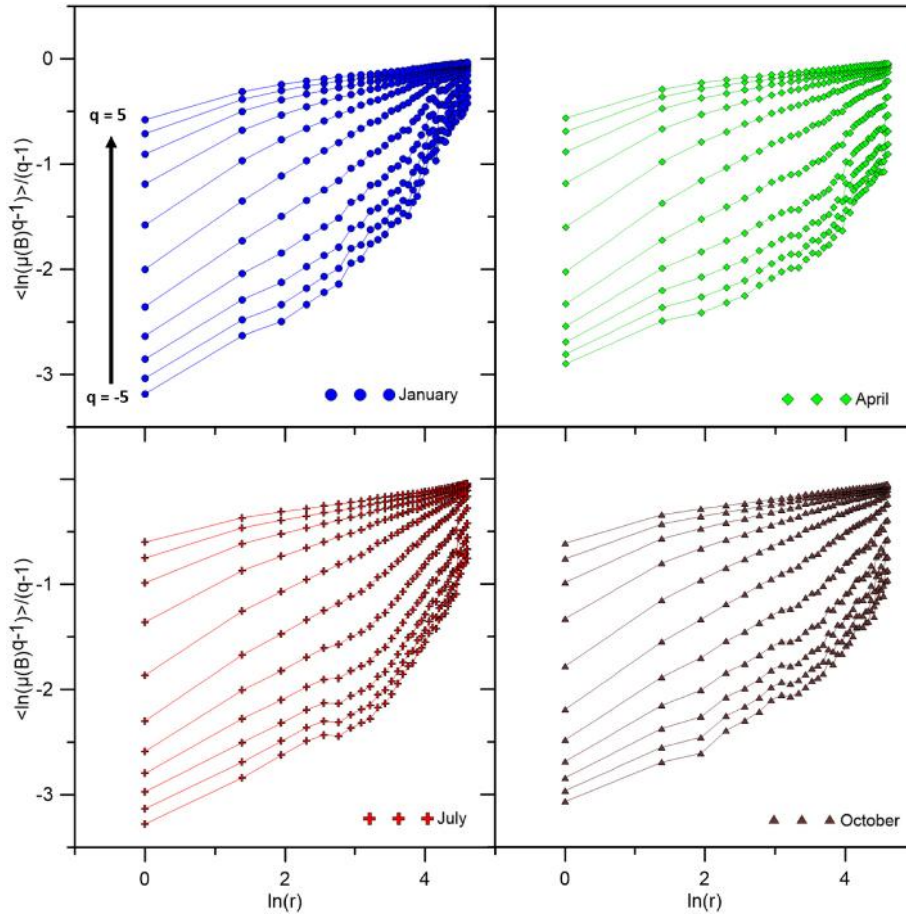
Fig. 4. Curves of the partition function against ln(r) obtained after applying the sandbox algorithm.

Being α known as the Lipschitz-Hölder exponent. In practice, those quantities are computed from the slope of $\sum_{i} \beta_i(q, r) \log[\beta_i(q, r)]$ over $\log r$ for $f(q)$; and $\sum_{i} \beta_i(q, r) \log[P_i(r)]$ over $\log r$ for $\alpha(q)$. This slope is determined by means of a linear regression in the same range of radii where the other fractal measures are computed.

2.4. Sandbox algorithm

The sandbox algorithm (SBA) was firstly introduced by Tél et al. (1989) and developed by Vicsek et al. (1990), as an improvement of the previously used fixed-size box-counting methods for computing the generalized fractal dimensions. The main advantage of this method with respect to other box-counting FSA is that it is capable of properly determine the side corresponding to the negative values of q from both the Rényi and singularity spectra.

The basic idea behind the SBA is that for each radius, a number of randomly placed boxes are selected, and they are always centered in a non-zero point of the system (a node). In that way the entire network (or the adjacency matrix in terms of the algorithm) is covered with those boxes by choosing a sufficiently high number of them. For each box (B), the probability measurement used is computed as shown in Eq. (9).

$$\mu(B) = \frac{M(B)}{M_0} \tag{9}$$

Once that quantity is computed for each sandbox for a given radius, the generalized fractal dimensions can be obtained as explained

previously in the multifractal measurement part. Applying it to Eq. (3), the following formula is obtained for D_q^{sb}:

$$D_q^{sb} = \frac{1}{q-1} \lim_{r \to 0} \frac{\ln \left\langle \mu(B)^{q-1} \right\rangle}{\ln r} \quad \forall q \neq 1 \tag{10}$$

And the expression when $q = 1$ for the SBA also can be adapted from Eq. (4):

$$D_1^{sb} = \lim_{r \to 0} \frac{\langle \ln \mu(B) \rangle}{\ln r} \tag{11}$$

With the aim of implementing the SBA on the VG, the steps to follow are:

1) The original time series is transformed into a VG, resulting on a matrix as described in the previous section.
2) A range of different radii is set in order to cover the entire network through the adjacency matrix. These radii are chosen between 1 and N, being N the total number of nodes. For this case, the typical values are $r \in [1, 100]$, because larger values have proven to give the same information, since the curves start to saturate.
3) Then, N_c centers are selected for each radius. When a node is chosen, in terms of the algorithm it means that a row of the matrix is picked to find the nodes that have visibility with that center (that have value one in the matrix) within the given radius. This number of centers N_c is inversely proportional to the radius and proportional to the length of the matrix: $N_c \propto N \cdot r^{-1}$; but always rounded to a positive integer bigger than one. This is done because the greater the box is, the smaller number of boxes will be needed to cover the matrix.
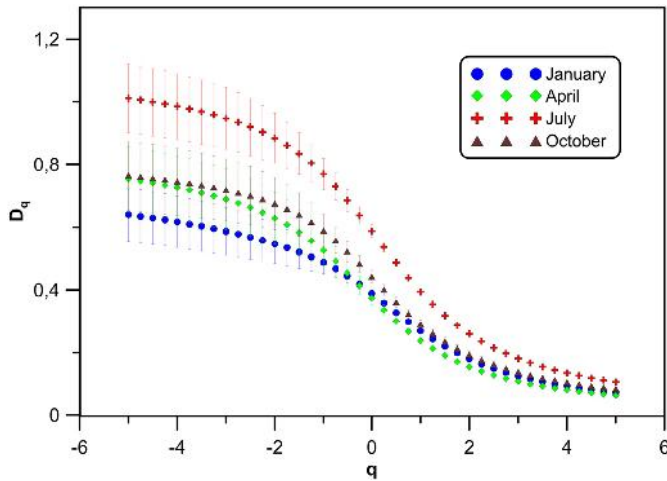
Fig. 5. Rényi dimensions for each month.

For instance, when the radius is $r = 100$, and the total number of nodes is $N = 4000$ ($N \times N$ adjacency matrix), it would probably be enough $N_c < 100$ to cover most of them. Whereas for smaller radii such as $r = 2$, this number should be much higher than before in order to ensure that the matrix is covered. The criterion to choose the minimum N_c is that this coverage is always at least 95%. The location of the centers is randomly chosen within the nodes of the graph.

4) For each sandbox, the amount of nodes inside the box connected to the center (M) are counted, giving N_c values of $M(B)$. From here, the quantity $\mu(B)$ (for every box) is computed by means of dividing the previous $M(B)$ by the total amount of nodes connected to the center M_0. Then $\mu(B)$ is used to compute the partition function for all the q values chosen, being the average in Eqs. (10) and (11) over all the sandboxes created.

5) The steps 3) and 4) are then repeated for all the radii considered, obtaining a value of the partition function for each one of them and q.

3. Results and discussion

3.1. Degree distribution

Firstly, a fast method to check the fractal nature of the ozone concentration time series, before performing a deeper study, is described here. This method consists on analyzing the degree distribution of the VGs.

The mentioned distribution is computed as the number of nodes that have a given degree and divided by the total number of them in the VG. It is clear that the greater the degree is, the less likely to be repeated within the network it will be; because large degrees are exclusive of nodes with the highest concentration (which we will refer as *hub*) due to their typically high visibility. As those hubs correspond normally to the extreme values of ozone concentration of each day, their likeliness will be lower, since the most repeated ones will be the ones close to the average. That leads to an expected distribution with a negative trend, as can be seen in Fig. 3.

Once the degree distributions of the different months are computed, a clear fractal behavior is observed since all of them follow a power law $P(k) \propto k^{-\gamma}$ as expected, with a linear part in the last part of the log-log plot. After computing the slope for the values of $k \geq 30$ in all the cases, the γ parameter can be determined, obtaining that the biggest and the lowest ones correspond to July and January respectively. Although this is a fast and direct method to determine whether the series is fractal or not, it does not give much detailed information about the differences between each moth, since all of them share a very similar degree distribution. Hence a deeper analysis devoted to the multifractal properties of the series has been done with that aim, as will be presented in the following section, in order to give some light to the usability of this degree distribution and some of its parameters.

3.2. Multifractality

In this part, the authors present two independent methods to study the multifractality of the signal, both previously described: the generalized fractal (Rényi) dimensions and the α spectra.

For the analysis of the Rényi dimensions of the samples, the quantities $\ln\langle\mu(B)^{q-1}\rangle/(q - 1)$ for $q \neq 1$ and $\langle\ln\mu(B)\rangle$ for $q = 1$ are plotted against $\ln r$, as seen in Fig. 4. The interval used for the radii goes from $r = 1$ to $r = 100$, as it was proven to be enough in order to reproduce the expected linear behavior (after several tests). It must be pointed out that the interval taken for the linear regression is always selected so that the Pearson correlation coefficient is $r \geq 0.99$. For April, July and October, the chosen range has been from $\ln r = 3$ up to end of the curve; whereas for January the range had to be taken such as $\ln r \in [2,4]$ because the linear part is larger and to avoid an artifact that appears for higher radii and made the results misleading (can be observed in Fig. 8). In all those cases, the values correspond with the higher radii used in the SBA, in accordance with previous studies that used the same technique (Ariza-Villaverde et al., 2015; De Bartolo et al., 2004; Jiménez-Hornero et al., 2013). After the linear regressions are performed for all values of q of each month, the Rényi spectrum can be constructed following Eqs. (10) and (11). In Fig. 5 the result of such procedure is shown in the form of the generalized fractal dimensions. After observing the graph, it can be inferred that the VG properly reproduces the multifractal behavior of the series, which was previously demonstrated to be so (He, 2017; Jiménez-Hornero et al., 2010a, 2010b; Pavon-Dominguez et al., 2013), since for all the studied months $D_0 > D_1 > D_2$ (see Table 1). The difference between the maximum value of D_q and the minimum (ΔD_q) is usually taken as a measure of the multifractality degree of the signal (ozone concentration in this case) (Ariza-Villaverde et al., 2013; Telesca et al., 2004). Looking at Table 1, this degree seems to be sensibly higher for July than the rest of the months, followed by April and October (with a very similar value) and then the last one is January.

Now, the fractal dimensions themselves from the Rényi spectrum (D_0, D_1 and D_2), are being discussed. Firstly, the capacity dimension or "box-counting" dimension D_0 is related with the number of boxes needed in order to cover the fractal object. The more boxes are needed, the more extended is the fractal object and then the grater would be D_0. As it was explained in SBA, each probability measure $\mu(B)$ is computed in terms of the total number of nodes connected to each center M_0,

Table 1
Relevant values for each month.

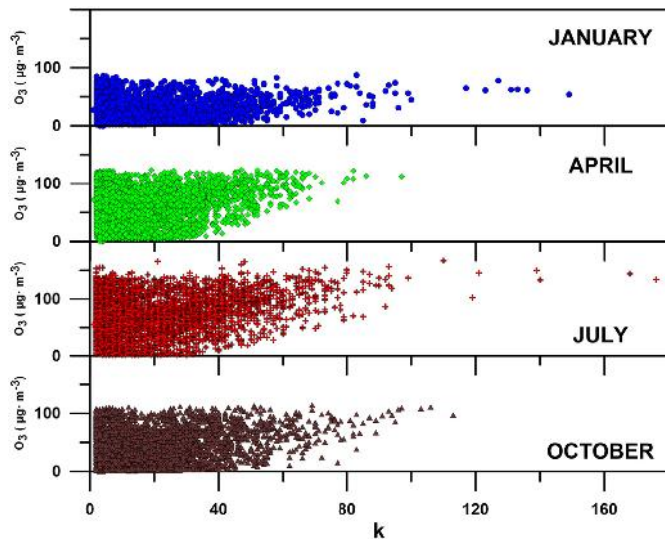| MONTH	Average Direct Radiation (MJ/m²)	γ	$\bar{k}$	σ_k	D_0	$D_0 - D_1$	$D_0 - D_2$	ΔD_q	W
January	310.2	3.59	14.78	15.00	0.389 ± 0.012	0.12 ± 0.03	0.21 ± 0.02	0.56 ± 0.10	0.779
April	577.8	3.62	15.12	13.98	0.37 ± 0.03	0.13 ± 0.04	0.22 ± 0.04	0.69 ± 0.12	0.934
July	991.4	3.92	19.94	17.69	0.59 ± 0.02	0.20 ± 0.04	0.33 ± 0.03	0.90 ± 0.12	1.152
October	419.3	3.69	17.90	15.91	0.44 ± 0.02	0.15 ± 0.04	0.25 ± 0.04	0.68 ± 0.12	0.841

Fig. 6. Plot that shows the relation between the values of the temporal series and the degree in the VG.

which is different for each iteration. The final behavior would be expected to be an average of the whole system. In the context of these complex networks, a high extension of the object would mean a bigger degree (then related to hubs). Following that logic, a higher value of D_0 is related with greater number of hubs in the system, meaning that one could expect the average degree $\bar{k}$ of the distribution increased. This can be observed in Fig. 6, where the relation between the values of ozone

concentration and the degree in the VG for each point are shown. This type of plot (v-k) was introduced by Pierini et al. (2012) for the mentioned purpose. In this case, it can be seen that the highest degrees are obtained for greater values of concentration. Although some high connectivity points (hubs) are found for wide range of values, it is clear that the mentioned relation is present on average. For instance, July and October have a greater amount of points with $k > 60$, meaning more hubs in the VG. Then the average degree $\bar{k}$ is computed and shown in Table 1, along with D_0. Values for the months of January and April are very close to each other, while the one for July is sensibly the highest and October has an intermediate value. This result is in accordance to the behavior of the ozone concentration, since for instance the month of July is the one with the greatest temperature and UV radiation rates and therefore elevated values are reached more often, producing a major number of hubs in the VG.

The second parameter is the entropy dimension D_1, which is usually known as the dimension that is related to the uniformity of the data and how different is the probability of certain regions to be visited by a randomly chosen box with respect to others. A way of measuring its uniformity is looking at the difference between D_0 and D_1, because when they are equal it means that the sample is uniform. The greater $(D_0 - D_1)$, the less uniform it is. Also, another reason to take the difference instead of the absolute value is because we are interested on comparing the curves and their behavior, and since the D_0 of each one is different, it is necessary to stablish a reference point. When it is translated to the context of VG, a greater uniformity would mean less difference between the degrees of the sample, hence the standard deviation of the degree would be decreased. In that case, what should be observed in the original ozone data is that the differences encountered in the data are less significative (what is indeed observed for the months with lower $D_0 - D_1$).
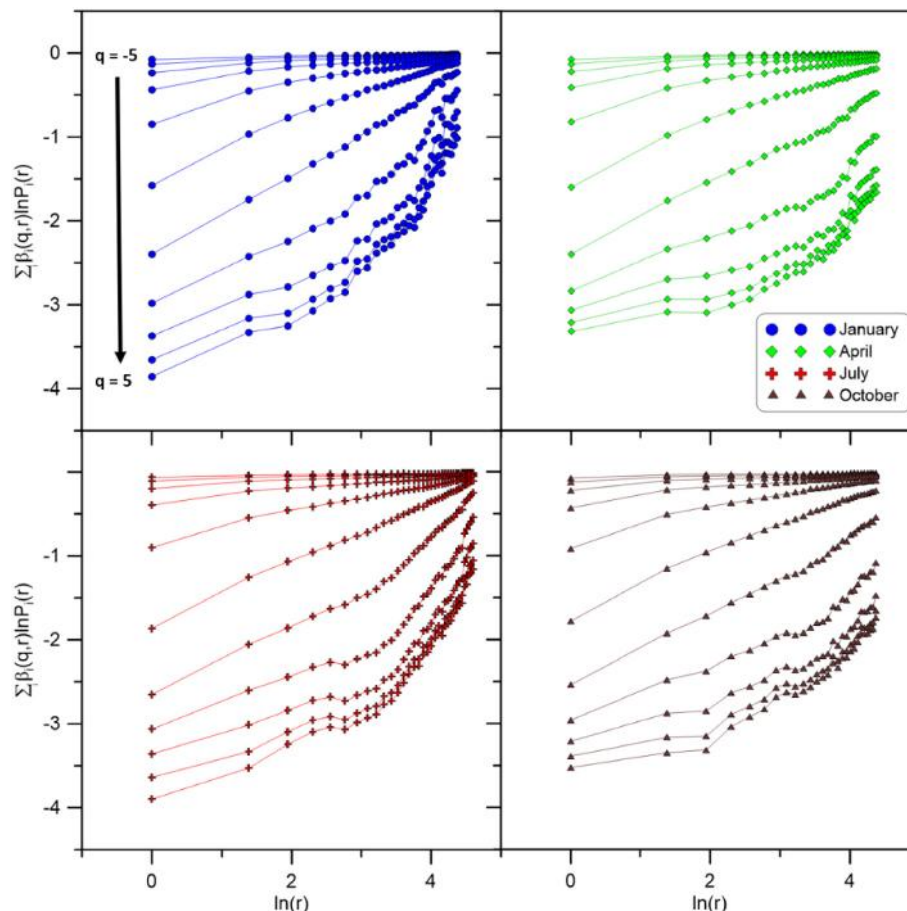

Fig. 7. Curves from where α is computed by regression.

The most uniform months would be January (minimum $D_0 - D_1$) and April (minimum σ_k), having both very similar results for those quantities in the two cases if the relative value with respect to July is considered, which is the month located in the opposite extreme. The reason for this is that in July the differences encountered between day and night concentrations of ozone are much greater than the other months, since during the night they remain at their minimum values for all months. In contrast, the maximum values are much higher in summer than the rest of the year because of the higher UV exposures and the opposite case for winter.

The last parameter that has been computed from the Rényi spectrum is the correlation dimension D_2. This dimension is based on the probability of finding a pair of points separated a distance lower than a given radius. In practice, D_2 can be used to see how often recurrence in a system can be expected: the higher the dimension is, the less recurrent and predictable will be the series (Ding et al., 1993). For the same comparison purposes as in $(D_0 - D_1)$, here $(D_0 - D_2)$ values are discussed as well, hence a greater value of this difference would mean a more recurrent VG. Once again, January and July exhibit the extreme cases whereas the other two months are located in between. That means that the VG of July is more recurrent and predictable than the one in January, with degrees that are repeated more often. Authors attribute this fact to the more unstable weather predominant in January, in contrast to the one in July, when the temperature and solar radiation is much more stable.

As commented before, another important feature that can be studied is the spectrum of α, following the Eqs. (6) and (7). Applying a similar approach to the one used to compute D_q, a linear regression is performed for each value of q to compute the corresponding magnitude (α in Fig. 7 and $f(\alpha)$ in Fig. 8). For each month, the range for the regression has

been chosen to be the same that was used for the Rényi dimensions, in order to conserve the scale used. In this case, the obtained curves have as well a positive trend that increases with q for the case of the curves of Fig. 7; and first increases and then decreases for Fig. 8 (as expected, because $f(\alpha)$ should have a maximum and then drop). As commented before, the artifact that made us change the chosen interval for the regression in January can be observed in Fig. 8 for high values of radius.

Once both α and $f(\alpha)$ are resolved for all the possible values of q, the spectrum of every month can be plotted, being the ones shown in Fig. 9. All the spectra as can be seen have their maximum at $f(\alpha) = 1$ and start at $\alpha = 0$, while the overall shape depends on each case. Several properties of the underlying signal can be extracted from the spectra. The width of the curve W is shown in Table 1, which is related with the multifractality degree of the signal (Ariza-Villaverde et al., 2013; Telesca et al., 2004).

According to the shapes of the curves, the spectra are strongly non-symmetric, being the right tail much more pronounced than the left one for all the different months. The right side is usually associated to the homogeneity of low values in a temporal distribution of data, whereas the left one is related with the same feature of the high values instead. In the context of visibility graphs, since the probability measure taken in the SBA corresponds to the number of nodes seen by each one selected as center, the homogeneity/heterogeneity of the highest values refer to the greatest degrees (hubs in the VG) and vice versa. Therefore, the behavior of the four months is very similar for high values (left tail) while it differs significantly for low ones (right tail). In Fig. 9, all of them show an heterogenous behavior since the α distribution is not uniform along the curve, being concentrated on the left and right extremes. Nevertheless, it is possible to extract from that figure that this heterogeneity is more pronounced in July. This fact suggests an influence of the higher
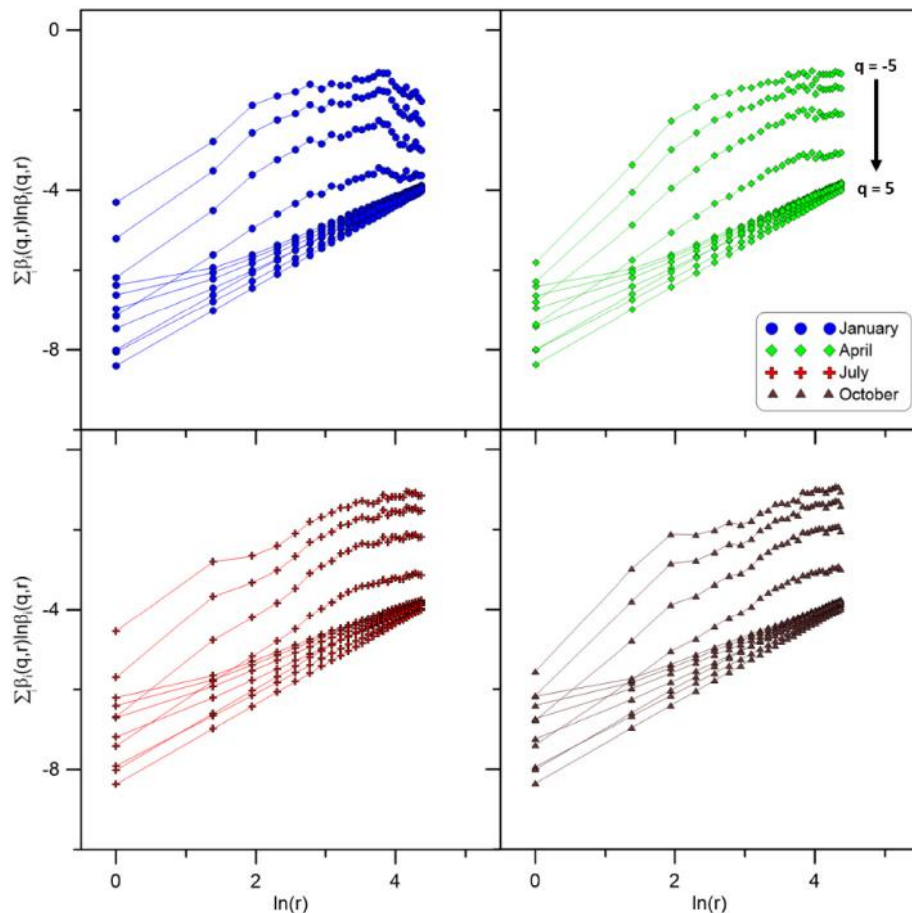

Fig. 8. Curves from where f(α) is computed using linear regressions.
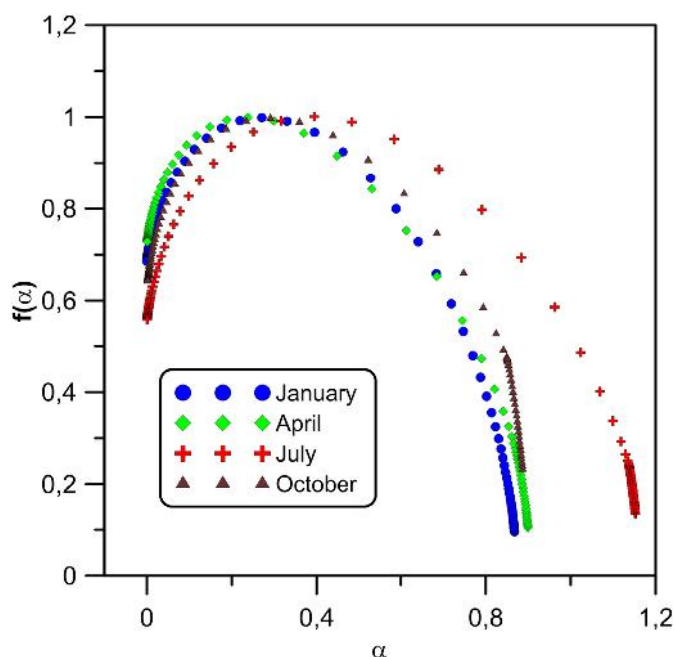
Fig. 9. α-spectra of all the months.

UV radiation that creates a greater difference between day and night ozone concentrations. This is in accordance with what was shown in the previous analysis described above for $(D_0 - D_1)$.

4. Conclusions

The results of this work show that the multifractal analysis of VGs from ground-level ozone concentration time series is a suitable tool to describe the seasonal dynamics of this air pollutant. VGs have proven to have advantages such as: i) their topology inherits the features of the associated time series, which ends up resulting on supplementary information through the degree distribution; ii) and also, this novel connection between time series and complex networks opens a broad range of possibilities within the study of complex signals.

When it comes to the multifractal analysis performed in this work, both the SBA and the Chhabra and Jensen method for the Rényi and singularity spectra respectively, where chosen based on their advantages with respect to other. The SBA overcomes the drawbacks of the box-counting algorithm for the computation of the generalized fractal dimensions for negative probability moment orders q. Furthermore, the Chhabra and Jensen method for the α-spectrum does not need a Legendre transform to be applied (with the associated error to it) and gives and independent measurement from the SBA as well, being more robust for comparison.

After applying the methods mentioned above, clear and coherent results that fits the expected behavior of the ozone dynamics were found for the different months. Furthermore, several properties can be directly inferred from the degree distribution almost at first sight, meaning a powerful tool for predicting results before any more complex data treatment is performed. The same relations between multifractal parameters such as capacity dimension, W of the α-spectrum amongst others are found in simple statistical parameters as the average or standard deviation of the degree distribution of the VG.

To conclude, authors would like to point out that this promising technique could be extended to other applications due to the many possibilities that complex networks have. One possible target would be the relation between ozone and its precursors, which could be looked by means of relating the VG of each one separately or using multi-layer networks as proposed by Lacasa et al. (2015).

References

Adame, J.A., Lozano, A., Bolívar, J.P., De la Morena, B.A., Contreras, J., Godoy, F., 2008. Behavior, distribution and variability of surface ozone at an arid region in the south of Iberian Peninsula (Seville, Spain). Chemosphere 70, 841–849. https://doi.org/10.1016/j.chemosphere.2007.07.009.

Ariza-Villaverde, A.B., Jiménez-Hornero, F.J., Gutiérrez de Ravé, E., 2013. Multifractal analysis applied to the study of the accuracy of DEM-based stream derivation. Geomorphology 197, 85–95. https://doi.org/10.1016/j.geomorph.2013.04.040.

Ariza-Villaverde, A.B., Jiménez-Hornero, F.J., Gutiérrez de Ravé, E., 2015. Influence of DEM resolution on drainage network extraction: a multifractal analysis. Geomorphology 241, 243–254. https://doi.org/10.1016/j.geomorph.2015.03.040.

Chhabra, A., Jensen, R.V., 1989. Direct determination of the f(α) singularity spectrum. Phys. Rev. Lett. 62, 1327–1330. https://doi.org/10.1103/PhysRevLett.62.1327.

De Bartolo, S.G., Gaudio, R., Gabriele, S., 2004. Multifractal analysis of river networks: sandbox approach. Water Resour. Res. 40. https://doi.org/10.1029/2003WR002760.

Ding, M., Grebogi, C., Ott, E., Sauer, T., Yorke, J.A., 1993. Estimating correlation dimension from a chaotic time series: when does plateau onset occur? Physica D 69, 404–424. https://doi.org/10.1016/0167-2789(93)90103-8.

Doherty, R.M., Heal, M.R., Wilkinson, P., Pattenden, S., Vieno, M., Armstrong, B., Atkinson, R., Chalabi, Z., Kovats, S., Milojevic, A., Stevenson, D.S., 2009. Current and future climate- and air pollution-mediated impacts on human health. Environ. Health 8 (S8). https://doi.org/10.1186/1476-069X-8-S1-S8.

Domínguez-López, D., Adame, J.A., Hernández-Ceballos, M.A., Vaca, F., De la Morena, B.A., Bolívar, J.P., 2014. Spatial and temporal variation of surface ozone, NO and NO2 at urban, suburban, rural and industrial sites in the southwest of the Iberian Peninsula. Environ. Monit. Assess. 186, 5337–5351. https://doi.org/10.1007/s10661-014-3783-9.

Feder, J., 1988. Fractals, Physics of Solids and Liquids. Springer, US.

Graedel, T.E., Crutzen, P.J., 1993. Atmospheric Change: An Earth System Perspective.

Guicherit, R., van Dop, H., 1977. Photochemical production of ozone in Western Europe (1971–1975) and its relation to meteorology. Atmos. Environ. 11, 145–155. https://doi.org/10.1016/0004-6981(77)90219-0.

Güsten, H., Heinrich, G., Weppner, J., Abdel-Aal, M.M., Abdel-Hay, F.A., Ramadan, A.B., Tawfik, F.S., Ahmed, D.M., Hassan, G.K.Y., Cvitaš, T., Jeftić, J., Klasinc, L., 1994. Ozone formation in the greater Cairo area. Sci. Total Environ. 155, 285–295. https://doi.org/10.1016/0048-9697(94)90507-X.

Halsey, T.C., Jensen, M.H., Kadanoff, L.P., Procaccia, I., Shraiman, B.I., 1986. Fractal measures and their singularities: the characterization of strange sets. Phys. Rev. A 33, 1141–1151. https://doi.org/10.1103/PhysRevA.33.1141.

He, H., 2017. Multifractal analysis of interactive patterns between meteorological factors and pollutants in urban and rural areas. Atmos. Environ. 149, 47–54. https://doi.org/10.1016/j.atmosenv.2016.11.004.

Jiménez-Hornero, F.J., Gutiérrez de Ravé, E., Ariza-Villaverde, A.B., Giráldez, J.V., 2010a. Description of the seasonal pattern in ozone concentration time series by using the strange attractor multifractal formalism. Environ. Monit. Assess. 160, 229–236. https://doi.org/10.1007/s10661-008-0690-y.

Jiménez-Hornero, Francisco J., Jiménez-Hornero, J.E., Gutiérrez de Ravé, E., Pavón-Domínguez, P., 2010b. Exploring the relationship between nitrogen dioxide and ground-level ozone by applying the joint multifractal analysis. Environ. Monit. Assess. 167, 675–684. https://doi.org/10.1007/s10661-009-1083-6.

Jiménez-Hornero, F.J., Ariza-Villaverde, A.B., De Ravé, E.G., 2013. Multifractal description of simulated flow velocity in idealised porous media by using the sandbox method. Fractals 21, 1350006. https://doi.org/10.1142/S0218348X13500060.

Kouvarakis, G., Tsigaridis, K., Kanakidou, M., Mihalopoulos, N., 2000. Temporal variations of surface regional background ozone over Crete Island in the southeast Mediterranean. J. Geophys. Res.-Atmos. 105, 4399–4407. https://doi.org/10.1029/1999JD900984.

Lacasa, L., Toral, R., 2010. Description of stochastic and chaotic series using visibility graphs. Phys. Rev. E 82. https://doi.org/10.1103/PhysRevE.82.036120.

Lacasa, L., Luque, B., Ballesteros, F., Luque, J., Nuño, J.C., 2008. From time series to complex networks: the visibility graph. Proc. Natl. Acad. Sci. 105, 4972–4975. https://doi.org/10.1073/pnas.0709247105.

Lacasa, L., Luque, B., Luque, J., Nuño, J.C., 2009. The visibility graph: a new method for estimating the Hurst exponent of fractional Brownian motion. EPL 86, 30001. https://doi.org/10.1209/0295-5075/86/30001.

Lacasa, L., Nicosia, V., Latora, V., 2015. Network structure of multivariate time series. Sci. Rep. 5, 15508. https://doi.org/10.1038/srep15508.

Mali, P., Manna, S.K., Mukhopadhyay, A., Haldar, P.K., Singh, G., 2018. Multifractal analysis of multiparticle emission data in the framework of visibility graph and sandbox algorithm. Physica A 493, 253–266. https://doi.org/10.1016/j.physa.2017.10.015.

Mandelbrot, B.B., 1974. Intermittent turbulence in self-similar cascades: divergence of high moments and dimension of the carrier. J. Fluid Mech. 62, 331–358.

Mandelbrot, B.B., 1982. The Fractal Geometry of Nature. W.H. Freeman, San Francisco.

Miao, W., Huang, X., Song, Y., 2017. An economic assessment of the health effects and crop yield losses caused by air pollution in mainland China. J. Environ. Sci. 56, 102–113. https://doi.org/10.1016/j.jes.2016.08.024.

Pavon-Dominguez, P., Jiménez-Hornero, F.J., Gutierrez de Rave, E., 2013. Multifractal analysis of ground–level ozone concentrations at urban, suburban and rural background monitoring sites in southwestern Iberian Peninsula. Atmos. Pollut. Res. 4, 229–237. https://doi.org/10.5094/APR.2013.024.

Pierini, J.O., Lovallo, M., Telesca, L., 2012. Visibility graph analysis of wind speed records measured in central Argentina. Physica A 391, 5041–5048. https://doi.org/10.1016/j.physa.2012.05.049.

Ribas, À., Peñuelas, J., 2004. Temporal patterns of surface ozone levels in different habitats of the North Western Mediterranean basin. Atmos. Environ. 38, 985–992. https://doi.org/10.1016/j.atmosenv.2003.10.045.

Sillman, S., 1999. The relation between ozone, NOx and hydrocarbons in urban and polluted rural environments. Atmos. Environ. 33, 1821–1845. https://doi.org/10.1016/S1352-2310(98)00345-8.

Tél, T., Fülöp, Á., Vicsek, T., 1989. Determination of fractal dimensions for geometrical multifractals. Physica A 159, 155–166. https://doi.org/10.1016/0378-4371(89)90563-3.

Telesca, L., Colangelo, G., Lapenna, V., Macchiato, M., 2004. Fluctuation dynamics in geoelectrical data: an investigation by using multifractal detrended fluctuation analysis. Phys. Lett. A 332, 398–404. https://doi.org/10.1016/j.physleta.2004.10.011.

Trainer, M., Parrish, D.D., Goldan, P.D., Roberts, J., Fehsenfeld, F.C., 2000. Review of observation-based analysis of the regional factors influencing ozone concentrations. Atmos. Environ. 17.

Veneziano, D., Moglen, G.E., Bras, R.L., 1995. Multifractal analysis: pitfalls of standard procedures and alternatives. Phys. Rev. E 52, 1387–1398. https://doi.org/10.1103/PhysRevE.52.1387.

Vicsek, T., Family, F., Meakin, P., 1990. Multifractal geometry of diffusion-limited aggregates. Europhys. Lett. 12, 217–222. https://doi.org/10.1209/0295-5075/12/3/005.

Zeleke, T.B., Si, B.C., 2006. Characterizing scale-dependent spatial relationships between soil properties using multifractal techniques. Geoderma 134, 440–452. https://doi.org/10.1016/j.geoderma.2006.03.013.

Publication 2
Can Complex Networks Describe the Urban and Rural Tropospheric O_3 Dynamics?

Journal

Chemosphere (Ed. Elsevier)

Position / Quartile

(29/265) / Q1 (Environmental Sciences)

Impact factor

5.778 (2019)

Can complex networks describe the urban and rural tropospheric O_3 dynamics?

Rafael Carmona-Cabezas[*], Javier Gómez-Gómez, Ana B. Ariza-Villaverde,
Eduardo Gutiérrez de Ravé, Francisco J. Jiménez-Hornero

Department of Graphic Engineering and Geomatic, University of Cordoba, Gregor Mendel Building (3rd floor), Campus Rabanales, Cordoba, 14071, Spain

GRAPHICAL ABSTRACT
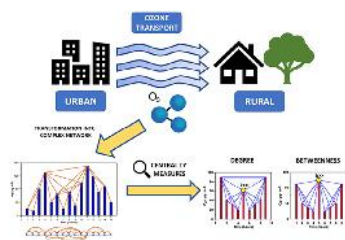
ARTICLE INFO

ABSTRACT

Tropospheric ozone (O_3) time series have been converted into complex networks through the recent so-called Visibility Graph (VG), using the data from air quality stations located in the western part of Andalusia (Spain). The aim is to apply this novel method to differentiate the behavior between rural and urban regions when it comes to the ozone dynamics. To do so, some centrality parameters of the resulting complex networks have been investigated: the degree, betweenness and shortest path. Some of them are expected to corroborate previous works in order to support the use of this technique; while others to supply new information.

Results coincide when describing the difference that tropospheric ozone exhibits seasonally and geographically. It is seen that ozone behavior is fractal, in accordance to previous works. Also, it has been demonstrated that this methodology is able to characterize the divergence encountered between measurements in urban environments and countryside.

In addition to that, the promising outcomes of this technique support the use of complex networks for the study of air pollutants dynamics. Particularly, new nuances are offered such as the identification and description of singularities in the signal.

© 2019 Elsevier Ltd. All rights reserved.

1. Introduction

In the last few decades, tropospheric ozone has been the focus of

many studies performed in different areas and scales around the world. This interest on ozone dynamics analysis and characterization has been awaken because it is one of the main photochemical oxidants on account of its abundance. When this irritant gas is found in high concentrations, severe impacts affect human health and harvest (Doherty et al., 2009). Those mentioned harms have a severe impact for economy, leading to losses of several billions of

dollars annually (Miao et al., 2017).

The gas object of the presented work is a secondary pollutant and it is known that its creation and destruction mechanisms are related to photochemical and nonlinear processes (Graedel and Crutzen, 1993; Trainer et al., 2000). The mentioned processes depend on meteorological features such as wind direction, temperature, and principally solar radiation (Graedel and Crutzen, 1993; Guicherit and van Dop, 1977). They have been analyzed for the Mediterranean basin by some authors previously (Cieslik and Labatut, 1997; Güsten et al., 1994; Kouvarakis et al., 2000; Ribas and Peñuelas, 2004). Besides, tropospheric ozone concentration levels depends also on the presence of other gases, such as nitrogen oxides and volatile organic compounds (called precursors) that are produced in the urban and industrial areas (Sillman, 1999). Because of all these considerations the analysis of the temporal dynamics of ozone becomes a very complex task. As a consequence, traditional statistical analysis of ozone may offer a limited view of more complex dynamics of signals where the level of variability is high (Pavón-Domínguez et al., 2015). To overcome such drawbacks, the multifractal approach has been successfully employed on this subject (He et al., 2017; Lee et al., 2006).

One of the focuses on this topic is the difference encountered between rural and urban areas (García-Gómez et al., 2016; Jirik et al., 2017; Kumar et al., 2015). Despite the fact that ozone is created mainly in urban nuclei, it is proven that higher concentrations of this contaminant are measured within rural regions, leading to reduction of crop yields among other environmental problems (Tai and Val Martin, 2017). One reason attributed to this phenomenon is the transport of ozone by the wind to the less industrialized regions. In those areas the destruction rate of ozone is less significative than in urban ones.

In the last few years, a new technique to analyze temporal series has been developed (Lacasa et al., 2008). It relies in the use of complex networks, which are defined as networks whose connections have patterns that are non-regular but non-random. They are obtained from the transformation of time series (ground-level ozone concentration in this case). It was given the name of *Visibility Graph* (VG). This method has been previously tested on other nonlinear time series (Lacasa et al., 2009; Pierini et al., 2012; Sannino et al., 2017), as well as ozone dynamics (Carmona-Cabezas et al., 2019); proving its potential and applicability in all of them. This transformation has proven to have several advantages such as: i) they inherit the features of the associated time series, which ends up resulting on additional feedback through the degree distribution (that will be defined later). ii) This technique offers non-trivial results as in the multifractal approach, but in this case they are unambiguous (i.e. a given time series will always be transformed into an unique VG). iii) In addition, they can be used for analyzing series of several variables simultaneously, which could be very helpful for finding correlations between tropospheric ozone and its precursors; and lastly, iv) this recent bridge between complex networks and temporal series opens a wide range of new opportunities within the study of complex signals.

In the presented work, the complex networks obtained from ozone time series through the VG are used to retrieve the centrality parameters. These parameters give information about the most important nodes in a system and will be further explained in subsection 2.3. Finally, the main purpose is to check the ability of this methodology to analyze the differences in the behavior of the tropospheric ozone between urban and rural environments. Besides, the possibility of obtaining new angles about this kind of series is explored.

2. Materials and methods

2.1. Data

The region that is the object of this analysis corresponds to the Guadalquivir Valley (South-western part of Spain) since the area has the proper orography, weather and anthropic conditions to be vulnerable to pollution by tropospheric ozone (Domínguez-López et al., 2014).

The data used here correspond to 1-h monthly ozone concentration values collected from 2013 to 2017. This is the maximum temporal coverage available in the studied region, supplied and validated by the European Environmental Agency (EEA). The measurements were performed at four different stations located in the province of Cadiz (see Fig. 1). Two of them are located in the southeastern part of the region (Algeciras and Alcornocales) and the others in the northwestern one (Cadiz and Prado del Rey). Algeciras and Cadiz correspond to urban areas, whereas the other two are situated within the natural reservoir named "Parque natural de Los Alcornocales", and so they have been labelled as rural.

These stations are part of the network that monitors the air pollution levels in the region of Andalusia, co-funded by the European Union and the Consejería de Medio Ambiente y Ordenación del Territorio (Regional Environmental and Territory Management Department). The data was collected and provided lastly by the EEA.

As an example, Fig. 2 shows ozone concentration time series for four months from 2015 for two different locations: Alcornocales (rural) and Algeciras (urban). One clear difference between the two of them is that in the urban station, the lowest values measured are in many occasions close to zero. On the other hand, in the rural one

Fig. 1. Location of the air quality control stations from which the data was retrieved. The image in the left-bottom corner shows the position of the studied area in the Iberian Peninsula. Green area indicates the natural reservoir "Parque natural de Los Alcornocales". (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)
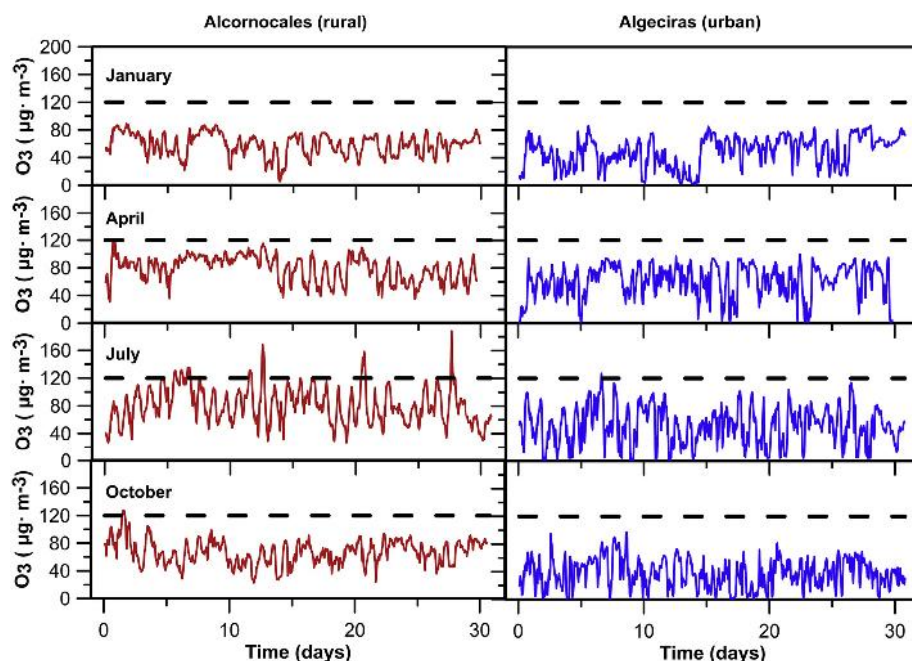
Fig. 2. Examples of ozone concentration time series for four months and a given year (2015) in two locations, one rural (red) and the other urban (blue). The dashed line indicates the value 120 μg/m³ stablished as a reference (World Health Organization, 2005). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

that is not indeed the case: the ozone concentration does not vanish in the whole month or barely does it. This behavior can be extended to all the rest of the years included in the presented study.

As it was clearly seen in Fig. 2, the ozone concentrations reach especially high values in summer, where the case in winter is the exact opposite. Those differences were observed for this area (Adame et al., 2008; Jiménez-Hornero et al., 2010), and the reason is that the most suitable conditions for the ozone creation are found around summer. The temperature and solar radiation progressively raise and reach their peak in July, which allows higher creation rates and therefore concentration. One of the reactions that governs this mechanism is the following (Graedel and Crutzen, 1993).

$$NO_2 + O_2 \leftrightarrow O_3 + NO \qquad (1)$$

This photochemical reaction is reversible and tends to the ozone creation (rightwards direction) when there is energy (light) available and vice versa when there is not. For that reason, the highest and lowest values of ozone are always measured during the day and night, respectively. The same happens summer and winter, as discussed previously.

Looking at the contrast between urban and rural environments, greater concentrations are usually found in the second ones. This well-known effect corresponds to the predominance of leftwards direction of the photochemical reaction described before in Eq. (1). After the ozone is created, it starts to react with the Nitrogen Oxide during night conditions (absence of light and lower temperatures). As can be observed in Figs. 2 and 4, the ozone values reach minima of zero quite often in the urban area of Algeciras for instance, which is not the case for the rural location of Alcornocales. That is directly related to the higher concentrations of NO from factories and vehicle emissions that can be found in an industrial area such as Algeciras; in contraposition to the natural reservoir of Parque natural de los Alcornocales (Sillman, 1999). Therefore, the ozone that is created and transported to the rural areas studied here cannot be transformed to other gases at the same rate as in the city, leading to higher concentrations values on average.
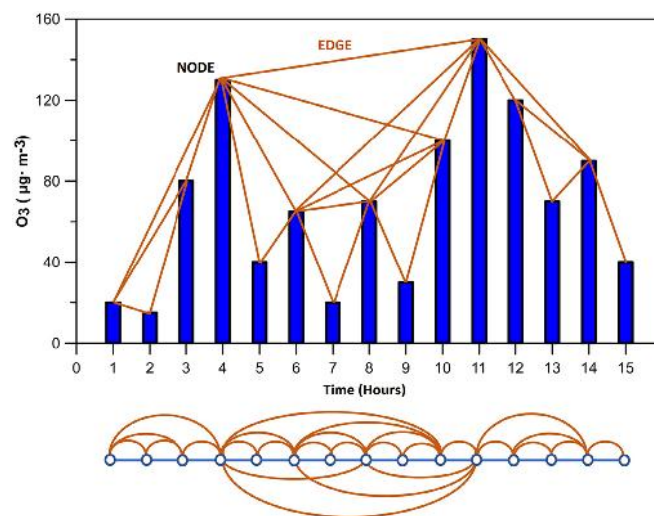

Fig. 3. Sample time series transformed into a complex network through the Visibility Graph algorithm. Below, all the connections are shown in a more visual way.

2.2. Visibility Graphs

The most common definition of a graph is a set of points or vertices that are connected to each other. The lines that connect them are called *edges,* while the interconnected points are usually known as *nodes* (see Fig. 3). As commented above, a tool that makes possible the transformation of a time series into a graph was introduced by Lacasa et al. (2008) and called Visibility Graph (VG). One of the main features that rises the interest of researchers relies on the fact that it inherits many properties of the original series.

The first thing that must be done for constructing the visibility adjacency matrix (which contains all the information of the new network), is to stablish a method to decide which points (or nodes)
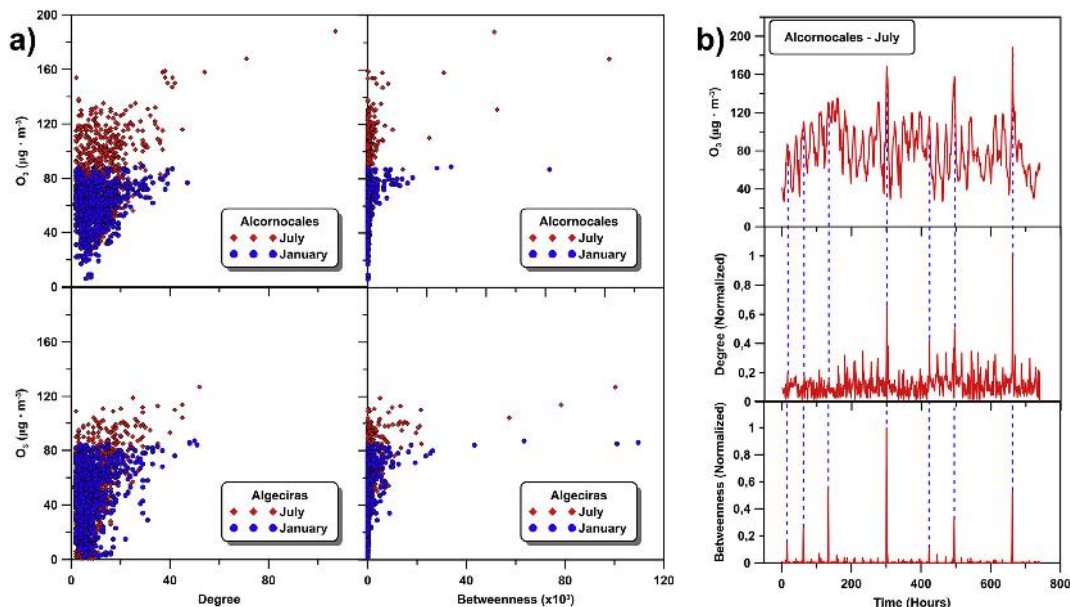
Fig. 4. a) Plots of the values of tropospheric ozone concentration against the degree and betweenness of each point. b) Temporal distribution of these two quantities and the ozone concentration for a given month and location. Blue dashed lines in b) are used to associate several peaks. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

in the system are connected to each other. Further in the text, "visibility" will be understood as connectivity in the VG. The criterion is the following: two arbitrary points from the time series (t_a, y_a) and (t_b, y_b) will have visibility (will be connected in the graph) if any given point (t_c, y_c) that is located in between ($t_a < t_c < t_b$) fulfills the following condition:

$$y_c < y_a + (y_b - y_a)\frac{t_c - t_a}{t_b - t_a} \qquad (2)$$

After checking this condition for every pair of points in the series, it is possible to stablish which pairs have visibility and which others do not. With this information, the VG (orange lines in Fig. 3) is constructed, being a mathematical object totally different from a time series (blue bars).

One example of how this method works can be seen in Fig. 3, where it is applied to a sample time series as an illustration. It can be observed how some pair of nodes fulfill the condition in Eq. (2) and have visibility, leading to a resulting complex network. This new graph will inherit the complexity of the original series (Lacasa et al., 2008; Lacasa and Toral, 2010), meaning that a regular graph would be created from a periodic time series, for instance.

The resulting complex network is represented by a $N \times N$ adjacency binary matrix, with N the total number of points in the studied time series. The information of the nodes is given by each row of the matrix, so that $a_{ij} = 1$ means that the node i and j have visibility; whereas $a_{ij} = 0$ means the opposite case (no edge connects those two nodes). The VG algorithm can be substantially simplified (reducing the computational cost of the process) if some considerations are taken. These can be done thanks to the properties that the adjacency matrix holds. The properties are listed below:

- Hollow matrix: Since there are no intermediate nodes to fulfill the condition, in the case of the diagonal, all the elements are zero ($a_{ii} = 0$). Hence a node does not have visibility with itself.
- Symmetric matrix: Due to the reciprocity of the visibility between two nodes, all the nodes in system fulfill $a_{ij} = a_{ji}$. This is a

property of all undirected graphs, which are those whose nodes are linked by edges that do not have direction.
- Nearest neighbors: The elements that surround the diagonal are always 1 ($a_{ij} = 1$ for $j = i \pm 1$). This is because each point always sees the closest previous and next node (there are no points in between to prevent the visibility).

With all these considerations, every visibility adjacency matrix has a general form as follows:

$$A = \begin{pmatrix} 0 & 1 & \dots & a_{1,N} \\ 1 & 0 & 1 & \vdots \\ \vdots & 1 & \ddots & 1 \\ a_{N,1} & \dots & 1 & 0 \end{pmatrix} \qquad (3)$$

2.3. Centrality measures

When trying to retrieve information from a given complex network, one of the most commonly used approaches is to discern which of them are the most important nodes in the system. To this purpose, centrality measures comes usually in handy. "Centrality" has the meaning of node importance in the context of complex networks. This concept was initially applied to the study of social networks and later transferred to other fields of knowledge (Agryzkov et al., 2019; Joyce et al., 2010; Liu et al., 2015). This work has been focused on two of them: the degree and betweenness centrality measures, which will be explained afterwards.

2.3.1. Degree centrality

The first and most widely used centrality measure is the so-called degree. A possible definition for the degree of a node (k_i) is the number of other nodes that have visibility with it ($k_i = \sum a_{ij}$). For instance, in Fig. 3, the degree of the three first nodes are $k_1^j = 3$, $k_2 = 2$ and $k_3 = 3$, respectively. A higher value of k_i would mean a greater centrality. Looking at Fig. 4a, thanks to the $v - k$ (i.e. ozone concentration vs degree) plots (Pierini et al., 2012), it is possible to

appreciate how the most central nodes from the VG are related to the highest values of ozone concentration. On the whole, it is possible to obtain the probability that corresponds to each degree, by simply counting how many times each value is repeated. From there, the degree distribution of the sample $P(k)$ can be retrieved.

By analyzing the degree distribution that is built from the VG it is possible to describe the nature of the time series, as previous works have shown (Lacasa et al., 2008; Mali et al., 2018; Pierini et al., 2012). It has been probed its capability to distinguish between fractal, random or periodic signals, for instance. Thus, by studying the degree distribution, a first insight of the behavior of the ozone concentration time series can be yielded as first step before getting into a more complex analysis. As some previous works explain (Lacasa et al., 2009; Lacasa and Toral, 2010), time series which have VGs whose degree distributions can be fitted to a power law $P(k) \propto k^{-\gamma}$ correspond to scale free due to the effect of *hub* repulsion (Song et al., 2006). The term *hub* refers to the nodes with unlikely highest number of links (highest degrees, see Figs. 4 and 5). The right tail of each degree distribution, governed by those *hubs*, can be represented in a log-log plot and fitted by a simple linear regression. The slope obtained by this regression provides an interesting parameter, the so-called γ exponent, which has already been used in some works (Lacasa and Toral, 2010; Mali et al., 2018).

2.3.2. Betweenness centrality

Before presenting this quantity, it is necessary to introduce a definition for the shortest path (SP). A definition of the SP for a pair of nodes (i, j) in a VG is the minimum possible number of edges to reach node j starting from node i. Consequently, the SP between two consecutive nodes $(i\,i+1)$ will be the unit.

The betweenness of a node (b_i) for an undirected graph (which was previously explained as a property of VG adjacency matrices) is defined as the total number of SPs which passes through this node and mathematically:

$$b_i = \sum_{\substack{j=1 \\ j \neq i}}^{N} \sum_{\substack{k=1 \\ k \neq i,j}}^{N} \frac{n_{jk}(i)}{2N_{jk}} \qquad (4)$$

where N_{jk} is the total number of SPs from node j to node k and $n_{jk}(i)$ is the number of SPs from node j to k, that contains the node i. It is divided by two in order not to repeat the same pair of nodes twice in an undirected graph.

The importance of such parameter lies on the fact that it measures the centrality of a node by considering whether it is between many of the nodes (Latora et al., 2017). In an equivalent way to the degree *hubs*, some points with a remarkably higher betweenness exist. For the sake of clarity, authors propose the term *skyline hubs* to refer to those with unlikely high betweenness. This name has been chosen because of its similarity to the skyline drawn by the skyscrapers in a city. Those important nodes are characterized by acting as bottlenecks in the network structure. This means that they accumulate a great number of SPs between nodes located at $t < t_{hub}$ and $t > t_{hub}$, respectively (Donner and Donges, 2012). It is possible to see how they are related to degree *hubs* in Fig. 4b and therefore to some local maxima of the time series. In short, *skyline hubs* constitute a more selective way to identify relevant nodes in the signal.

3. Results and discussion

The first step was to transform the ozone concentration time series from all the locations for all the months (from 2013 to 2017) into complex networks with the VG algorithm. In this section, the

results that are shown and discussed were obtained from the direct analysis of these networks.

3.1. Degree centrality

Following the definition of degree (k) given previously (subsection 2.3.1), the number of edges connected to each node in the different VGs was computed. From these values, it has been possible to construct the degree distribution of the networks, as shown in Fig. 5. In this plot, the degree distribution of the whole year (2015 shown as reference) was computed; with the final aim of discerning whether this distribution could give some insight on the difference between ozone dynamics in rural and urbanized areas. This first guess was motivated by previous works that use it in order to analyze the nature of the time series for several quantities (Lacasa and Toral, 2010; Mali et al., 2018).

The results for all the years and months are quite similar indeed, in accordance with previous studies (Carmona-Cabezas et al., 2019); and for that reason, only one year (2015) is used for the sake of clarity in Fig. 5. As can be seen in the cited plot, the tail of the distributions follows a power law of the form $P(k) \propto k^{-\gamma}$, that leads to a linear part of the curve when plotted in logarithmic scale both k and $P(k)$. This behavior points to the fractal nature of the signal, which was expected looking at some prior analyses (Jiménez-Hornero et al., 2010; Pavon-Dominguez et al., 2013). The slope of the linear portion in absolute value leads to the computation of this γ parameter. It is clear that the trend is negative, since the nodes with biggest degrees and known as *hubs* correspond usually to the high values of the distribution (see Fig. 4), whose likeliness is very low. In all the cases studied here, the distributions are almost overlapped and finally the exponent $\gamma \sim 3.4$ roughly for all of them (see Table 1), as depicted in Fig. 5. For that reason, this parameter is not able to distinguish the dynamics of the tropospheric ozone in the different regions on which this study focuses (urban - rural). Nevertheless, it does give useful information about the nature of
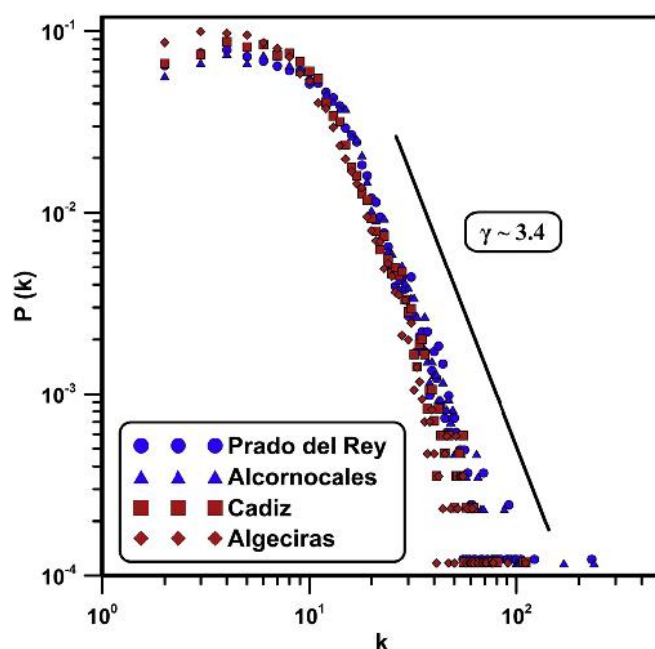
Fig. 5. Annually degree distribution for the four places studied. The shown year corresponds to 2015 as an example, since all the other years have been checked to give equivalent information. The red and blue colors refer to urban and rural environments, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

Table 1
Mean concentration and gamma exponent for each location (averaged for all the years).

	Location	$\overline{O_3}(\mu g \cdot m^{-3})$	$\overline{\gamma}$
Northwestern coast	Prado del Rey	81±4	3.50±0.13
	Cadiz	69±3	3.37±0.25
Southeastern coast	Alcornocales	72±3	3.39±0.15
	Algeciras	54±5	3.40±0.12

the time series as discussed and validates the data, since equivalent studies for different years and geographical area gave similar values of γ (Carmona-Cabezas et al., 2019).

Looking at the average degree ($\overline{k}$) of all the nodes from the VGs of each month, some information can be drawn. In Fig. 6a), this averaged value is plotted for each month and the first thing that can be commented is that the shape of the curves changes along the year. For all the studied locations, there is an increasing tendency towards summer that then decays typically after August. That is in the end due to the more suitable conditions for ozone formation that exist in summer with respect to the other season (specially winter). As seen, this quantity drops around April and November, being a possible explanation the weather of spring and autumn, unstable by nature, that favors the dispersion of gases and particles in the air (tropospheric ozone amongst them), as discussed in other studies (Dueñas et al., 2002). This has been encountered as well in a recent study performed at a different location and year with the same methodology (Carmona-Cabezas et al., 2019).

It can be deducted that a greater $\overline{k}$ would mean a higher number of *hubs* in the signal and those are related to the greatest concentrations of ozone as in fact it is shown in Fig. 4. Thus, this quantity can give equivalent information to the one retrieved by the study of ozone magnitude from the descriptive statistics.

Also, looking at Fig. 6a, a clear difference between the curves of the rural (blue) and urban (red) areas is found. Having all the same behavior mentioned before, the values of Prado del Rey and Alcornocales (both rural) are sensibly higher than those of Algeciras and Cadiz (urban). The contrast between summer and winter is as well more pronounced in the rural locations. This fact is attributed to the already known transport of ozone with the wind (Dueñas et al., 2004). Added to this is the process of destruction of this secondary contaminant in the urban areas in opposite to the rural ones where it is not significative (Kumar et al., 2015). This can be corroborated by regarding Table 1, where the total average ozone concentration of each place is shown.

Another parameter from the computed degrees that can be seen

(Fig. 6b) is the standard deviation of k (σ_k). As it was discussed in a previous work on this topic (Carmona-Cabezas et al., 2019), it is related to the differences encountered between daily and night values of ozone concentration. Hence, this σ_k can be associated to the variability of the time series. As expected, maxima are found again for summer, seemingly due to the fact that the biggest differences in UV radiation and temperature in this area are found in that season between day and night. Again, drops of this quantity σ_k are observed in both spring and autumn, which can be imputed to the same reason as in the case of $\overline{k}$.

Now the curves of σ_k in rural and urban areas are more similar than in the case $\overline{k}$, but still a difference can be observed. This lower distinction with respect to the curves of $\overline{k}$ can be explained using the same effect discussed before: the destruction of ozone is lower in the rural areas.

Based on the exposed facts, authors point out that the degree centrality is able to describe ozone dynamics differences from urban to rural areas, in an equivalent manner to those of the descriptive statistics. As seen, it gives information about the magnitude and variability. Therefore, it is confirmed that it is possible to employ this complex network methodology to such purpose, with the advantages that it implies, as anteriorly commented.

3.2. Betweenness centrality

After the analysis of the degree results, authors checked the suitability of the next centrality parameter to describe ozone temporal dynamics: betweenness (b). It was motivated by the seek of new insights that it could possible give. Moreover, the average SP quantity was obtained with the same purpose, since it is directly related to b (see Eq. (4)). Both of them were computed for every node in the VGs (from each month). After the information of each node was retrieved, the average was taken for SP, whereas in the case of the betweenness centrality, the median has been chosen. This decision was inspired by the fact that the distribution of the betweenness is much more skewed than those of the degree centrality and SP. In addition to that, the vast majority of values of b are zero or very close to it (see Fig. 4). For those reasons, authors consider median from a month as a more representative measure of the overall behavior rather than the mean. Averaged results over all years for monthly b and SP can be observed in Fig. 7, where the different locations studied are included.

On the one hand, in Fig. 7a) the betweenness centrality shows a seasonal pattern as well as the degree, being more pronounced in the rural areas than in the urban ones. As can be easily seen, the
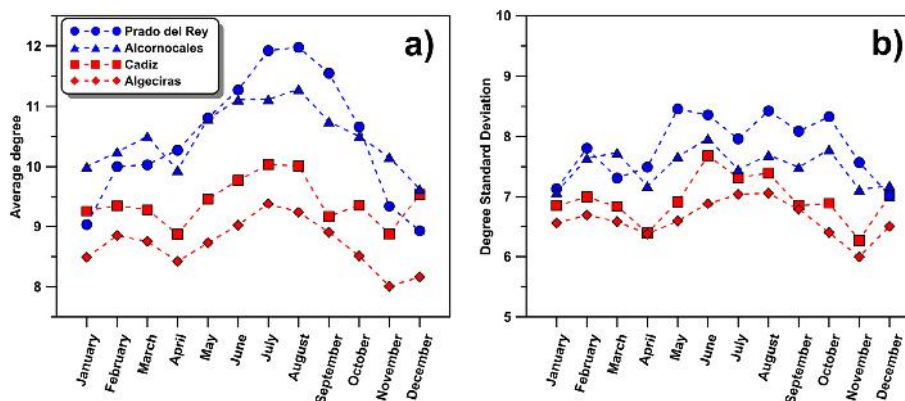

Fig. 6. Computed average degree and standard deviation from the degree distribution of each month in the four locations considered. Each monthly value is the average of the computed ones from all the years available.
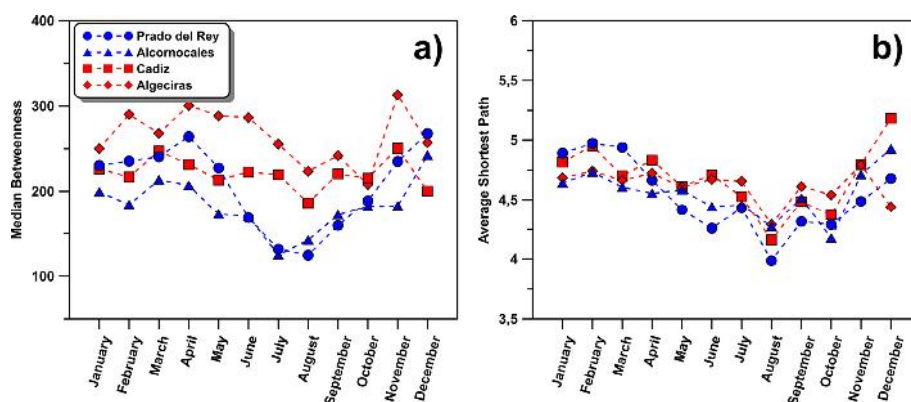
Fig. 7. a) Median betweenness and b) average SP computed from the VG of each month in the four locations. Blue values indicate rural area, while the red ones are urban. Each monthly value is the average of the computed ones from all the years available. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

minimum values are reached for late spring, summer and early autumn (from May to October), in contrast to the degree centrality, which was maximum for this period.

Authors attribute this seasonal pattern to the degeneracy of the SP. The definition of b is based on this degeneracy N_{jk} (see Eq. (4)). The larger N_{jk}, the lower will be the resulting value of betweenness for each node, leading to a final lower median for all months and vice versa. To authors mind, the degeneracy can be related to the seasonal behavior of tropospheric ozone. More precisely, since the SPs use mainly the highest values (*skyline hubs*) to cover most of the distance between two nodes, this feature provides information about diurnal maxima concentrations. Consequently, the degeneracy N_{jk} could be used to characterize the behavior of the high singularities in the signal.

As a result, a lower degeneracy (higher b) can be interpreted as an irregular maxima tendency with more *skyline hubs* and the opposite. This is due to the bottleneck effect for edges that lessens the degeneracy around those hubs. For the same reasons, the SPs are also minimalized.

Physically speaking, it can be inferred from these results that the behavior of the diurnal maxima reflects a more regular envelope (the upper curve joining the successive peaks in the signal) for the summer months in contrast to winter. It suggests that in this period (summer) it is more difficult to find abnormal diurnal peaks of ozone concentration.

On the other hand, differences have been encountered in b between rural and urban environments (Fig. 7a). In the case of rural ones, the betweenness profiles are lower than the those from the urban areas. Nonetheless, this divergence is not shown in the SP patterns, which are almost overlapped (Fig. 7b). The reason behind is that degeneracy is not taken into account in the computation of the average of this quantity. Only the length is used and not the number of possible SP between two given nodes, which is not the case for the betweenness, as explained before.

Following the previous reasonings for the seasonal behavior, results seem to indicate that the rural locations exhibit a more regular envelope of diurnal maxima. As a consequence, the urban areas would have more unexpected peaks in general, that could be related to a greater anthropogenic activity in those regions.

4. Conclusions

On the whole, results show that the use of complex networks for analyzing temporal series of tropospheric ozone is also suitable to distinguish the dynamics in rural and urban areas. The probability distribution of the degree centrality $P(k)$ identifies the nature of the signals, being fractal for all the cases, as it was previously known (Jiménez-Hornero et al., 2010; Pavon-Dominguez et al., 2013). Moreover, by looking at the values of $\bar{k}$ and σ_k a seasonal behavior has been observed. Besides, clear differences between rural and urban locations can be appreciated from those values, especially in the case of the average degree. Additionally, betweenness centrality might be considered a supplementary source of information for understanding the diurnal maxima behavior (upper envelope of the signals). All these outcomes support the capability of complex network analysis to describe ozone dynamics and transport, being an interesting auxiliary technique to the ones that already exist.

Finally, the advantages of using VGs for the analysis of time series must be remarked. In the last years, advances in the field of complex networks have made them a very convenient tool for several reasons: their computation efficiency, suitability for big data series and wide range of application, among others. In addition to that, when applied to time series, complex networks provide unambiguous results that can be compared to the multifractal formalism. Since VG is a state-of-the-art methodology, it opens multiple possibilities for future works. Authors consider appropriate to focus on the use of *Multiplex Visibility Graphs* (Lacasa et al., 2015) to study multivariate time series. Furthermore, the concept of *skyline hubs* could be employed to identify relevant points in a time series, exploring the betweenness centrality parameter applications on time series.

References

Adame, J.A., Lozano, A., Bolívar, J.P., De la Morena, B.A., Contreras, J., Godoy, F., 2008. Behavior, distribution and variability of surface ozone at an arid region in the south of Iberian Peninsula (Seville, Spain). Chemosphere 70, 841—849. https://doi.org/10.1016/j.chemosphere.2007.07.009.

Agryzkov, T., Tortosa, L., Vicent, J.F., 2019. A variant of the current flow betweenness centrality and its application in urban networks. Appl. Math. Comput. 347, 600—615. https://doi.org/10.1016/j.amc.2018.11.032.

Carmona-Cabezas, R., Ariza-Villaverde, A.B., Gutiérrez de Ravé, E., Jiménez-Hornero, F.J., 2019. Visibility graphs of ground-level ozone time series: A multifractal analysis. Sci. Total Environ. 661, 138—147. https://doi.org/10.1016/j.

scitotenv.2019.01.147.

Cieslik, S., Labatut, A., 1997. Ozone and heat fluxes over a Mediterranean pseudosteppe. Atmos. Environ. 31, 177—184. https://doi.org/10.1016/S1352-2310(97)00084-8.

Doherty, R.M., Heal, M.R., Wilkinson, P., Pattenden, S., Vieno, M., Armstrong, B., Atkinson, R., Chalabi, Z., Kovats, S., Milojevic, A., Stevenson, D.S., 2009. Current and future climate- and air pollution-mediated impacts on human health. Environ. Health 8, S8. https://doi.org/10.1186/1476-069X-8-S1-S8.

Domínguez-López, D., Adame, J.A., Hernández-Ceballos, M.A., Vaca, F., De la Morena, B.A., Bolívar, J.P., 2014. Spatial and temporal variation of surface ozone, NO and NO2 at urban, suburban, rural and industrial sites in the southwest of the Iberian Peninsula. Environ. Monit. Assess. 186, 5337—5351. https://doi.org/10.1007/s10661-014-3783-9.

Donner, R.V., Donges, J.F., 2012. Visibility graph analysis of geophysical time series: Potentials and possible pitfalls. Acta Geophys. 60, 589—623. https://doi.org/10.2478/s11600-012-0032-x.

Dueñas, C., Fernández, M.C., Cañete, S., Carretero, J., Liger, E., 2004. Analyses of ozone in urban and rural sites in Málaga (Spain). Chemosphere 56, 631—639. https://doi.org/10.1016/j.chemosphere.2004.04.013.

Dueñas, C., Fernández, M.C., Cañete, S., Carretero, J., Liger, E., 2002. Assessment of ozone variations and meteorological effects in an urban area in the mediterranean coast. Sci. Total Environ. 299, 97—113. https://doi.org/10.1016/S0048-9697(02)00251-6.

García-Gómez, H., Aguillaume, L., Izquieta-Rojano, S., Valiño, F., Àvila, A., Elustondo, D., Santamaría, J.M., Alastuey, A., Calvete-Sogo, H., González-Fernández, I., Alonso, R., 2016. Atmospheric pollutants in peri-urban forests of Quercus ilex: Evidence of pollution abatement and threats for vegetation. Environ. Sci. Pollut. Res. 23, 6400—6413. https://doi.org/10.1007/s11356-015-5862-z.

Graedel, T.E., Crutzen, P.J., 1993. Atmospheric change: An earth system perspective. J. Chem. Educ. https://doi.org/10.1021/ed070pA252.2.

Guicherit, R., van Dop, H., 1977. Photochemical production of ozone in Western Europe (1971–1975) and its relation to meteorology. Atmos. Environ. 11, 145—155. https://doi.org/10.1016/0004-6981(77)90219-0.

Güsten, H., Heinrich, G., Weppner, J., Abdel-Aal, M.M., Abdel-Hay, F.A., Ramadan, A.B., Tawfik, F.S., Ahmed, D.M., Hassan, G.K.Y., Cvitaš, T., Jeftić, J., Klasinc, L., 1994. Ozone formation in the greater Cairo area. Sci. Total Environ. 155, 285—295. https://doi.org/10.1016/0048-9697(94)90507-X.

He, H., Qiao, Z., Pan, W., Lu, W.-Z., 2017. Multiscale multifractal properties between ground-level ozone and its precursors in rural area in Hong Kong. J. Environ. Manag. 196, 270—277. https://doi.org/10.1016/j.jenvman.2017.02.024.

Jiménez-Hornero, F.J., Gutiérrez de Ravé, E., Ariza-Villarverde, A.B., Giráldez, J.V., 2010. Description of the seasonal pattern in ozone concentration time series by using the strange attractor multifractal formalism. Environ. Monit. Assess. 160, 229—236. https://doi.org/10.1007/s10661-008-0690-y.

Jirik, V., Brezna, B., Machaczka, O., Honkysova, S., Miturova, H., Janout, V., 2017. Associations between air pollution in the industrial and suburban parts of Ostrava city and their use. Environ. Monit. Assess. 189. https://doi.org/10.1007/s10661-017-6094-0.

Joyce, K.E., Laurienti, P.J., Burdette, J.H., Hayasaka, S., 2010. A new measure of centrality for brain networks. PLoS One 5. https://doi.org/10.1371/journal.pone.0012200.

Kouvarakis, G., Tsigaridis, K., Kanakidou, M., Mihalopoulos, N., 2000. Temporal variations of surface regional background ozone over Crete Island in the southeast mediterranean. J. Geophys. Res. Atmos. 105, 4399—4407. https://doi.org/10.1029/1999JD900984.

Kumar, A., Singh, D., Singh, B.P., Singh, M., Anandam, K., Kumar, K., Jain, V.K., 2015. Spatial and temporal variability of surface ozone and nitrogen oxides in urban and rural ambient air of Delhi-NCR, India. Air Qual. Atmos. Health 8, 391—399.

https://doi.org/10.1007/s11869-014-0309-0.

Lacasa, L., Luque, B., Ballesteros, F., Luque, J., Nuño, J.C., 2008. From time series to complex networks: The visibility graph. Proc. Natl. Acad. Sci. Unit. States Am. 105, 4972—4975. https://doi.org/10.1073/pnas.0709247105.

Lacasa, L., Luque, B., Luque, J., Nuño, J.C., 2009. The visibility graph: A new method for estimating the Hurst exponent of fractional Brownian motion. Europhys. Lett. 86, 30001. https://doi.org/10.1209/0295-5075/86/30001.

Lacasa, L., Nicosia, V., Latora, V., 2015. Network structure of multivariate time series. Sci. Rep. 5, 15508. https://doi.org/10.1038/srep15508.

Lacasa, L., Toral, R., 2010. Description of stochastic and chaotic series using visibility graphs. Phys. Rev. E 82, 036120. https://doi.org/10.1103/PhysRevE.82.036120.

Latora, V., Nicosia, V., Russo, G., 2017. Complex Networks. Cambridge University Press.

Lee, C.-K., Juang, L.-C., Wang, C.-C., Liao, Y.-Y., Yu, C.-C., Liu, Y.-C., Ho, D.-S., 2006. Scaling characteristics in ozone concentration time series (OCTS). Chemosphere 62, 934—946. https://doi.org/10.1016/j.chemosphere.2005.05.046.

Liu, C., Zhan, X.-X., Zhang, Z.-K., Sun, G.-Q., Hui, P.M., 2015. Events determine spreading patterns: Information transmission via internal and external influences on social networks. New J. Phys. 17, 113045. https://doi.org/10.1088/1367-2630/17/11/113045.

Mali, P., Manna, S.K., Mukhopadhyay, A., Haldar, P.K., Singh, G., 2018. Multifractal analysis of multiparticle emission data in the framework of visibility graph and sandbox algorithm. Phys. Stat. Mech. Its Appl. 493, 253—266. https://doi.org/10.1016/j.physa.2017.10.015.

Miao, W., Huang, X., Song, Y., 2017. An economic assessment of the health effects and crop yield losses caused by air pollution in mainland China. J. Environ. Sci. 56, 102—113. https://doi.org/10.1016/j.jes.2016.08.024.

Pavón-Domínguez, P., Jiménez-Hornero, F.J., Gutiérrez de Ravé, E., 2015. Joint multifractal analysis of the influence of temperature and nitrogen dioxide on tropospheric ozone. Stoch. Environ. Res. Risk Assess. 29, 1881—1889. https://doi.org/10.1007/s00477-014-0973-5.

Pavon-Dominguez, P., Jimenez-Hornero, F.J., Gutierrez de Rave, E., 2013. Multifractal analysis of ground—level ozone concentrations at urban, suburban and rural background monitoring sites in Southwestern Iberian Peninsula. Atmos. Pollut. Res. 4, 229—237. https://doi.org/10.5094/APR.2013.024.

Pierini, J.O., Lovallo, M., Telesca, L., 2012. Visibility graph analysis of wind speed records measured in central Argentina. Phys. Stat. Mech. Its Appl. 391, 5041—5048. https://doi.org/10.1016/j.physa.2012.05.049.

Ribas, À., Peñuelas, J., 2004. Temporal patterns of surface ozone levels in different habitats of the North Western Mediterranean basin. Atmos. Environ. 38, 985—992. https://doi.org/10.1016/j.atmosenv.2003.10.045.

Sannino, S., Stramaglia, S., Lacasa, L., Marinazzo, D., 2017. Visibility graphs for fMRI data: Multiplex temporal graphs and their modulations across resting state networks. Netw. Neurosci. 208—221. https://doi.org/10.1101/106443.

Sillman, S., 1999. The relation between ozone, NOx and hydrocarbons in urban and polluted rural environments. Atmos. Environ. 33, 1821—1845. https://doi.org/10.1016/S1352-2310(98)00345-8.

Song, C., Havlin, S., Makse, H.A., 2006. Origins of fractality in the growth of complex networks. Nat. Phys. 2, 275—281. https://doi.org/10.1038/nphys266.

Tai, A.P.K., Val Martin, M., 2017. Impacts of ozone air pollution and temperature extremes on crop yields: Spatial variability, adaptation and implications for future food security. Atmos. Environ. 169, 11—21. https://doi.org/10.1016/j.atmosenv.2017.09.002.

Trainer, M., Parrish, D.D., Goldan, P.D., Roberts, J., Fehsenfeld, F.C., 2000. Review of observation-based analysis of the regional factors influencing ozone concentrations. Atmos. Environ. 34, 2045—2061.

World Health Organization, 2005. WHO Air Quality Guidelines for Particulate Matter, Ozone, Nitrogen Dioxide and Sulfur Dioxide.

Publication 3
A Sliding Window-Based Algorithm for Faster Transformation of Time Series into Complex Networks

Journal

Chaos (Ed. American Institute of Physics)

Position / Quartile

(16/261) / Q1-D1 (Applied Mathematics)

Impact factor

2.832 (2019)

A sliding window-based algorithm for faster transformation of time series into complex networks

Rafael Carmona-Cabezas, Javier Gómez-Gómez, Eduardo Gutiérrez de Ravé, and Francisco José Jiménez-Hornero

View Online   Export Citation   CrossMark

AIP Author Services
English Language Editing

A sliding window-based algorithm for faster transformation of time series into complex networks

View Online   Export Citation   CrossMark

Rafael Carmona-Cabezas,[a)] Javier Gómez-Gómez, Eduardo Gutiérrez de Ravé,
and Francisco José Jiménez-Hornero

AFFILIATIONS

Department of Graphic Engineering and Geomatic, University of Cordoba, Gregor Mendel Building, 3rd Floor,
Campus Rabanales, 14071 Cordoba, Spain

[a)]Author to whom correspondence should be addressed: f12carcr@uco.es

ABSTRACT

A new alternative method to approximate the Visibility Graph (VG) of a time series has been introduced here. It exploits the fact that most of the nodes in the resulting network are not connected to those that are far away from them. This means that the adjacency matrix is almost empty, and its nonzero values are close to the main diagonal. This new method is called Sliding Visibility Graph (SVG). Numerical tests have been performed for several time series, showing a time efficiency that scales linearly with the size of the series [$O(N)$], in contrast to the original VG that does so quadratically [$O(N^2)$]. This fact is noticeably convenient when dealing with very large time series. The results obtained from the SVG of the studied time series have been compared to the exact values of the original VG. As expected, the SVG outcomes converge very rapidly to the desired ones, especially for random and stochastic series. Also, this method can be extended to the analysis of time series that evolve in real time, since it does not require the entire dataset to perform the analysis but a shorter segment of it. The length segment can remain constant, making possible a simple analysis as the series evolves in time.

The Visibility Graph is a tool used for the transformation of a time series into a complex network that preserves the features of the original. Nevertheless, as the time series gets larger, the traditional method becomes too slow. In this work, the authors propose an alternative way that exploits a property of the network to reduce drastically the time required for very large datasets.

I. INTRODUCTION

In the current days, investigation on large time series or signals is needed to comprehend a lot of phenomena which appears in many and various fields, from nature to market researches or technological applications. Some examined aspects refer to hidden behaviors or trends and patterns, such as seasonality, which can be expected or not.[1–3] Forecasting, characterization of long-range correlation, chaotic properties, and scale invariance are also supplementary spheres, where time series were widely studied and promising results were obtained.[4–8]

Lately, a new point of view on the study of time series has been born, thanks to the transformation of them into complex networks. A recent method named as Visibility Graph algorithm (VG) has been released for this purpose.[9] It is demonstrated that the characteristics of the time series remain in the corresponding complex networks, adding the possibility of analyzing other parameters. Among them, one can find those which measure the centrality of these resulting networks such as the degree distribution,[10,11] which will be explained below.

The computation of Visibility Graphs through the basic algorithm requires a time complexity of quadratic order [$O(N^2)$].[12,13] Another technique which is based on the strategy of "divide and conquer" has been developed in the last years and needs a time computation of order {$O[N \log(n)]$}.[14]

Additionally, in recent studies, some sliding window-based algorithms have been released for very different fields.[15–18] In some cases, these methods have proven to be more efficient in computation time and memory allocation than their respective competitors, especially for large datasets and real-time data. Moreover, some

information from datasets is mostly not fully employed, carrying to unnecessary computations.

In this paper, the authors have developed a new method which transforms time series into complex networks based on a sliding window algorithm derived from the basic VG principles. It essentially splits the signal into different fixed-size portions, computes the visibility graph basic algorithm for the first window, and checks the visibility criterion for each next point. This algorithm has been tested for diverse sizes of windows and several time series as benchmarks. In addition, it has been employed to analyze a real dataset from tropospheric ozone pollution.

The main purpose of this study is to extent the variety of techniques which can be used for converting signals into complex networks, based on the fact that all the information of the time series is not required to build an associated complex network. This can save superfluous computations and time for large datasets by approximating the main characteristics of the corresponding complex networks. The idea is to retrieve acceptable results for small windows sizes compared to the total size of the dataset, with a very narrow confidence interval.

II. MATERIALS AND METHODS

A. Visibility graphs

A graph is known to be a set of points (nodes) that are connected by lines called edges. In the recent years, a tool that makes possible the transformation from a time series to a graph has gain the interest of the scientific community. This method was introduced by Lacasa et al.[9] and called Visibility Graph (VG), due to its resemblance with the one used to connect points in an Euclidean plane taking into account the possible obstacles.[19] One of its main features is that it inherits many properties from the original series.

In order to build the resulting complex network, it is necessary first to stablish a criterion to determine whether two points in the time series will become connected or not. That condition reads as follows: two arbitrary points from the time series (t_a, y_a) and (t_b, y_b) will have visibility (and will be connected in the graph) if any given point (t_c, y_c) that is located between them $(t_a < t_c < t_b)$ fulfills

$$y_c < y_a + (y_b - y_a)\frac{t_c - t_a}{t_b - t_a}. \tag{1}$$

In Fig. 1, an illustrative time series is depicted in order to show how the points would become connected after applying the VG method. Therefore, the time series is then transformed into a complex network for a posterior analysis. The new network, as it has been said, inherits the complexity of the original signal, meaning that, for instance, a periodic series would result on a regular graph.[9,20]

The resulting complex network has several properties, which all the transformed signals have in common, that can be used in order to simplify the algorithm:

- Every node does not have visibility with itself (there are no self-loops in the complex network).
- The resulting graph is undirected, meaning that the visibility between two points is reciprocal and the edges have no direction.
- Every point always has visibility with at least two other points: its closest neighbors, because there are no intermediate nodes to prevent them from fulfilling the condition for visibility.

It can be easily seen in the adjacency matrix A_{ij}, which is a representation of the complex network where the elements are $a_{ij} = 1$ if nodes i and j have visibility and 0 if the opposite is true. Taking the previous considerations into account, a VG adjacency matrix has the
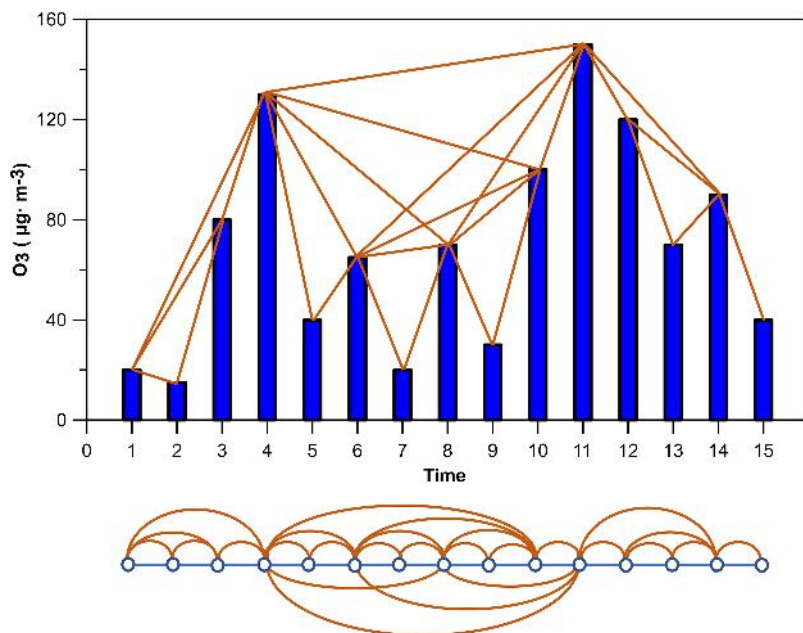

FIG. 1. Example time series transformed into a complex network by means of the VG method.

general form,

$$A_{ij} = \begin{pmatrix} 0 & 1 & \cdots & a_{1,N} \\ 1 & 0 & 1 & \vdots \\ \vdots & 1 & \ddots & 1 \\ a_{N,1} & \cdots & 1 & 0 \end{pmatrix}. \qquad (2)$$

As a simplification to the VG, another technique was introduced by Luque *et al.* in 2009[21] to obtain a complex network from a time series with a more restrictive criterion. It was called Horizontal Visibility Graph (HVG). The only difference in the implementation of this method is that the visibility criterion that needs to be fulfilled is

$$y_c < \min(y_a, y_b). \qquad (3)$$

B. Sliding visibility graph

One of the main problems when it comes to the computation of Visibility Graphs is the time complexity of the basic algorithm $[O(N^2)]$, which is quadratic with the size of the time series, N.[12,13] This is due to the fact that the condition needs to be checked $N(N-1)/2$ times at least in order to evaluate the visibility of every pair of points. When the size of the time series becomes substantially large, the time required to compute the VG rises considerably. Some examples of applications where one can find such great sizes of temporal series are, for instance, datasets that could be studied with a much higher resolution in order to give more light on searched behaviors, such as financial or meteorological time series that are usually regarded from a daily point of view, although much higher resolutions are available.[5,22] Also, there are time series which have a huge amount of points that need to be split for the convenience of the computation. One example can be found in time series of musical compositions, where only a small piece is usually taken, while it would be of interest to analyze the whole series.[23,24]

Here, the authors propose an alternative to the original VG that has a time complexity of $O(N)$, which ends up being very convenient for the computation of very large time series. The name given to this new algorithm is Sliding Visibility Graph (SVG), and the idea behind lies in the fact that most of the adjacency matrix of such complex networks are almost empty as one goes further from the main diagonal. This means that the vast majority of the iterations are considering pairs of nodes that will not have visibility.

The basic principle of this method is to take a set of points of length W (which we will refer to as "window" in the text) to check the visibility criterion inside of it and then start moving that window toward the end of the series. In practice, it is desirable to initialize the routine with a regular VG for the first W points, in order to have all the information from those first W points. The main advantage of this technique is that each time that the window is displaced, it is not necessary to compute again the visibility of all the points included within. Since most of them were already computed in the previous iteration, only the new point needs to check the criterion with the prior W ones (W times). It is easy to check that the number of total iterations n_i needed will be as shown in Eq. (1),

$$n_i = (N - W)W + W(W - 1)/2. \qquad (4)$$

The last term of the sum comes from the initial iterations required for the computation of the first VG with the length of the chosen window. It is derived from the number of iterations needed for the basic VG algorithm, as explained before. This term will be negligible for $W \ll N$, while it will be responsible for most of the iterations as $W \to N$.

In Fig. 2, the exposed procedure can be graphically observed, for the sake of clarity. Figure 2(a) illustrates the cited initial iteration that requires a regular VG.

Obviously, this method constitutes an approximation to the original VG as long as the size of the window is smaller than the total one of the time series. Nevertheless, due to the difficulty for two points to become connected nodes as the distances between them get higher, the results derived converge very rapidly to the original ones, as will be shown in the later sections. When the size of the window and the time series are equal, it is trivial that both SVG and VG are equivalent.

C. Degree centrality

When trying to retrieve information from a given complex network, one of the most commonly used approaches is discerning which of them are the most important nodes in the system. To this purpose, centrality measures come usually in handy. This concept was initially applied to the study of social networks and later transferred to other fields of knowledge.[25–27] This work has been focused on one of them: the degree centrality measure, which will be explained next.

The degree of a node (k_i) in a graph is the number of other nodes which are connected to it $\left(k_i = \sum_j a_{ij}\right)$ and in this case, the one that it has visibility with. For instance, in Fig. 1, the degree of the three first nodes are $k = 3$, $k = 2$, and $k = 3$, respectively. After computing the degree of the whole network, it is possible to obtain the probability that corresponds to each value, by simply counting how many times each possible one is repeated. From there, the degree probability distribution of the sample $P(k)$ can be obtained.

One possible way to describe the nature of the time series is to analyze the degree distribution that is built from the VG as previous works have shown,[9,10,28] being able to distinguish between fractal, random, or periodic signals, for instance. As some former studies explain,[13,20] time series which have VGs whose degree distributions can be fitted to a power law $P(k) \propto k^{-\gamma}$ correspond to scale free ones, which can be explained due to the effect known as "hub" repulsion.[29] The term hub refers to a node with an unlikely high number of links with respect to the rest of the network (the hubs have the highest degrees). The right tail of each degree distribution, governed by those hubs, can be represented in a log-log plot and fitted by a simple linear regression when the series obeys a power law. The slope obtained by this regression provides an interesting parameter, the so-called γ exponent, which has already been used in some works[20,28] to retrieve some useful information about the signal.

D. Complementary centrality parameters

There are several other centrality measurements that have been considered to check the correct suitability of the proposed method
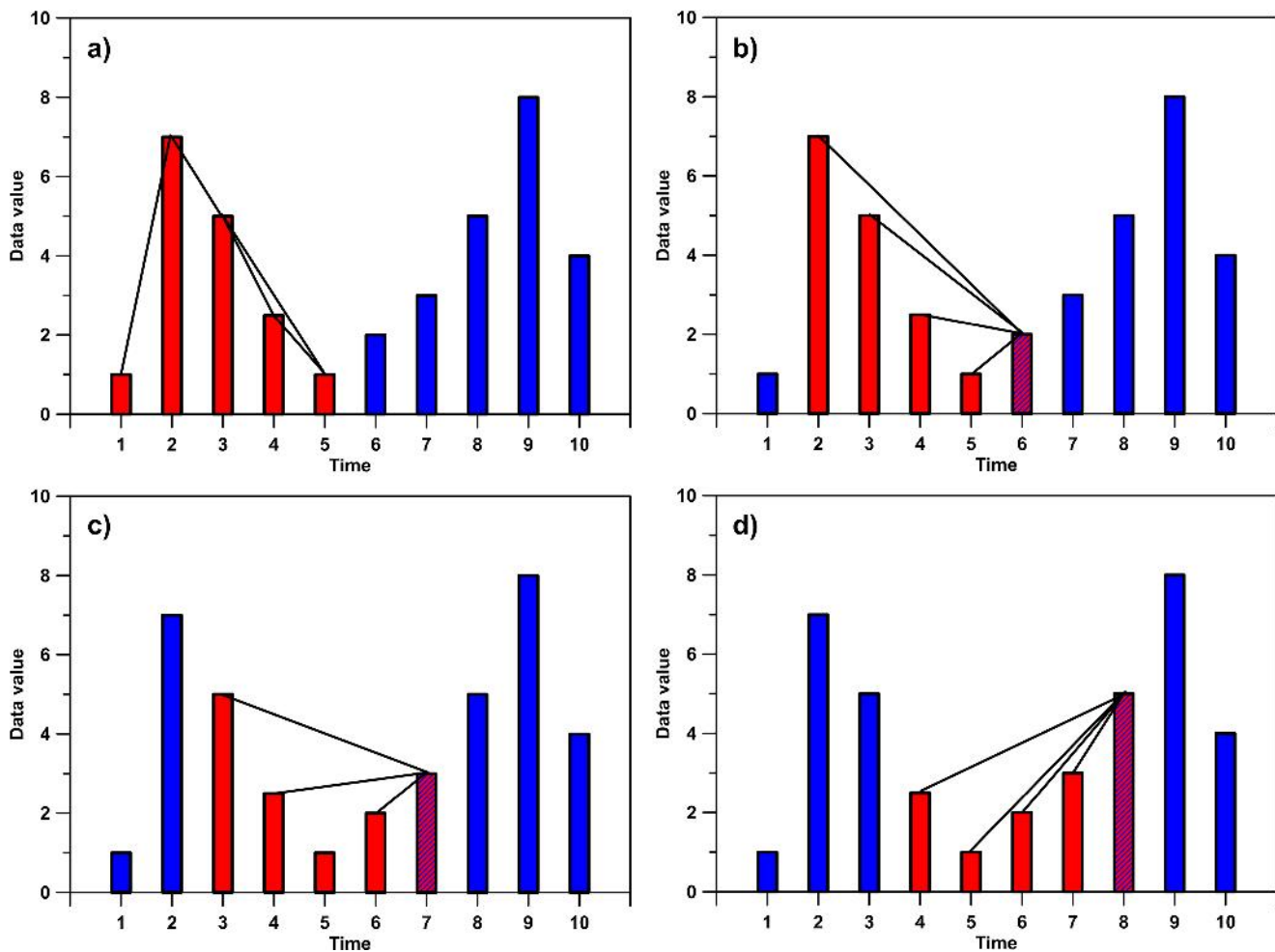
FIG. 2. Illustration of some example iterations of the SVG method. The plot (a) shows the first basic VG that is performed, while (b)–(d) are the posterior three iterations.

in order to describe the same complex network. One of them is the so-called shortest path (SP), which can be understood as the minimum number of edges that connects two arbitrary points. More precisely, two distant nodes (i, j) will have different numbers of edges and paths (distinct configurations of edges that link both) between them, but there will be some of these paths where the number of edges will be minimum (minimal path is degenerated); this quantity is named as the SP.

After introducing this parameter, one can define the closeness centrality. If all pairs of nodes are considered, it is possible to obtain a matrix, which is called distance matrix D. In this matrix, each element $d_{i,j}$ refers to the SP from node i to j. Diagonal elements are set to zero. For an undirected graph, this matrix is symmetric, as in the adjacency matrix case. It is possible to define the named closeness centrality of a node i as the inverse of the sum of distances from this node to the others, i.e.,

$$c_i = \frac{1}{\sum_{j=1}^{N} d_{i,j}},\tag{5}$$

where $d_{i,j}$ is the element (i, j) from the corresponding distance matrix of the graph.

Finally, the betweenness parameter has been added to these, which is a measurement of how a node is between many others. That is, how much a node is passed through by shortest paths of other pairs of nodes. Therefore, this quantity can be defined for a node i as

$$b_i = \sum_{\substack{j=1 \\ j \neq i}}^{N} \sum_{\substack{k=1 \\ k \neq i,j}}^{N} \frac{n_{jk}(i)}{n_{jk}},\tag{6}$$

where n_{jk} is the number of SP's from j to k (remember that these paths can be degenerated), whereas $n_{jk}(i)$ is the number of SP's that contains the node i.

III. RESULTS AND DISCUSSION

A. Performance tests

In this section, the proposed algorithm is tested by using several time series with different natures. All the studies in this work have been performed using MATLAB 2018b on an Intel(R)Core(TM) i7-8700 CPU @ 3.20 GHz, with a RAM of 8 GB and the O.S. Windows 10 Education x64.

In Fig. 3, the time series used for testing the algorithm are shown in the upper part. The first one (a.1.) corresponds to a random series (from 0 to 1) with $N = 5 \cdot 10^5$; the second one (b.1.), to a Brownian motion time series with a Hurst exponent of $H = 0.5$ and $N = 2 \cdot 10^4$; while the last one (c.1.) is a stochastic one of hourly ozone concentrations recorded in the southern part of the Iberian Peninsula in 2013 (nearly $N \approx 8500$). On the other hand, the bottom

plots show the algorithm computation times on each time series, for both SVG and original VG. It must be noted the fact that the window size (W) in the x axis of the last mentioned plots are normalized to the total size (N) of each example series. It has been done in order to compare the different tests, since all the chosen values of N are unequal.

As it can be regarded in the lower part of Fig. 3, the behavior of the computation time with W is parabolic, having its maximum when $N = W$. To check that quadratic trend, a second-degree polynomial fit was performed to the curves and in all the cases the Pearson coefficient was greater than 0.9999. At the maximum of the curve, the computation times of both methods are the same since they are equivalent when those two quantities are equal (as explained before). Before that maximum is reached, the computation times of the SVG are always lower than those of the original VG. That is because, as it can be seen in Eq. (1), the number of iterations of SVG can never be greater than the amount needed for the VG, since by definition $W \leq N$.

In the case of the ozone concentrations, the measured computation times have a bigger noise due to the lower size of the time series,
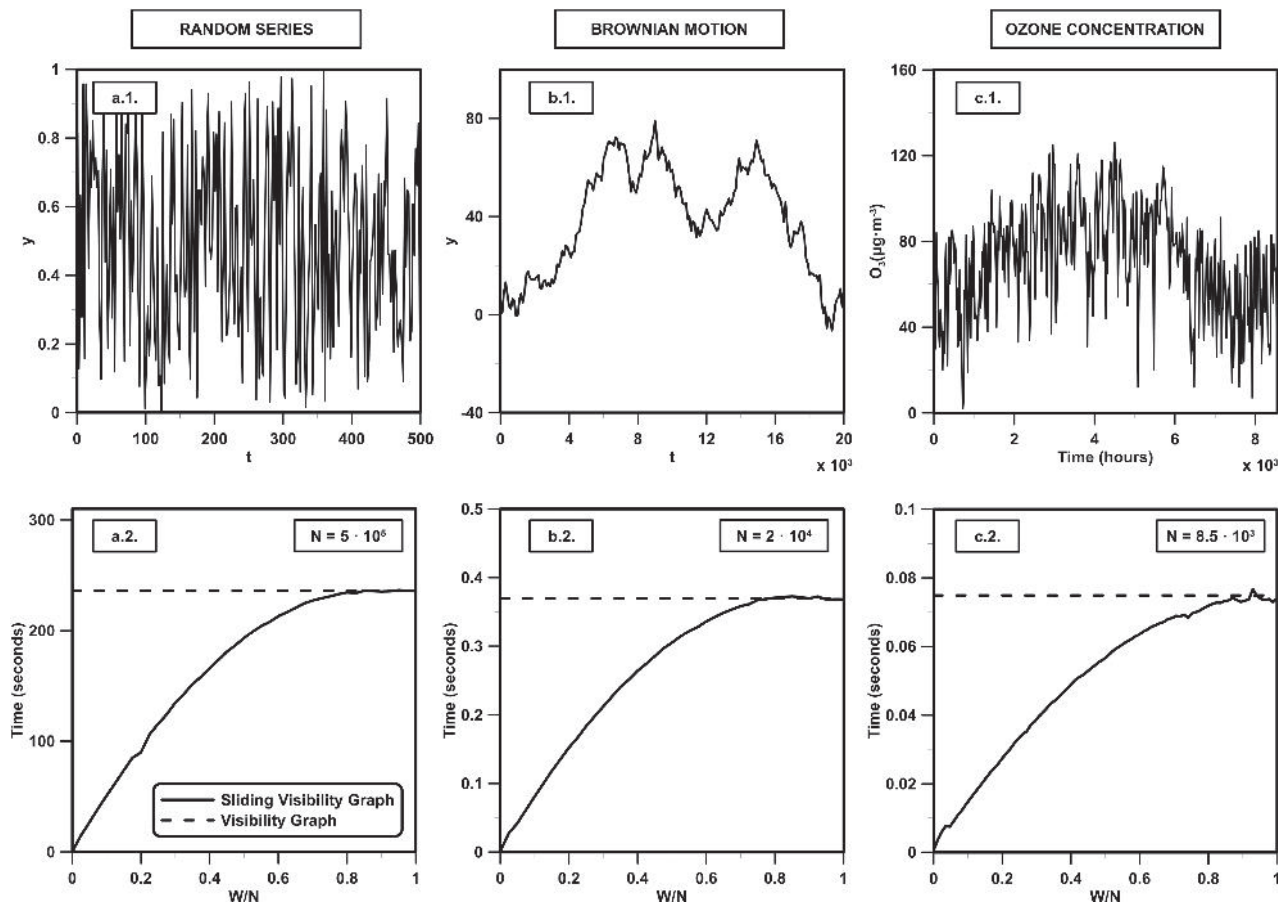

FIG. 3. Time series used for the analysis (up) and computation times vs window size (down). Dashed line indicates the time required for the basic VG method. In a.1., only a portion of the total series is shown, for clarity reasons.
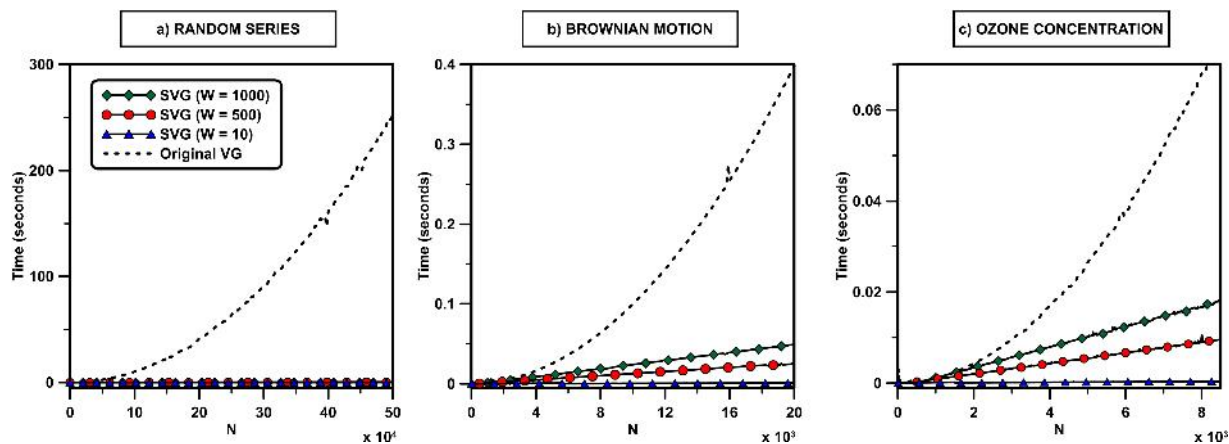
FIG. 4. Computation time vs the size of the time series for several values of window size.

for which the computation times are considerably reduced, compared to the other two.

Once the scaling with the size of the window W has been discussed, the equivalent analysis with N is necessary as well. Indeed, it is where one of the main advantages of SVG with respect to the original can be observed, since the scaling changes from $O(N^2)$ in VG to $O(N)$ in the newly proposed one.

In Fig. 4, the computation time required to perform the SVG method is plotted against the size of the analyzed time series. The same series as in the previous case were used but in this case varying their size from 10 to the original one. The result for all of them is the same: the original VG has a time efficiency that scales quadratically with the size of the time series [$O(N^2)$], while the SVG behavior with respect to N is always linear [$O(N)$]. In these plots, the absolute values of W have been used instead of the ones relative to the value of N as in Fig. 3 because its value is now variable and it could be misleading. Instead, three different lengths of window have been shown (10, 500, and 1000). As it is clearly seen, the larger is the window size, the greater the slope of the computation time over N. The same tests were performed for higher values of W, obtaining the same linear behaviors, despite having a shorter curve (because N can only be evaluated from a W size onwards). These results are in accordance to what was expected from Eq. (1).

B. Degree distribution results

After the computation efficiency of this method has been tested, the next step is to check whether the results of the SVG merge the ones of the VG and when this approximation is valid. To see this, in Fig. 5, the degree probability distributions of the three series used before are shown. There, the distribution of each series is computed using the regular VG, the HVG, and then with the SVG, with several increasing sizes of window. The same relative values of W/N are used in the three cases for clarity reasons. The series sizes N in all of them is fixed to the original values used for discussion in Fig. 3, hence the study conditions remain the same as in Sec. III A. A log-log plot is used because one of the main uses of this degree

probability distribution is the acquisition of the γ exponent. That is done by fitting the linear tail in the last part for scale free networks using the logarithmic scale, as explained. Some works have already shown that those series degree distributions can be fitted to a power law $P(k) \propto k^{-\gamma}$.[10,20] Therefore, in Figs. 5(b) and 5(c), it was interesting to check whether this linear behavior of the tail is hold with the SVG as well.

For the three series, it can be observed that the degree distribution differs evidently from the original for very low values of W, especially for the Brownian motion case and the ozone time series. In those cases, it is seen that the largest degrees of the distribution (hubs) are always sensibly lower than the actual hubs degrees in the original VG. That is due to the fact that the chosen window ($W = N/500$ in the plots) is still too narrow to allow hubs to check the visibility with most of the nodes they are connected to. Thus, the biggest values of k for those windows will have a limit given by twice the size of the window ($2W$), since a node i will check its visibility with others only from $i - W$ to $i + W$. This is in accordance with the results in Fig. 5(b), where for $W = 40$ ($N = 2 \cdot 10^4$), there are no values of k higher than 80 ($2W$), for instance. Hence, this can be used as criterion for choosing a minimum value of W that allows the convergence $P(k)$.

On the other hand, when W increases, the degree distributions of SVGs start to resemble the original ones. So that at a value of $W = N/100$, they are almost overlapped and big changes on the length of the window (for instance $W = N/10$) give almost equivalent results on the random and ozone series. In the case of the Brownian motion, this convergence is slower, being necessary larger sizes of window. The reason behind this phenomenon is the roughness of the series that depends on the Hurst exponent.[30] Nevertheless, it must be pointed out that the computation time remains lower than in the original case no matter the value of W, as discussed previously. This saturation can be seen more in detail in Fig. 6. In every case, networks from the HVG in Fig. 5 give distributions that are very different from the original VG, having SVG a better approximation of the distribution for all the selected windows.

To test the convergence of some actual numerical values rather than the shape of the curves, the average degree $\bar{k}$ of each distribution
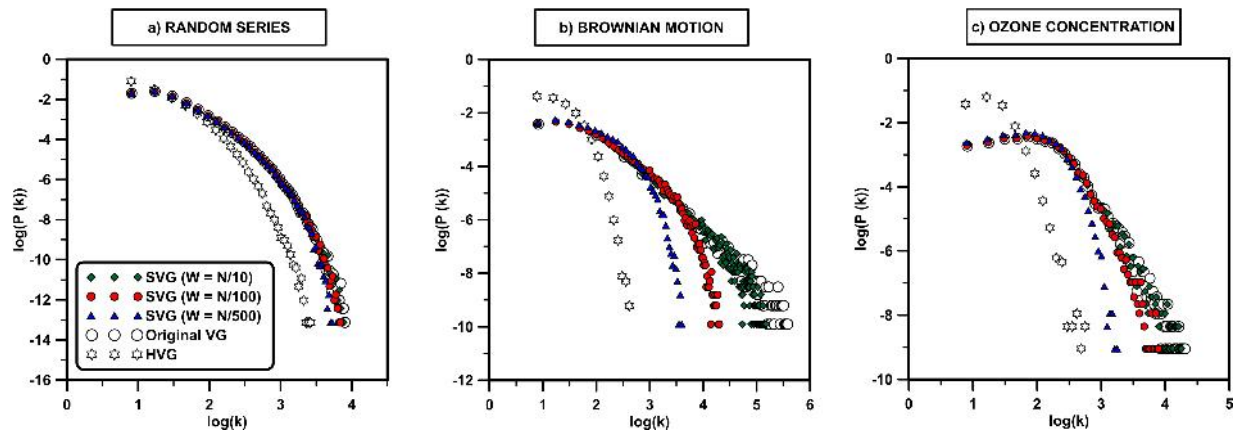
FIG. 5. Degree probability distributions of the networks obtained by the VG and SVG are shown here. Three window sizes from $W = N/1000$ to $N/10$ are used for comparison in the SVG.
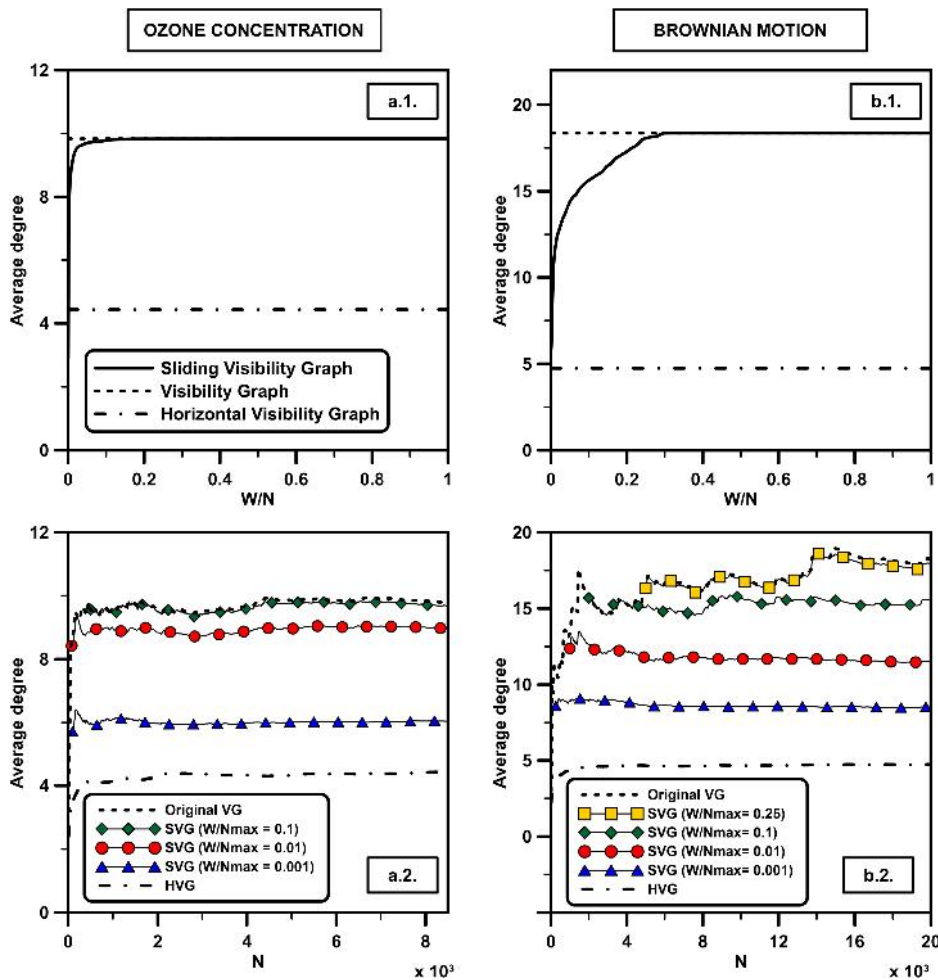

FIG. 6. Average degree obtained through VG and SVG vs the window size (up) and vs the total series size (down).

has been computed (Fig. 6) as well as the γ exponent from the linear regression (Fig. 7) and other parameters described before. Now, it is only shown for the case of the Brownian motion and ozone concentration time series, since the random series does not have a power-law behavior and for the value of $\bar{k}$ will converge much faster than the others, as can be deduced from Fig. 5. Hence, the two series that fit the original case worse are looked at in more detail.

Regarding the top part of Fig. 6, the value of $\bar{k}$ obtained through the SVG against the VG and HVG can be seen. As it was expected from the previous commented figure, the merging in the case of the Brownian series is slower than that in the case of the stochastic ozone one. Nonetheless, its shape converges faster than the number of iterations (quadratic) even in the Brownian case, as discussed in relation to the time performance of the method (see Fig. 3).

On the other hand, when the size of the series itself is increased, $\bar{k}$ obtained with the SVG and a fixed value of W converges as well. This is shown in the lower part of Fig. 6. In this case, for the ozone time series a window of $W = 0.1 \cdot N_{max}$ is enough for the convergence in all the sizes tested, while for the case of the Brownian motion, it is necessary to reach a value of $W = 0.25 \cdot N_{max}$. On the contrary, it is seen in this same figure how the values of the HVG

TABLE I. Results obtained from the optimal window criterion for three time series.

Time series	W_{Opt}/N (%)	t_{SVG}/t_{VG} (%)	ε_k (%)
Random	0.8	1.7	0.03
Brownian ($H = 0.5$)	27.3	45.3	1.08
Annual ozone concentration	5.5	10.2	1.28

are always lower than those of VG, being the SVG ones greater even for very small windows. This fact could be anticipated by looking at Fig. 5.

To provide an adequate criterion to decide a proper optimal window, the authors propose a technique that will be further explained. After several test with time series of different natures and lengths, it has been determined that the SVG results always saturate when they approach the value of the original VG. This leads to the difference between results from consecutive windows decaying to zero. This can be used to determine a saturation point by computing the slope of the curve for each window. When this slope is considered to be close enough to zero, the optimal window is taken. The criterion
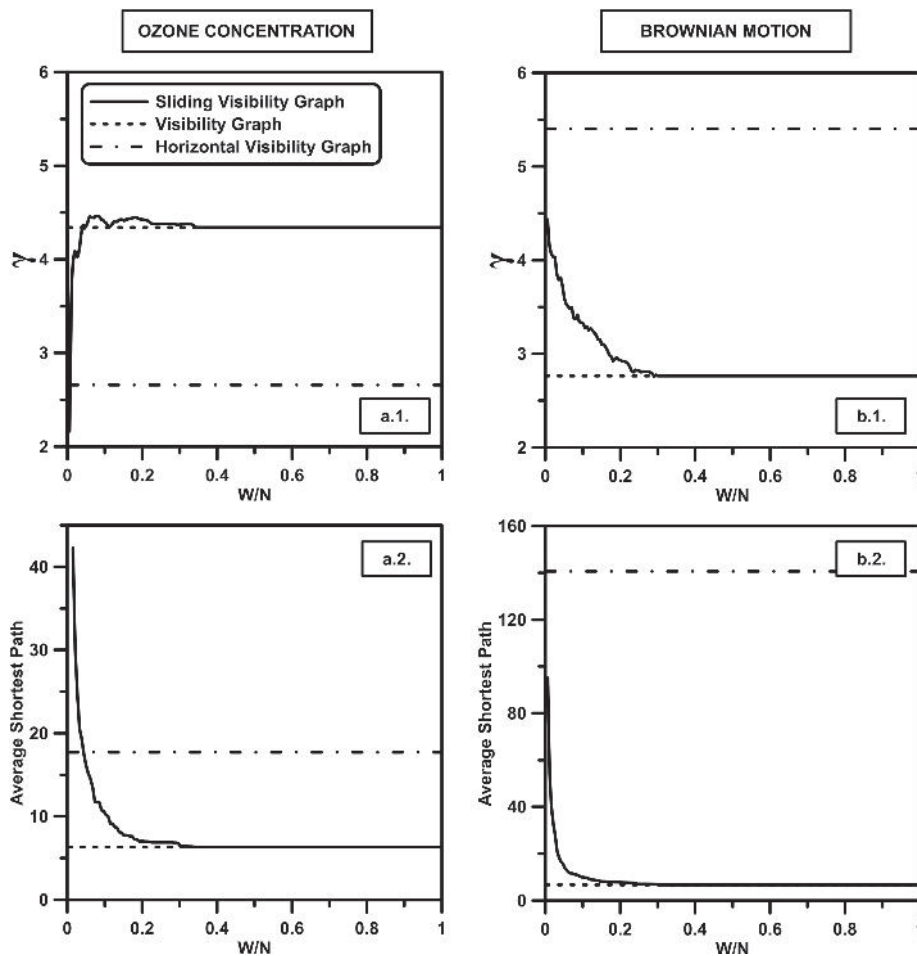

FIG. 7. γ-exponent (up) and average SP (bottom) computed by using the original VG method (dashed line), the HVG (dashed-dotted line) and the SVG for different values of window length.

for this consideration is that the slope between the two last points (W_i and W_{i-1}) falls below a percentage of a characteristic value automatically set for a given time series ($\overline{k_i}/N$). The authors have detected that a 5% of this value is suitable to an optimal window. The results for the average degree computed can be seen in Table I, where the ratio between the optimal window and the total size (W_{Opt}/N) is shown in the second column, the ratio of the required times in the third one (t_{SVG}/t_{VG}), and the relative error (ε_k) from the computation of the average degree in the last one. Since all of them are relative values, they are expressed as percentages for the sake of clarity.

In order to test the applicability of the SVG, four more different parameters (γ exponent, SP, betweenness, and closeness) will be described in the following part of the presented document. Those have been previously employed to study both visibility graphs and complex networks.[22,31] The results in Figs. 7 and 8 show the values of these quantities after applying the SVG for several windows (as it was done before with the degree) and their comparison with the VG and HVG methods.

The results depicted in Fig. 7 show a similar behavior to those in the upper part of Fig. 6 but this time with the slope of the distribution tail (γ exponent) and the SP. The γ exponent of both time series (upper part) tends to the value obtained through the basic VG as the size of the window increases. Again, the Brownian series needs a larger size of window to have a negligible error from the SVG, due to its greater roughness, as discussed. Nevertheless, a good fit is already achieved for windows of around 25% of the entire size of the series, meaning that the information of the hubs in the system can be retrieved by using this method. Regarding the SP (bottom part of the same figure), a fast convergence is as well observed. For very low values of W, the SP is higher in comparison to VG, because such small window does not allow fast connections between distant time points that would lead to the real SP, and therefore the distances are greater. As this window is increased, those fast links are possible, and the SP is decreased.

Finally, Fig. 8 shows now the two next computed parameters: betweenness (up) and closeness (bottom). Both are in accordance with the results discussed previously. The reason for the convergence
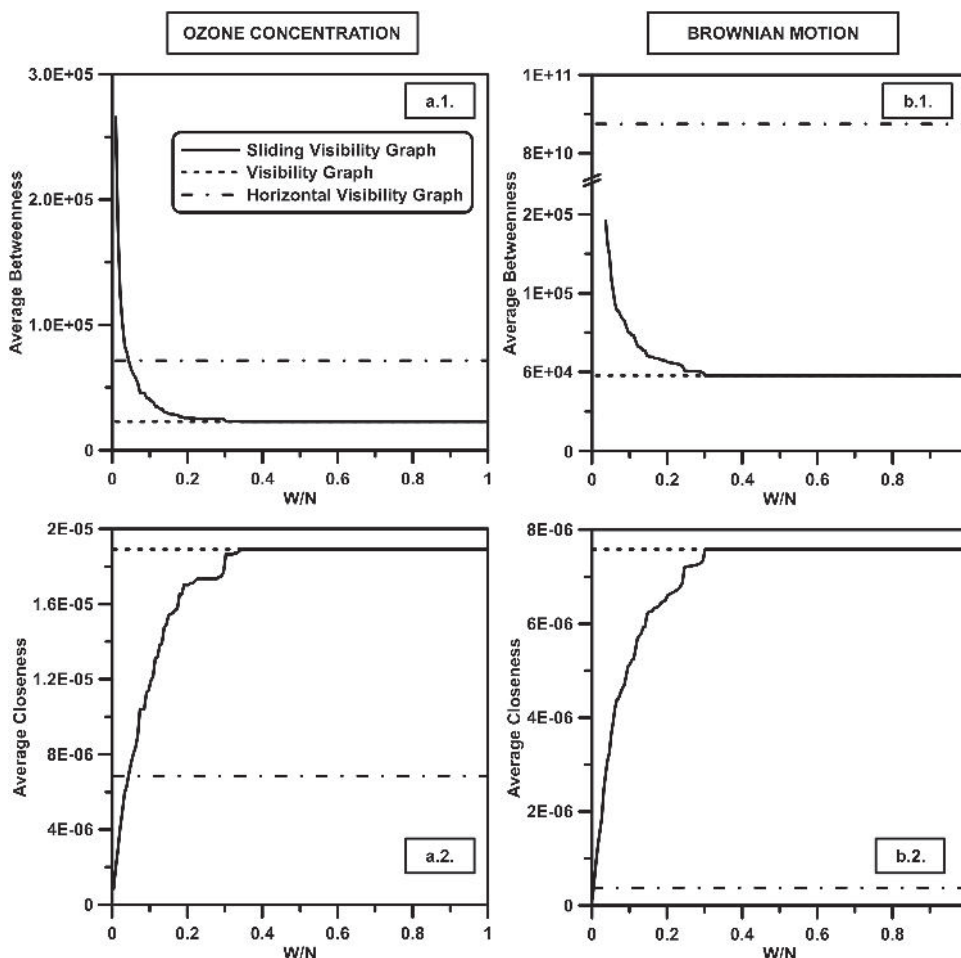

FIG. 8. Average betweenness (up) and closeness (bottom) computed by using the original VG method (dashed line), the HVG (dashed-dotted line) and the SVG for different values of window length.

can be understood in the same way as what was explained above for the case of SP.

Once again, outcomes from HVG differ substantially from the VG ones, being the SVG results closer to them. This fact is even more pronounced in the case of the Brownian motion time series (in both last figures). The authors attribute this effect to its higher irregularity that leads to a worse overall performance of the HVG algorithm. Since the HVG algorithm excludes nodes once a higher value than the initial one is found, it can omit important nodes that VG and SVG (with a proper window) would not. This, in the sense of connections between very distant points in the time series, is very important for the computation of the three last parameters mentioned. For this reason, the approximation of the HVG can even differ several orders of magnitude from both VG and SVG ones [see Fig. 8(b.1)].

IV. CONCLUSIONS

Throughout this work, a new method to compute the Visibility Graph out of a time series has been introduced and tested. This new approach, the Sliding Visibility Graph, approximates the original VG and has its basis on the fact that the adjacency matrix of the network built is almost empty and all the values tend to be as close as possible from the main diagonal.

The performance was tested on several time series with different natures (random, fractional Brownian motion, and real stochastic measurements) and the results show that the time efficiency has a parabolic trend with respect to the window size W and linear with the size of the series N. By definition, the number of iterations needed by the SVG is going to be always lower than those of the original VG, leading to a faster performance for every time series. In the limit case of $W = N$, SVG converges to the basic VG, bringing obviously to the same time performance.

When it comes to the results obtained from this alternative method, as expected from the properties of the adjacency matrix, the SVG outcomes rapidly converge to the ones obtained by the VG for low sizes of the window. This has been demonstrated for several distinct parameters. The main advantage of this is that once a proper window is chosen for a kind of time series, the correctness of the approximation appears to hold for larger sizes of the series, making it very convenient for great data series. The authors propose a technique that can be automatized in order to find the optimal windows and has been tested to provide satisfactory results. It has been compared to HVG results as well, showing that SVG outperforms it for almost every chosen window.

To conclude, it must be pointed out that SVG constitutes an alternative approximation to the widely used VG and HVG that could have a big potential for several cases. In particular, authors would like to mark two main scenarios: (i) very large time series, where the computation times would be huge and (ii) real-time analysis, where the size of the system gets larger with every measurement. This will be studied in future works, in order to check whether the parameters derived from these complex networks can be used to predict changes in the behavior of a temporal variable, for instance.

REFERENCES

[1] R. W. Macdonald, T. Harner, and J. Fyfe, Sci. Total Environ. **342**, 5 (2005).
[2] P. Pavón-Domínguez, S. Serrano, F. J. Jiménez-Hornero, J. E. Jiménez-Hornero, E. Gutiérrez de Ravé, and A. B. Ariza-Villaverde, Phys. Stat. Mech. Appl. **392**, 4466 (2013).
[3] J. Yang and J. Leskovec, in *Proceedings of the Fourth ACM International Conference on Web Search and Data Mining—WSDM 11, Hong Kong, China* (ACM Press, 2011), p. 177.
[4] J. D. Hamilton, *Time Series Analysis* (Princeton University Press, Princeton, NJ, 1994).
[5] H. E. Stanley, V. Plerou, and X. Gabaix, Phys. Stat. Mech. Appl. **387**, 3967 (2008).
[6] P. Xu, R. Zhang, and Y. Deng, Chaos Solitons Fractals **117**, 201 (2018).
[7] R. Zhang, B. Ashuri, Y. Shyr, and Y. Deng, Phys. Stat. Mech. Appl. **493**, 239 (2018).
[8] F. Liu and Y. Deng, IEEE Access **7**, 102554 (2019).
[9] L. Lacasa, B. Luque, F. Ballesteros, J. Luque, and J. C. Nuño, Proc. Natl. Acad. Sci. U.S.A. **105**, 4972 (2008).
[10] R. Carmona-Cabezas, A. B. Ariza-Villaverde, E. Gutiérrez de Ravé, and F. J. Jiménez-Hornero, Sci. Total Environ. **661**, 138 (2019).
[11] C. Zhou, L. Ding, M. J. Skibniewski, H. Luo, and S. Jiang, Saf. Sci. **98**, 145 (2017).
[12] L. Lacasa, A. Nuñez, É Roldán, J. M. R. Parrondo, and B. Luque, Eur. Phys. J. B **85**, 217 (2012).
[13] L. Lacasa, B. Luque, J. Luque, and J. C. Nuño, Europhys. Lett. **86**, 30001 (2009).
[14] X. Lan, H. Mo, S. Chen, Q. Liu, and Y. Deng, Chaos Interdiscip. J. Nonlinear Sci. **25**, 083105 (2015).
[15] F. Cao, J. Liang, L. Bai, X. Zhao, and C. Dang, IEEE Trans. Fuzzy Syst. **18**, 872 (2010).
[16] S. K. Tanbeer, C. F. Ahmed, B.-S. Jeong, and Y.-K. Lee, Inf. Sci. **179**, 3843 (2009).
[17] U. Yun, D. Kim, H. Ryang, G. Lee, and K.-M. Lee, J. Intell. Fuzzy Syst. **30**, 3605 (2016).
[18] Y. Wen, J. Hao, X. Guo, X. Xiao, C. Wu, S. Wang, T. Yang, H. Shen, X. Chen, L. Tan, Q. Tian, H.-W. Deng, and F. Zhang, J. Biomed. Inform. **57**, 38 (2015).
[19] T. Lozano-Pérez and M. A. Wesley, Commun. ACM **22**, 560 (1979).
[20] L. Lacasa and R. Toral, Phys. Rev. E **82**, 036120 (2010).
[21] B. Luque, L. Lacasa, F. Ballesteros, and J. Luque, Phys. Rev. E **80**, 046103 (2009).
[22] A. O. Bielinskyi and V. N. Soloviev, "Complex network precursors of crashes and critical events in the cryptocurrency market," in *Computer Science & Software Engineering, Kryvyi Rih, Ukraine, November 2018* (Kryvyi Rih, 2018), Vol. 2290, pp. 37–45.
[23] D. d. F. P. Melo, I. d. S. Fadigas, and H. B. De Barros Pereira, Appl. Netw. Sci. **2**, 32 (2017).
[24] N. Wang and D. Wei, in *26th Chinese Control and Decision Conference 2014 CCDC, Changsha, China* (IEEE, 2014), pp. 3647–3650.
[25] T. Agryzkov, L. Tortosa, and J. F. Vicent, Appl. Math. Comput. **347**, 600 (2019).
[26] K. E. Joyce, P. J. Laurienti, J. H. Burdette, and S. Hayasaka, PLoS One **5** e12200 (2010).
[27] C. Liu, X.-X. Zhan, Z.-K. Zhang, G.-Q. Sun, and P. M. Hui, New J. Phys. **17**, 113045 (2015).
[28] P. Mali, S. K. Manna, A. Mukhopadhyay, P. K. Haldar, and G. Singh, Phys. Stat. Mech. Appl. **493**, 253 (2018).
[29] C. Song, S. Havlin, and H. A. Makse, Nat. Phys. **2**, 275 (2006).
[30] B. Mandelbrot and J. W. van Ness, SIAM Rev. **10**, 422 (1968).
[31] R. Carmona-Cabezas, J. Gómez-Gómez, A. B. Ariza-Villaverde, E. Gutiérrez de Ravé, and F. J. Jiménez-Hornero, Chemosphere **230**, 59 (2019).

5. Conclusions

"Me encontré en la posición de ese niño de un cuento que ve un trozo de cuerda y, por curiosidad, tira de ella para descubrir que era solo la punta de una cuerda muy larga y cada vez más gruesa ... y seguí mostrando maravillas más allá del cálculo de cuentas."

(Original: "I found myself in the position of that child in a story who noticed a bit of string and - out of curiosity - pulled on it to discover that it was just the tip of a very long and increasingly thick string...and kept bringing out wonders beyond reckoning.")

Benoît Mandelbrot

Individual conclusions

Along the development of the research that has been the focus of this thesis, several conclusions have been drawn, and those can be separated into the different papers where each one has been yielded. From the study performed in the **<u>first paper</u>**, the following inferences may be extracted:

1) Multifractal analysis of VGs from ground-level ozone through SB method is a suitable tool for the description of the seasonal patterns found for this pollutant.

2) The relation between the generalized fractal dimensions can give some light into the behavior of singularly high episodes of concentration of this pollutant. For instance, $(D_0 - D_1)$ is considered to be related to the uniformity of those episodes.

3) Some properties can be directly inferred by just looking at the degree distribution of the mapped complex network, which can turn into a powerful tool for preliminary analyses, before performing a more complex one.

4) Results from the generalized fractal dimensions (from the Rényi spectrum) and the singularity spectrum, which are computed by two separate methods, are consistent. This is a confirmation of the correctness of both SB, and Chhabra and Jensen algorithms.

Moreover, after looking at the **<u>second one</u>** of the introduced manuscripts, some verdicts can be gleaned as well after all the considerations and results obtained are considered. These are shown below:

5) Results prove that complex networks derived from O_3 time series are also suitable to distinguish between the dynamics of that pollutant in urban and rural areas.

71

Conclusions

6) In every location, the nature of the series is fractal and the γ-exponents (the power law exponent of the degree distributions) is very similar for all of them.

7) Apart from the seasonal pattern, the average and standard deviation of the degree centrality also show a difference between the rural and urban stations. This distinction is especially pronounced in summer and clearer by looking at the mean degree.

8) Betweenness centrality might be used as a supplementary source of information regarding the variability of the diurnal maxima behaviour, in relation to the introduced skyline hubs.

Finally, once the conclusions of the two first works have been listed, only the **third article** remains. Its conclusions are shown in the next paragraphs:

9) Regular VG adjacency matrices are almost empty, and one can take advantage of that fact to greatly reduce the number of iterations in the calculation of the complex network.

10) The computation time required to obtain a complex network from a time series is greatly reduced when using the SVG instead of the original VG, and this difference increases rapidly as the size of the series is enlarged.

11) When the size of the selected window is set to the total number of points in the time series, the SVG is equivalent to the VG.

12) The convergence of several tested complex network parameters to the VG (degree, betweenness, shortest path and closeness) values is very fast and it is achieved for low computation times.

13) A proper window can be selected by identifying when the approximation of the parameters, such as the degree computed from the SVG, starts to saturate. Once an optimal window is set, the correctness appears to hold for larger sizes of the time series. This validates the SVG as an approximation of the VG with a great potential when dealing with very large time series or real time scenarios, where the size is constantly increasing.

General conclusions

Throughout this work, the capability of complex networks for the description of pollutant dynamics (in particular O_3) has been tested by using the Visibility Graph algorithm. And what is more, it has been shown that it is not only possible but also new insights can be obtained by choosing this new approach. Furthermore, the new supplementary information that is gained from these new complex networks might serve as a feedback for databases used by new techniques such as machine learning and data mining. Therefore, in addition to the already existing studies from descriptive statistics and multifractal analysis, it can be adjoined in order to validate theoretical models and improve predictive strategies.

As a parallel thought, and not directly derived from this thesis, it is worth noticing the next considerations regarding the tropospheric O_3 problem (and all air pollution in general). No matter how much research and time is devoted to this matter, at the end of the day the most important action must be taken by governments and by all of us as citizens. We must realize that it is a major health issue that costs millions of deaths annually to our society and immense economical losses. Sadly, the light given the public and action taken by the administrations in this affair is far from being enough, given the severity of the problem. Ironically, one of the few positive aftereffects of the Covid-19 crisis is that it has been demonstrated that it is possible to improve the air quality. The techniques presented and evaluated in this work might help (combined with others) evaluating future tropospheric ozone mitigation policies. The structural reduction of carbon-based transport and industrial activity (and therefore the emission of O_3 precursors) have been validated to be crucial for the reduction of O_3 concentration in major cities.

Finally, I would like to highlight as a corollary to this thesis the fact that all the developed research has allowed me advance and grow as a scientist. There has been no shortage of problems encountered and new paths started that finally led nowhere.

Conclusions

Nevertheless, there is no wasted time in science when a dead end is encountered, since it is the experience gained in the way what actually matters. As long as we continue walking forward, I believe there is no barrier that will stand in the way of humanity, not even atmospheric pollution.

Conclusiones individuales

A lo largo del desarrollo de esta tesis, ciertas conclusiones han sido obtenidas; las cuales pueden separarse según el artículo del cual se ha desprendido cada una. Del estudio realizado en el **<u>primer artículo</u>**, se pueden extraer las siguientes inferencias:

1) El análisis multifractal de los Grafos de Visibilidad (GV) del ozono troposférico es una herramienta adecuada para la descripción de los patrones estacionales de este contaminante.

2) La relación entre las dimensiones fractales generalizadas puede arrojar cierta luz sobre el comportamiento de los episodios de concentraciones singularmente altas de O_3. Por ejemplo, $(D_0 - D_1)$ se considera relacionado con la uniformidad de estas singularidades.

3) Algunas propiedades pueden ser deducidas directamente mirando a la distribución del grado de las redes complejas obtenidas. Esto puede resultar una herramienta muy útil de cara a análisis preliminares en un futuro, antes de acometer otros más complejos.

4) Los resultados de las dimensiones fractales generalizadas (del espectro de Rényi) y del espectro de singularidades, los cuales son obtenidos por dos métodos separados, son consistentes. Esto puede considerarse como una comprobación de la corrección del método Sand-Box (SB) y del algoritmo de Chhabra y Jensen.

Además, teniendo en cuenta los resultados y consideraciones obtenidos en el **<u>segundo artículo</u>** presentado en esta tesis, se pueden extraer más conclusiones individuales. Éstas son mostradas a continuación:

5) Los resultados prueban que las redes complejas derivadas de las series temporales de O_3 son adecuadas para distinguir las dinámicas de este contaminante en zonas urbanas y rurales.

Conclusions

6) En cada localización, la naturaleza de la serie es fractal como cabía esperar y los exponentes γ (exponentes de la ley potencial) son muy similares para todos ellos.

7) Además del patrón estacional, la media y la desviación estándar muestra una diferencia entre las estaciones urbanas y rurales. Esta distinción es más pronunciada en verano y aún más clara para el grado medio.

8) La intermediación podría ser usada como una fuente suplementaria de información acerca de la variabilidad del comportamiento de los máximos diurnos, en relación a los *skyline hubs* introducidos.

Finalmente, tras listar las conclusiones de los dos primeros trabajos, solo quedan las del **tercer artículo**. Sus conclusiones se muestran en los siguiente párrafos:

9) Las matrices de adyacencia de los GV están típicamente casi vacías, por lo que se puede aprovechar este hecho para reducir en gran medida el número de iteraciones necesarias para calcular la red compleja.

10) El tiempo de cálculo necesitado para obtener un red compleja de una serie temporal se reduce considerablemente con el uso del Grafo de Visibilidad Deslizante (GVD) en lugar del GV. Esta diferencia aumenta rápidamente conforme el tamaño de la serie crece.

11) Cuando el tamaño de la ventana seleccionado abarca el total de la serie temporal, el GVD coincide exactamente con el GV.

12) La convergencia de diversos parámetros del GV (grado, intermediación, camino mínimo y cercanía) es muy rápida y se obtiene con tiempos de cálculo muy bajos.

13) Una ventana adecuada se puede seleccionar identificando cuándo la aproximación de los parámetros se empieza a saturar. Una vez que dicha ventana es seleccionada, la corrección se mantiene para mayores tamaños de la serie temporal. Esto valida el potencial del GVD para tratar con series muy grandes o escenarios en tiempo real, donde el tamaño de la serie aumenta con el tiempo.

Conclusiones generales

A lo largo de este trabajo, se ha evaluado la capacidad de las redes complejas para la descripción de la dinámica de contaminantes (en concreto O_3), mediante el uso del algoritmo del Grafo de Visibilidad. Se ha demostrado que no sólo es posible, sino que se puede obtener información nueva eligiendo este método. Además, la nueva información suplementaria adquirida podría servir como apoyo para bases de datos usadas por nuevas técnicas, como el aprendizaje automático o la minería de datos. Por ello, adicionalmente a los estudios ya existentes basados en la estadística descriptiva y el análisis multifractal, esta metodología puede ser usada complementariamente para validar modelos teóricos y mejorar estrategias predictivas.

A título de reflexión paralela, aunque no directamente derivado de los resultados de esta tesis, merece la pena llamar la atención sobre el problema del O_3 troposférico (y la contaminación en general). No importa cuánto tiempo se dedique a investigar este asunto, pues al final la acción más importante debe realizarse por parte de los gobiernos y de todos nosotros como ciudadanos. Debemos darnos cuenta de que es un problema para la salud pública de gran relevancia, que cuesta millones de vidas anualmente e inmensas pérdidas económicas a nuestra sociedad. Desgraciadamente, la información dada a la población y las acciones tomadas por las administraciones en este asunto están aún muy lejos de ser suficientes, dada la gravedad del problema. Irónicamente, uno de los pocos efectos positivos de la crisis derivada de la Covid-19 es que se ha demostrado que es posible mejorar la calidad del aire. Las técnicas presentadas y evaluadas en esta tesis podrían ayudar (combinadas con otras) a evaluar futuras políticas para mitigar niveles de ozono troposférico. La reducción estructural del transporte basado en hidrocarburos y de la actividad industrial (con la consecuente emisión de precursores de O_3) ha sido validada como crucial para la reducción de la concentración de O_3 en las grandes ciudades.

Finalmente, a modo de valoración personal y corolario de esta tesis, me gustaría destacar el hecho de que la investigación realizada me ha permitido crecer como persona

Conclusions

y científico. No han sido pocos los problemas encontrados y los nuevos caminos comenzados, que finalmente no han conducido a ningún sitio. Sin embargo, creo que no hay tiempo desperdiciado en ciencia cuando se encuentra un callejón sin salida, puesto que es la experiencia ganada en el trayecto lo que de verdad importa. Siempre y cuando continuemos caminando hacia delante, pienso que no hay barrera que se interponga al desarrollo de la humanidad, ni siquiera la contaminación atmosférica.

6. Other publications

> *"La manzana no puede ser vuelta a poner de nuevo en el árbol del conocimiento; una vez que empezamos a ver, estamos condenados y enfrentados a buscar la fuerza para ver más, no menos."*
>
> *(Original: "The apple cannot be stuck back on the Tree of Knowledge; once we begin to see, we are doomed and challenged to seek the strength to see more, not less.")*
>
> Arthur Miller

Publication 4
Joint Multifractal Analysis of Air Temperature, Relative Humidity and Reference Evapotranspiration in the Middle Zone of the Guadalquivir River Valley
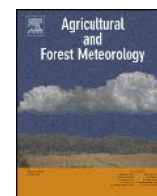
Journal

Agricultural and Forest Meteorology (Ed. Elsevier)

Position / Quartile

(3/91) / Q1-D1 (Agronomy)

Impact factor

4.651 (2019)

Joint multifractal analysis of air temperature, relative humidity and reference evapotranspiration in the middle zone of the Guadalquivir river valley

A.B. Ariza-Villaverde[a,*], P. Pavón-Domínguez[b], R. Carmona-Cabezas[a], E. Gutiérrez de Ravé[a], F.J. Jiménez-Hornero[a]

[a] GEPENA (TEP-957) Research Group, Department of Graphic Engineering and Geomatics, University of Córdoba, Gregor Mendel Building (3rd floor), Campus of Rabanales, 14071 Córdoba, Spain
[b] GEPENA (TEP-957) Research Group, Department of Mechanical Engineering and Industrial Design, University of Cádiz, University Avenue, Puerto Real University Campus, 11519, Spain

ABSTRACT

Previous works have analysed the relationship existing between reference evapotranspiration (ET_0) and other climatic variables under a one-at-a-time perturbation condition. However, due to the physical relationships between these climatic variables is advisable to study their joint influence on ET_0. The box-counting joint multifractal algorithm describes the relations between variables using relevant information extracted from the data singularities. This work investigated the use of this algorithm to describe the simultaneous behaviour of ET_0, calculated by means of Penman–Monteith (PM) equation, and the two main climatic variables, relative humidity (RH) and air temperature (T), influencing on it in the middle zone of the Guadalquivir river valley, Andalusia, southern Spain. The studied cases were grouped according to the fractal dimension values, obtained from the global multifractal analysis, which were related to their probability of occurrence. The most likely cases were linked to smooth behaviour and weak dependence between variables, both circumstances were detected in the local multifractal analysis. For these cases, the rest of Penman Monteith (PM) equation variables, neither the T nor the RH, seemed to influence on ET_0 determination, especially when low T values were involved. By contrast, the least frequent cases were those with variables showing high fluctuations and strong relationship between them. In these situations, when T is low, the ET_0 is affected by the rest of PM equation variables. This fact confirmed T as main driver of ET_0 because the higher T values the lesser influence of other climate variables on ET_0. This condition could not be extended to RH because the variability in ET_0 singularities was not significantly influenced by low or high values of this variable. These results show that the joint multifractal analysis can be regarded as a suitable tool for describing the complex relationship between ET_0, T and RH, providing additional information to that derived from descriptive statistics.

Although, joint multifractal analysis shows some limitations when it is applied to large number of variables, the results reported are promising and suggest the convenience of exploring the relationships between ET_0 and other climatic variables not considered here with this framework such as wind speed and net radiation.

1. Introduction

Evapotranspiration (*ET*) is one of the most important components of the hydrological cycle, and its estimation is essential for scheduling irrigation systems, preparing input data for hydrological water balance models, computing actual *ET* for watersheds, regional water resource planning and analysing climate change effects (Tabari and Talaee,

2014; Tanasijevic et al., 2014; Gong et al., 2006; Villagra et al., 1995). *ET* is a combination of two separate processes whereby water is lost from the soil by evaporation and crop transpiration and evaporation. Commonly, *ET* is modelled by separating the effects of meteorological conditions from the nature of crop and soil available water content (Doorenbos and Pruitt, 1977). For this reason, reference evapotranspiration (*ET₀*) was introduced to quantify the evaporative demand

---

of the actual water-state of soil of the atmosphere. *ET* is affected by several factors, including weather parameters, crop characteristics and management and environmental aspects. However, ET_0 measures the evaporative demand of the atmosphere independently of crop type, crop development and management practices. As water is abundantly available at the ET_0 surface, soil factors do not affect it. Thus, the only factors affecting ET_0 are climatic parameters. Consequently, ET_0 is a climatic parameter and can be computed from weather data.

The most accurate manner in which ET_0 can be quantified is using weighing lysimeters or micrometeorological methods; however, these procedures are not practical because they are time-consuming and expensive (Gavilán et al., 2007). To this end, numerous approaches have been proposed to estimate ET_0, which are grouped in the following categories: i) water budgets (Guitjens, 1982), ii) combined energy and mass balance methods (Monteith, 1965; Penman, 1948), iii) temperature-based methods (Hargreaves and Samani, 1985; Blaney and Criddle, 1950) iv) radiation-based methods (Priestley and Taylor, 1972;), v) mass transfer-based methods (Trabert, 1896; Papadakis, 1966; World Meteorological Organization, 1966), and vi) pan evaporation-based models (Almedeij, 2012; Liu et al., 2004). The use of one method over another is based on the number of atmospheric variables, such as air temperature, wind speed, relative air humidity and solar or net radiation, required as input. These approaches have been evaluated under different climatic conditions (Jabloun and Sahli, 2008; Berengena and Gavilán, 2005; Smith et al., 1996; Allen et al., 1989). The results of ET_0 estimation methods vary with climatic conditions (Eslamian et al., 2011), except for the Penman–Monteith (PM) equation, which demonstrates its superiority for estimating ET_0 over a wide range of climates (Jensen et al., 1990). Consequently, the PM equation is recommended as the standard equation by the United Nations Food and Agriculture Organization (FAO) and the American Society of Civil Engineers (ASCE) for estimating ET_0 (ASCE-EWRI, 2005). Nevertheless, a major drawback of employing the PM equation is its relatively high data demand. For the calculation of daily ET_0, apart from site location, the PM equation requires data for daily maximum and minimum air temperatures, relative humidity (*RH*), solar radiation and wind speed. The number of meteorological stations where all these parameters are observed is limited in many areas worldwide. The number of stations where reliable data for these parameters exist is even smaller, particularly in developing countries (Droogers and Allen, 2002).

Numerous authors have studied ET_0 to improve the understanding of its relationship with climatic variables. Most of these studies employed sensitivity analysis to assess the variation in ET_0 with climatic variables. Gong et al. (2006), explored the influence of *RH*, air temperature, shortwave radiation and wind speed on the PM equation's ET_0 variation in the Changjiang Basin, China. They observed that the most sensitive variable was *RH* and the four studied climatic variables generally varied with season and region. Estévez et al. (2009) assessed the impact of climatic variables on the PM equation's ET_0 in southern Spain and showed a high degree of daily and seasonal variability, particularly for air temperature (*T*) and *RH*. Tabari and Talae (2014) observed the sensitivity of ET_0 to climate change in four types of climates and found that the sensitivity of ET_0 to wind speed and air temperature decreased from arid to humid climates. In all cases, sensitivity analysis was examined under a one-at-a-time perturbation condition, i.e. sensitivity analysis studies the effect of change of one factor on another (McCuen, 1974). It is well established in sensitivity studies that significant effects can be produced by a pair of variables acting in concert (Burgman et al., 1993). Such combined effects can be larger than the sum of the individual effects of two variables (Gong et al., 2006). Nevertheless, the application of sensitivity analysis to evapotranspiration studies to date has been limited to one-at-a-time cases. In real world, the perturbation of more than one variable is likely to happen at the same time. The joint multifractal analysis studies the relations between variables when all of them coexist at the same time. The only condition considered in this regard is that the study variables have multifractal nature. Numerous

authors have demonstrated the multifractal nature of ET_0 (Wang et al., 2014; Xie et al., 2008; Liu et al., 2006), describing the behaviour of the distribution of this variable by means of this formalism. The multifractal theory (Mandelbrot, 1982; Feder, 1988) implies that the complex and heterogeneous behaviour of a self-similar measure (i.e. statistically similar on any scale) can be represented as a combination of interwoven fractal sets (Kravchenko et al., 2009), each of which is characterised by its strength singularity and fractal dimension. This approach reveals certain levels of complexities that are overlooked by traditional statistical tools and fractal analyses (Zeleke and Si, 2004). On the other hand, it transforms irregular data into a more compact form and amplifies slight differences among the variables' distribution (Lee, 2002). The main advantage of this formalism is that its parameters are independent over a range of scales. Additionally, no assumption that the data follow any specific distribution is required. An extension of this procedure is joint multifractal theory, which is proposed by Meneveau et al. (1990). This approach examines the correlations of several multifractal variables that coexist at the same time and quantifies the relations between the studied measures. Joint multifractal analysis has previously been successfully employed to determine the relation between two variables in fields, including soils (Li et al., 2011; Zeleke and Si, 2004, 2005;), pollution (Jiménez-Hornero et al., 2011, 2010) and agronomy (Zeleke and Si, 2004; Kravchenko et al., 2000). This methodology can be extended to study the relationships between three variables. Meneveau et al. (1990) reported that the extension of this method to more than two multifractal distributions is straightforward. Considering the promising possibilities of this formalism, this research explores the simultaneous behaviour of climatic variables, *RH*, *T* and ET_0 in the middle zone of Guadalquivir valley, southern Spain. Regarding the semi-arid regions of southern Spain, these climatic variables are strongly related to each other, as reported by Estévez et al. (2009).

2. Study area

This study was conducted in the middle zone of the Guadalquivir river valley located in Córdoba province, Andalusia, southern Spain (Fig. 1). Specifically, the data were collected in an experimental station located in the vicinity of Córdoba city, (37°51′34.9″ latitude and 04°47′48.9″ longitude; altitude: 110 m). According to the Köppen climate classification, the study area's climate is defined as Csa, with a warm average temperature and dry and hot summers. The annual mean temperature is 17.8 °C, and the annual average rainfall is 621 mm. The weather conditions considerably vary year on year. Moreover, a continental effect is reflected in specific thermal and rainfall regimes: low rainfall, low *RH* and wide ranges for daily and annual temperature are distinctive of the study area (Domínguez-Bascón, 2002).

3. Materials and methods

3.1. Materials

The relationships between three climatic variables were studied in this work: *T*, *RH* and ET_0. To conduct this analysis, the times-series data were collected for 17 years (2001–2017) from the daily database of the Agroclimatic Information Network of Andalusia (Spain), belonging to the Ministry of Agriculture, Livestock, Fisheries and Sustainable Development. The agroclimatic station are situated on the latitude 37° 51′ 25″ N, longitude 04° 48′ 10″ W and altitude 117 above mean sea level (Cordoba, Andalusia). Temperature-humidity probe and wind sensors are placed 1.5 and 2 m above the soil surface respectively.

The three times-series datasets corresponded to daily time resolution, with a total length of 6205 data, as shown in Fig. 2. These time series had 3% of missing data due to maintenance problems or erroneous records. Days for which observations were not available were excluded in the analysis, as Gavilán et al. (2007) and Espadafor et al.
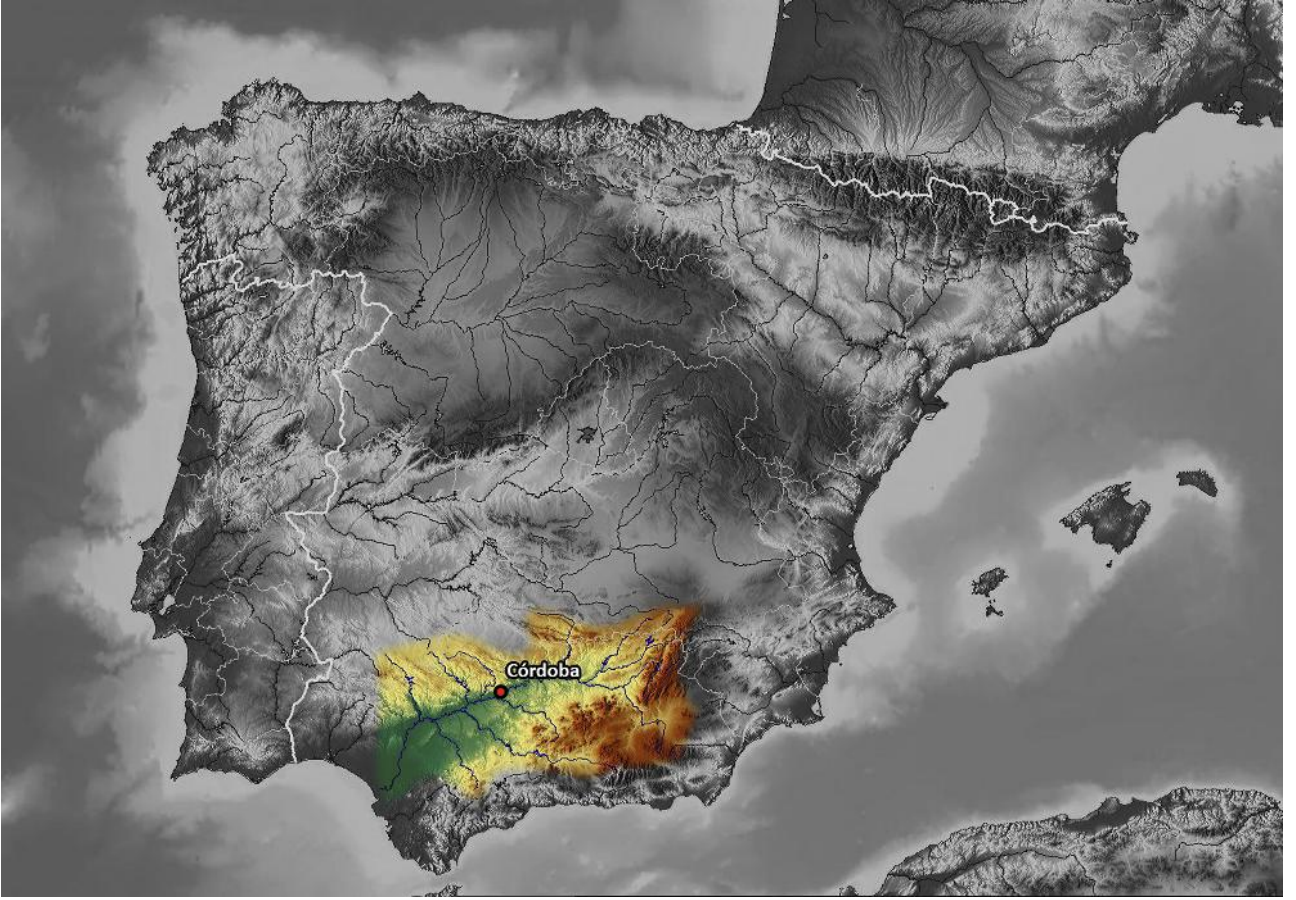
Fig. 1. Location of the study area.

(2011) did in their studies. Gaps were no longer than 3 days in a row. The corresponding averages, coefficients of variation (*CV*), skewness and kurtosis are listed in Table 1. The mean value of ET_O was 3.47 mm/ day, with the maximum and minimum values of 0.1 and 5.80 mm/day, respectively, and *CV* of 0.79. These variations were notably influenced by *T* and *RH*, as can be seen in Fig. 3, where the Pearson's coefficient between *T* and ET_O is greater than 0.427 at a significance level of 0.01. For the case of ET_O and *RH*, they are inversely correlated to a high Pearson's coefficient (-0.694).

3.2. Methodology

3.2.1. Joint multifractal analysis

This formalism was employed herein to describe the relationship between *T*, *RH* and ET_O. Proposed by Meneveau et al. (1990), joint multifractal formalism is based on strange attractor formalism (Halsey et al., 1986; Grassberger, 1983; Hentschel and Procaccia, 1983), which deals with the fractal dimensions of geometric sets associated with the singularities of the measures. In this study, the proposed approach was extended to study the relationships between ET_O, T and *RH*. Thus, to employ this formalism, the time series of ET_O, T and *RH* were divided into non-overlapping intervals of an initial time resolution, n_{ini} ($\delta_{ini} = 2^1$, data = 1 day), in such a way that all of them contained at least one sample of the measure. Thus, the measures $(ET_0 ini)_j$, $(T_{ini})_j$ and $(RH_{ini})_j$ in any initial interval *j* were set to be equal to the sample measurement or to the average when there was more than one sample. Subsequently, the domain was successively divided into *n* non-overlapping intervals for a time resolution ranging from δ_{ini} to δmax. When the analysed time series was split into *n* non-overlapping intervals of time resolution $\delta > \delta_{ini}$, the probability mass functions $c_i ET_0 (\delta)$, $c_i T (\delta)$ and $c_i RH (\delta)$ were defined in each grid size *i* as follows:

$$c_{i[ET_0]}(\delta) = \frac{ET_0 i}{\sum_{j=1}^{n_{ini}} (ET_0 ini)_j}$$

$$c_{i[T]}(\delta) = \frac{T_{ini}}{\sum_{j=1}^{n_{ini}} (T_{ini})_j}$$

$$c_{i[RH]}(\delta) = \frac{RH_i}{\sum_{j=1}^{n_{ini}} (RH_{ini})_j} \tag{1}$$

where $ET_0 i$, T_i and RH_i were calculated as the sum of the $(ET_0 ini)_j$, $(T)_j$ and $(RH)_j$ values, respectively, and included in the interval *i* for a specific time resolution (δ). The distribution of the probability mass function was analysed using the method of moments (Evertsz and Mandelbrot, 1992), and the joint partition function $\chi (q_{ET_0}, q_T, q_{RH}, \delta)$, where q_{ET_0}, q_T and q_{RH} are the statistical moments for ET_O, *T* and *RH*, was calculated from $c_i ET_0 (\delta)$, $c_i T (\delta)$ and $c_i RH (\delta)$ as follows:

$$\chi (q_{ET_0}, q_T, q_{RH}, \delta) = \sum_{i=1}^{n} [c_{i_{ET_0}} (\delta)]^{q_{ET0}} [c_i T (\delta)]^{q_T} [c_i RH (\delta)]^{q_{RH}} \tag{2}$$

The three exponents (q_{ET_0}, q_T, q_{RH}) ranged from -5 to 5 at 0.25 increments. The higher positive values of the statistical moments amplified the greater values of the time series *i*, while the higher negative values of statistical moments amplified the lower values of the time series *i*. This study considered this range of *q* values to avoid instabilities in the multifractal analysis because higher and lower moment orders might magnify the influence of outliers in the measurements as stated by Zeleke and Si (2005).

The joint partition function has the following scaling property for multifractal measures:

$$\chi (q_{ET_0}, q_T, q_{RH}, \delta) \approx \delta^{\tau (q_{ET_0}, q_T, q_{RH})} \tag{3}$$

where $\tau (q_{ET_0}, q_T, q_{RH})$ is known as the joint mass exponent function and

ORIGINAL TIME SERIES
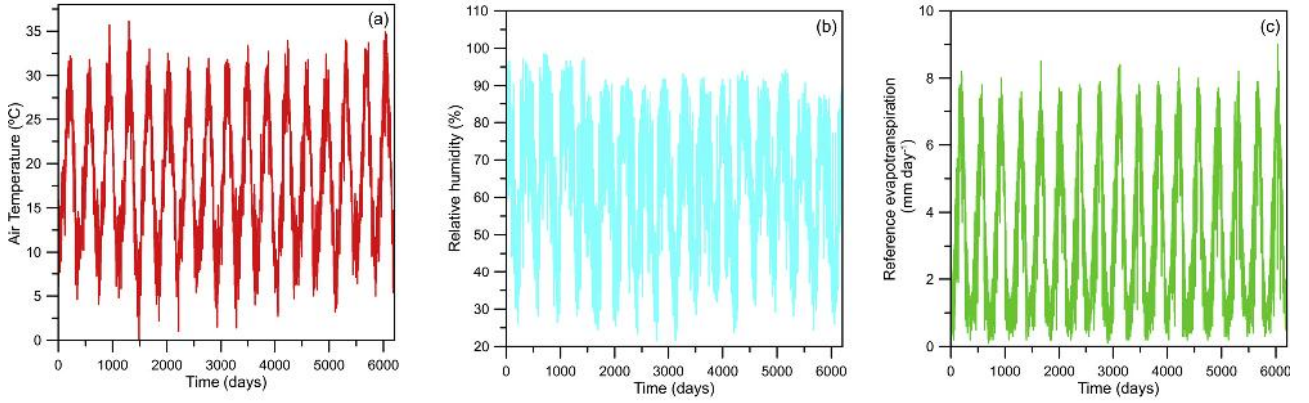
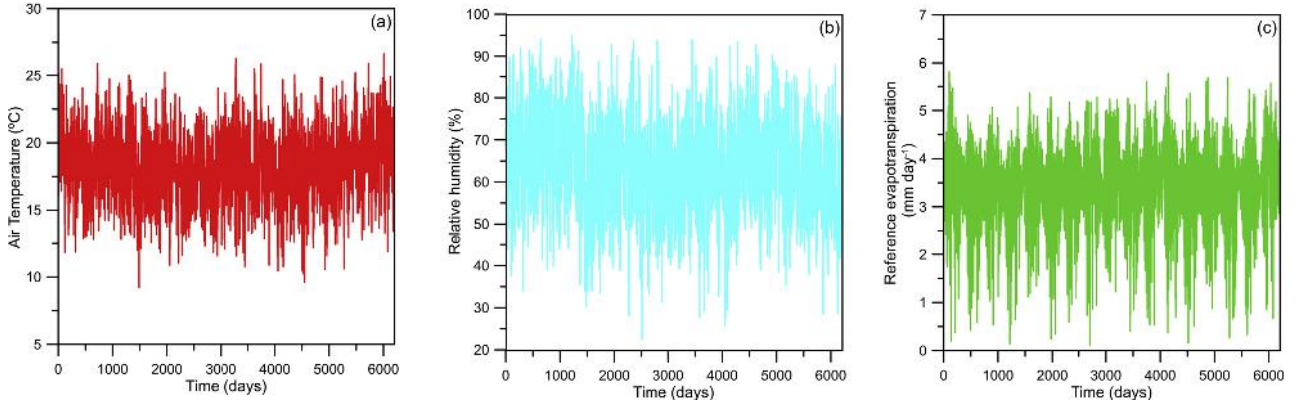**DESEASONALIZED TIME SERIES**

Fig. 2. Time-series data of the studied climatic variables.

Table 1
Statistical parameters of the studied climatic variables.

Climatic Variables	N	Minimum	Maximum	Average	CV	Skewness	kurtosis
T (°C)	6205	9,2	26,65	18,31	2,54	-,106	-,123
RH (%)	6205	22,6	95,00	63,28	10,70	-,086	,033
ET$_o$ (mm/day)	6205	0,0	5,80	3,47	0,79	-,860	1,880

depends only on the exponents q_{ET_0}, q_T and q_{RH}. This exponent can be obtained from the slope of the linear segment of a log–log plot of $\chi(q_{ET_0}, q_T, q_{RH}, \delta)$ vs. δ. At $q_{ET_0} > 0$, $q_T > 0$ and $q_{RH} > 0$, the value of the joint partition function was mainly determined by the high values of ET_0, T and RH, while the influence of the low ET_0, T and RH contributed mostly to the partition function for $q_{ET_0} < 0$, $q_{T_1} < 0$ and $q_{RH} < 0$. The linear fits from the logarithmic plots of the partition function versus the time resolution show the range of temporal scales, $\delta_{min} - \delta max$, for which the multifractal nature is exhibited.

The Hölder exponents α_{ET_0}, α_T and α_{RH}, characterised the singularities contained in the signal, which were inversely proportional to the strength of the singularity that described an abrupt change in the time series (Hampson and Malen, 2011). α_{ET_0}, α_T and α_{RH} are also known as local fractal dimensions and can be determined by Legendre transformations of the $\tau(q_{ET_0}, q_{T_1}, q_{RH})$ function (Kravchenko et al., 2000):

$$\alpha_{ET_O}(q_{ET_0}, q_T, q_{RH}) = d\tau(q_{ET_0}, q_{T_1}, q_{RH})/dq_{ET_0}$$
$$\alpha_T(q_{ET_0}, q_{T_1}, q_{RH}) = d\tau(q_{ET_0}, q_T, q_{RH})/dq_T$$
$$\alpha_{RH}(q_{ET_0}, q_T, q_{RH}) = d\tau(q_{ET_0}, q_{T_1}, q_{RH})/dq_{RH} \qquad (4)$$

The main purpose of any multifractal algorithm is to determine the distribution of singularities. The Hölder exponent can be mathematically defined as a point where conventional mathematical modelling breaks down (Cheng, 2007). In classical methods, a difficulty arises in the computation of the derivative of a noisy and discrete signal. These methods involve smoothing of discrete time-series data, and the gradient is then computed by differentiating the smoothed signal (Levy-Vehel and Berroir, 1942). In contrast, the multifractal approach directly uses the initial discrete time-series data and the information is straight extracted from singularities. The main advantage of this idea is that no information is lost or introduced by the smoothing process.

For time series (one-dimensional signal), when the Hölder exponent has an approximate value of 1, the measure is regular; therefore, a large change does not occur. When the Hölder exponents are less or greater than 1 ($\alpha \neq 1$), the variable in the time series shows a high gradient or discontinuity of the signal, and for α equal to 1 indicates smooth behaviour or considerably small activity (Levy-Vehel et al., 1992).

Let $N(\alpha_{ET_0}, \alpha_T, \alpha_{RH}, \delta)$ be the number of intervals of a given grid size (δ), where a given combination of α values is found. Now, let us define
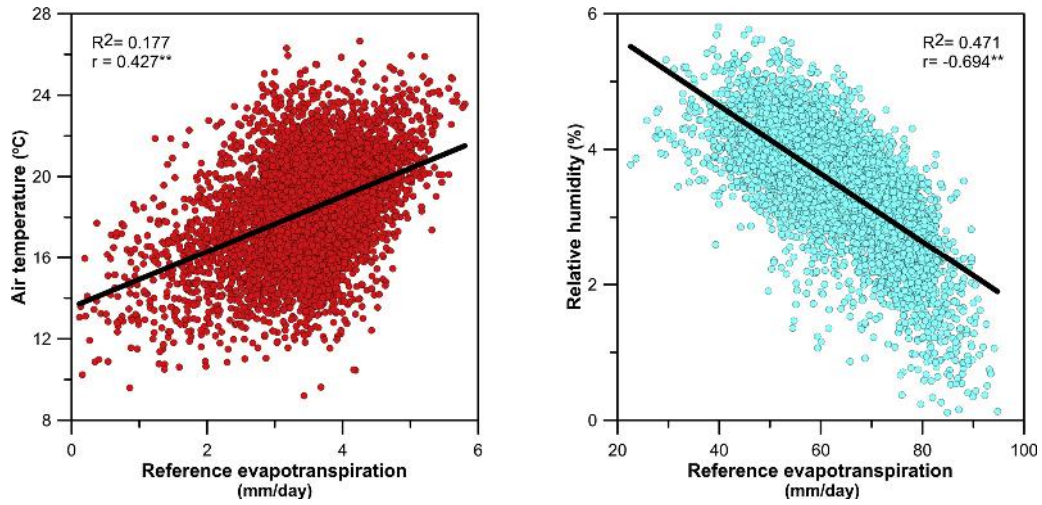
Fig. 3. Scatter plots and linear fits of the studied climatic variables.

$f(\alpha_{ET_0}, \alpha_T, \alpha_{RH})$ from the scaling relation:

$$N(\alpha_{ET_0}, \alpha_T, \alpha_{RH}, \delta) = \delta^{-f(\alpha_{ET_0}, \alpha_T, \alpha_{RH})}. \tag{5}$$

$f(\alpha_{ET_0}, \alpha_T, \alpha_{RH})$ can be considered as a fractal dimension of a set of intervals that correspond to the singularities α_{ET_0}, α_T and α_{RH}, respectively. A plot of $f(\alpha_{ET_0}, \alpha_T, \alpha_{RH})$ vs. α_{ET_0}, α_T and α_{RH} is referred to as the joint multifractal spectrum. It can be calculated from the following equation (Meneveau et al., 1990; Chhabra et al., 1989; Chhabra and Jensen, 1989):

$$f(\alpha_{ET_0}, \alpha_T, \alpha_{RH}) = q_{ET_0}\alpha_{ET_0} + q_T\alpha_T + q_{RH}\alpha_{RH} - \tau(q_{ET_0}, q_T, q_{RH}). \tag{6}$$

The highest value of $f(\alpha_{ET_0}, \alpha_T, \alpha_{RH})$ corresponds to the fractal capacity dimension, D_0, that is equal to the Euclidean dimension (i.e. 1 when dealing with time series) in the box-counting joint multifractal analysis. D_0 is reached when q_{ET_0}, q_T and q_{RH} are equal to zero. $f(\alpha_{ET_0}, \alpha_T, \alpha_{RH})$ represents the frequency of the occurrence of a certain value of α_{ET_0}, α_T and α_{RH} (Biswas et al., 2012). Generally, the joint multifractal spectrum is used to represent the joint dimensions $f(\alpha_{ET_0}, \alpha_T, \alpha_{RH})$ of the analysed variables. When one statistical moment is equal to zero, the joint multifractal spectrum is identical to that for two variables. Likewise, a single multifractal spectrum is obtained when two statistical exponents are equal to zero at the same time.

3.2.2. Reference evapotranspiration

ET_0 is defined as the theoretical ET from an extensive surface of actively growing green grass of uniform height completely shading the ground and not short of water, i.e. without water restrictions (Allen et al., 1998). In this study, the ET_0 values were calculated using the FAO-56 PM equation (Allen et al., 1998), which is a simplification of the original PM equation (Monteith, 1965). For the grass reference surface and daily time step, this equation is expressed as follows:

$$ET_0 = \frac{0.408\Delta(R_n - G) + 100/T + 273U_2(e_s - e_a)}{\Delta + \gamma(1 + 0.34U_2)}, \tag{7}$$

where ET_0 is reference evapotranspiration (mm day^{-1}); Rn is the net radiation at the crop surface (MJ m^{-2} day^{-1}); G is the soil heat flux (MJ m^{-2} day^{-1}), assumed to be zero for a daily time step (Allen et al., 1998) because step soil heat flux is small compared to net radiation when the soil is covered by vegetation; T is the mean daily air temperature (°C); U_2 is the wind speed at a height of 2 m (m s^{-1}); e_s is the saturation vapour pressure (kPa); e_a is the actual vapour pressure (kPa); $(e_s - e_a)$ is the saturation vapour pressure deficit (kPa); Δ is the slope of the saturated vapour–pressure curve (kPa °C^{-1}) and γ is the psychometric constant (kPa °C^{-1}). The unit for the coefficient 0.408 was MJ^{-1} m^2 mm. All daily calculations were performed following the FAO-

56 (Jabloun and Sahli, 2008) equation. Therefore, grass height and bulk canopy resistance were assumed to be 0.12 m and 70 m s^{-1}, respectively.

3.2.3. Seasonality

The analysed time series exhibited seasonal fluctuations (Fig. 2) that correspond to a nonstationary process, typical behaviour for most hydro-meteorological variables at sub-annual time scales. However, the multifractal formalism applies to stationary processes (see e.g., Mandelbrot, 1982). The periodicity in hydro-meteorological time series affects their nonlinear properties and the width of the multifractal spectrum (Krzyszczak et al., 2017; Livina et al., 2011). Thus, it is necessary to filter the time series out before studying fractal properties. With this aim the seasonal decomposition procedure provided by IBM SPSS Statistics (v.25.0, IBM, 2017) was used to identify and remove the variations associated with the seasonal effects. Therefore, the original time series were decomposed into trend, seasonal and irregular components (Fig. 4).

4. Results and discussion

The present work carried out the study of the simultaneous behaviour of the climatic variables T, RH and ET_0 using the joint multifractal analysis. This method assumes that an individual variable is multifractal. Therefore, it was necessary to check the multifractal nature of the study climatic variables before employing this algorithm. This behaviour was confirmed by previous research, which found the fluctuations of T, RH and ET_0, showing self-similar structures for several ranges of time resolutions (Baranowski et al., 2015; Wang et al., 2014; Li-Hao and Zun-Ta, 2013; Xie et al., 2008; Liu et al., 2006; Király and Jánosi, 2005). Fig. 5 shows the single multifractal spectra of the climatic variables considered herein at daily time scale. The shapes of these spectra (looking like inverted parabolas) evidencing their multifractal nature.

Fig. 6 shows the joint multifractal spectrum obtained for the climatic variables examined in this study. From the logarithmic plots of the partition function versus the time resolution, linear fits ($R^2 > 0.999$, for all cases) were obtained for $\delta = \delta_{ini} = 8$ days to $\delta max = 512$ days. Koutsoyiannis (2003) suggested that aggregation scales should range from the basic scale to a maximum value δmax chosen so that sample moments can be estimated from at last 10 data values. This analysis revealed that the climatic variables exhibited multifractal nature between time resolutions ranging from 8 to 512 days.

The joint multifractal spectrum (Fig. 6) represents the relations between the Hölder exponents of the three variables (α_{ET_0}, α_T, α_{RH}).
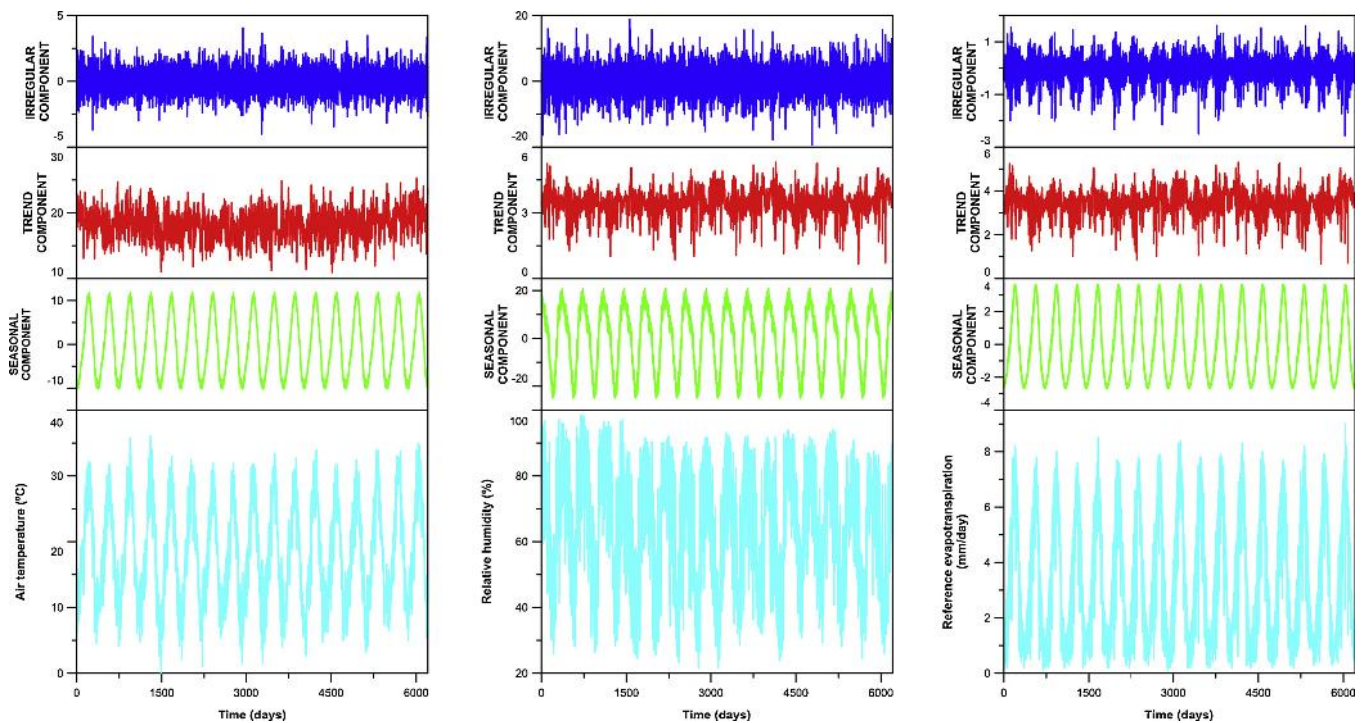
Fig. 4. Decomposition of the studied variables time series into seasonal, trend and irregular components for daily series.
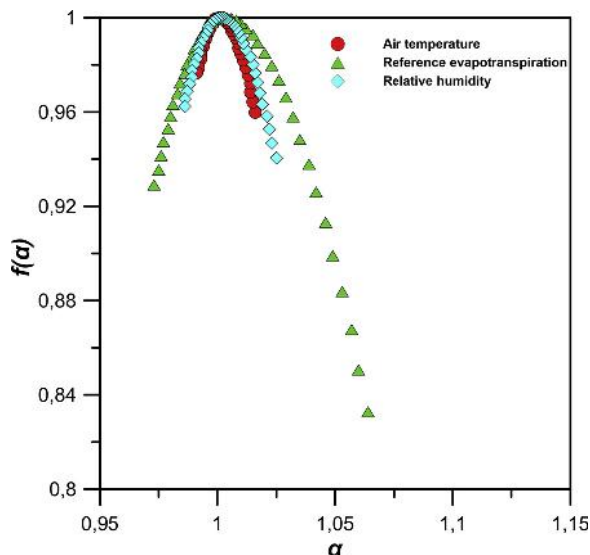

Fig. 5. Single multifractal spectra of the studied variables.

Each singularity combination corresponded to a fractal dimension value $f(\alpha_{ET_O}, \alpha_T, \alpha_{RH})$. When the three variables were analysed, the joint multifractal spectrum was characterised by a volume. The current study investigated the relations between T, RH and ET_O under different scenarios, each of which was determined by the range of orders (q) selected for the statistical moments. Table 2 lists the eight studied cases. Here, $q > 0$ and $q < 0$ mean high and low variable values, respectively.

According to Fig. 7, which displays the results obtained from the global multifractal analysis, the cases listed in Table 2 can be clustered in three groups based on the median (M_e) value of the fractal dimension values, $f(\alpha_T, \alpha_{RH}, \alpha_{ET_O})$, as it is shown in the box-whisker plots. Thus, the first group includes Cases 1, 2 and 3 which were the most frequent situations found in the time series. A second group is composed by Cases 4, 5 and 6 which exhibited a lower probability of occurrence than those contained in the first group. Finally, the last group can be set with

Cases 7 and 8 which were less likely.

The results yielded from the applying the local multifractal analysis to the components of each group are shown in Fig. 8 (Cases 1, 2 and 3), Fig. 9 (Cases 4, 5 and 6) and Fig. 10 (Cases 7 and 8). Regarding Fig. 8 it can be checked that T, RH and ET_O behaviour was regular because the corresponding Hölder exponents, α_T, α_{RH}, α_{ETO}, were around 1. In addition, α_T, α_{RH}, α_{ETO} distributions were similar revealing low association between the studied variables. Both circumstances suggested that the rest of variables included in PM equation influenced on ET_O determination, especially in Case 2 where low T values were involved provoking wider ET_O box-whisker plot than Cases 1 and 3. The local behaviour detected is related to frequent situations according to the fractal dimensions shown in Fig. 7 for Cases 1, 2 and 3.

Box-whisker plots displayed in Fig. 9 exhibits higher association between variables for Cases 4, 5 and 6, due to the differences in α_T, α_{RH}, α_{ETO} distributions, than those included in the first group. Therefore, the rest of the PM equation variables has less influence on these cases. The fluctuations reported for T, RH and ET_O were more relevant than those detected for the previous cases 1, 2 and 3 because there were some situations in which the Hölder exponents differed to 1. Cases 5 and 6, with low T values, had wider ET_O box-whisker plots suggesting the influence of PM equation variables as happened in Case 2. Local multifractal analysis results obtained for the second group of cases agreed with the fact of being less recurrent likely than those included in the first group (Fig. 7).

Local multifractal analysis results for Cases 7 and 8, belonging to the least frequent group (Fig. 7), are shown in Fig. 10. The α_T, α_{RH}, α_{ETO} distributions exhibited more relevant differences between them compared to the cases included in the other groups. This situation is related with stronger association between T, RH and ET_O because the Hölder exponents differed to 1 for all variables. It can be verified that low T values (Case 8) were linked to wider ET_O box-whisker plots indicating the influence of PM equation variables as it was commented before.

A last comment about global multifractal analysis results (Fig. 7) should be done because the stronger association between T, RH and ET_O the wider range for $f(\alpha_T, \alpha_{RH}, \alpha_{ET_O})$ in each group of cases, i.e. higher variability probability of occurrence. This fact was more relevant when
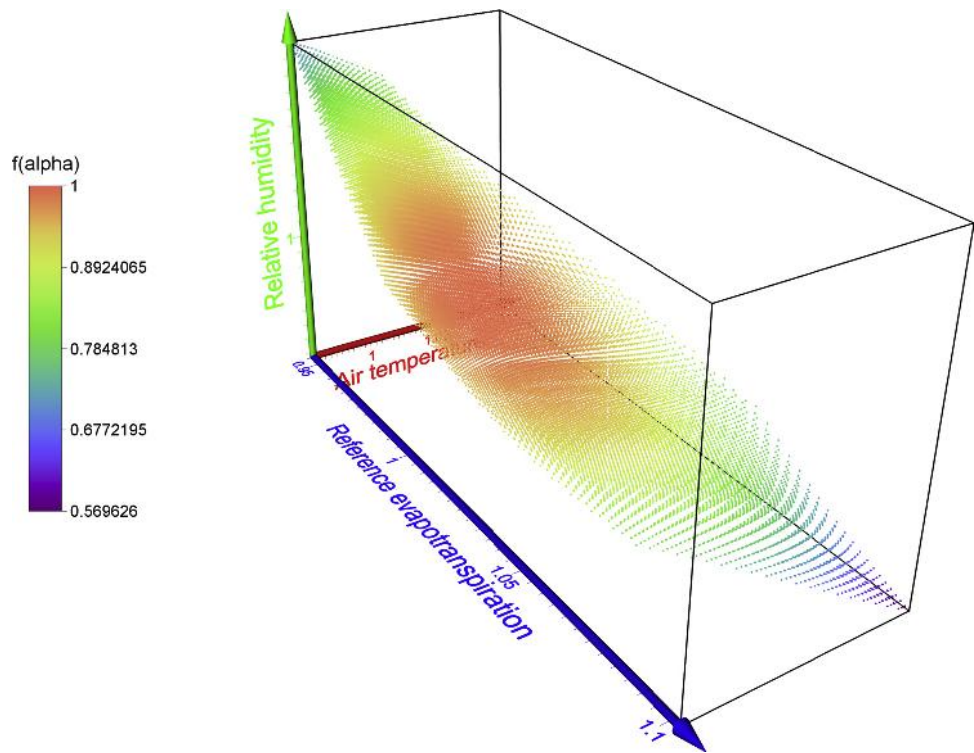
Fig. 6. Joint multifractal spectrum.

Table 2
Studied sceneries description.

	T	RH	ET0
CASE 1	q > 0	q > 0	q < 0
CASE 2	q < 0	q < 0	q > 0
CASE 3	q > 0	q > 0	q > 0
CASE 4	q > 0	q < 0	q < 0
CASE 5	q < 0	q < 0	q < 0
CASE 6	q < 0	q > 0	q > 0
CASE 7	q > 0	q < 0	q > 0
CASE 8	q < 0	q > 0	q < 0

$q > 0$ and $q < 0$ mean high and low variable values, respectively.

low T values were involved (see Case 2 in the first group; Cases 5 and 6 in the second group; and Case 8 in the last group).

5. Conclusion

Previous works suggested physical interaction between T and RH as the most relevant process influencing on ET_0 in locations placed in middle zone of the Guadalquivir river Valley (Andalusia, southern Spain). Joint multifractal analysis performed here confirmed this fact according to the results yielded by extracting time series information from singularities. This information allowed to group the studied cases according to their probability of occurrence determined by the fractal dimension values from the global multifractal analysis. The most likely cases were related to smooth behaviour and weak association between T, RH and ET_0, both circumstances detected in the local multifractal analysis. For these cases, it was suggested that the rest of PM equation variables influenced on ET_0 determination, especially when low T values were involved. By contrast, the least frequent cases were those with high T, RH and ET_0 fluctuations and strong relationship between them. In these situations, the other PM equation variables effects on ET_0 were only relevant with low T values again. Thus, T can be regarded as main driver of ET_0 because the higher T values the lesser influence of the rest of the PM equation variables acting on ET_0. However, for cases with low
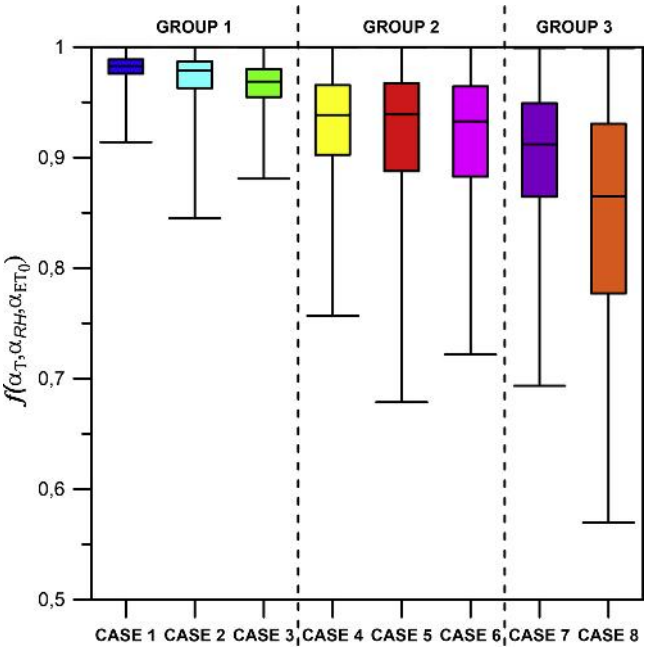

Fig. 7. Box–whisker plots of the fractal dimensions considered in each study case.

T values, the variability in ET_0 singularities was higher and not significantly being influenced by low or high RH values implying the action of other PM equation variables.

To date, the relationships between ET_0 and other climatic variables have been analysed under a one-at-a-time perturbation condition. However, this study has explored the links between T, RH and ET_0 acting in concert using box-counting joint multifractal analysis. The results obtained herein are promising and expand the existing description of the complex interactions between these variables.
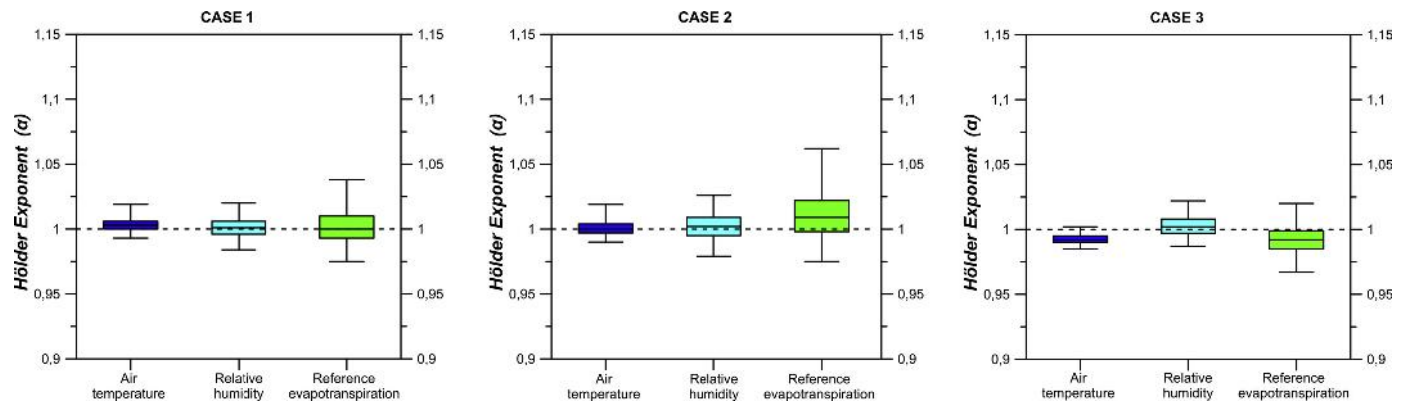
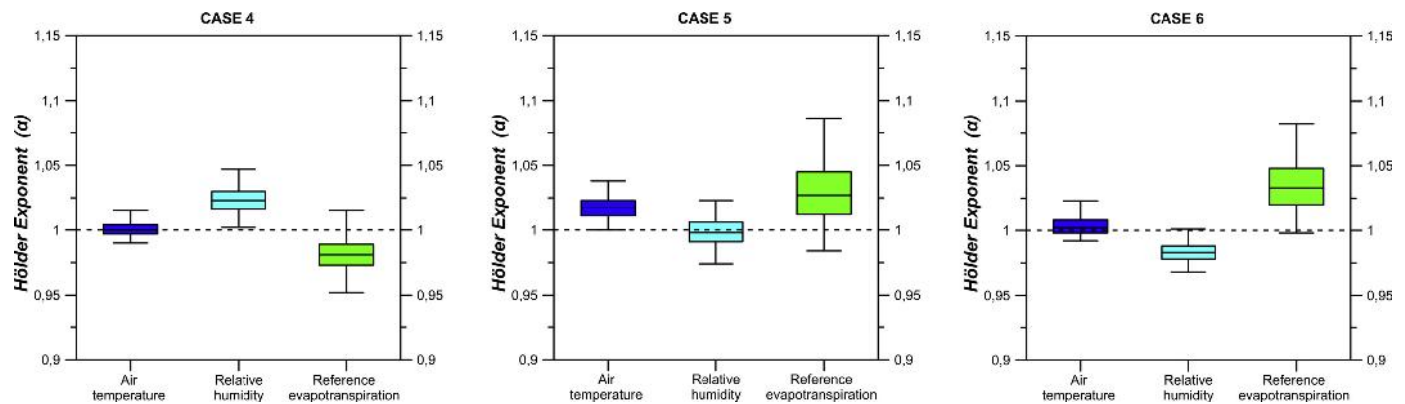**Fig. 8.** Box–whisker plots of the singularities considered in Cases 1, 2 and 3.

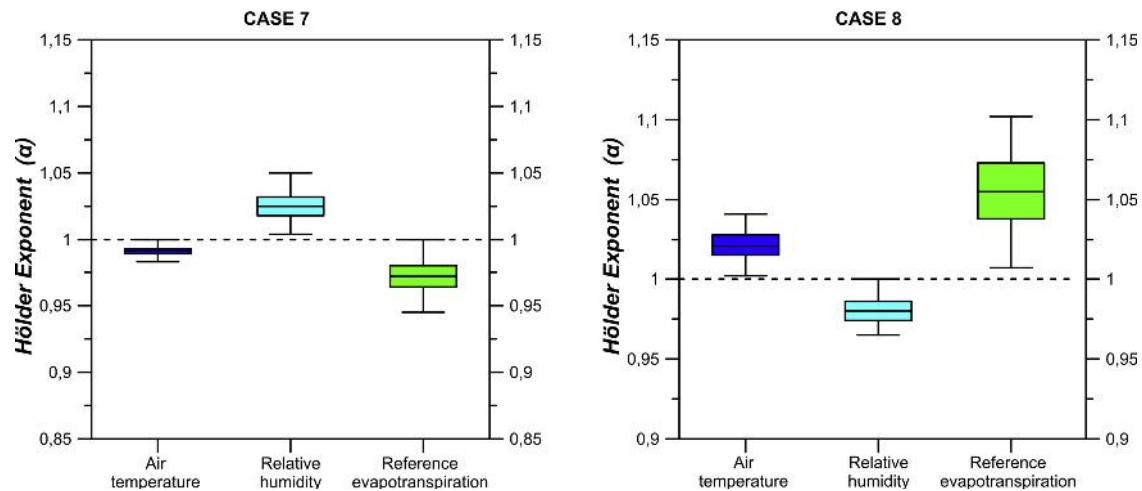**Fig. 9.** Box–whisker plots of the singularities considered in Cases 4, 5 and 6.

Fig. 10. Box–whisker plots of the singularities considered in Cases 7 and 8.

Alternative approach to descriptive statistics, such as multifractal analysis, applied to study the ET_O links to climate drivers are needed due to the increase of freshwater resources demand under global climate change.

From the theoretical point of view, joint multifractal analysis can be trivially extended to the study of more than three variables. However, some limitations, related to the computational load and graphic representation of the results, prevent its use in practice. Therefore, the joint multifractal study of ET_O and some variables appearing in PM equation, such as net radiation and wind velocity, is pending as future work. Thus, new insights might be found specially for situations in which air temperature is low allowing the influence of other climate drivers on ET_O different from *RH*.

References

Allen, R.G., Jensen, M.E., Wright, J.L., Burman, R.D., 1989. Operational estimates of reference evapotranspiration. Agron. J. 81, 650–662.

Allen, R.G., Pereira, L.S., Raes, D., Smith, M., 1998. Crop Evapotranspiration: Guidelines for Computing Crop Water Requirements, FAO Irrigation and Drainage, first ed. Food and Agriculture Organization of the United Nations, Rome, Italy.

Almedeij, J., 2012. Modeling pan evaporation for Kuwait by multiple linear regression. Sci. World J ID 574742.

ASCE-EWRI, 2005. The ASCE standardized reference evapotranspiration equation. (Environmental and Water Resources Institute of the American Society of Civil Engineers, Task Committee on Standardization of Reference Evapotranspiration Calculation, first ed. pp. 70 Washington DC, USA.

Baranowski, P., Krzyszczak, J., Slawinski, C., Hoffmann, H., Kozyra, J., Nieróbca, A., Siwek, K., Gluz, A., 2015. Multifractal analysis of meteorological time series to assess climate impacts. Clim. Res. 65, 39–52.

Berengena, J., Gavilán, P., 2005. Reference evapotranspiration estimation in a highly advective semiarid environment. J. Irrig. Drainage Eng. ASCE 131 (2), 147–163.

Biswas, A., Cresswell, H.P., Bing, C.S., 2012. Application of Multifractal and Joint Multifractal Analysis in Examining Soil Spatial Variation: a Review, in Fractal Analysis and Chaos in Geosciences, Ed. Quadfeul S, pp. 109–138.

Blaney, H.F., Criddle, W.D., 1950. Determining Water Requirements in Irrigated Areas From Climatologically and Irrigation Data. USDA Soil Conservation Service SCS, pp. 44.

Burgman, M.A., Ferson, S., Akcakaya, H.R., 1993. Risk Assessment in Conservation Biology. Chapman & Hall, London.

Cheng, Q., 2007. Mapping singularities with stream sediment geochemical data for prediction of undiscovered mineral deposits in Gejiu, Yunnan Province, China. Ore Geol. Rev. 32, 314–334.

Chhabra, A.B., Meneveau, C., Jensen, R.V., Sreenivasan, K.R., 1989. Direct determination of the f (α) singularity spectrum and its application to fully developed turbulence. Phys. Rev. A 40, 5284–5294.

Chhabra, A.B., Jensen, R.V., 1989. Direct determination of the f(α) singularity spectrum. Phys. Rev. Lett. 62, 1327–1330.

Domínguez-Bascón, P., 2002. Clima regional y microclimas urbanos en la provincial de Córdoba. Servicio de Publicaciones de la Universidad de Córdoba, Córdoba, Spain.

Doorenbos, J., Pruitt, W.O., 1977. Crop Water Requirements. FAO Irrigation and Drainage, 24. Food and Agriculture Organization, Rome.

Droogers, P., Allen, R.G., 2002. Estimating reference evapotranspiration under inaccurate data conditions. Irrig. Drain Syst. 16, 33–45.

Eslamian, S., Khordadi, M.J., Abedi-Koupai, J., 2011. Effects of variations in climatic parameters on evapotranspiration in the arid and semi-arid regions. Glob. Planet. Change 78 (3–4), 188–194.

Espadafor, M., Lorite, I.J., Gavilán, P., Berengena, J., 2011. An analysis of the tendency of reference evapotranspiration estimates and other climate variables during the last 45 years in Southern Spain. Agric. Water Manag. 98 (6), 1045–1061.

Estévez, J., Gavilán, P., Berengena, J., 2009. Sensitivity analysis of a Penman–Monteith type equation to estimate reference evapotranspiration in southern Spain. Hydrol. Process. 23, 3342–3353.

Evertsz, C.J.G., Mandelbrot, B.B., 1992. Multifractal Measures. Chaos and Fractals: New Frontiers in Science. Springer, New York.

Feder, J., 1988. Fractals. Plenum Press, New York.

Gavilán, P., Berengena, J., Allen, R.G., 2007. Measuring versus estimating net radiation and soil heat flux: impact on Penman–monteith reference ET estimates in semiarid regions. Agric. Water Manage. 89, 275–286.

Gong, L., Xu, C., Chen, D., Halldin, S., 2006. Sensitivity of the Penman–Monteith reference evapotranspiration to key climatic variables in the Changjiang (Yangtze River) basin. J. Hydrol. 329, 620–629.

Grassberger, P., 1983. Generalized dimensions of strange attractors. Phys. Lett. A 97, 227–230.

Guitjens, J.C., 1982. Models of alfalfa yield and evapotranspiration. J. Irrig. Drain Div. 108 (3), 212–222.

Halsey, T.C., Jensen, M.H., Kadanoff, L.P., Procaccia, I., Shraiman, B.I., 1986. Fractal measures and their singularities: the characterization of strange sets. Phys. Rev. 33, 1141–1151.

Hampson, K.M., Malen, E.A.H., 2011. Multifractal nature of ocular aberration dynamics of the human eye. Biomed. Opt. Express 1, 464–477.

Hargreaves, G.H., Samani, Z.A., 1985. Reference crop evapotranspiration from temperature. Appl. Eng. Agric. 1 (2), 96–99.

Hentschel, H.E., Procaccia, I., 1983. The infinite number of generalized dimensions of fractals and strange attractors. Phys. D 8, 435–444.

Jabloun, M., Sahli, A., 2008. Evaluation of FAO-56 methodology for estimating reference evapotranspiration using limited climatic data Application to Tunisia. Agr. Water Manage. 95, 707–715.

Jensen, M.E., Burman, R.D., Allen, R.G., 1990. Evapotranspiration and Irrigation Water Requirements, ASCE Manuals and Reports on Engineering Practice 70 American Society of Civil Engineers, New York.

Jiménez-Hornero, F.J., Jiménez-Hornero, J.E., Gutiérrez de Ravé, E., Pavón-Domínguez, P., 2010. Exploring the relationship between nitrogen dioxide and ground-level

ozone by applying the joint multifractal analysis. Environ. Monit. Assess. 167, 675–684.

Jiménez-Hornero, F.J., Pavón-Domínguez, P., Gutiérrez de Ravé, E., Ariza-Villaverde, A.B., 2011. Joint multifractal description of the relationship between wind patterns and land surface air temperature. Atmos. Res. 99 (3–4), 366–376.

Király, A., Jánosi, I.M., 2005. Detrended fluctuation analysis of daily temperature records: geographic dependence over Australia. Meteorol. Atmos. Phys. 88, 119–128.

Koutsoyiannis, D., 2003. Climate change, the Hurst phenomenon, and hydrological statistics. Hydrol. Sci. J. 48 (1), 3–24.

Kravchenko, A.N., Bullock, D.G., Boast, C.W., 2000. Joint multifractal analysis of crop yield and terrain slope. Agron. J. 92 (6), 1279–1290.

Kravchenko, A.N., Martin, M.A., Smucker, A.J.M., Rivers, M.L., 2009. Limitations in determining multifractal spectra from pore–solid soil aggregate images. Vadose Zone J. 8, 220–226.

Krzyszczak, J., Baranowski, P., Zubik, M., Hoffmann, H., 2017. Temporal scale influence on multifractal properties of agro-meteorological time series. Agric. For. Meteorol. 239, 223–235.

Lee, C.K., 2002. Multifractals characteristics in air pollutant concentration times series. Water Air Soil Pollut. 135, 389–409.

Levy-Vehel, J., Berroir, J.P., 1942. INRIA, London, United KingdomImage Analysis Through Multifractal Description, Presented at the Fractal's 93 Conference1942. Image Analysis Through Multifractal Description, Presented at the Fractal's 93 Conference.

Levy-Vehel, J., Mignot, P., Berroir, J.P., 1992. Multifractal, texture and image analysis. Computer vision and pattern recognition. Proceedings 1992 IEEE Computer Society Conference on Computer Vision and Pattern Recognition.

Li, Y., Li, M., Horton, R., 2011. Single and joint multifractal analysis of soil particle size distributions. Pedosphere. 21 (1), 75–83.

Li-Hao, G., Zun-Ta, F., 2013. Multi-fractal behaviors of relative humidity over China. Atmos. Oceanic Sci. Lett. 6 (2), 74–78.

Liu, B., Shao, D., Shen, X., 2006. Research on temporal fractal features of reference evapotranspiration. J. Irrig. Drain. 25 (5), 9–13.

Liu, B., Xu, M., Henderson, M., Gong, W., 2004. A spatial analysis of pan evaporation trends in China. J. Geophys. Res. 109 (D15102), 1955–2000.

Livina, V.N., Ashkenazy, Y., Bunde, A., Havlin, S., 2011. In: Kropp, J., Shellnhuber, H.J. (Eds.), Seasonality Effects on Nonlinear Properties of Hydrometeorological Records. Springer, Berlin.

Mandelbrot, B.B., 1982. The Fractal Geometry of Nature. W.H. Freeman, New York.

McCuen, H.R., 1974. A sensitivity and error analysis of procedures used for estimating evapotranspiration. Water Resour. Bull. 10 (3), 486–498.

Meneveau, C., Sreenivasan, K.R., Kailasnath, P., Fan, M.S., 1990. Joint multifractal measures: theory and applications to turbulence. Phys. Rev. A 41, 894–913.

Monteith, J.L., 1965. Evaporation and environment. The State and Movement of Water in Living Organisms, in XIX Symposia of the 1965 Society for Experimental Biology. Cambridge University Press, Swansea, UK.

Papadakis, J., 1966. Climates of the World and Their Agricultural Potentialities. Papadakis, Buenos Aires.

Penman, H.L., 1948. Natural evaporation from open water, bare soil and grass. Proceedings Royal Society A, (London, United Kingdom 193. pp. 120–145.

Priestley, C.H.B., Taylor, R.J., 1972. On the assessment of surface heat flux and evaporation using large-scale parameters. Mon. Weather Rev. 100, 81–92.

Smith, M., Allen, R.G., Pereira, L.S., 1996. Revised FAO methodology for crop water requirements. Evapotranspiration and Irrigation Scheduling, in Proceedings International Conference 116–123.

Tabari, H., Talaee, P.H., 2014. Sensitivity of evapotranspiration to climatic change in different climates. Global Planet Change 115, 16–23.

Tanasijevic, L., Todorovic, M., Pereira, L., Pizzigalli, C., Lionello, P., 2014. Impacts of climate change on olive crop evapotranspiration and irrigation requirements in the Mediterranean region. Agric. Water Manage. 144, 54–68.

Trabert, W., 1896. Neue Beobachtungen uber Verdampfungsgeschwindigkeiten. Meteorologische Zeitschrift 13, 261–263.

Villagra, M.M., Bacchi, O.O.S., Tuon, R.L., Reichardt, K., 1995. Difficulties of estimating evapotranspiration from the water balance equation. Agr. Forest Meteorol. 72, 317–325.

Wang, W., Zou, S., Luo, Z., Zhang, W., Chen, D., Kong, J., 2014. Prediction of the reference evapotranspiration using a chaotic approach. Sci. World J. 347625, 13.

World Meteorological Organization, 1966. Measurement and Estimation of Evaporation and Evapotranspiration. Secretariat of the World Meteorological Organization, Geneva.

Xie, X., Cui, Y., Zhou, Y., 2008. Long-term correlation and multifractality of reference crop evapotranspiration time series. J. Hydraul. Eng. 39 (12), 1327–1333.

Zeleke, T.B., Si, B.C., 2004. Scaling properties of topographic indices and crop yield: multifractal and joint multifractal approaches. Agron. J. 96, 1082–1090.

Zeleke, T.B., Si, B.C., 2005. Scaling relationships between saturated hydraulic conductivity and soil physical properties. Soil Sci. Soc. Am. J. 69, 1691–1702.

Publication 5
Checking Complex Networks Indicators in Search of Singular Episodes of the Photochemical Smog

Journal

Chemosphere (Ed. Elsevier)

Position / Quartile

(29/265) / Q1 (Environmental Sciences)

Impact factor

5.778 (2019)

Checking complex networks indicators in search of singular episodes of the photochemical smog

Rafael Carmona-Cabezas[*], Javier Gómez-Gómez, Eduardo Gutiérrez de Ravé, Francisco J. Jiménez-Hornero

Complex Geometry, Patterns and Scaling in Natural and Human Phenomena (GEPENA) Research Group, University of Cordoba, Gregor Mendel Building (3rd floor), Campus Rabanales, 14071, Cordoba, Spain

HIGHLIGHTS

- Visibility Graphs can be used to identify singularities in pollutant series.
- Peaks from complex network indicators coincide with the multifractal ones.
- Hölder and singularity exponents give ambiguous results due to parameter selection.
- Betweenness and degree give the clearest signal for identifying singular points.
- Among VG indicators, closeness centrality gives the least accordance.

GRAPHICAL ABSTRACT
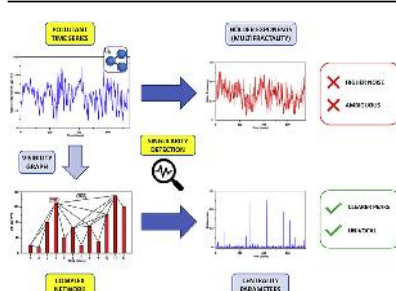
ABSTRACT

A set of indicators derived from the analysis of complex networks have been introduced to identify singularities on a time series. To that end, the Visibility Graphs (VG) from three different signals related to photochemical smog (O_3, NO_2 concentration and temperature) have been computed. From the resulting complex network, the centrality parameters have been obtained and compared among them. Besides, they have been contrasted to two others that arise from a multifractal point of view, that have been widely used for singularity detection in many fields: the Hölder and singularity exponents (specially the first one of them).

The outcomes show that the complex network indicators give equivalent results to those already tested, even exhibiting some advantages such as the unambiguity and the more selective results. This suggest a favorable position as supplementary sources of information when detecting singularities in several environmental variables, such as pollutant concentration or temperature.

© 2019 Elsevier Ltd. All rights reserved.

1. Introduction

Photochemical smog is a severe problem that has gain attention of the scientific community in the last years. It is compounded by several gases and particles that have complex interactions. It becomes especially dangerous in highly populated and warm cities.

* Corresponding author.
E-mail address: f12carcr@uco.es (R. Carmona-Cabezas).

One of the most recently studied gases is the tropospheric ozone due to its abundance, which makes it one of the main photochemical oxidants. It is a secondary pollutant, which in high concentrations, can affect human health and crops harshly (Doherty et al., 2009), as well as having a great impact on economy (Miao et al., 2017). It has been demonstrated that it does not only affect big cities, but also rural areas (Domínguez-López et al., 2014). Another interesting component of the photochemical smog is the nitrogen dioxide, which is a precursor for the mentioned ozone. It is a primary pollutant derived directly from the anthropogenic activity that arises in urban areas. It also has serious impacts on human health (Yue et al., 2018).

The World Health Organization (2005) stablished references for this kind of pollutants, in order to warn against dangerous effects. For instance, the level at which the ozone concentration is considered to be hazardous is $120 \, \mu g/m^3$. For that reason, the identification of singularly high episodes of pollutant concentration gains importance.

Multifractal analysis has been previously used both for global behavior of the system and singularity detection in relation to pollutant dynamics (Pavón-Domínguez et al., 2015). In order to identify singularities in a general signal, one of the most commonly used techniques is the so-called pointwise Hölder exponents (Shang et al., 2006). It gives an estimation of how singular a given point is within a series, although its implementation has several disadvantages, as the dependence on parameters chosen by the user and numerical instability. Another one called the singularity exponent (Dai et al., 2014) will be used as well, in order to support the results from multifractal analysis.

In the last decade, a new approach designed to analyze time series was introduced by (Lacasa et al., 2008) and named Visibility Graph (VG). It is based on the transformation of those signals into a completely different mathematical object: a complex network. For the description of such new entities, the centrality parameters are very useful, as will be shown later in the text. Among the advantages of VGs, authors would like to remark the following: i) The characteristics of the original time series are inherited by the resulting network, leading to the possibility of describing the system through it. ii) They allow the analysis of various variables simultaneously, which can be used to find correlations.

In the introduced work, the VGs of three different time series related to the photochemical smog are computed. From them, the centrality measurements have been obtained and compared to multifractal indicators that are known to be useful for singularity detection. Finally, the searched purpose is to discern whether the complex network indicators can be used for the same applications that these multifractal parameters, giving equivalent results and overcoming some of their disadvantages.

2. Materials and methods

2.1. Data

This manuscript has employed real data from a 1 hourly ozone and nitrogen dioxide concentration (chemical factors from photochemical smog) and temperature (physical factor) time series, all measured in 2017. They were collected at the urban station in San Fernando (36°27′ N, 6°12′ W), in the province of Cádiz belonging to the southern part of the Iberian Peninsula. The reason behind choosing this location was that the area presents characteristics to be potentially vulnerable to the accumulation of photochemical smog (Domínguez-López et al., 2014). These are orographic (the Guadalquivir Valley), anthropic (two relevant industrial centers such as the chemical focus of Huelva and the Bay of Algeciras, and

four capitals) and weather conditions (high solar radiation and temperature). The cited station is part of the network in charge of controlling the air quality in the region of Andalusia, which is administered by the Consejería de Medioambiente (Regional Environmental Department) and co-financed by the European Union.

To work with uncorrelated data (in order to obtain independent results), authors have selected different periods of time for each time series. For the case of ozone, the amount of data corresponds to the month of July; for NO$_2$, January has been chosen and, finally, for temperature data, October has been picked. Apart from the uncorrelation of the data, the choice was motivated by several reasons: in the case of ozone, the month of July presents the most suitable (and stable) conditions for the creation of this pollutant. For the nitrogen dioxide, January is the month where the photochemical reaction activity is lower and therefore its concentration depends more on the sources. Finally, October was chosen for the last series in order to see singular episodes of this quantity in a region where most of the year the oceanic influence stabilizes the temperature. This physical factor is unstable in autumn by nature in this area, as discussed in previous works (Dueñas et al., 2002). All real time series data have been represented in Fig. 1.

2.2. Visibility Graphs

As mentioned before, the VG is introduced by (Lacasa et al., 2008) and is defined as a tool that makes possible to transform a time series into a graph, i.e. it converts a signal into a set of nodes connected through lines called edges.

To obtain the VG, which is associated to the time series, it is necessary to determine a criterion to establish which points (or nodes) are linked to each other, that is, have visibility. Let (t_a, y_a) and (t_b, y_b) be two arbitrary points from the time series which are chosen in order to check the mentioned criterion. One of the most commonly used is to consider that both have visibility (are connected in the graph) if any given point (t_c, y_c) that is situated between the first two $(t_a < t_c < t_b)$ satisfies the following:

$$y_c < y_a + (y_b - y_a) \frac{t_c - t_a}{t_b - t_a} \tag{1}$$

This visibility algorithm concludes by repeating the previous step for every pair of points in the signal. As an example, one can observe this procedure applied to a sample time series in Fig. 2.

In practice, it is more useful to obtain a matrix representation of the graph that contains the information of the complex network: the visibility adjacency matrix. It is a $N \times N$ binary matrix, with N the total number of nodes. Each element of the matrix a_{ij} takes the value of unit if nodes i and j have visibility; otherwise, it is null and this means that nodes are not linked to each other.

The algorithm can be simplified after some factors are taken into account, leading to a visibility adjacency matrix with a general form as follows:

$$A = \begin{pmatrix} 0 & 1 & \dots & a_{1,N} \\ 1 & 0 & 1 & \vdots \\ \vdots & 1 & \ddots & 1 \\ a_{N,1} & \dots & 1 & 0 \end{pmatrix} \tag{2}$$

Visibility Graphs are undirected networks, since the visibility criterion is reciprocal (one node has visibility with other and vice versa). However, there are some ways of mapping a directed network where the time order can be considered and this has been used previously for reversibility studies of time series (Lacasa et al., 2012; Xie et al., 2019). Nonetheless, when selecting a direction in order to account for the time order, there is a problem that arises
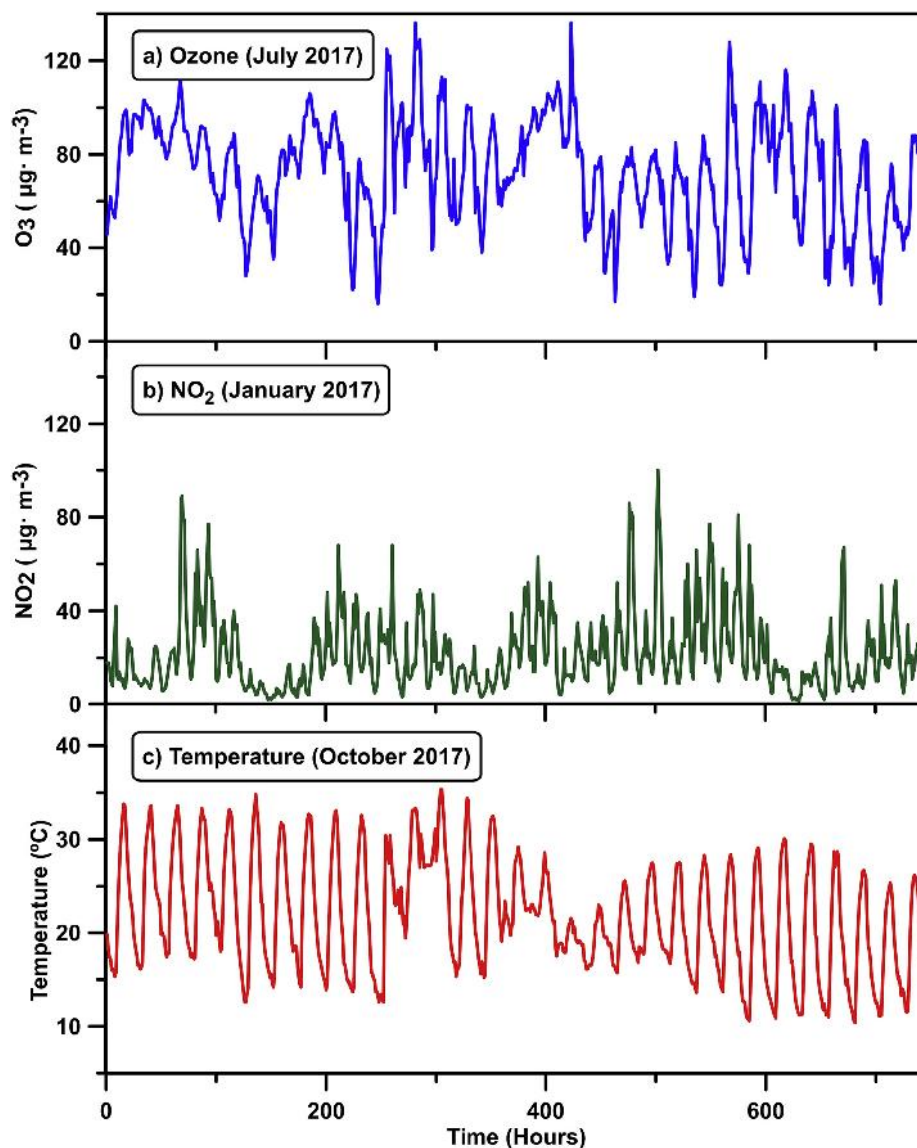
Fig. 1. Ozone, nitrogen dioxide and temperature time series for each selected month.
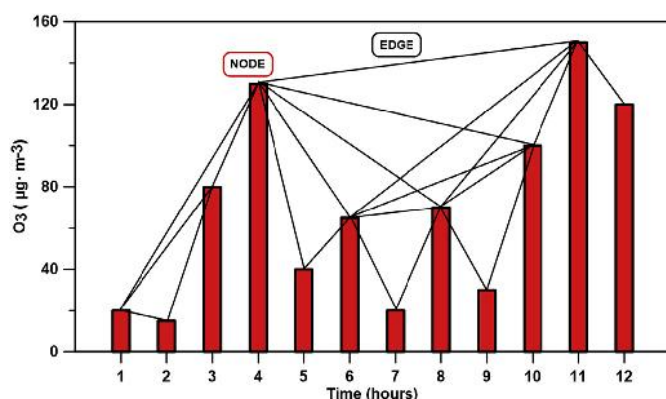

Fig. 2. Visibility Graph obtained from a sample time series by the visibility algorithm. The nodes of the graph are the data points (red bars), while the links among them are illustrated as solid lines. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

regarding the mapped complex network. Since the size of the series is by definition finite, the (ingoing or outgoing) degree depends on the position of the point with regard to the end and beginning of the series. Ingoing degree refers to the number of links that enter into a node, while the opposite is for the outgoing degree. This means that the first points in the series will have more outgoing degree than the last ones, and vice versa. Due to this artifact, it not suitable for being used when a pointwise description (for singularity detection, for instance) is desired.

2.3. Complex networks indicators

Once the new complex network is retrieved, there are some parameters which can be studied to characterize its nodes importance, such as centrality measures. This concept was firstly used in the study of social networks to turn out to be introduced into other fields of knowledge (Agryzkov et al., 2019; Joyce et al., 2010; Liu et al., 2015). Some of these centrality measures that are used in this work will be further explained next.

2.3.1. Degree

The degree of a node (k_i) in a graph or complex network is defined as the number of other nodes with which it is linked ($k_i = \sum_j a_{ij}$). As authors exposed in section 2.2, in the context of the visibility algorithm, two points having visibility means that they are linked because they fulfill the given criterion (see Equation (1)). After computing the degree for each node, one can retrieve the probability or distribution for each result by using a histogram. This outcome is called degree distribution of the sample $P(k)$. There are some points with a singularly high degree, called *hubs*, that are of great importance in this distribution.

As it is shown in previous works, the degree distribution obtained from the VG can characterize the nature of the time series involved (Lacasa et al., 2008; Mali et al., 2018). For example, it is possible to make a distinction among fractal, random or periodic signals.

2.3.2. Closeness

In the previous sections, centrality parameters have been defined by considering the number of edges and the adjacency matrix properties. Nevertheless, it is necessary to specify the meaning of another property within a graph in order to carry on with next definitions, which is the so-called shortest path (SP). In a network, one can observe a different number of edges (as a measurement of length) passing through any (in general, distant) pair of nodes. Two distant nodes (i, j) will have different number of edges and paths between them, but there will be some of these paths where the number of edges will be minimum; this quantity is named as the SP.

If one takes all pairs of nodes, it is possible to obtain a matrix, the so-called distance matrix D, where each element $d_{i,j}$ contains the SP from node i to j. One usually sets diagonal elements as zero. For an undirected graph, this matrix will be symmetric, as in the adjacency matrix case (see Section 2.2).

After the explanation of this graph property, it is possible to define the closeness centrality of a node i as the inverse of the sum of distances from this node to the others:

$$c_i = \frac{1}{\sum_{j=1}^{N} d_{i,j}} \tag{3}$$

where $d_{i,j}$ is the element (i, j) from the corresponding distance matrix of the graph.

2.3.3. Betweenness

The main idea behind betweenness parameter is to focus on the centrality as a measurement of how a node is between many others. That is, how much a node is passed through by shortest paths of other pairs of nodes. Therefore, the equation that defines this quantity for a node i is the following:

$$b_i = \sum_{\substack{j=1 \\ j \neq i}}^{N} \sum_{\substack{k=1 \\ k \neq i,j}}^{N} \frac{n_{jk}(i)}{n_{jk}} \tag{4}$$

where n_{jk} is the number of SP's from j to k (notice that these paths can be degenerated), whereas $n_{jk}(i)$ is the number of SP's that contains the node i.

2.4. Multifractal indicators

2.4.1. Pointwise Hölder exponent method

The Hölder exponent is a measure used in the context of multifractal analysis to characterize the singularities which are present in a signal. A function or a time series is fractal when it exhibits some local properties as self-similarity, irregularity, fine structure and fractional dimension; if they are variable at different points, then this function is multifractal. Consequently, multifractal analysis describes the singularities implicated in a time series.

A commonly used method for multifractal analysis of a signal is to compute the pointwise Hölder exponent. This parameter is defined as a local characteristic of a function which is computed at every point of its domain. It refers to the decay rate of the amplitude of the function fluctuations in the neighborhood of the data point when the size of the neighborhood shrinks to zero, that is, the function singularities. The Hölder exponent at a point t of $f(t)$ can be expressed as:

$$\alpha_f(t) = \liminf_{h \to 0} \frac{\log|f(t+h) - f(t)|}{\log|h|} \tag{5}$$

In practice, one can only obtain discrete time series and so some different methods for computing the Hölder exponent has been elaborated to solve this problem (Peng-Jian and Jin-Sheng, 2007; Shang et al., 2006). developed an algorithm for numerical evaluation of Hölder exponent based on the previous equation. This method takes $n + 1$ points from a signal equally spaced, $\{y_0, y_1, \ldots, y_n\}$ and calculates the intensity of its Hölder exponent at a specific point y_i. This computation considers a number of preceding and following points, s, which is named as the window width and is set by the user (in total, $2s$ values around the point are taken). Each value in the window has got a different weight controlled by other parameter, λ, known as regression coefficient and whose values are in the interval $[0, 1]$. This last parameter reduces the weight for each point in the window as they are further from the center. Their influence over the ultimate computation of the Hölder exponent must be greater for close points than those which are more distant.

Once values for λ and s are selected, it is necessary to obtain the next quantity for each integer $k \neq 0$ from $-s$ to s in each window:

$$R_{i,k} = \frac{\log\left(\frac{y_{i+k} - y_i}{C_1}\right)}{\log\left(\frac{|k|}{C_2}\right)} \tag{6}$$

where C_1 and C_2 are the normalizing parameters which are chosen for the convenience of computation (Peng-Jian and Jin-Sheng, 2007) and k must fulfill $0 \leq i + k \leq n$. For points which are at the very beginning or end of the signal, where the k index falls out of the domain, the window width s is properly shrunk. Next, for each integer j, $0 \leq j \leq s$, it must be computed:

$$h_{i,j} = \min\{R_{i,k} : |k| \leq j\} \tag{7}$$

This last equation is related to take the lim inf as in Equation (5) for the case of a continuous function. Finally, the local Hölder exponent, α_i, is retrieved by the computation of the weighted average of the approximations $h_{i,j}$:

$$\alpha_i = \frac{1 - \lambda}{1 - \lambda^s}\left(h_{i,1} + \lambda h_{i,2} + \lambda^2 h_{i,3} + \ldots + \lambda^{s-1} h_{i,s}\right) \tag{8}$$

where the most important factors are those which are close enough to ith point.

2.4.2. The singularity exponent method

In an effort to get better results for the singularities of traffic data from a highway (Dai et al., 2014), proposed another approach

to analyze this property through the "singularity exponents", as they named in the paper.

The algorithm takes a given time series data of length n, $\{y_1, y_2, ..., y_n\}$, uniformly spaced, and requires choosing two window widths r_i ($i = 1, 2$). Next, one must compute the following quantities:

$$\overline{y} = \frac{1}{n} \sum_{k=1}^{n} y_k \tag{9}$$

$$\overline{y_{r_i}(t)} = \frac{1}{2r_i + 1} \sum_{k=-r_i}^{r_i} y_{t+k} \tag{10}$$

$$|\Delta y(t)|_{r_i} = |\overline{y_{r_i}(t)} - \overline{y}| \tag{11}$$

where $\overline{y}$ is the average of the whole signal, whereas $\overline{y_{r_i}(t)}$ and $|\Delta y(t)|_{r_i}$ denotes the average volume in the field and the absolute errors between $\overline{y}$ and $\overline{y_{r_i}(t)}$ in $[-r_i, r_i]$, respectively. It must be noticed that Equation (10) must satisfy $0 \leq (t + k) \leq n$.

Finally, the singularity exponents at each point $\beta(t)$ are obtained by computing the fluctuations in each of the previously defined scales:

$$\varepsilon_{r_1}(t) = \frac{1}{2r_1} \sum_{t'=t-r_1}^{t'=t+r_1} |\Delta y(t')|_{r_1} = c(t) r_1^{\beta(t)} \tag{12}$$

$$\varepsilon_{r_2}(t) = \frac{1}{2r_2} \sum_{t'=t-r_2}^{t'=t+r_2} |\Delta y(t')|_{r_2} = c'(t) r_2^{\beta(t)} \tag{13}$$

Assuming that $c(t) \approx c'(t)$, then, one can get:

$$\beta(t) \approx \frac{\ln \varepsilon_{r_1}(t) - \ln \varepsilon_{r_2}(t)}{\ln r_1 - \ln r_2} \tag{14}$$

Finally, a usual way of checking the Hölder exponent method is to take the generalized Weierstrass function as a test function. In this work, authors have proven the implemented algorithm of both multifractal indicators. The selected Weierstrass function is the following:

$$f(t) = \sum_{k=0}^{\infty} 3^{-ks(t)} \sin\left(3^k t\right) \tag{15}$$

where $s(t)$ is the seed function, whose values are contained in the interval $[0, 1]$. As (Daoudi et al., 1998) shows, $s(t) = \alpha_f(t)$ for all t and so, one can compare theorical values of the Hölder exponent (given by $s(t)$) with the numerical results.

This analysis can be observed in Fig. 3 for both algorithms, where the chosen seed function is $s(t) = |\sin(10 \sin(\pi x))|$. As expected, the Hölder exponent method fits well the theorical value, while the singularity exponent method does it as well. However, in the last case, shown in Fig. 3c), the level of singularity of the data can be understood as the deviation from 0, which is independent of the sign. For this reason, authors have decided to plot the absolute value of the results.

3. Results and discussion

In this section, the parameters previously introduced in the last part are tested on different real time series of environmental variables related to photochemical smog (see Section 2.1). As it was
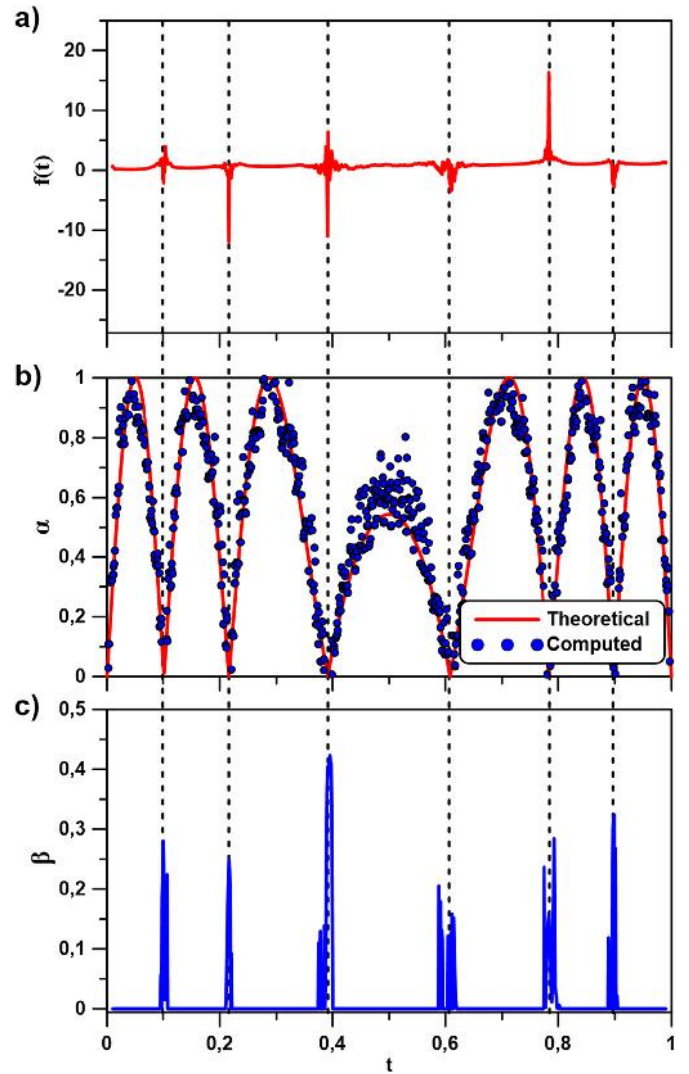

Fig. 3. Time series obtained from the Weierstrass function for an interval $t \in [0, 1]$ (a) and the Hölder and singularity exponents α and β (b and c, respectively).

mentioned before, these three time series have been chosen based on the fact that they are different in nature, in order to test the proposed indicators for distinct scenarios. All the complex network indicators shown in the plots are normalized to the maximum value of each one of them.

In order to evaluate these quantities, the widely used Hölder exponents have been obtained following the approach described in the former section. The parameters set for the calculation have been in all the cases $\lambda = 0.8$, $C_1 = 70$, $C_2 = 10^5$ and $s = 10$. Moreover, the other exponent proposed in order to overcome some of the shortcomings of the Hölder exponents (Dai et al., 2014) has been analyzed (the singularity exponent). For this indicator, the chosen parameters were $r_1 = 1$ and $r_2 = 7$.

In Fig. 4, all these indicators are shown for the case of the ozone concentration time series. For the other series, the results are similar and therefore the same considerations are taken into account. It must be pointed out that the complex network ones are univocal for a given time series, since the VG constructed does not depend on any numerical parameter. On the other hand, the multifractal indicators depend substantially on the chosen parameters when running the algorithm (see Sections 2.4.1 and 2.4.2).

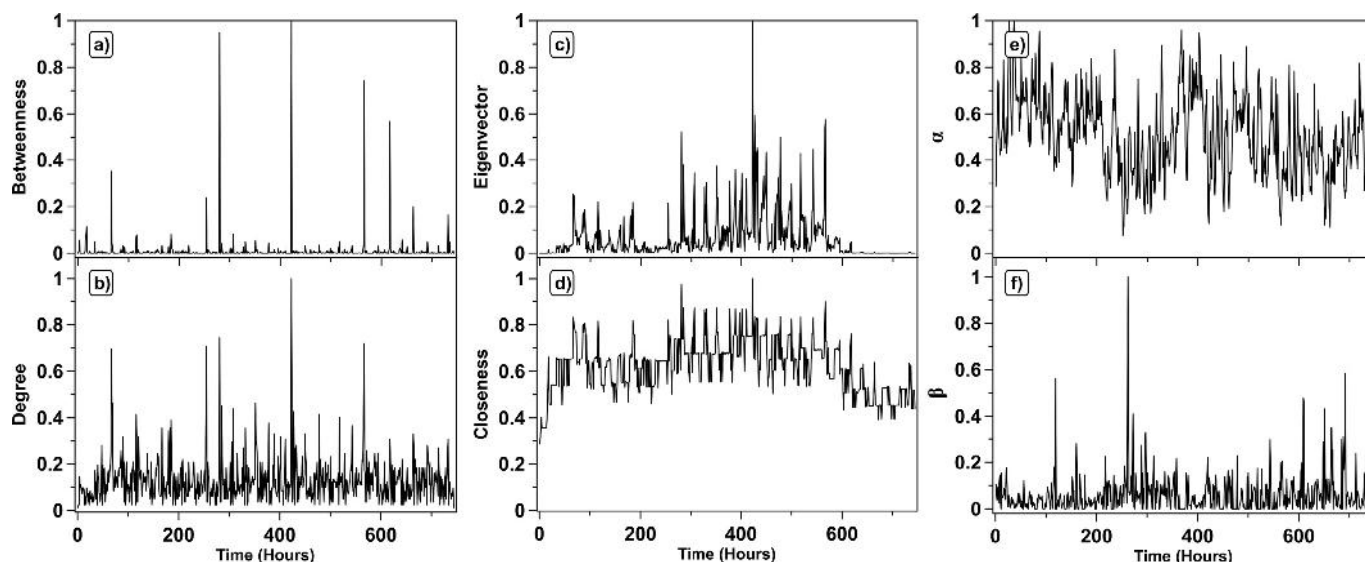In order to make Fig. 4 more intelligible, the followed process is

Fig. 4. Different indicators computed from the whole ozone concentration time series: (a−d) Centrality parameters from the Visibility Graph and (e−f) Hölder and singularity exponents.

explained next. What has been done in practice is computing firstly the betweenness centrality of the total time series. Afterwards, a dynamic criterion has been stablished in order to select the most important central nodes from this quantity. It consisted on searching the relative maxima which are above a given percentage of the absolute one. After several tests, it has been found that the identification of singularities holds down to a 5%, which could be used as a rule of thumb for future works. Nonetheless, with this value, a considerable amount of peaks are chosen and for practical reasons, from this point only the five most central nodes will be shown in the figures. These points correspond to the five main *skyline hubs* (Carmona-Cabezas et al., 2019b) in the series (i.e. the nodes with the highest singular values of betweenness). The reason for choosing this term was the similarity to the skyline drawn by the skyscrapers in a city from the point of view of other nodes.

Once those nodes are selected, authors have investigated the values of the rest of the indicators around them. The reason for choosing the betweenness centrality as a first reference is the more distinguished and smoother results that it provides for some specific important nodes from the VG, as depicted in Fig. 4a). Furthermore, the multifractal measurements are built upon parameters chosen by the user for the convenience of the computation. This might lead to problems and ambiguities in the outcomes.

Firstly, the results obtained for the ozone concentration time series can be regarded in Fig. 5, where five different betweenness peaks (*skyline hubs*) from b) were closely studied (numbers 1−5). The different complex network indicators have positive pronounced peaks in the same temporal points of the series. When it comes to the Hölder and singularity exponents from multifractal analysis, they present minima and maxima values respectively at those same points as well.

In order to understand the relation of this parameter to the photochemical pollution (ozone in this case), authors would like to point out a recent study (Carmona-Cabezas et al., 2019b). In that work, betweenness centrality peaks have been related with the values of the series that store most of the information about the upper envelope. Considering this, a skyline hub can be regarded as a singular episode of ozone concentration in the sense that it indicates a change of the tendency with respect to the previous maxima. For instance, if the maximal concentration of ozone of

several days is steadily increasing and then starts to decrease, that critical point is identified by a betweenness peak.

In Fig. 5c1), the peaks selected by the authors for their higher betweenness are magnified and superposed in the same plot. This has been done in order to compare with the rest of parameters. As it could be seen as well in Fig. 4, betweenness centrality gives the clearest signal and therefore its peaks are sharper.

The next indicator (see Fig. 5c2) from the VG is the degree of the nodes, which has been widely used in many studies of this kind (Carmona-Cabezas et al., 2019a; Pierini et al., 2012; Zhou et al., 2017). Before studying its usefulness for ozone description, it is easily regarded at a glance that all the positions selected correspond as well to very pronounced peaks in the degree (*hubs*). It is known that a high degree implies a high concentration of ozone (Carmona-Cabezas et al., 2019b), while the opposite is not always true. Therefore, when a point is identified as a singularity from the degree, it means that the ozone concentration at that particular time is especially high in magnitude within a time interval around it. It suggests that the conditions for its production would be optimal at that time. The correspondence with the betweenness peaks is due to the fact that when ozone concentration reaches a singular maximum before a downwards trend, this peak magnitude will be as well singularly high, locally.

The last of the VG indicators is the closeness centrality. This quantity has been used before to describe VGs with interesting results, although in theoretical point of view (Bianchi et al., 2017; Iacovacci and Lacasa, 2019). Closeness centrality of a given time point is related to the values at its left and right (Donner and Donges, 2012). Thus, authors attribute a singular peak in the closeness to a rarely high episode surrounded by a concave up tendency (since it favors connectivity) in the surrounding concentrations of this pollutant. As depicted in Fig. 5c3), the betweenness peaks coincide as well with singular high values of closeness, in every one of the selected points. Nevertheless, the obtained curves are rather noisy (compared to the degree and betweenness), although the peaks can be inferred without problem.

Finally, the multifractal indicators are shown in Fig. 5c4) and c5). On the one hand, the Hölder exponent plot displays minimal values around the positions used as reference. It was expected, since highly irregular points exhibit Hölder exponents closer to zero,
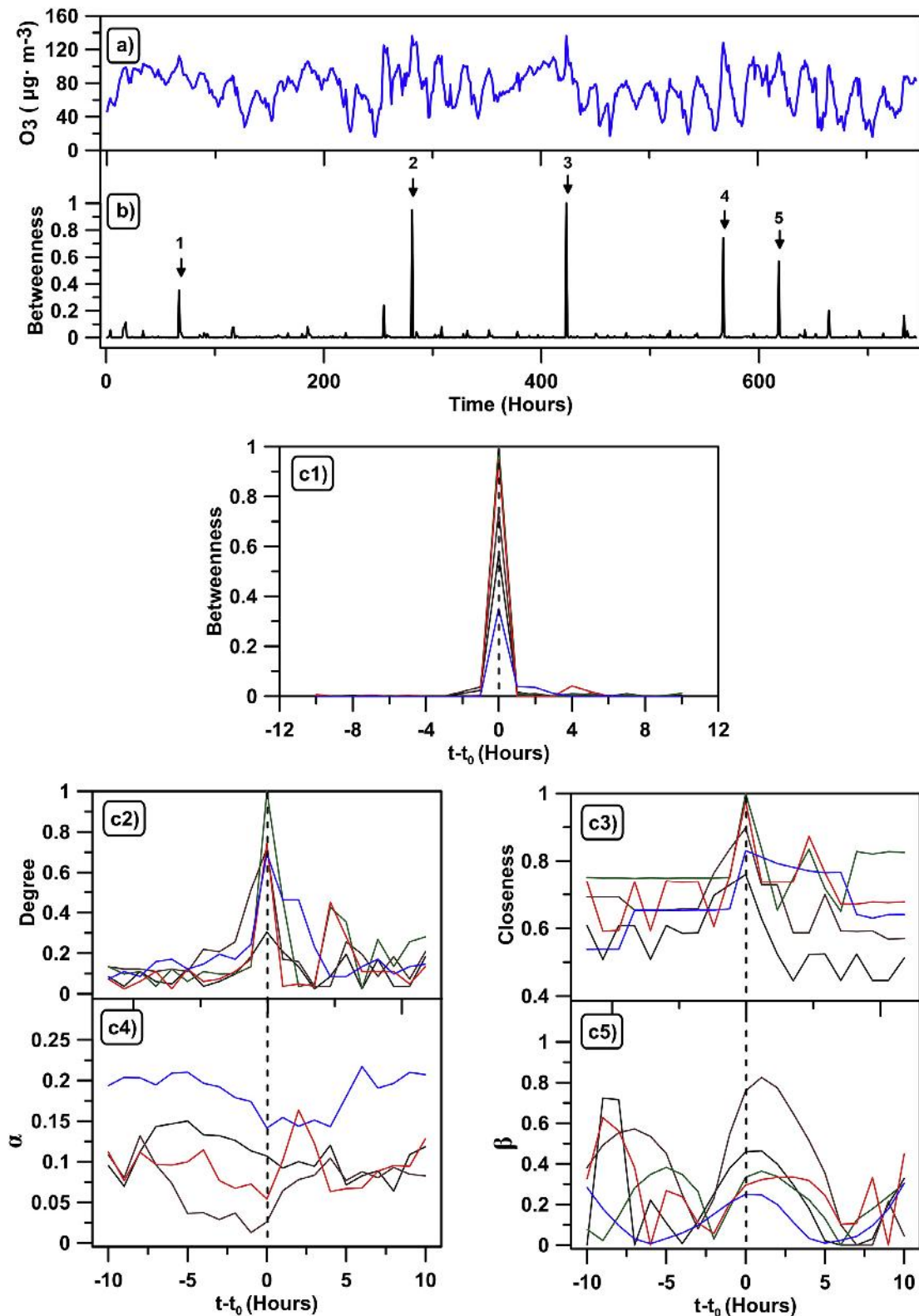
Fig. 5. Ozone concentration time series (a) with the betweenness values computed from it below (b). Plots from c1) to c3) show the complex network indicators: betweenness, degree, closeness (in appearance order). c4) and c5) display the Hölder and singularity exponents (respectively).

whereas smoother regions present greater values (Jaffard, 1997; Safonov et al., 2002). On the other hand, the singularity exponents, as explained before in Section 2.4.2, show maximal values on singular points. That is as well corroborated in the reference peaks

positions, in a clearer way than the Hölder exponents in this case. In any case, neither of the multifractal indicators show peaks as acute as the complex network ones. It should be pointed out as well the noise embedded within the singularity exponent curves, that

makes more difficult to identify the searched maxima for some of the peaks (for instance, peak 1 and 2).

When it comes to the underlying tropospheric ozone pollution concentration, all of these singularities correspond to unusual episodes of especially steep accumulation of that gas. More precisely, the selected peaks coincide with 3rd, 12th, 18th, 24th and 26th of July, all of them located between 2 p.m. and 6 p.m. (GMT+2), which is the time period when the radiation and
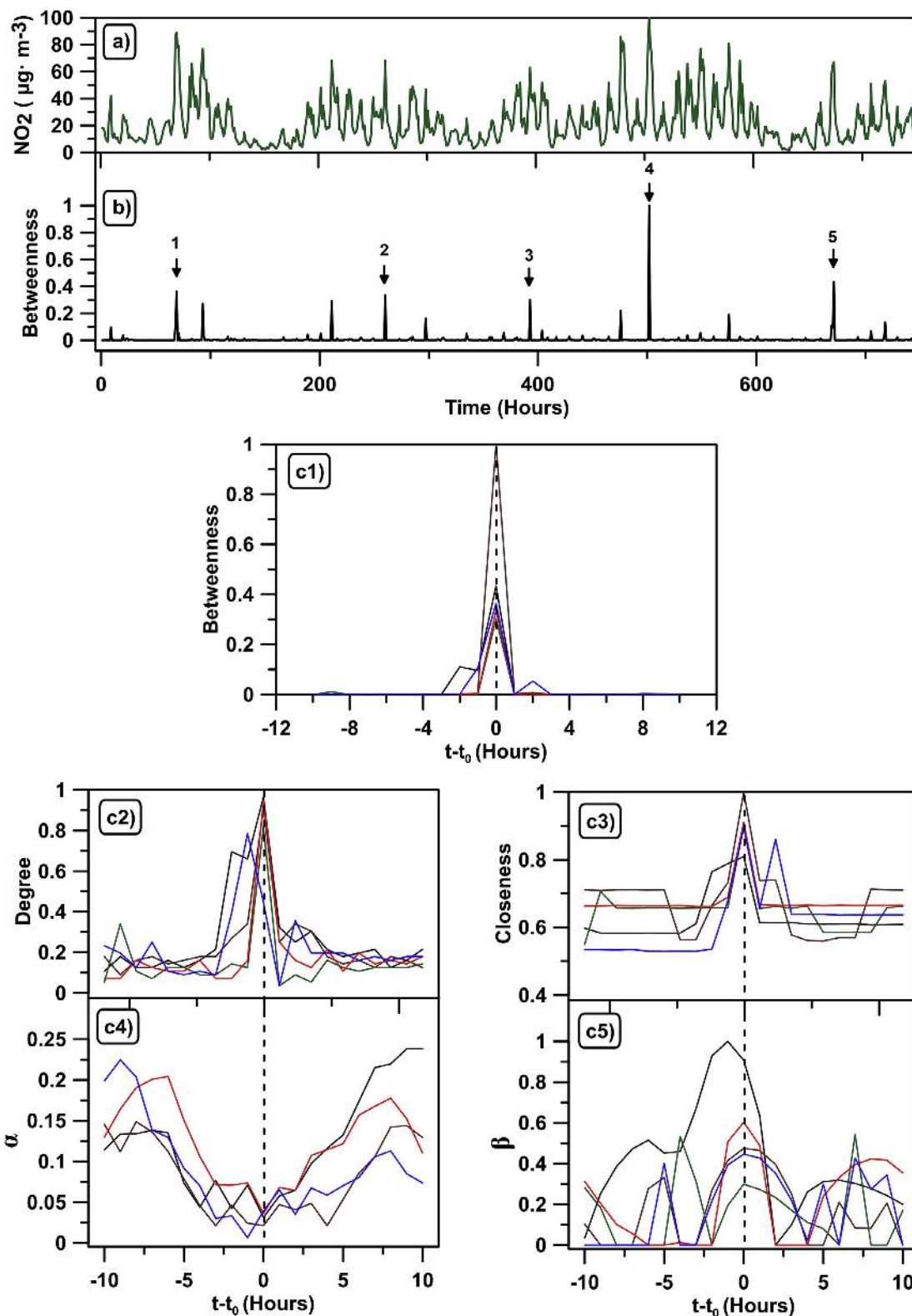

Fig. 6. NO_2 concentration time series (a) with the betweenness values computed from it below (b). Plots from c1) to c3) show the complex network indicators: betweenness, degree, closeness (in appearance order). c4) and c5) display the Hölder and singularity exponents (respectively).

temperature reach their maxima. I every case, the corresponding concentrations are above or reaching the dangerous threshold stablished by the World Health Organization. What is significant about the selected ones is that, as can be regarded in Fig. 5a), they

are placed before a change in the tendency of the previous maxima or daily concentrations of ozone. It corresponds to what was previously discussed for betweenness, as the peaks are selected by looking at this quantity. The utility of this is that once the highly
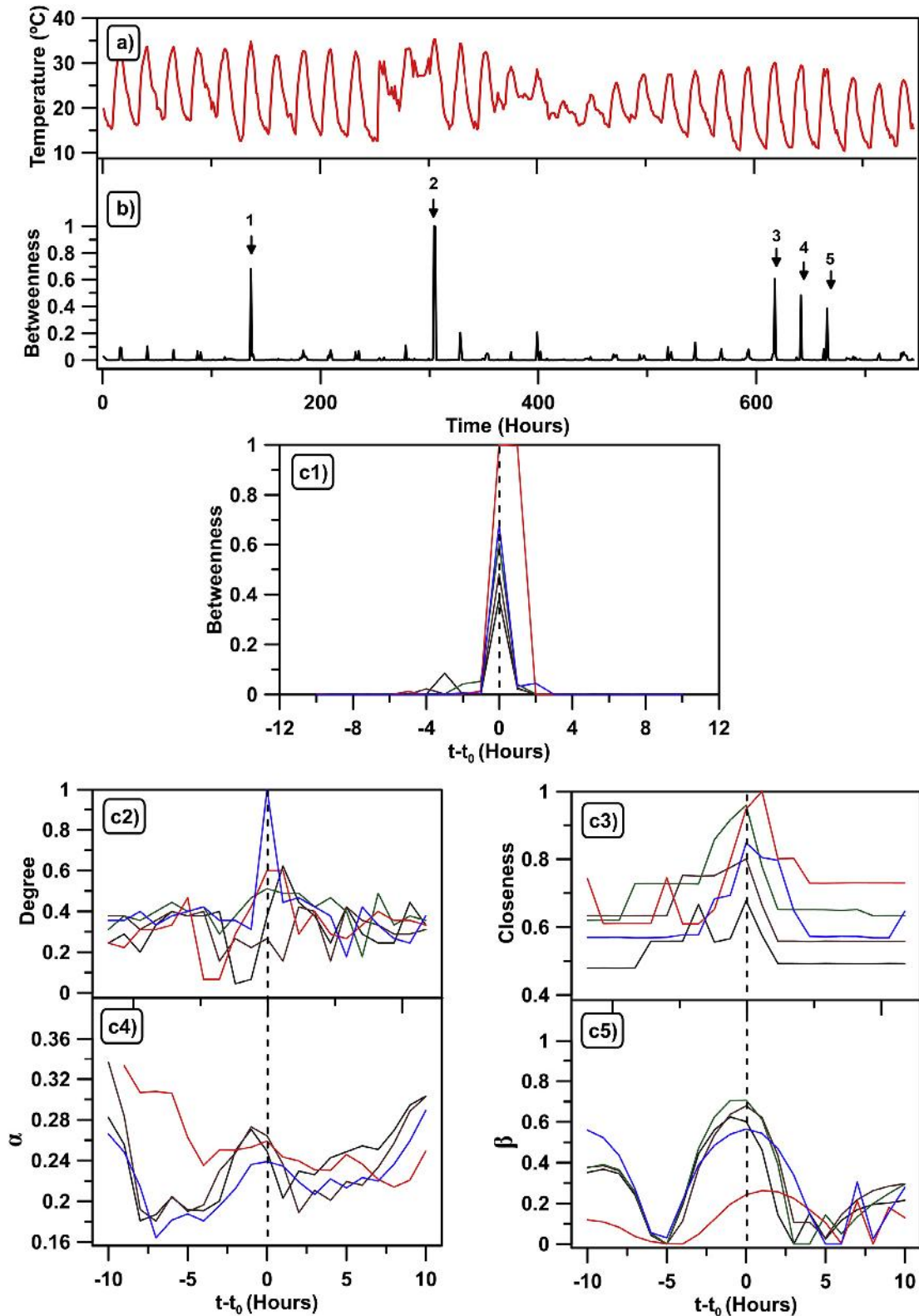

Fig. 7. Temperature time series (a) with the betweenness values computed from it below (b). Plots from c1) to c3) show the complex network indicators: betweenness, degree, closeness (in appearance order). c4) and c5) display the Hölder and singularity exponents (respectively).

irregular O_3 concentration that diverge abruptly from the trend are detected, it would be possible to perform a deeper study in order to find the origin of it and act accordingly to prevent future similar episodes.

In Fig. 6, the same parameters as in the previous case are studied for the NO_2 concentration time series. The same procedure was followed in order to obtain the plots. It can be regarded that equivalent behaviors are present here: all peaks from the complex network indicators identify quite well the singularities found in betweenness; whereas the multifractal ones (which coincide as well) have wider shapes, with less accuracy. Nonetheless, the Hölder exponent minima (see Fig. 6c4) are much clearer than above. The curves corresponding to the singularity exponent exhibit again a considerable level of noise.

As in the previous case, looking at the physical meaning of the series, the peaks where authors focus accord with unusual maxima of NO_2 throughout the month. In detail, the days corresponding to these pollutant concentration singularities are 3rd, 11th, 17th, 21st and 28th of January. In this case, the singular hours are not as consistent as in the ozone. Authors attribute this irregularity to the fact that the main source for NO_2 is human activity, which in many cases might vary. Hence this could be used to identify singular acute activity from industry or traffic, as well as meteorological unexpected events, that could lead to unanticipated concentrations derived from transport of this pollutant.

Finally, a different type of time series has been analyzed through all the indicators shown up to this point. In this case, this series corresponds to the hourly average temperature measured, which as expected, displays a more regular behavior (see Fig. 7a). The complex networks parameters identify in a similar way the same singular points. Moving to the multifractal parameters, the singularity exponent behaves as in the previous series (see Fig. 7c5), in contrast to the Hölder exponent, that shows maxima instead of the expected minima (see Fig. 7c4). Authors attribute this anomaly to some of the disadvantages of the computation of this parameter by the algorithm, which under some circumstances may provide non-finite or misleading values.

As mentioned for the previous variables, the actual meaning of these peaks resides on unexpected values of high temperature that occur on the 6th, 13th, 26th, 27th and 28th of October. Now, the singularities encountered on temperature might be associated to unpredicted meteorological events.

4. Conclusion

After discussing the proposed indicators, authors consider that these complex network parameters can be properly used to identify relevant points in environmental time series, such as the ones analyzed here. It has been demonstrated how indicators that are different in nature, can obtain complementary results that can be employed to characterize the behavior of experimental signals from pollutants and temperature in the context of photochemical smog. They have been compared to widely known singularity indicators from multifractal analysis, showing some advantages as well. This opens a bridge between complex networks and multifractal studies for local singular behavior of time series.

Finally, it can be argued that some of the shortcomings of the Hölder and singularity exponents are overcome with the proposed methodology. On the one hand, the multifractal indicators depend on a number of parameters that rely on the nature of the series, giving different ambiguous results. By contrast, complex network ones are univocal for a time series. No matter how one runs the algorithm, the result would be the same. On the other hand, the way Hölder exponents are defined gives complications for certain

cases derived from the logarithm in its expression. Conversely, the centrality parameters do not have such problems, since their computation is based on simple arithmetic calculations from graph theory. Also, based on the properties of each parameter, they seem to be able to describe different properties of the pollutant dynamics at the selected times. For instance, betweenness is found to be related to a change in the tendency of the upper envelope of the signal; degree identifies singularly high concentrations or temperature episodes; while closeness characterizes the behavior of the concentrations surrounding the detected singularity. Taking all these facts into account, it is possible to consider the proposed indicators as a future additional information source.

For future works, it remains open a wide range of possible applications for these local studies. For instance, in the field of environmental analysis, authors would like to stress the possibility to employ these indicators to relate singular events of different variables, such as the ozone and some of its precursors (both chemical and physical). Also, based on the use of the Hölder exponent for predictive purposes (Shang et al., 2006), it might be investigated the usefulness of VGs for the same aim.

Declaration of competing interest

None.

Acknowledgements

References

Agryzkov, T., Tortosa, L., Vicent, J.F., 2019. A variant of the current flow betweenness centrality and its application in urban networks. Appl. Math. Comput. 347, 600–615. https://doi.org/10.1016/j.amc.2018.11.032.

Bianchi, F.M., Livi, Lorenzo, Alippi, Cesare, Jenssen, Robert, 2017. Multiplex visibility graphs to investigate recurrent neural network dynamics. Sci. Rep. 13 https://doi.org/10.1038/srep44037.

Carmona-Cabezas, R., Ariza-Villaverde, A.B., Gutiérrez de Ravé, E., Jiménez-Hornero, F.J., 2019a. Visibility graphs of ground-level ozone time series: a multifractal analysis. Sci. Total Environ. 661, 138–147. https://doi.org/10.1016/j.scitotenv.2019.01.147.

Carmona-Cabezas, R., Gómez-Gómez, J., Ariza-Villaverde, A.B., Gutiérrez de Ravé, E., Jiménez-Hornero, F.J., 2019b. Can complex networks describe the urban and rural tropospheric O3 dynamics? Chemosphere 230, 59–66. https://doi.org/10.1016/j.chemosphere.2019.05.057.

Dai, M., Zhang, C., Zhang, D., 2014. Multifractal and singularity analysis of highway volume data. Phys. A Stat. Mech. Appl. 407, 332–340. https://doi.org/10.1016/j.physa.2014.04.005.

Daoudi, K., Lévy Véhel, J., Meyer, Y., 1998. Construction of continuous functions with prescribed local regularity. Constr. Approx. 14, 349–385. https://doi.org/10.1007/s003659900078.

Doherty, R.M., Heal, M.R., Wilkinson, P., Pattenden, S., Vieno, M., Armstrong, B., Atkinson, R., Chalabi, Z., Kovats, S., Milojevic, A., Stevenson, D.S., 2009. Current and future climate- and air pollution-mediated impacts on human health. Environ. Health 8, S8. https://doi.org/10.1186/1476-069X-8-S1-S8.

Domínguez-López, D., Adame, J.A., Hernández-Ceballos, M.A., Vaca, F., De la Morena, B.A., Bolívar, J.P., 2014. Spatial and temporal variation of surface ozone, NO and NO2 at urban, suburban, rural and industrial sites in the southwest of the Iberian Peninsula. Environ. Monit. Assess. 186, 5337–5351. https://doi.org/10.1007/s10661-014-3783-9.

Donner, R.V., Donges, J.F., 2012. Visibility graph analysis of geophysical time series: potentials and possible pitfalls. Acta Geophys. 60, 589–623. https://doi.org/10.2478/s11600-012-0032-x.

Dueñas, C., Fernández, M.C., Cañete, S., Carretero, J., Liger, E., 2002. Assessment of ozone variations and meteorological effects in an urban area in the Mediterranean Coast. Sci. Total Environ. 299, 97–113. https://doi.org/10.1016/S0048-9697(02)00251-6.

Iacovacci, J., Lacasa, L., 2019. Visibility graphs for image processing. IEEE Trans. Pattern Anal. Mach. Intell. https://doi.org/10.1109/TPAMI.2019.2891742, 1–1.

Jaffard, S., 1997. Multifractal formalism for functions Part I: results valid for all functions. SIAM J. Math. Anal. 944–970. https://doi.org/10.1137/

S0036141095282991.

Joyce, K.E., Laurienti, P.J., Burdette, J.H., Hayasaka, S., 2010. A new measure of centrality for brain networks. PLoS One 5. https://doi.org/10.1371/journal.pone.0012200.

Lacasa, L., Luque, B., Ballesteros, F., Luque, J., Nuño, J.C., 2008. From time series to complex networks: the visibility graph. Proc. Natl. Acad. Sci. 105, 4972–4975. https://doi.org/10.1073/pnas.0709247105.

Lacasa, L., Nuñez, A., Roldán, É., Parrondo, J.M.R., Luque, B., 2012. Time series irreversibility: a visibility graph approach. Eur. Phys. J. B 85. https://doi.org/10.1140/epjb/e2012-20809-8.

Liu, C., Zhan, X.-X., Zhang, Z.-K., Sun, G.-Q., Hui, P.M., 2015. Events determine spreading patterns: information transmission via internal and external influences on social networks. New J. Phys. 17, 113045. https://doi.org/10.1088/1367-2630/17/11/113045.

Mali, P., Manna, S.K., Mukhopadhyay, A., Haldar, P.K., Singh, G., 2018. Multifractal analysis of multiparticle emission data in the framework of visibility graph and sandbox algorithm. Physica A 493, 253–266. https://doi.org/10.1016/j.physa.2017.10.015.

Miao, W., Huang, X., Song, Y., 2017. An economic assessment of the health effects and crop yield losses caused by air pollution in mainland China. J. Environ. Sci. 56, 102–113. https://doi.org/10.1016/j.jes.2016.08.024.

Pavón-Domínguez, P., Jiménez-Hornero, F.J., Gutiérrez de Ravé, E., 2015. Joint multifractal analysis of the influence of temperature and nitrogen dioxide on tropospheric ozone. Stoch. Environ. Res. Risk Assess. 29, 1881–1889. https://doi.org/10.1007/s00477-014-0973-5.

Peng-Jian, S., Jin-Sheng, S., 2007. Multi-fractal analysis of highway traffic data. Chin. Phys. 16, 365–373. https://doi.org/10.1088/1009-1963/16/2/016.

Pierini, J.O., Lovallo, M., Telesca, L., 2012. Visibility graph analysis of wind speed records measured in central Argentina. Phys. A Stat. Mech. Appl. 391, 5041–5048. https://doi.org/10.1016/j.physa.2012.05.049.

Safonov, L.A., Tomer, E., Strygin, V.V., Ashkenazy, Y., Havlin, S., 2002. Delay-induced chaos with multifractal attractor in a traffic flow model. Europhys. Lett. 57, 8. https://doi.org/10.1209/epl/i2002-00555-0.

Shang, P., Lu, Y., Kama, S., 2006. The application of Hölder exponent to traffic congestion warning. Phys. A Stat. Mech. Appl. 370, 769–776. https://doi.org/10.1016/j.physa.2006.02.032.

World Health Organization, 2005. WHO Air Quality Guidelines for Particulate Matter, Ozone, Nitrogen Dioxide and Sulfur Dioxide.

Xie, F., Nian, D., Fu, Z., 2019. Differential temporal asymmetry among different temperature variables' daily fluctuations. Clim. Dyn. 53, 585–600. https://doi.org/10.1007/s00382-018-04603-1.

Yue, H., Yan, W., Ji, X., Zhang, Y., Li, G., Sang, N., 2018. Maternal exposure to NO2 enhances airway sensitivity to allergens in BALB/c mice through the JAK-STAT6 pathway. Chemosphere 200, 455–463. https://doi.org/10.1016/j.chemosphere.2018.02.116.

Zhou, C., Ding, L., Skibniewski, M.J., Luo, H., Jiang, S., 2017. Characterizing time series of near-miss accidents in metro construction via complex network theory. Saf. Sci. 98, 145–158. https://doi.org/10.1016/j.ssci.2017.06.012.

Publication 6
Multiplex Visibility Graphs as a Complementary Tool for Describing the Relation Between Ground Level O_3 and NO_2

Journal

Atmospheric Pollution Research (Ed. Elsevier)

Position / Quartile

(80/265) / Q2 (Environmental Sciences)

Impact factor

3.527 (2019)

Multiplex Visibility Graphs as a complementary tool for describing the relation between ground level O$_3$ and No$_2$

Rafael Carmona-Cabezas[*], Javier Gómez-Gómez, Ana B. Ariza-Villaverde,
Eduardo Gutiérrez de Ravé, Francisco J. Jiménez-Hornero

*Complex Geometry, Patterns and Scaling in Natural and Human Phenomena (GEPENA) Research Group, University of Cordoba, Gregor Mendel Building (3rd Floor),
Campus Rabanales, 14071, Cordoba, Spain*

ARTICLE INFO

ABSTRACT

The usage of multilayer complex networks for the analysis of correlations among environmental variables (such as O_3 and NO_2 concentrations from the photochemical smog) is investigated in this work. The mentioned technique is called *Multiplex Visibility Graphs (MVG)*. By performing the joint analysis of those layers, the parameters named Average Edge Overlap and Interlayer Mutual Information are extracted, which accounts for the microscopical time coherence and the correlation between the time series behavior, respectively.

These parameters point to the possibility of using them independently to describe the correlation between atmospheric pollutants (which could be extended to environmental time series). More precisely the first one of them is considered to be a potential new approach to determine the time required for the correlation of NO_2 and O_3 to be observed, since it is obtained from the correlation of the pollutants at the smallest time scale. As for the second one, it has been checked that the proposed technique can be used to describe the variation of the correlation between the two gases along the seasons. In short, MVGs parameters are introduced and results show that they could be potentially used in a future for correlation studies, supplementing already existing techniques.

1. Introduction

In the last years, many studies have been conducted to give some light on formation and dynamics of ground-level ozone. The important of these analyses lie on the fact that it is one of the main photochemical oxidants (due to its abundance) and it can lead to serious damage for human health and harvest for high concentrations (Doherty et al., 2009). According to Miao (Miao et al., 2017), its impact can be quantified in losses of billions of dollars from the economical point of view.

The formation and destruction of this secondary pollutant is known to be governed by photochemical and nonlinear processes (Graedel and Crutzen, 1993; Trainer et al., 2000) that depend highly on meteorological conditions such as temperature, wind direction and mainly solar radiation (Trainer et al., 2000). In addition to all that, O$_3$ concentration also depends on the behavior of its chemical precursors. These are mainly nitrogen oxides (amongst them, NO$_2$ is studied here) and volatile organic compounds produced from the urban and industrial activity (Sillman, 1999). It is because of all those factors that the analysis of the temporal evolution of ozone is a very complex task. Hence, some questions remain open such as the influence of the

working time scale on these studies and the relevance of NO$_2$ as precursor depending on the season of the year.

A very recent method (called *Multiplex Visibility Graph*, MVG) to analyze nonlinear multivariate timeseries (Lacasa et al., 2015) is used in this work to see the relation between O$_3$ and its precursor NO$_2$ to answer the questions posed above. This method consists basically on turning the time series into complex networks and then forming multilayered structures that can be analyzed afterward, thanks to the last advances in this field. MVGs have shown to be useful for several applications already, from economics to neurology (Bianchi et al., 2017; Lacasa et al., 2015; Sannino et al., 2017). Those works have used last developments in multilayer networks (Bianconi, 2013; Boccaletti et al., 2014; De Domenico et al., 2013; Kivela et al., 2014) to obtain information to describe and compare the signals.

The reason to use this approach instead of other ways to construct functional networks is because those usually require performing a preprocessing or symbolization, associated with loss of information (Kantz and Schreiber, 2004). Also, one of the advantages of using complex networks for analyzing time series is that they are becoming powerful tools when one seeks to construct feature vectors that can be used to

automatically feed classifiers with low computational cost (Lacasa et al., 2015).

2. Materials and methods

2.1. Experimental data

The data of pollutant used for this analysis correspond to ozone and nitrogen oxide concentration values collected from 2010 to 2017, with a frequency of 10 min between each measure. Then, these data were separated into the different months in order to perform the analysis. The measurements were performed at the urban station located in Lepanto, Córdoba (37.53° N, 4.47° W). The cited station belongs to the regional network in charge of controlling the air quality in Andalusia, co-financed by the Consejería de Medioambiente (Regional Environmental Department) and the European Union. This station is located at 117 m of altitude and the average temperature and solar radiation is maximum on July and minimum on January every year. The region where it is placed is the western part of Andalusia (Spain). Since as exposed previously (Domínguez-López et al., 2014), this area meets the weather conditions (high temperatures and solar radiation), orographic (the valley of the Guadalquivir river) and anthropic ones to be potentially vulnerable to pollution by ground-level ozone and nitrogen oxide. The climate of the zone of study, according to the Köppen-Geiger classification, is defined as Csa, with warm average temperatures and hot and dry summer. Furthermore, the city of Córdoba is surrounded by two main industrial parks. One of them is located at southwest, and the other is at east. Moreover, there is a highway at southeast with frequently high traffic volume from both directions.

Authors have also employed the temperature, wind direction and average solar radiation in this work. They are shown further in the text (see Fig. 7), were they are plotted along with the results in question, in order to illustrate their apparent relationship. These meteorological quantities have been provided by the Andalusian Agency of Energy.

2.2. Visibility graph

A graph can be defined as a set of vertices, points or nodes connected to each other by lines that are usually called *edges*. A tool to transform time series into a graph was presented in the last decade (Lacasa et al., 2008). This new complex network receives the name of *Visibility Graph* (VG) and has been proven to inherit many of the properties of the original signal (Lacasa and Toral, 2010). This means that, for instance, a periodic time series would lead to a regular graph and a fractal series to a scale-free one.

In order to construct the visibility matrix which contains the information of all the nodes in the system, it is necessary to stablish a criterion to discern whether two points would be connected or not. This criterion reads as follows: two arbitrary data from the time series (t_a, y_a) and (t_b, y_b) have visibility (and would become two connected nodes in the graph) if any other data point (t_c, y_c) between them $(t_a < t_c < t_b)$ fulfills the following condition:

$$y_c < y_a + (y_b - y_a)\frac{t_c - t_a}{t_b - t_a}$$

(1)

The result of applying this visibility method is a NxN adjacency binary matrix, being N the number of points in the set. Each row of the matrix contains the information of a different node. For example, an element as $a_{ij} = 1$ means that the node i and j have visibility; whereas $a_{ij} = 0$ means that there is no edge between them. The resulting matrix has several properties that can be used to simplify the algorithm and thus reduce the computational required time: it is a hollow matrix ($a_{ii} = 0$), symmetric ($a_{ij} = a_{ji}$) and all the nearest neighbors have visibility between each other ($a_{ij} = 1$ for $j = i \pm 1$). Its general form is shown below:

$$A = \begin{pmatrix} 0 & 1 & \dots & a_{1,N} \\ 1 & 0 & 1 & \vdots \\ \vdots & 1 & \ddots & 1 \\ a_{N,1} & \dots & 1 & 0 \end{pmatrix}$$

(2)

In Fig. 2, the application of the VG to two arbitrary time series is shown, highlighting the connections of two given time points (nodes in the graph) for the sake of clarity.

2.3. Degree centrality

To study the main properties of a complex network, centrality parameters become convenient mathematical tools to take into account. This kind of parameters measure the node importance within the graph in relation to the others by different approaches (Latora et al., 2017).

A very frequently used centrality parameter and an important feature of graphs in general is the degree. The degree of a node (k_i) can be defined as the number of nodes that have reciprocal visibility (in an undirected graph) with the first one ($k_i = \sum_j a_{ij}$). In Fig. 2, the degree of the node that is highlighted is $k = 6$ for O_3 and $k = 3$ for NO_2.

From the degree of each one of the nodes present in the VG, it is possible to obtain the degree distribution of the sample $P(k)$, which is nothing but the probability that every degree has within the graph. This distribution can be analyzed to get a deep insight of the intrinsic nature of the time series, as previously demonstrated by (Lacasa et al., 2008; Mali et al., 2018; Pierini et al., 2012). The degree distribution of VGs whose right tails can be fitted to a power law in the way $P(k) \propto k^{-\gamma}$, are associated to fractal time series (Lacasa et al., 2008). The right tails are related to hubs, which are unlikely highly-connected nodes in the graph and therefore points with large values of degree. In a log-log plot, one can fit $P(k)$ to a simple linear regression, obtaining the so-called γ coefficient, which has been directly related to the Hurst exponent of the time series in the Brownian motion (Lacasa et al., 2009).

2.4. Multiplex visibility graph

In the case of a multivariate time series of M variables, it is possible to construct a M-dimensional network from the VG of each one for the description of the signals (Lacasa et al., 2015). This multilayer network is called *Multiplex Visibility Graph* and each one of its M layers corresponds to the VG of one of the variables from the underlying time series (see Fig. 2). The MVG is represented by a vector of adjacency matrices $\Omega = \{A^{[1]}, A^{[2]}, ..., A^{[M]}\}$, being $A^{[\alpha]}$ the matrix corresponding to the VG of the α-dimension (or layer in the multiplex) from the multivariate time series.

When it comes to analyzing the information that lies within these complex multilayer networks, there are several measures that can be used (Nicosia and Latora, 2015). Here, two quantities have been chosen. The first one is the *Average Edge Overlap (ω)*, that measures the number of layers on which a given edge between two nodes is found, on average. The other one, that captures the presence of inter-layer correlations of the degree distributions between two layers α and β, is the so-called *Interlayer Mutual Information* ($I_{\alpha,\beta}$). These layers in the presented study correspond to the VGs of O_3 and its precursor NO_2 concentration time series, hence the notation I_{O_3,NO_2} will be used in this work.

The formula for the calculation of ω is presented in Equation (3), where $\delta_{0,\sum_\alpha a_{ij}^{[\alpha]}}$ corresponds to a Kronecker Delta and the other quantities are already defined.

$$\omega = \frac{\sum_i \sum_{j>i} \sum_\alpha a_{ij}^{[\alpha]}}{M \sum_i \sum_{j>i} \left(1 - \delta_{0,\sum_\alpha a_{ij}^{[\alpha]}}\right)}$$

(3)

The maximum possible value of this quantity is $\omega = 1$, and corresponds to the case where all the layers are identical. On the other hand,
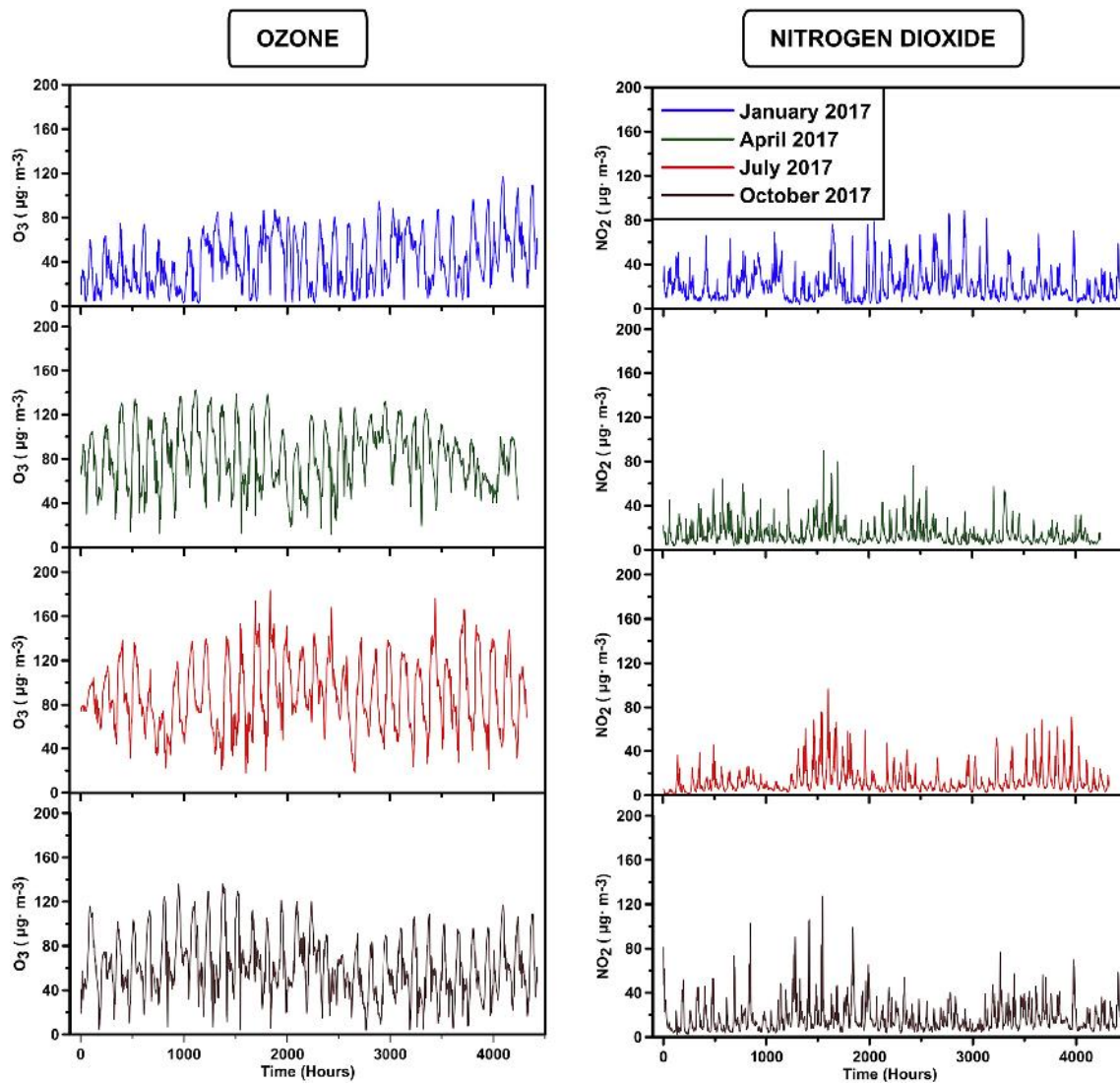
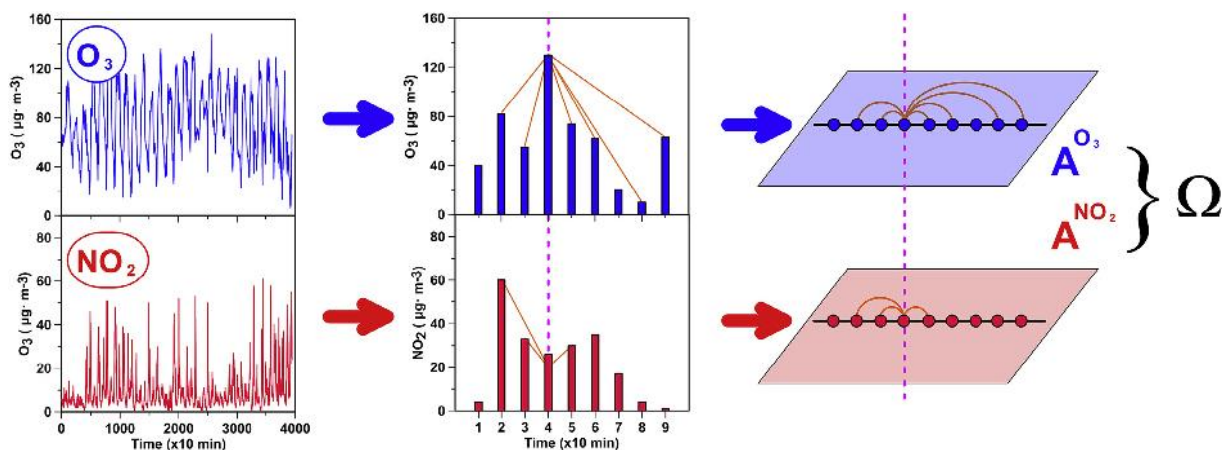**Fig. 1.** Sample time series of O_3 and NO_2 for four different months (year 2017).

Fig. 2. Time series of ozone and nitrogen dioxide concentrations (left) are transformed into complex networks through the VG algorithm (center), which is defined by an adjacency matrix (A^{O_3} and A^{NO_2}). Finally, the two of them are combined to create a two-layered MVG, called Ω (right image).
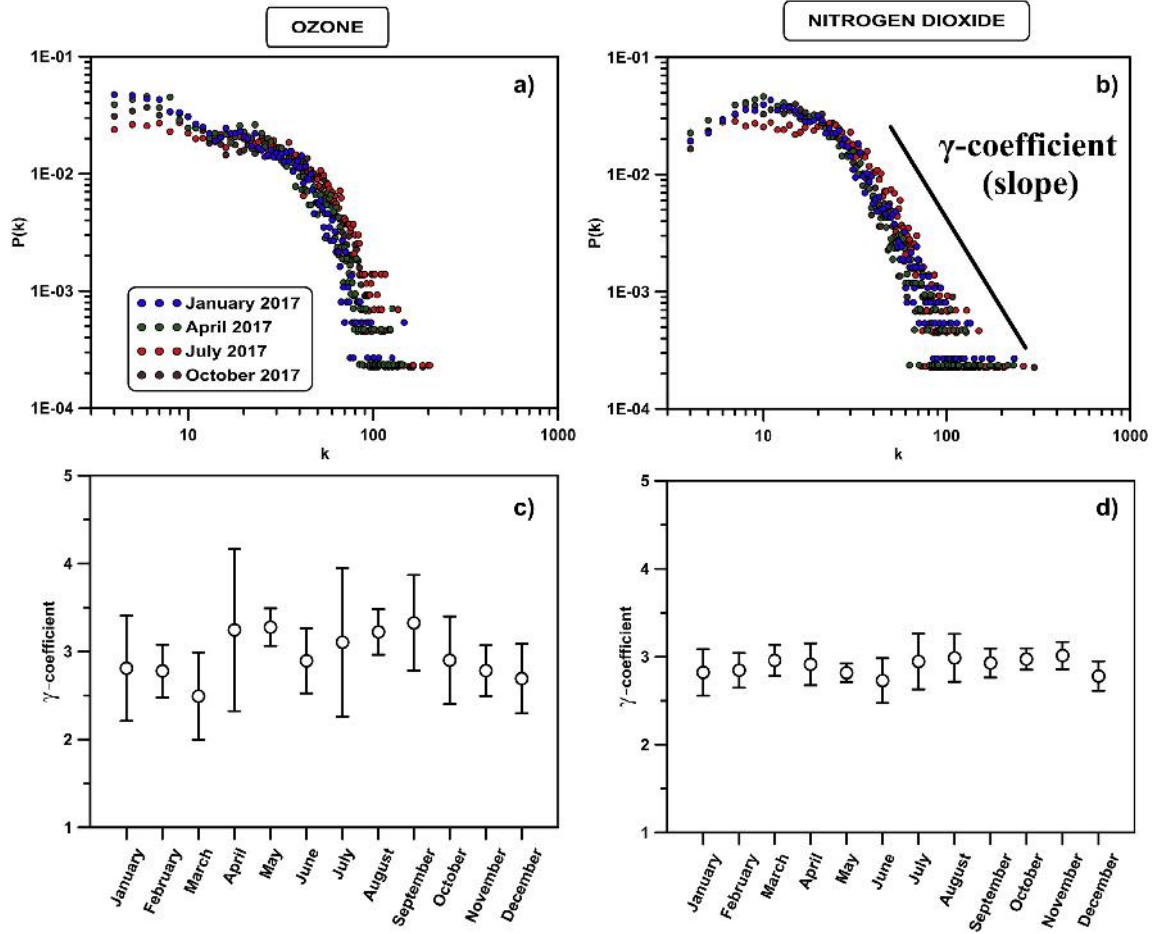
Fig. 3. Degree distributions of four example months in 2017 for O$_3$ (a) and NO$_2$ (b) and the actual values of the γ coefficient obtained for every studied month and averaged from 2010 to 2017 (c and d). The error bars come from the standard deviation of the values obtained for all the years.

the minimum value is $\omega = 1/M$ (being M the number of layers), meaning that a case with each edge in the multiplex existing just in one layer.

In Equation (4), $I_{\alpha,\beta}$ is defined:

$$I_{\alpha,\beta} = \sum_{k^{[\alpha]}} \sum_{k^{[\beta]}} P(k^{[\alpha]}, k^{[\beta]}) \log \frac{P(k^{[\alpha]}, k^{[\beta]})}{P(k^{[\alpha]})P(k^{[\beta]})} \tag{4}$$

where $P(k^{[\alpha]}, k^{[\beta]})$ is the joint probability of finding a node having a degree of $k^{[\alpha]}$ in the layer α and $k^{[\beta]}$ in the layer β. This joint probability is computed as follows:
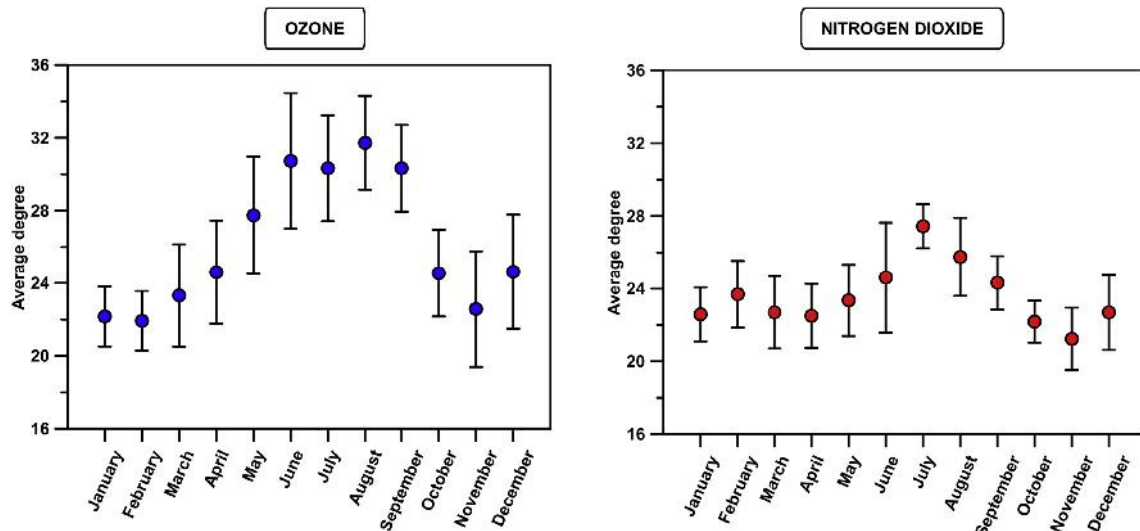

Fig. 4. Monthly average degree values for each pollutant from 2010 to 2017. Again, the standard deviation along the different years is reflected through the error bars.
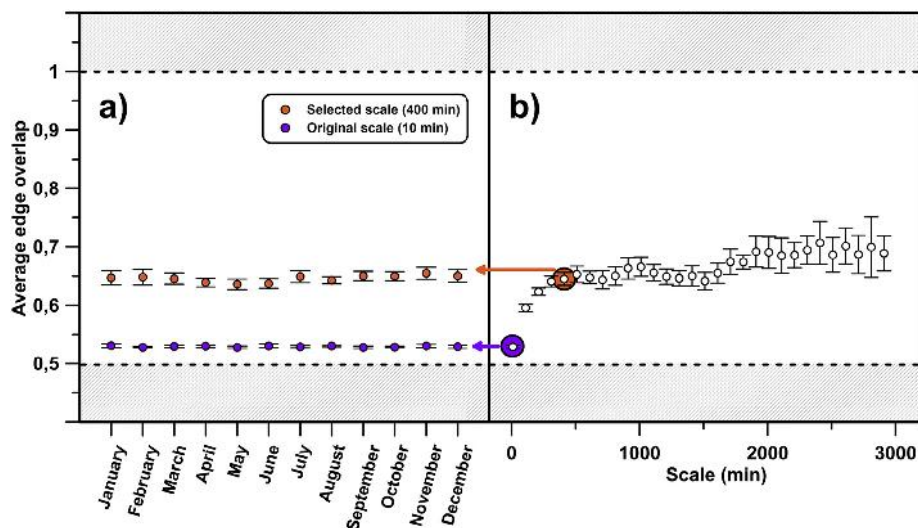
Fig. 5. a) Parameter ω is computed for each month and then averaged over all the analyzed years (2010–2017), first with the original time scale (purple) and finally with a bigger one (orange) derived from the analysis shown in b). In this one, ω is recalculated with several scales for constructing the MVGs. The selected scale is taken from the point where the saturation starts. The white area between the grey zones corresponds to all the possible values that ω can have for MVGs with two layers.

$$P(k^{[\alpha]}, k^{[\beta]}) = \frac{N_{k^{[\alpha]}, k^{[\beta]}}}{N} \qquad (5)$$

with $N_{k^{[\alpha]}, k^{[\beta]}}$ being the number of nodes that have the corresponding degree of $k^{[\alpha]}$ and $k^{[\beta]}$ in layers α and β, respectively. Since N is the total amount of nodes, it must be fulfilled that it is equal to the sum over all the possible $N_{k^{[\alpha]}, k^{[\beta]}}$ values.

3. Results and discussion

3.1. Exploratory analysis

Before applying the MVG methodology, authors have performed a preliminary analysis of the data employed in this work. To do so, the first feature to consider has been the distribution of the degree of the independent VGs obtained from each time series. Some examples of these distributions can be regarded at Fig. 3a and b. In these plots, only four months (the same ones for both) are depicted for illustrative purposes: January, April, July and October. The reason for choosing these
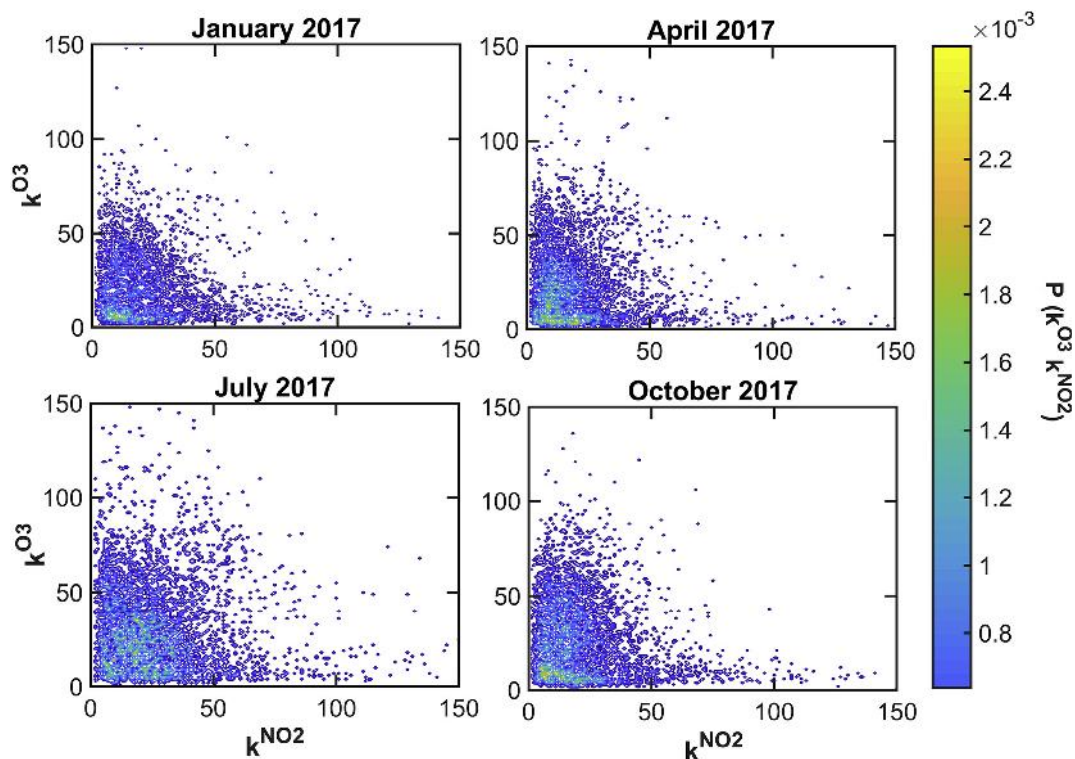

Fig. 6. Graphical representation of the joint probability distribution of the degrees of both layers for four sample months in 2017, the most recent year. Each point represents the probability that degree is exactly k^{O3} in the ozone VG, while is k^{NO2} in the nitrogen dioxide case, at the same time node.
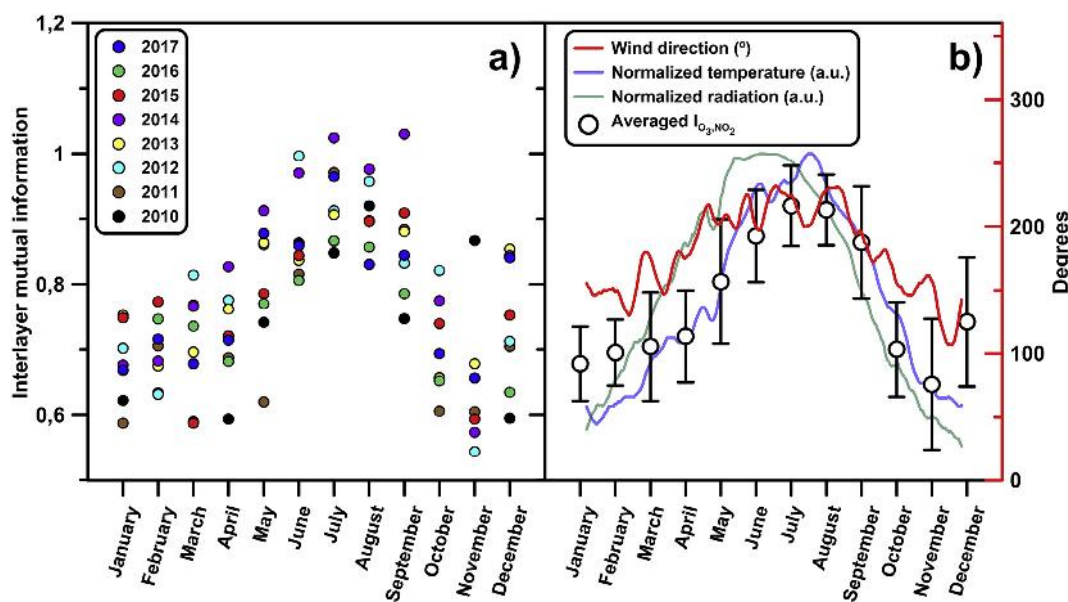
Fig. 7. a) Seasonal pattern obtained for I_{O_3,NO_2} for each one of the studied months and years. b) Averaged I_{O_3,NO_2} over the whole period (2010–2017) for every month and normalized temperature, solar radiation and wind direction for comparison. 0, 90, 180 and 240° correspond to winds coming from the East, North, West and South directions, respectively.

months is that they are equally spaced through the year, each one re-presents a different season and they have been used in previous works in the same location (Carmona-Cabezas et al., 2019a; Jiménez-Hornero et al., 2010a). The year shown in this case is 2017, the most recent one that has been used here. All data used has not undergone any desea-sonalizing because, as it has been previously discussed, VGs are not suitable for this kind of preprocessing approaches (Lange et al., 2018).

Looking at the mentioned figures, it can be noticed that both pol-lutants present a power-law behavior in the tail of the degree dis-tribution obtained from their respective VGs, which points to fractal behavior of the time series. As commented in Section 0, from the linear regression in the log-log plot of this tail, one is able to obtain the γ coefficient. At plain sight, it can be already seen Fig. 3a and b how the distribution of the degree for ozone varies more along the year than those of the nitrogen dioxide in the given examples. Authors have computed the γ coefficient of each month from 2010 to 2017 and shown their average monthly value in Fig. 3c and d; where the previous statement is checked. The error bars correspond to the standard de-viation of that quantity along the studied years. Therefore, the ozone concentration has a different behavior along the seasons, being the mean values of the coefficients between 2 and 4.25. Moreover, it can be appreciated that those are also irregular from one year to another, as the standard deviation is higher in comparison to the second gas by looking at Fig. 3c and d. This was already seen in a previous work by the authors (Carmona-Cabezas et al., 2019a). On the other hand, NO_2 coefficients do not vary as much as the previous one, being always its value between 2.5 and 3.25. These results seem to indicate a different trend in the likeliness of hubs coming from VGs of O_3 and NO_2 and so distinct unlikely large values variation. Although in some cases the physical meaning of those coefficients has been related to fractal parameters such as the Hurst exponent (see Section 0 for more details), authors have employed these as a preliminary study before going deeper into the analysis with MVGs.

From the degree computed for each node of a given VG, it is possible to obtain also its total average value, which has been used in previous papers to describe the behavior of the maxima in the series (Carmona-Cabezas et al., 2019b; Donner and Donges, 2012). As in previous stu-dies, a clear seasonal pattern is observed for both pollutants in Fig. 4. The values of this figure correspond to the average obtained over the whole period from 2010 to 2017, while the error bars come from their

standard deviation. This seasonality is less intense in the case of NO_2, which is in accordance to what was discussed in Fig. 3.

In both cases, the maxima of the average degrees correspond to summer months. It was expected in ozone, since those months have the most favorable conditions for its creation and therefore there will be a higher number of maximal values (hubs) that increase the average degree. Nevertheless, the behavior of nitrogen dioxide is not as acute during summer. This difference might be due to the different factors that influence on both pollutants. Results point to the possibility that this quantity could be used to identify the known correlations between the two pollutants, and hence, authors have tested this hypothesis by using the MVG parameters. More precisely, $I_{\alpha,\beta}$ that is directly based on the degree of the two time series.

3.2. MVG analysis

After applying the VG algorithm to transform the O_3 and NO_2 concentration time series into complex networks, the MVGs for each month were built. With these multilayer networks, it was possible to compute ω and I_{O_3,NO_2} for all the months considered. Fig. 5a shows that ω values obtained are very similar for all the months and during the different studied years. This is clearly seen, as it has been averaged over the different studied years (2010–2017) and the standard deviation is as well remarkably low.

As ω accounts for the overall coherence of the original time series, high values of it indicates a strong correlation in the microscopic structure and vice versa (Lacasa et al., 2015). In the context of VGs applied to time series, the microscopic structure can be understood as the most likely scale of the edges of the graph. This scale can be con-sidered as the time resolution of the recorded signal. For that reason, a study was performed to see how the ω parameter behaves when the minimal size of the edges of the concentration of O_3 and NO_2 is changed (see Fig. 5b).

The result is that the edge overlap increases rapidly as the number of points in the sample is decreased (the VGs become smaller and their edges bigger) and then it starts to saturate around $\omega \approx 0.65$ for scales greater than 400 min per measured point (about 6 h). On the other hand, for the original scale of the measure $\omega \approx 0.53$, which is very close to the minimum value that quantity can hold ($\omega = 0.5$). From that, this result can be understood as follows: the correlation between the O_3 and

the NO_2 concentrations is very weak at a scale of 10 min (the order of magnitude of the smallest edges in the graphs). Nevertheless, this correlation becomes greater as the minimal size of these edges (the microscopic scale) is increased, until this rising saturates. Authors attribute this effect to the time needed for the two systems (gases) to reach coupling, so that it affects the correlation between their concentrations. The posterior saturation would correspond to the fact that they are already coupled for higher scales, so there is no increase on ω due to physical processes. The slight increase observed would be due to the reduction of the number of nodes of the VGs (mathematical artifact). Hence, this approach to determine the time scale at which the saturation starts could be used to determine the effective time of a given reaction. This could be useful for later works describing the relationship between some other pollutants.

As a previous step before introducing the computed I_{O_3,NO_2}, authors have depicted in Fig. 6 the quantity $P(k^{O3}, k^{NO2})$. Again, the same months as before are used in these plots as an example, for the sake of clarity. Also, only one year is taken (2017, the most recent), since the results have been found to be equivalent and the same conclusions could be drawn for different years. The meaning of these figures can be interpreted as combined degree distributions, where the colors indicate the probability of the two VGs having degree k^{O3} and k^{NO2}, simultaneously. It is regarded how the most likely combinations of values of k are those of the lowest values of the degree ($k \in [0, 50]$). By contrast, as the degree increases, the joint probability becomes less and less significative.

It must be pointed out that the probability approaches asymptotically to both X and Y axis. It means that as the degrees increase, the probability of encountering relatively similar both k^{O3} and k^{NO2}, decreases exponentially. This translates into the alternation between the hubs of the two time series. The reason behind this is the time shift that exists between both NO_2 and O_3 maximal concentrations throughout the day (previously mentioned). One of the reactions that governs the ozone creation and destruction is $NO_2 + O_2 \leftrightarrow O_3 + NO$ (Graedel and Crutzen, 1993). According to this one, when the ozone reaches a maximum, the concentration of nitrogen dioxide decreases in general, leading to what has been argued here.

Another characteristic of the plot is that as the year advances, the distribution of the joint probability changes, being more concentrated around the k^{NO2} axis for January, while in July it is more equally distributed (April and October present an intermediate behavior). Since the value of the degree and concentration are related (Carmona-Cabezas et al., 2019b; Pierini et al., 2012), the interpretation can be seen as follow:

- In January, $P(k^{O3}, k^{NO2})$ is more concentrated to specific combinations of degrees of the gases, specially, low k^{O3} and higher k^{NO2}. This translates into fewer values of the temporal series that are considerably correlated and thus, the overall correlation will decrease (Fig. 7). As a result, high concentrations of NO_2 will not necessarily lead to greater production of ozone, as can be regarded in Fig. 1; where there are many days with extreme values of nitrogen dioxide concentration, whereas the ozone levels remain at minima with respect to the rest of the year. The reason is that although high concentrations of NO_2 are available, there is not enough solar radiation to make the optimal interaction possible.
- In July, the concentration of ozone rises, as it widely known. In Fig. 6, $P(k^{O3}, k^{NO2})$ is in this case more homogeneous and non-null for the values of k where the vast majority of points are located: $k \in [0, 50]$. The result of this will be an increase in the I_{O_3,NO_2} that is shown on average in Fig. 7. In contrast to January, now the extreme concentrations of NO_2 coincide with those of O_3, for instance in Fig. 1 for July, from $1.5 \cdot 10^4$ to $2 \cdot 10^4$ minutes both reach their highest concentrations (taking into account that there exist a time delay between both quantities).
- Lastly, April and October are intermediate cases, where the

distribution of $P(k^{O3}, k^{NO2})$ is neither as acute as in January, nor as regular as July. In both cases, the probability is higher for finding low k^{O3} when k^{NO2} is greater.

The monthly computed values of I_{O_3,NO_2} are presented in Fig. 7a, where the different years are indicated by several colors. Fig. 7b shows the monthly average value (over all the mentioned years) and the standard deviation along with the temperature, the average global radiation (normalized for the sake of clarity) and wind direction.

It can be appreciated that there is a sensible increase of the mutual information between O_3 and NO_2 from May to September, coinciding with the period of highest temperature and radiation. By definition, $I_{\alpha,\beta}$ describes the average correlation of the degree distributions and these can be used to describe the nature of the time series (Lacasa et al., 2008; Mali et al., 2018). From that, it could be inferred that the greater this parameter is, the more similarly are expected to behave the signals underlying the VGs studied. Therefore, in this case the degree distributions of the two pollutants are more correlated during the summer months and the opposite during winter. It might indicate that they will have a more similar behavior in the first case. For instance, O_3 is known to have multifractal nature (Carmona-Cabezas et al., 2019a; He et al., 2017; Jiménez-Hornero et al., 2010a; Pavon-Dominguez et al., 2013), hence NO_2 could be expected to have higher multifractality in the summer months, being in accordance with the findings reported in a prior work for the same area (Jiménez-Hornero et al., 2010b). A higher multifractality is understood as a greater degree of multifractal behavior, which is usually related to complex systems with fluctuations appearing at different scales (Kantelhardt, 2011).

The reason for this clear seasonal pattern seems obvious looking at the distribution of solar radiation and temperature, which change greatly along the seasons in the study area (both maximal in summer and vice versa in winter). The first one of them (solar radiation) is the main source of energy for the photochemical reaction needed for the formation of O_3 in the troposphere (Graedel and Crutzen, 1993), since nitrogen dioxide (NO_2) and oxygen (O_2) are recombined and produce ozone gas in the presence of ultraviolet light. The second one has been found to have relation as well with the production of ozone from nitrogen oxides in previous works in the same area (Pavón-Domínguez et al., 2015). Many of the previous studies are based on multifractal approaches to analyze time series, which have some disadvantages such as the need for choosing a scale interval where the searched behavior holds, leading to errors in the computation and some undesired ambiguities of the results. On the other hand, VGs are univocal for a given time series and their computation do not have any error associated, since their results are based on basic graph theory arithmetic.

Furthermore, regarding the average wind direction that can be seen in Fig. 7b, it is clearly seen how the direction is roughly ~220° when the correlation between NO_2 and O_3 reaches it maximal value for summer, as commented before. This corresponds to wind coming as an average from the South-West. This fact has been previously described for the Guadalquivir Valley (Guardans and Palomino, 1995), where the pressure and temperature differences make more predominant the wind coming from South-West direction when the temperature is maximum (summer). The reason behind this phenomenon is that the air mass moves from the plain areas towards the upper parts of the valley (North-East). The opposite case occurs in winter following an equivalent reasoning.

It must be pointed out the fact that in that predominant direction, the main sources of NO_2 in the vicinities of Cordoba are located, since there are two of the main industrial parks of the city. Also, the most populated capital of the region (Sevilla) is situated in that direction, as well as the main highway that connect the two cities, which is one of the most transited roadways in Spain. This fact further corroborates the adequation of the correlation results, as was seen with the temperature and solar radiation previously.

4. Conclusion

All the stated results confirm the capability of the two parameters provided by the MVGs for describing the interaction between ozone and nitrogen dioxide in the troposphere. On the one hand, authors consider that the first studied parameter (ω) may be used to infer the time shift in the coupling of the two systems, represented by the two layers. Given that the value of this parameter does not vary significantly neither in the different seasons, nor along the years, it means that ω does not depend on any external factors, such as meteorological features. Authors believe that this quantity might be used to check theoretical models of O_3–NO_2 interaction, although further investigation will be needed.

On the other hand, I_{O_3,NO_2} and $P(k^{O3}, k^{NO2})$ could be used to have an insight of the correlation in the behavior of the series (from the degree, which is associated to the concentration itself). Regarding the first one, one can obtain an overall look of how the NO_2 and O_3 are correlated depending on the concentration. For instance, it is possible to see how great values of NO_2 correlate with large or low ones of O_3, and vice versa. Furthermore, the second quantity corresponds to a numerical value that measures the correlation of the whole set of series that are transformed into the different layers of the MVG.

To authors' mind, the outcomes of this research support the capability of multilayer complex networks for looking at the relation between several variables such as atmospheric pollutants. By using this approach, one can take advantage of some of its assets, as the computation efficiency or the univocity of the results. All this situates MVG as a suitable complementary technique for tackling analyses within the environmental problem. It as well compatible with others that have demonstrate to give satisfactory results, such as multifractal algorithms, as it has been demonstrated (Carmona-Cabezas et al., 2019a; Mali et al., 2018; P. Pavón-Domínguez et al., 2017). Hence one possible application of MVGs in this context could be the enlargement of databases used for predictive techniques that rely on data mining and machine learning. For this aim, it remains open the study of more pollutants and other variables through the methodology analyzed here. Since it is possible to focus MVGs on only two variables or to see the correlation of many of them at the same time, it allows a flexible analysis. All these considerations could come in handy for describing the many mechanisms involved in the dynamics of the photochemical smog for future works.

Declaration of interests

None.

Acknowledgements

References

Bianchi, F.M., Livi, Lorenzo, Alippi, Cesare, Jenssen, Robert, 2017. Multiplex visibility graphs to investigate recurrent neural network dynamics. Sci. Rep. 7, 44037. https://doi.org/10.1038/srep44037.

Bianconi, G., 2013. Statistical mechanics of multiplex networks: entropy and overlap. Phys. Rev. E 87, 062806. https://doi.org/10.1103/PhysRevE.87.062806.

Boccaletti, S., Bianconi, G., Criado, R., del Genio, C.I., Gómez-Gardeñes, J., Romance, M., Sendiña-Nadal, I., Wang, Z., Zanin, M., 2014. The structure and dynamics of multi-layer networks. Phys. Rep. 544, 1–122. https://doi.org/10.1016/j.physrep.2014.07.001.

Carmona-Cabezas, R., Ariza-Villaverde, A.B., Gutiérrez de Ravé, E., Jiménez-Hornero, F.J., 2019a. Visibility graphs of ground-level ozone time series: a multifractal analysis. Sci. Total Environ. 661, 138–147. https://doi.org/10.1016/j.scitotenv.2019.01.147.

Carmona-Cabezas, R., Gómez-Gómez, J., Ariza-Villaverde, A.B., Gutiérrez de Ravé, E., Jiménez-Hornero, F.J., 2019b. Can complex networks describe the urban and rural tropospheric O3 dynamics? Chemosphere 230, 59–66. https://doi.org/10.1016/j.chemosphere.2019.05.057.

De Domenico, M., Solé-Ribalta, A., Cozzo, E., Kivelä, M., Moreno, Y., Porter, M.A., Gómez, S., Arenas, A., 2013. Mathematical formulation of multilayer networks. Phys. Rev. X 3, 041022. https://doi.org/10.1103/PhysRevX.3.041022.

Doherty, R.M., Heal, M.R., Wilkinson, P., Pattenden, S., Vieno, M., Armstrong, B., Atkinson, R., Chalabi, Z., Kovats, S., Milojevic, A., Stevenson, D.S., 2009. Current and future climate- and air pollution-mediated impacts on human health. Environ. Health 8, S8. https://doi.org/10.1186/1476-069X-8-S1-S8.

Domínguez-López, D., Adame, J.A., Hernández-Ceballos, M.A., Vaca, F., De la Morena, B.A., Bolívar, J.P., 2014. Spatial and temporal variation of surface ozone, NO and NO2 at urban, suburban, rural and industrial sites in the southwest of the Iberian Peninsula. Environ. Monit. Assess. 186, 5337–5351. https://doi.org/10.1007/s10661-014-3783-9.

Donner, R.V., Donges, J.F., 2012. Visibility graph analysis of geophysical time series: potentials and possible pitfalls. Acta Geophys. 60, 589–623. https://doi.org/10.2478/s11600-012-0032-x.

Graedel, T.E., Crutzen, P.J., 1993. Atmospheric change: an earth system perspective. J. Chem. Educ. 70, A252. https://doi.org/10.1021/ed070pA252.2.

Guardans, R., Palomino, I., 1995. Description of wind field dynamic patterns in a valley and their relation to mesoscale and synoptic-scale meteorological situations. J. Appl. Meteorol. 34, 49–67.

He, H., Qiao, Z., Pan, W., Lu, W.-Z., 2017. Multiscale multifractal properties between ground-level ozone and its precursors in rural area in Hong Kong. J. Environ. Manag. 196, 270–277. https://doi.org/10.1016/j.jenvman.2017.02.024.

Jiménez-Hornero, F.J., Gutiérrez de Ravé, E., Giráldez, J.V., Giráldez, J.V., 2010a. Description of the seasonal pattern in ozone concentration time series by using the strange attractor multifractal formalism. Environ. Monit. Assess. 160, 229–236. https://doi.org/10.1007/s10661-008-0690-y.

Jiménez-Hornero, F.J., Jimenez-Hornero, J.E., Gutiérrez de Ravé, E., Pavón-Domínguez, P., 2010b. Exploring the relationship between nitrogen dioxide and ground-level ozone by applying the joint multifractal analysis. Environ. Monit. Assess. 167, 675–684. https://doi.org/10.1007/s10661-009-1083-6.

Kantelhardt, J.W., 2011. Fractal and multifractal time series. In: Meyers, R.A. (Ed.), Mathematics of Complexity and Dynamical Systems. Springer New York, New York, NY, pp. 463–487. https://doi.org/10.1007/978-1-4614-1806-1_30.

Kantz, H., Schreiber, T., 2004. Nonlinear Time Series Analysis, second ed. Cambridge University Press, Cambridge, UK; New York.

Kivela, M., Arenas, A., Barthelemy, M., Gleeson, J.P., Moreno, Y., Porter, M.A., 2014. Multilayer networks. J. Complex Net. 2, 203–271. https://doi.org/10.1093/comnet/cnu016.

Lacasa, L., Luque, B., Ballesteros, F., Luque, J., Nuño, J.C., 2008. From time series to complex networks: the visibility graph. Proc. Natl. Acad. Sci. 105, 4972–4975. https://doi.org/10.1073/pnas.0709247105.

Lacasa, L., Luque, B., Luque, J., Nuño, J.C., 2009. The visibility graph: a new method for estimating the Hurst exponent of fractional Brownian motion. Europhys. Lett. 86, 30001. https://doi.org/10.1209/0295-5075/86/30001.

Lacasa, L., Nicosia, V., Latora, V., 2015. Network structure of multivariate time series. Sci. Rep. 5, 15508. https://doi.org/10.1038/srep15508.

Lacasa, L., Toral, R., 2010. Description of stochastic and chaotic series using visibility graphs. Phys. Rev. E 82, 036120. https://doi.org/10.1103/PhysRevE.82.036120.

Lange, H., Sippel, S., Rosso, O.A., 2018. Nonlinear dynamics of river runoff elucidated by horizontal visibility graphs. Chaos 28, 075520. https://doi.org/10.1063/1.5026491.

Latora, V., Nicosia, V., Russo, G., 2017. Complex Networks: Principles, Methods and Applications, first ed. Cambridge University Press https://doi.org/10.1017/9781316216002.

Mali, P., Manna, S.K., Mukhopadhyay, A., Haldar, P.K., Singh, G., 2018. Multifractal analysis of multiparticle emission data in the framework of visibility graph and sandbox algorithm. Physica A 493, 253–266. https://doi.org/10.1016/j.physa.2017.10.015.

Miao, W., Huang, X., Song, Y., 2017. An economic assessment of the health effects and crop yield losses caused by air pollution in mainland China. J. Environ. Sci. 56, 102–113. https://doi.org/10.1016/j.jes.2016.08.024.

Nicosia, V., Latora, V., 2015. Measuring and modelling correlations in multiplex networks. Phys. Rev. E 92, 032805. https://doi.org/10.1103/PhysRevE.92.032805.

Pavón-Domínguez, P., Ariza-Villaverde, A.B., Rincón-Casado, A., Gutiérrez de Ravé, E., Jiménez-Hornero, F.J., 2017. Fractal and multifractal characterization of the scaling geometry of an urban bus-transport network. Comput. Environ. Urban Syst. 64, 229–238. https://doi.org/10.1016/j.compenvurbsys.2017.03.003.

Pavón-Domínguez, P., Jiménez-Hornero, F.J., Gutiérrez de Ravé, E., 2015. Joint multifractal analysis of the influence of temperature and nitrogen dioxide on tropospheric ozone. Stoch. Environ. Res. Risk Assess. 29, 1881–1889. https://doi.org/10.1007/s00477-014-0973-5.

Pavon-Dominguez, P., Jimenez-Hornero, F.J., Gutierrez de Rave, E., 2013. Multifractal analysis of ground–level ozone concentrations at urban, suburban and rural background monitoring sites in Southwestern Iberian Peninsula. Atmos. Pollut. Res. 4, 229–237. https://doi.org/10.5094/APR.2013.024.

Pierini, J.O., Lovallo, M., Telesca, L., 2012. Visibility graph analysis of wind speed records measured in central Argentina. Phys. A Stat. Mech. Appl. 391, 5041–5048. https://doi.org/10.1016/j.physa.2012.05.049.

Sannino, S., Stramaglia, S., Lacasa, L., Marinazzo, D., 2017. Visibility Graphs for fMRI Data: Multiplex Temporal Graphs and Their Modulations across Resting State Networks. Network Neurosci, pp. 208–221. https://doi.org/10.1101/106443.

Sillman, S., 1999. The relation between ozone, NOx and hydrocarbons in urban and polluted rural environments. Atmos. Environ. 33, 1821–1845. https://doi.org/10.1016/S1352-2310(98)00345-8.

Trainer, M., Parrish, D.D., Goldan, P.D., Roberts, J., Fehsenfeld, F.C., 2000. Review of observation-based analysis of the regional factors influencing ozone concentrations. Atmos. Environ. 34, 2045–2061. https://doi.org/10.1016/S1352-2310(99)00459-8.

<u>Publication 7</u>
Improving Graph-Based Detection of Singular Events for Photochemical Smog Agents

<u>Journal</u>

Chemosphere (Ed. Elsevier)

<u>Position / Quartile</u>

(29/265) / Q1 (Environmental Sciences)

<u>Impact factor</u>

5.778 (2019)

Improving graph-based detection of singular events for photochemical smog agents

Rafael Carmona-Cabezas [a, *], Javier Gómez-Gómez [a], Eduardo Gutiérrez de Ravé [a], Elena Sánchez-López [a], João Serrano [b], Francisco José Jiménez-Hornero [a]

[a] Complex Geometry, Patterns and Scaling in Natural and Human Phenomena (GEPENA) Research Group, University of Cordoba, Gregor Mendel Building (3rd floor), Campus Rabanales, 14071, Cordoba, Spain
[b] Mediterranean Institute for Agriculture, Environment and Development (MED), Departamento de Engenharia Rural, Escola de Ciências e Tecnologia, Universidade de Évora, P.O. Box 94, Évora, 7002-554, Portugal

HIGHLIGHTS

- Detection of singularities using graphs is improved by taking the inverted series.
- Maxima and minima of pollutant series are identified by VG and UDVG respectively.
- Asymmetries in the distribution of NO_2 might be caused by reaction with VOCs.
- NO_2 singularity identification is more difficult due to its more complex dynamics.
- A more complete analysis tool is obtained by combining both approaches.

GRAPHICAL ABSTRACT
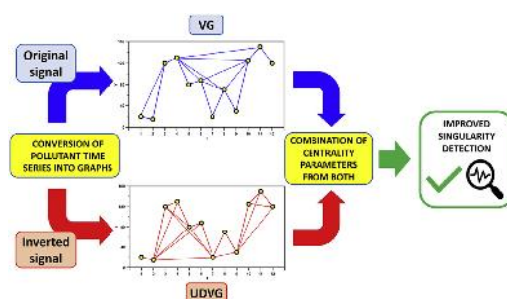
ABSTRACT

Recently, a set of graph-based tools have been introduced for the identification of singular events of O_3, NO_2 and temperature time series, as well as description of their dynamics. These are based on the use of the Visibility Graphs (VG). In this work, an improvement of the original approach is proposed, being called Upside-Down Visibility Graph (UDVG). It adds the possibility of investigating the singular lowest episodes, instead of the highest. Results confirm the applicability of the new method for describing the multifractal nature of the underlying O_3, NO_2, and temperature. Asymmetries in the NO_2 degree distribution are observed, possibly due to the interaction with different chemicals. Furthermore, a comparison of VG and UDVG has been performed and the outcomes show that they describe opposite subsets of the time series (low and high values) as expected. The combination of the results from the two networks is proposed and evaluated, with the aim of obtaining all the information at once. It turns out to be a more complete tool for singularity detection in photochemical time series, which could be a valuable asset for future research.

© 2020 Elsevier Ltd. All rights reserved.

1. Introduction

Among the problems related to atmospheric pollution, there is a matter of special concern studied by environmental scientists in the

recent years, the so-called photochemical smog. Also known as "Los Angeles smog", since it was firstly noticed in that city in 1944, as a result of the observed damage on the vegetation (NAPCA, 1970). It can be defined as the accumulation of gases and aerosols as a result of reactions between nitrogen oxides (NO_x), certain volatile organic compounds (VOCs) and oxygen under the influence of solar radiation. A wide range of chemicals (ozone, aldehydes or hydrogen peroxides among them) are created in the process (Guicherit and van Dop, 1977). Typically, this phenomenon is more prominent when a city is more populated and warmer. Among the gases involved, there are two which are extensively researched due to the many harms associated to them and their quantitative importance: the tropospheric ozone (O_3) and the nitrogen dioxide (NO_2), being the second a precursor of the first one. It must be stressed that both of them (O_3 and NO_2) have a serious impact on human health (Cheng et al., 2020; Kampa and Castanas, 2008; Yue et al., 2018). Furthermore, a recent study has demonstrated that O_3 produces harsh effects on the economy due to a reduction of the crop yield (Miao et al., 2017).

During the last decades, investigation on complex networks and their applications has been carried out in many works (Boccaletti et al., 2006; Gan et al., 2014; Newman, 2003; Stam, 2010). A complex network can be understood as a graph (a set of nodes and edges as will be further explained) which exhibits nontrivial topological properties and is often used to model and describe real systems. Furthermore, in the recent years there have been a considerable amount of works seeking ways to represent nonlinear time series as complex networks (Zou et al., 2019). This includes manuscripts based on recurrence networks, transition networks and visibility graphs. The main potential of these approaches is the vast number of tools that there exist to analyze networks from a computational perspective. Authors highlight the centrality parameters, since they are essential to this work. They are used to quantify the importance of the nodes within a graph and will be introduced and used later in the text.

Among the new methodologies previously described, there is one that has been recently used to investigate environmental time series (Carmona-Cabezas et al., 2019b; Donner and Donges, 2012; Pierini et al., 2012). This methodology received the denomination of Visibility Graph (VG) algorithm (Lacasa et al., 2008). As it has been demonstrated several times, the complex networks obtained through this method inherit the main features of the original time series and therefore can be used to describe them (Lacasa et al., 2009; Lacasa and Toral, 2010).

Besides describing the nature and main features of the time series, another possibility implies the detection of singularities within these signals. For that purpose, many techniques have been used. One example is the Hölder exponent, which is based on multifractal properties of the system (Loutridis, 2007; Shang et al., 2006). By looking at the information retrieved from the transformed complex network, it is also possible to detect singular points, as it has been explored in several works recently (Bielinskyi and Soloviev, 2018; Carmona-Cabezas et al., 2019b, 2020). In particular, the unusually large values of the cited centrality parameters associated to each node, can provide much of the information that could be derived from the time series.

In the presented work, a new approach is introduced to improve this detection of singular points in a time series from photochemical smog variables (pollutant concentration and temperature), using the VG. The motivation behind it was the fact that regular VG criterion associates the highest connectivity to the points with largest concentration. Therefore, singular events that have low value are overlooked by the original technique. The proposed improvement analyzes the original and inverted series and combines their parameters for a wider point of view.

The pursued aim with this work is to test the application range and possible advantages or pitfalls of the proposed improvement. By doing that, authors intend to explore how this advance could complement the identification of singular episodes of pollutant time series (which could be potentially extended to others apart from O_3 and NO_2). Being that the case, future researchers will benefit from a more thorough technique for detecting unusual low and high gas concentrations, with different criteria, as a result.

2. Materials and methods

2.1. Data

For this work, measurements of tropospheric ozone (O_3), nitrogen dioxide (NO_2) and temperature have been used. All of them correspond to hourly time series, being recorded in 2017. In the last part of this manuscript, months corresponding to different seasons are selected. The reason for this lies in the fact that, as explained before, this work seeks improving a previous one (Carmona-Cabezas et al., 2020), and therefore the same months have been used for clearer comparison. The station where they were collected is called San Fernando (36°27′ N, 6°12′ W), which is located in the province of Cádiz (southern Iberian Peninsula) and administered by the Consejería de Medioambiente (Regional Environmental Department) of Andalusia and the European Union.

According to the Köppen-Geiger classification, the zone where the data is collected is labelled as "Csa", as it is most of the Mediterranean basin. "Csa" regions are characterized by warm temperatures with summers that are regularly hot and dry. Furthermore, two of the most important industrial centers in the region (Huelva and Bay of Algeciras) are located relatively close to the study area. As a result of the mentioned conditions, this selected place is propense to accumulation of tropospheric ozone (O_3) and nitrogen dioxide (NO_2) (Domínguez-López et al., 2014).

2.2. Visibility graph

As it was introduced before, in the last decade, a new method to analyze one dimensional series was introduced (Lacasa et al., 2008). This technique transforms these series into a different mathematical entity: a graph or network. Therefore, it was given the name Visibility Graph, because of its resemblance to the original one used in architecture for space analysis (Lozano-Pérez and Wesley, 1979; Turner et al., 2001). One of the main features of the VG is that it has been demonstrated that it inherits properties of the original time series that it is obtained from (Lacasa et al., 2009, 2008; Lacasa and Toral, 2010). For instance, a periodic series would result on a regular graph after applying it.

In general, a graph can be understood as a set of *nodes* and *edges* that link them. In the context of VG, the nodes correspond to the points in the time series. Thus, it is necessary to stablish the criterion for linking them and so stablishing the *edges*. The basic idea is that two nodes are connected to each other if a line between them can be drawn and it does not pass below any other point in the signal. That is, two points (t_a, y_a) and (t_b, y_b) are connected in the graph (have visibility) if any point (t_c, y_c) between them ($t_a < t_c < t_b$) fulfills:

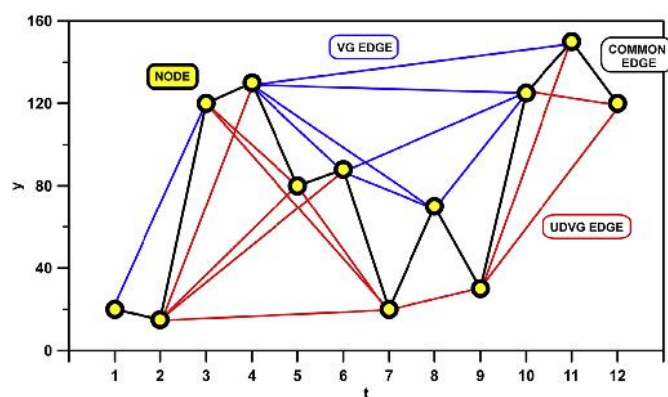$$y_c < y_a + (y_b - y_a)\frac{t_c - t_a}{t_b - t_a} \tag{1}$$

Fig. 1. Example of computation of the regular VG (blue lines) and the UDVG (red lines) to a sample time series and resulting graphs. Black lines indicate the common edges. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

From the VG method described, it is easy to see that the nodes with highest connectivity (also known as *hubs*) will be usually the ones with the unusual greatest values in the time series. This approach comes in handy in order to describe these points with higher magnitude; nevertheless, if one is interested on what happens with the opposite case (i.e. minimal unlikely values), the indicated technique is not suitable for describing them. That is indeed one disadvantage of employing VG for detecting singular points in a time series.

In a recent work, a variation of VG was presented (Soni, 2019) in order to explore new approaches to gain information about a time series. There, the concept of a signed complex network is introduced. The basic idea behind that method is that some of the edges will have a positive sign, while some other will be negative. The regular VG computed as explained before corresponds to the positive edges of this signed graph. On the other hand, the negative connections are obtained also from the regular VG but performed this time over the "upside-down" time series. That is, instead of using the original series $f(t)$, the converted series $-f(t)$ is used. This new graph was employed as a whole, in order to obtain series of clusters from the network and to analyze multivariate correlations, as an extension of previous works (Lacasa et al., 2015; Sannino et al., 2017). Nevertheless, the purpose of the work introduced here is to investigate the possibility of applying this idea for improving the detection of singular points in a time series, such as O_3 and NO_2 concentration, or temperature. For that reason, the positive and negative parts need to be obtained separately, as some of the parameters that will be further explained cannot be retrieved from a signed network (e.g. the betweenness centrality). For clarity reasons, the "positive" network will be given the name of regular VG in the text; while for the "negative" one, the term Upside-Down VG (UDVG) will be used. In Fig. 1, an example of the two types of network is shown.

It must be highlighted as well that all the edges of the two graphs are different, except for those connecting each node to it nearest neighbors in the time series. In Fig. 1, the first statement is seen by looking at the blue and red edges, while the second one reflects in the black ones that both regular VG and UDVG have in common. In other words, the elements of the adjacency matrices fulfill: $a_{ij}^{VG} + a_{ij}^{UDVG} \leq 1$; $\forall j \neq i \pm 1$. This means that the elements surrounding the main diagonal are equal and the others cannot be $a_{ij} = 1$ simultaneously in the two matrices.

2.3. Centrality parameters

One of the most widely used approaches to characterize graphs and complex networks is based on the analysis of the most important nodes within. It is done by employing the so-called centrality parameters, which are evaluated at each node, giving an idea about how "central" each one is, in relation to the rest of them. This concept was firstly used for studying social networks and transferred to other fields of research afterwards (Agryzkov et al., 2019; Joyce et al., 2010; Liu et al., 2015). The actual meaning of a central node may vary depending on the actual parameter used to evaluate the network. Here, authors focus on three of them: the degree, betweenness and closeness centrality, which have been used to describe physical systems in previous works (Carmona-Cabezas et al., 2020; Donner and Donges, 2012; Mali et al., 2018).

The first centrality parameter that will be explained is the degree. In a graph, the number of edges which are connected to a given node i is defined as the degree of that node (k_i), i.e., $k_i = \sum_j a_{ij}$, being a_{ij} the elements of the adjacency matrix. Once the degree for each node is obtained, the degree distribution $P(k)$ can be computed. This quantity has been proven to be able to characterize the nature of the studied signal (Lacasa et al., 2008; Mali et al., 2018). In fact, degree distributions that can be adjusted to a power law $P(k) \propto k^{-\gamma}$ correspond to scale free networks which comes from fractal series, as it was discussed by (Lacasa et al., 2009; Lacasa and Toral, 2010). The reason for this is the effect of hub repulsion (Song et al., 2006). A hub is a node from a graph with unlikely greater number of links, and so, higher degree. Therefore, the right tail of degree distributions is dominated by these nodes and, after being represented in a log-log plot, they can be fitted by a simple linear regression.

The other two employed parameters cannot be understood without defining the shortest path (SP) quantity first. SP is a measurement of the number of different edges that connect two distant nodes. Given a pair of nodes (i,j), different possible paths between them are available. Some of them (not necessarily unique) will have the minimum possible number of edges and, thus, they will be the minimal paths known as SP. It must be regarded that it has an important presence in the definition of the betweenness and closeness centrality. The betweenness of a node i can be computed by the following expression:

$$b_i = \sum_{\substack{j=1 \\ j \neq i}}^{N} \sum_{\substack{k=1 \\ k \neq i,j}}^{N} \frac{n_{jk}(i)}{n_{jk}} \tag{2}$$

where n_{jk} is the number of SP's from node j to k, whereas $n_{jk}(i)$ is the number of those SP's that contain the node i. A high betweenness can be interpreted as a node which is passed through by SP's connecting the rest of nodes.

Lastly, the closeness centrality is obtained as shown in the following expression:

$$c_i = \frac{1}{\sum_{j=1}^{N} d_{i,j}} \tag{3}$$

There, the closeness of each node c_i is computed from the so-called distance matrix D, where each element $d_{i,j}$ corresponds to the SP from node i to j. Therefore, this quantity accounts for how close a given node is to the rest of the network, in terms of edges needed for other nodes to be reached.

3. Results and discussion

3.1. Degree distributions

After the proposed methodology has been explained, authors have analyzed firstly how the UDVG degree distribution differs from the regular VG with the same time series. It served as a preliminary study, before tackling the identification of relevant points in the signal, which is the main objective of this work.

Fig. 2 represents two theoretical time series have been employed to test the method. The first one of them is obtained from a fractional Brownian motion with Hurst exponent $H = 0.5$ and 10^4 points. The second one corresponds to a random series with 10^5 points. The reason for choosing them is that they are standard well-known series that are frequently used within this type of studies with VGs. This figure shows the series (a and b) and their respective degree distribution computed for both approaches (c and d). It can be inferred that the distributions that arise from using the UDVG are almost identical to the VG ones. Therefore, at least for these types of series, the VG and UDVG degree distributions describe the same properties of the underlying time series.

In the case of the fractional Brownian motion, they also present curves which can be adjusted to the same power law. Thus, this might indicate that UDVG would be also suitable for describing scale-free networks, such as those extracted from these type of series, which are fractal (Lacasa et al., 2009). For the random time series, the result is a distribution with a tail that follows an exponential trend, as expected (Lacasa et al., 2008).

Once it was observed that UDVG and VG obtain the same results for the theoretical time series, authors have tested the photochemical time series, which are the focus of this study. These correspond to three different signals: two from O_3 and NO_2 concentration, and the other one temperature, all of them from the same year (2017), as previously stated. The reason behind choosing one complete year in this particular part of the study is that for a reliable comparison of degree distributions, a considerable amount of points in the time series is required (Carmona-Cabezas et al., 2019a). Since the resolution of the measurements is 1 h, it has
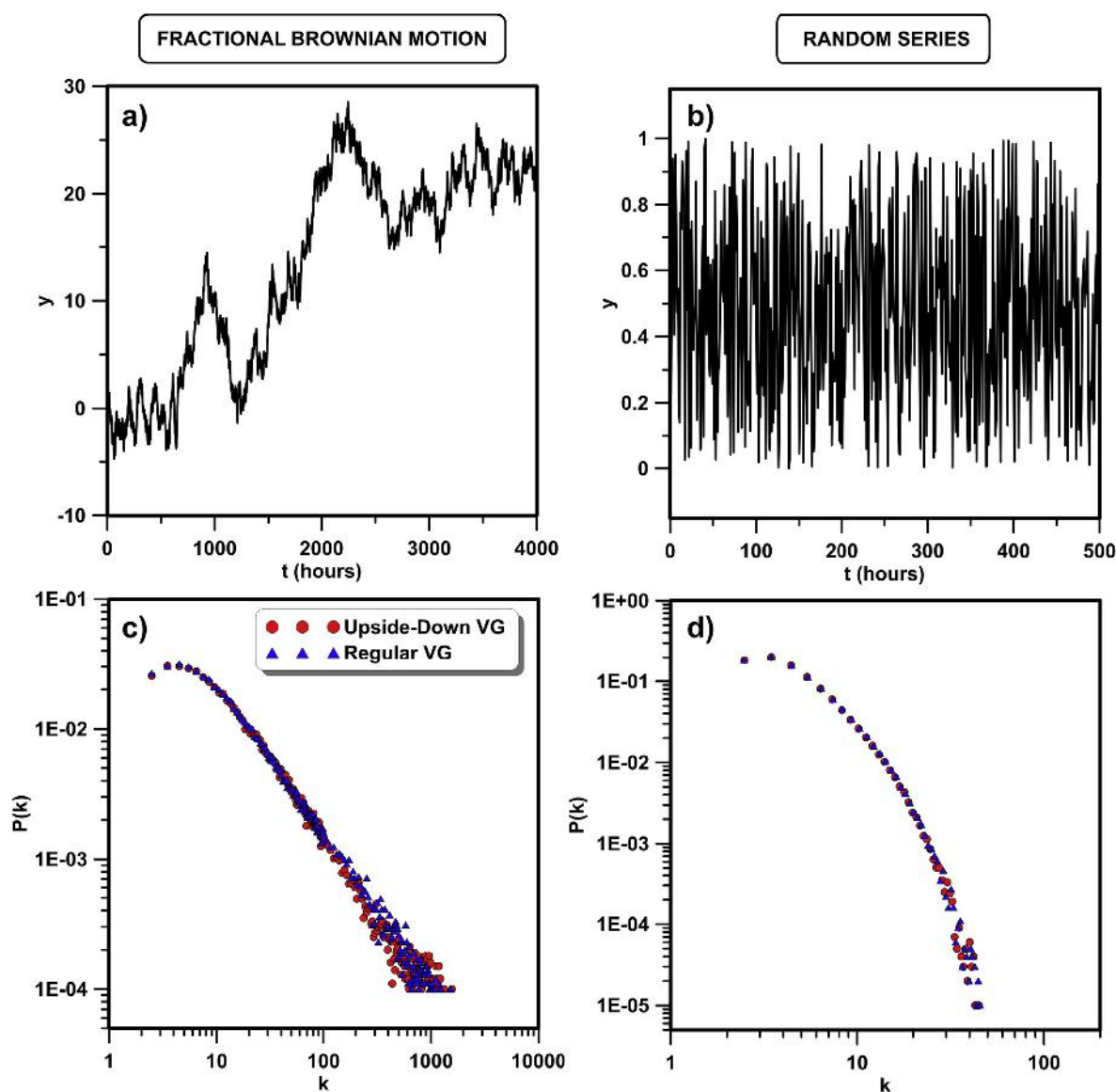

Fig. 2. Top: fractional Brownian motion signal with Hurst exponent $H = 0.5$ and 10^4 points (a) and 500 points from a random series (b). Bottom: The degree distribution computed from the complex networks obtained for both series, by employing the VG and UDVG.
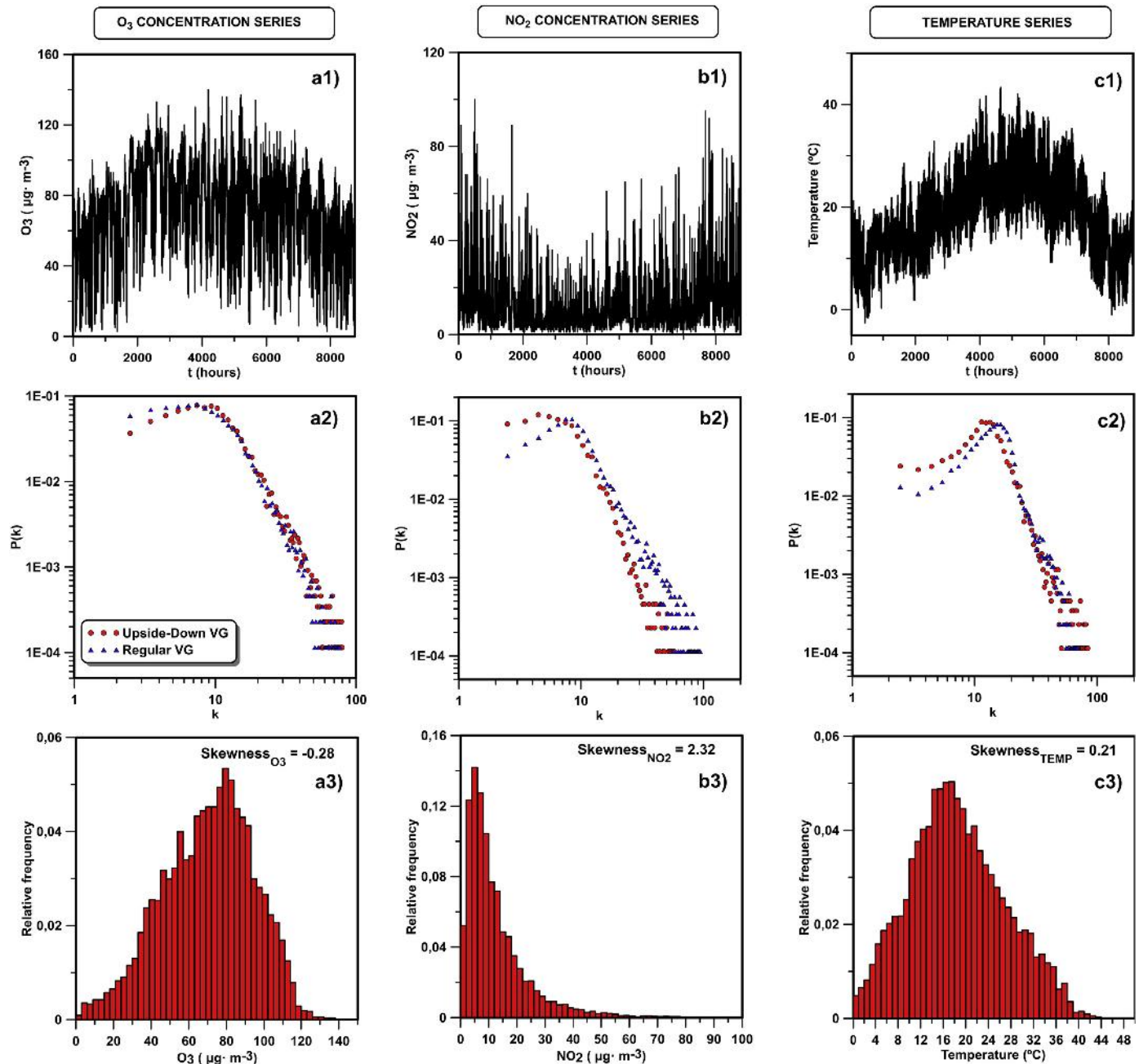
Fig. 3. Top: Ozone (O_3), nitrogen dioxide (NO_2) concentration and temperature annual temporal series (a1, b1 and c1). Middle: Degree distributions computed with regular VG and UDVG (a2, b2 and c2). Bottom: frequency distributions of the pollutant concentrations and temperature time series (a3, b3 and c3).

been observed after several tests that taking only monthly samples for comparison could give misleading results. It should be underscored that the same is not true for the later analysis of singular episodes, since that is a local study and does not depend in the extension of the pollutant time series.

The three signals are depicted in the upper part of Fig. 3 (a1, b1 and c1). Conversely to what was observed in the previous case, now Fig. 3 (a2, b2 and c2) display slight differences between the degree distributions of the classic VG and the UDVG. Nevertheless, a clear power law behavior is observed in every case. This is in accordance with previous works, where ozone (O_3) and nitrogen dioxide (NO_2) time series exhibited scale-free behavior, as a consequence of the multifractal nature of its dynamics (He, 2017; Pavón-Domínguez

et al., 2015). The observed contrasts are more pronounced in the case of the nitrogen dioxide (NO_2) concentration time series, clearly showing a marked difference in the slope of the distribution tail (the γ -exponent). Authors attribute this effect to the difference between the three underlying time series. The ground-level ozone signal exhibits a pattern that equally presents singular minima and maxima, and the same can be said about the temperature. Therefore, the frequency distributions of their concentrations will have roughly symmetric shapes. On the other hand, the same cannot be argued for the nitrogen dioxide (NO_2) concentration. Minima and maxima values are not distributed evenly along the time series, which is clear in Fig. 3c. The maxima are rather infrequent and singular in comparison to the minima, which are much more

common, as most of the values are very close to zero. Therefore, one could expect the probability distribution of the concentration of nitrogen dioxide (NO_2) to be non-symmetric. To investigate that, the most suitable method is to inspect the skewness (S) for each time series. This quantity describes the asymmetry of the probability distribution of a given real measure around its mean. When skewness is equal to zero, it means that the distribution is symmetric respect to its mean, being the opposite case ($S \neq 0$) for non-symmetric distributions. In Fig. 3, probability distributions of concentrations and temperature are depicted with their respective skewness value. In the case of O_3 and temperature (Fig. 3 a3 and c3), the distributions are almost symmetric, as mentioned before, with skewness close to zero ($S_{O_3} = -0.28$ and $S_{temp} = 0.21$). Despite this, a mild deviation between the regular and inverted distributions can be observed, leading to the low negative skewness that is observed. On the contrary, a positive skewness value ($S_{NO_2} = 2.32$) of nitrogen dioxide (NO_2) concentration is clearly seen (Fig. 3b3), i.e. low values with respect to the mean are highly frequent. Therefore, the concentration of nitrogen dioxide (NO_2) in San Fernando reaches low peaks many times during the month, while the high accumulations of this noxious gas are much rarer.

For the case of temperature, this symmetry can be interpreted as the relatively regular behavior of day and night values, meaning that the appearances of singular episodes of low and high temperature will be linked during the year, depending on the meteorological conditions of each season. On the other hand, one could expect this difference between nitrogen dioxide (NO_2) and tropospheric ozone (O_3), regarding the symmetry of the degree distribution. Both gases are correlated through the simplified photochemical reaction $NO_2 + O_2 \leftrightarrow O_3 + NO$. Production and destruction of ozone will occur during day and night times respectively. The photolysis that leads to the ozone accumulation and the reach of the photostationary state regularly happens during mid-day, when there is radiation available. The sense of the reaction is reverted during nighttime in the absence of light. Although the quantitative concentration levels may vary depending on factors such as wind speed, temperature or mixing height, the distributions of maxima and minima could be expected to be symmetric as it is seen for the O_3. However, the nitrogen dioxide (NO_2) intervenes in other reactions that could lead to the appearance of singular minima in its concentration. One example is the aldehyde production through interaction with VOCs, which results on a lower rate of NO_2 - NO reaction. For a deeper understanding of this, further analysis with NO and VOCs time series would be necessary.

Authors would like to point out that, for the previous theoretical series this relation is also observed, being their computed skewness values very close to zero ($S_{brownian} = 0.13$ and $S_{random} = 3.10 \cdot 10^{-4}$), as expected since their distribution where almost perfectly coincident for VG and UDVG.

3.2. Identification of hubs

After the preliminary study of probability distributions has been carried out, a pointwise study of NO_2 and O_3 concentrations and temperature is undertaken here. Fig. 4 depicts a comparison between the hubs computed by applying the VG on the unvaried time series and those of the inverted one. Now only one real time series is shown, because the actual interest here is to observe the differences between UDVG and regular VG when detecting the singular extremes. In this case, only one month (July) from the ozone concentration time series was chosen for the sake of clarity (see Fig. 4a). This month was chosen because, in this location, July is the period of the year were the most severe episodes of ozone pollution

occur. In the next two figures (Fig. 4b and c), the normalized betweenness and degree values are shown for both networks (blue is for the original VG, while red for the UDVG). Only these two centrality parameters were chosen in this case because they have clearer signals. The three centrality parameters presented in the methodology section of this work will be used in later discussions.

It can be regarded in Fig. 4 the fact that both networks (the regular and inverted one) are able to identify extrema in the time series in a complementary manner, as anticipated. While the regular VG hubs correspond maximal episodes of tropospheric ozone concentration (which has been already used), the UDVG obtained ones do the same with minima of the concentration. These latter correspond to the nighttime, when the photochemical reaction is unbalanced towards NO_2 formation in the absence of radiation. The actual physical interpretation of the different centrality parameters can be observed in the previous related work (Carmona-Cabezas et al., 2020) for the regular VG hubs. Additionally, it will be explained for the UDVG case in the last figures.

Moreover, it must be noticed how the hubs from betweenness coincide with those of the degree, while the opposite case is not always true. Therefore, the first one may be a more selective approach to identify singular nodes in a signal, as it has been discussed in a previous work (Carmona-Cabezas et al., 2019b). This filtering feature might be useful for the use of this technique on environmental series where the density of extrema is considerably high.

Once the difference between the VG and UDVG hubs has been discussed, authors propose an approach for combining the information given by both networks. The aim is to yield a more complete technique for future investigations to analyze pollutant time series. The combined parameters tested here simply consist on adding each VG centrality parameter to the opposite of the one computed using UDVG. To make it clearer, for the betweenness, degree and closeness:

$$\begin{cases} b_i^{comb} = b_i^{VG} - b_i^{UDVG} \\ k_i^{comb} = k_i^{VG} - k_i^{UDVG} \\ c_i^{comb} = c_i^{VG} - c_i^{UDVG} \end{cases} \quad (4)$$

This transformation is useful for the identification of singularities or extreme values, considering both the minima and maxima values. It is based on the fact that when the VG maps a hub, the UDVG will not, since they are complementary, as exposed in the Methodology section. Thus, the hubs information is not lost by this procedure, because their values will not be canceled out for the case of extremes. Consequently, it improves the differentiation between regular and singular values. This results in the derived combined degree signal being smoother and clearer than in the separated case. For the combined betweenness, there will be almost no difference in the smoothness, since the values that do not correspond to skyline hubs are practically zero in any case.

For clarity reasons, the same structure as in a previous work (Carmona-Cabezas et al., 2020) has been followed for the plots. Hence, the combined betweenness is computed first and from it, the five most pronounced peaks are chosen automatically. Equivalent results can be yielded by selecting a greater number of peaks, as it has been tested. A criterion for it was indicated in the mentioned previous work (Carmona-Cabezas et al., 2020). Afterwards, the remaining centrality measures were analyzed in the positions where the first peaks are located. It must be highlighted that all the plotted parameters are normalized to the maximum absolute value of each one, for the sake of comparison.

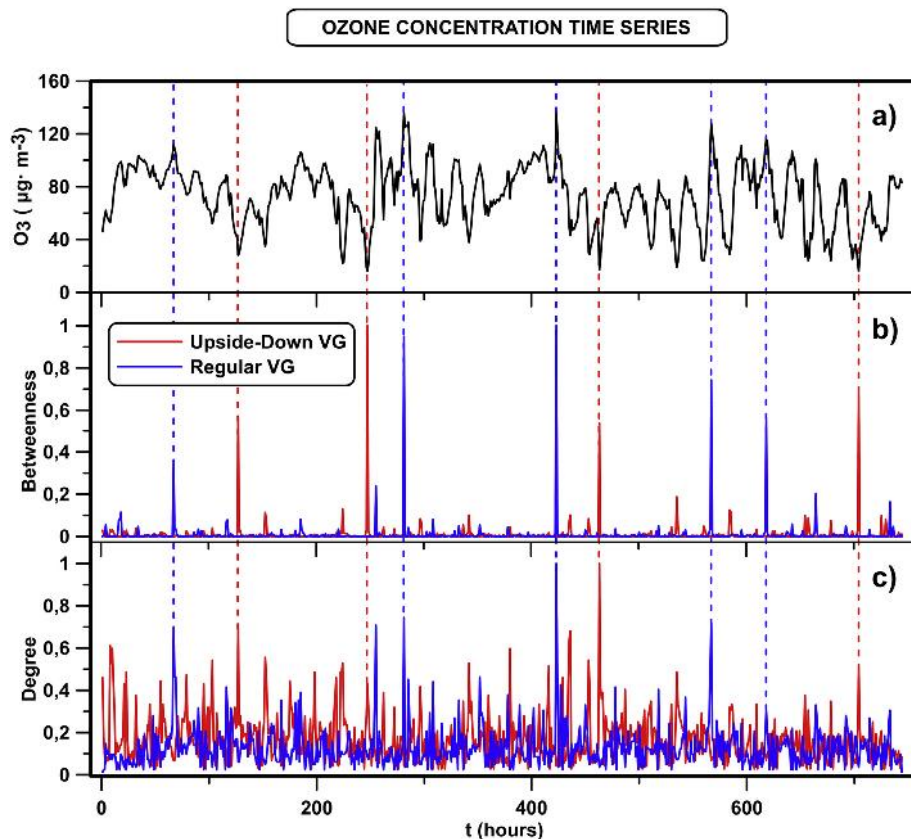In Fig. 5, this explained procedure is performed using the series

Fig. 4. a) Ozone concentration time series from a selected month (July). b) and c) Normalized betweenness and degree centrality values, respectively, for each point in the time series, from the two graphs studied (VG in blue and UDVG in red). The dashed lines are used to highlight the hubs positions and compare them in the three plots. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

of tropospheric ozone previously introduced. As it is easily seen, the accordance between the different studied parameters is adequate, as it was expected. The combination of the VG and UDVG still preserve the capability to identify extrema by the different centrality parameters. The smoothest series corresponds to that of the betweenness as previously explained, followed by the degree and finally by the closeness. It is in accordance to what was observed using the less complete method in the previous paper (Carmona-Cabezas et al., 2020). It must be stressed that the order of the magnitude of the different peaks is not conserved in the different combined centrality parameters. For instance, in Fig. 5c, the peak 2 is the most negative one, while in Fig. 5d and e are the peaks 3 and 5 respectively. This is due to the different physical meanings of each parameter related to the concentration time series. Therefore, this should be taken in consideration if different parameters are used to compare ozone (or other pollutant) extreme concentration episodes in future studies.

The first one of them is the combined betweenness (Fig. 5c). In order to understand the usefulness of this parameter to the photochemical pollution, it must be pointed out that in a previous study (Carmona-Cabezas et al., 2019b) skyline hubs were related to values of the series which can give more information about its upper envelope. In short, one of the detected O_3 singularities may be considered as an unlikely high episode of ozone concentration in relation to other maxima in the same series. This means that if ozone daily maximal concentrations were raising for several consecutive days, a peak in the betweenness indicates that after this encountered, the trend is likely to change to a downwards one. Conversely, translating this to the inverted series (and the resulting

UDVG), the same could be inferred about minimal night O_3 concentrations. Environmentally speaking, a change in the tendency could be a pointer to an alteration of the previous ambient conditions that would lead to an abnormal shift in the height of the mixing layer, for instance. Therefore, the combined betweenness can serve as a more complete warning tool, pointing changes in the conditions that affect the temporal evolution of pollutants concentration, while the previous approach would only yield insight on the upper one, limiting the analysis.

Regarding the next complex network indicator, the combined degree (Fig. 5d), many works have been devoted to its study (Pierini et al., 2012; Zhou et al., 2017). It is known that a degree hub is associated to a specially high ozone concentration episode (Carmona-Cabezas et al., 2019a). At the position where the hubs are encountered, the gas has reached a peculiarly high concentration. This condition is less restrictive, as every day it is fulfilled. As a result, the number of this type of peaks is greater, compared to the betweenness. In this case, a peak would not be necessarily associated to a change in the prior tendency of the O_3 concentrations. When the regular VG results are combined with the UDVG, the unlikely low values of concentration can be identified as well. Again, the identification of rare concentrations of O_3 is improved by this combination, getting at the same time the information from low and high values from one single parameter. Here, the sense of "singularity" in the ozone is referred only to its magnitude, and not to the trend of the previous and posterior days, as in the betweenness.

Fig. 5e) illustrates the closeness centrality results. In previous research, this quantity was mainly used for theoretical purposes.
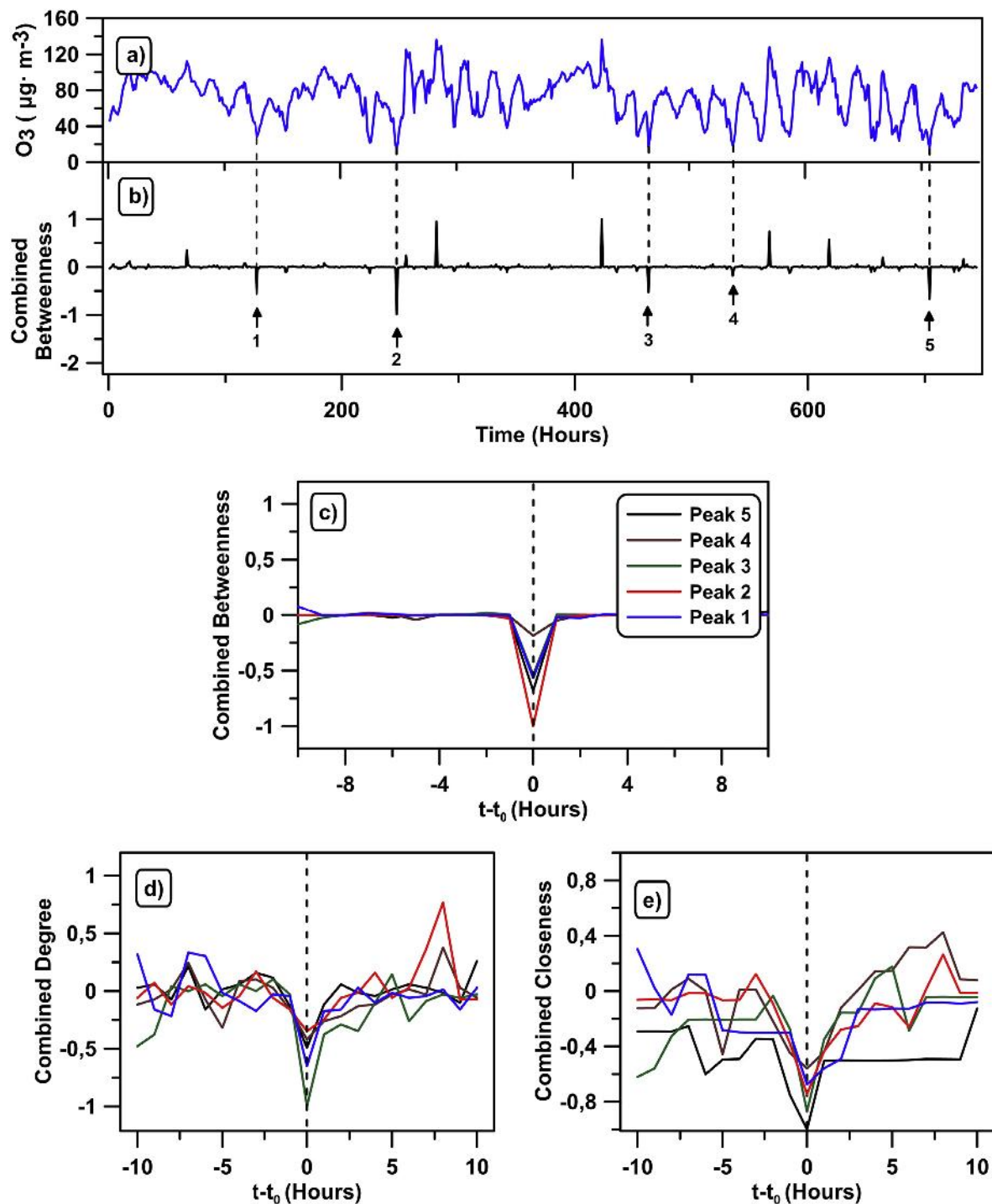
Fig. 5. Ozone concentration time series (a) with the combined betweenness computed from the UDVG and VG (b). Plots from c) to e) show the complex network indicators: betweenness, degree and closeness in the selected five negative peaks.

Nonetheless, it was demonstrated recently that it could identify singularities as the previous ones, but with a different criterion (Carmona-Cabezas et al., 2020). The peaks of this magnitude were related to high concentrations of ozone episodes surrounded by concave up tendency. This quantity was found to be noisier than betweenness and degree, and so it is as well here. As in the previous parameters, now the combined quantity (more specifically the negative part), gives additionally information about the points where a minimal rare concentration value is found, surrounded by

a concave down accumulation of values (the reversed shape with respect to the regular VG). In the context of photochemical pollution, it would mean that this could be used to identify daily high concentrations (during the photostationary state) that somehow drop, due for instance to unexpected atmospheric conditions.

The selected minima correspond to the 6th, 11th, 20th, 23rd and 30th of July, all of them occurring between 6:00 a.m. and 7:00 a.m. (UTC+1) as could be expected. The photochemical reaction is reverted during the nighttime and most of the tropospheric ozone
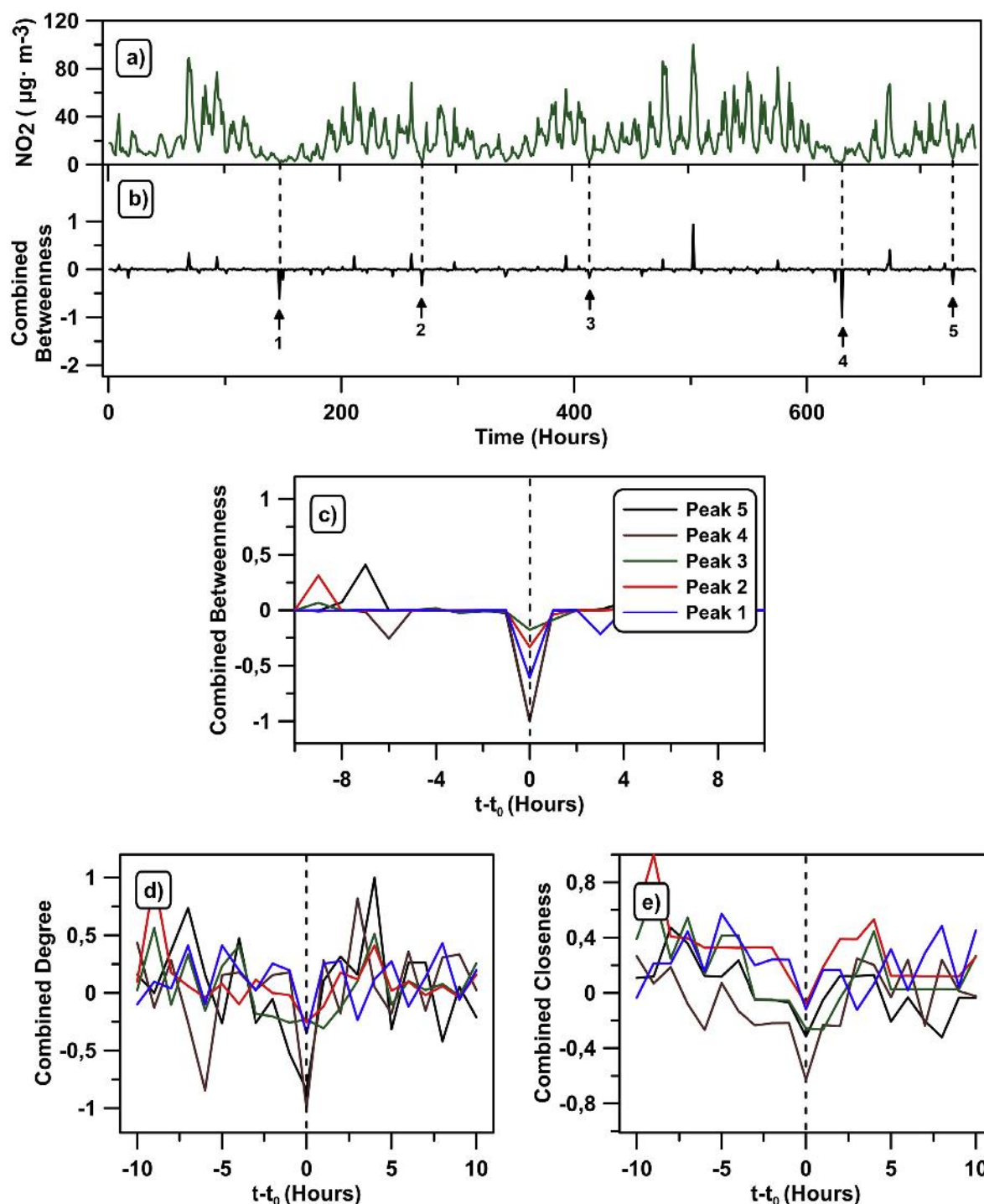
Fig. 6. Nitrogen dioxide (NO_2) concentration time series (a) with the combined betweenness computed from the UDVG and VG (b). Plots from c) to e) show the complex network indicators: betweenness, degree and closeness in the selected five negative peaks.

(O_3) (produced during the previous day) is recombined with NO to yield NO_2 in the absence of light. After this time, there is radiation available and its concentration has an upward trend. In the previous work, the singular high episodes between 2:00 p.m. and 6:00 p.m. (UTC+1), which corresponds to the opposite case.

The next graph (Fig. 6) shows the results obtained for the nitrogen dioxide (NO_2) concentration time series, which as seen before, has a different minima and maxima behavior. For this study, the studied month is January as in the previous work, since in this region that is period of the year when less photochemical activity takes place. Therefore, reactions with other chemicals (such as VOCs to yield aldehydes) could play a more important role, leading to more singular extrema.

Once again, there is a good fit between the different parameters, although in this case, the combined degree is noisier and not as clear as before. This might be caused by the accumulation of low values of nitrogen dioxide (NO_2) concentration that make the distribution to be more asymmetric, as discussed (see Fig. 3). The

greater number of reactions that involve NO_2 might increase the number of singularities, being the degree noisier as a result.

In this case, the selected concentrations of NO_2 are in the 7th, 12th, 18th, 27th and 31st of January, between 2:00 a.m. and 5:00 a.m. (UTC+1). Regarding the high singularities investigated in the previous work, there was no consistent time frame where it could be encountered. Also, it is well known that there is a marked difference between concentrations during weekends and weekdays (Qin, 2004), which could be another possible cause for this

uncertainty.

Finally, in Fig. 7 the temperature time series is studied locally as in the previous two cases. Now the selected month is October, in order to observe singular episodes of this quantity. Due to the oceanic influence, the temperature is stable throughout almost all the year. Nevertheless, it is more unstable in autumn in this area, as discussed in previous works (Dueñas et al., 2004).

It is clearly seen that for temperature there is also concordance between the combined betweenness and the rest of combined
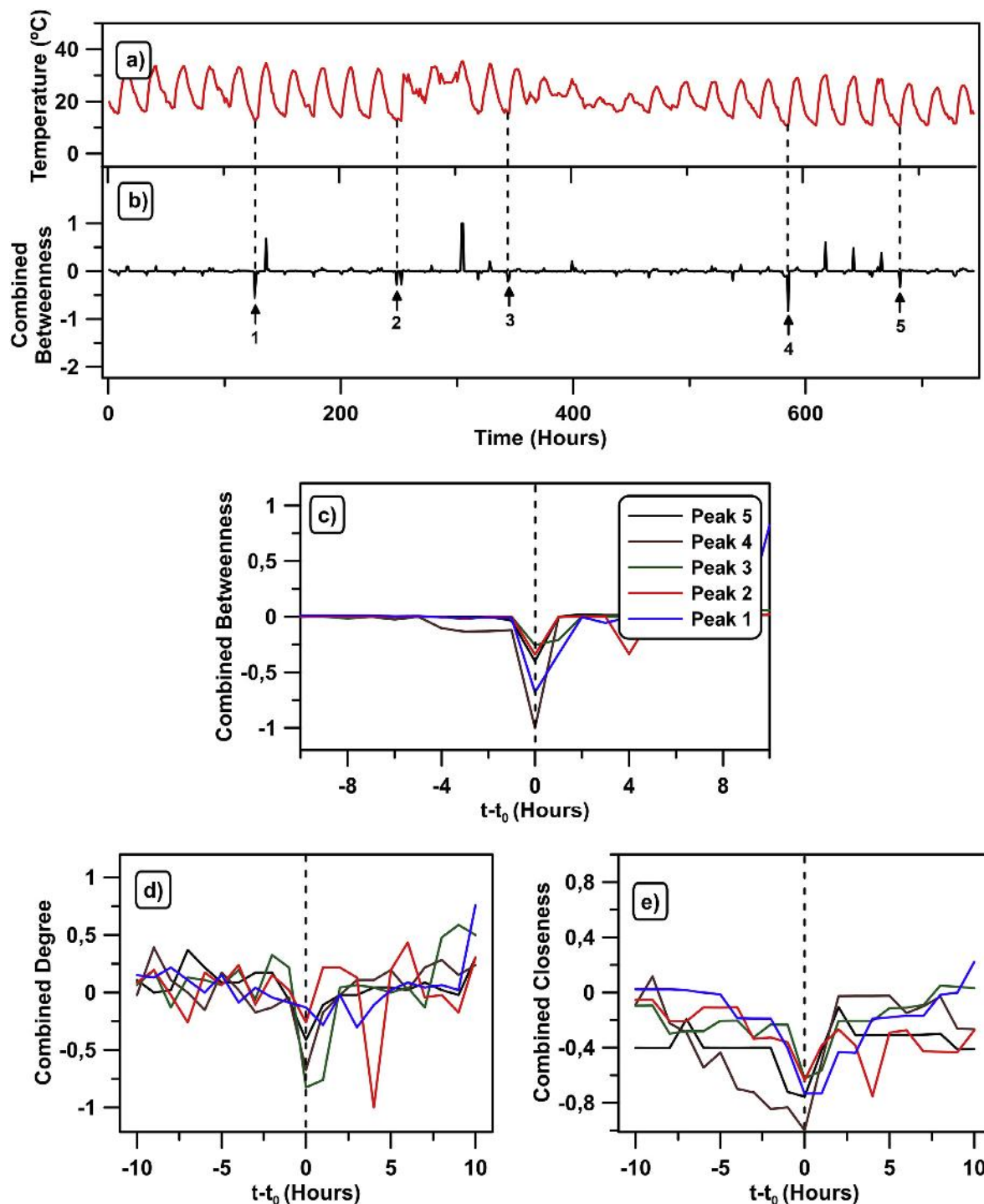

Fig. 7. Temperature time series (a) with the combined betweenness computed from the UDVG and VG (b). Plots from c) to e) show the complex network indicators: betweenness, degree and closeness in the selected five negative peaks.

centrality parameters. Now the combined degree signal has less noise than in the case of NO_2, except for Peak 2. This one is more difficult to identify due to the fact that there are two betweenness peaks very close to each other (see Fig. 7 b).

The temperature singular minima that have been selected, following the previous criterion, correspond to the 6th, 11th, 15th, 25th and 29th of October, between 5:00 a.m. and 8:00 a.m. (UTC+1). It could be easily expected, since it is the time when the minimum temperature is reached every day. Even during the days in which the temperature becomes more unpredictable (around the middle part of the month), these minima can be observed with a relatively constant frequency.

4. Conclusions

An improvement of a singularity detection technique is tested for its application on photochemical time series in this manuscript. It adds the possibility of describing singular minima and maximal singular values at the same time, making it a more complete tool. Authors believe that it may have a great potential for monitoring and analyzing pollutant and atmospheric time series in the future.

The degree distributions obtained have been compared, proving that UDVG inherits the nature of the original NO_2, O_3 and temperature time series. Moreover, different theoretical series have been tested, proving the suitability of both VG and UDVG. It has been found that those distribution are coincident for tropospheric ozone (O_3) and temperature, while they are not for the nitrogen dioxide (NO_2). Their disparity has been related to the greater number of reactions that involve NO_2, such us its interaction with VOCs to yield aldehydes. This must be investigated more in detail in a future study, applying different complex networks tools developed to series of NO_x, VOCs and O_3 at the same time.

Furthermore, the usefulness of UDVG for singular minima detection has been successfully proven on the NO_2, O_3 and temperature series. The combination of VG and UDVG parameters (degree, betweenness and closeness) is proposed as a more exhaustive method, compared to only employing VG. Due to their complementary nature, these combinations store the original information of the most central nodes, showing all the relevant information at a glance. To authors' mind, this can widen the range of the research applications of complex networks for photochemical pollution in a future.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

Rafael Carmona-Cabezas: Conceptualization, Methodology, Software, Validation, Formal analysis, Data curation, Investigation, Writing - original draft, Writing - review & editing. **Javier Gómez-Gómez:** Software, Investigation, Writing - original draft. **Eduardo Gutiérrez de Ravé:** Project administration, Funding acquisition, Supervision, Writing - review & editing. **Elena Sánchez-López:** Visualization. **João Serrano:** Resources. **Francisco José Jiménez-Hornero:** Project administration, Funding acquisition, Supervision, Writing - review & editing.

Acknowledgements

The FLAE approach for the sequence of authors is applied in this

References

Agryzkov, T., Tortosa, L., Vicent, J.F., 2019. A variant of the current flow betweenness centrality and its application in urban networks. Appl. Math. Comput. 347, 600–615. https://doi.org/10.1016/j.amc.2018.11.032.

Bielinskyi, A.O., Soloviev, V.N., 2018. Complex network precursors of crashes and critical events in the cryptocurrency market. In: Presented at the Computer Science & Software Engineering 2018. Kryvyi Rih, Ukraine, pp. 37–45.

Boccaletti, S., Latora, V., Moreno, Y., Chavez, M., Hwang, D., 2006. Complex networks: structure and dynamics. Phys. Rep. 424, 175–308. https://doi.org/10.1016/j.physrep.2005.10.009.

Carmona-Cabezas, R., Ariza-Villaverde, A.B., Gutiérrez de Ravé, E., Jiménez-Hornero, F.J., 2019a. Visibility graphs of ground-level ozone time series: a multifractal analysis. Sci. Total Environ. 661, 138–147. https://doi.org/10.1016/j.scitotenv.2019.01.147.

Carmona-Cabezas, R., Gómez-Gómez, J., Ariza-Villaverde, A.B., Gutiérrez de Ravé, E., Jiménez-Hornero, F.J., 2019b. Can complex networks describe the urban and rural tropospheric O3 dynamics? Chemosphere 230, 59–66. https://doi.org/10.1016/j.chemosphere.2019.05.057.

Carmona-Cabezas, R., Gómez-Gómez, J., Gutiérrez de Ravé, E., Jiménez-Hornero, F.J., 2020. Checking complex networks indicators in search of singular episodes of the photochemical smog. Chemosphere 241, 125085. https://doi.org/10.1016/j.chemosphere.2019.125085.

Cheng, W., Li, H., Zhang, X., Sun, W., Chong, K.C., Lau, S.Y.-F., Yu, Z., Liu, S., Ling, F., Pan, J., Chen, E., 2020. The association between ambient particulate matters, nitrogen dioxide, and childhood scarlet fever in Hangzhou, Eastern China, 2014–2018. Chemosphere 246, 125826. https://doi.org/10.1016/j.chemosphere.2020.125826.

Domínguez-López, D., Adame, J.A., Hernández-Ceballos, M.A., Vaca, F., De la Morena, B.A., Bolívar, J.P., 2014. Spatial and temporal variation of surface ozone, NO and NO2 at urban, suburban, rural and industrial sites in the southwest of the Iberian Peninsula. Environ. Monit. Assess. 186, 5337–5351. https://doi.org/10.1007/s10661-014-3783-9.

Donner, R.V., Donges, J.F., 2012. Visibility graph analysis of geophysical time series: potentials and possible pitfalls. Acta Geophys. 60, 589–623. https://doi.org/10.2478/s11600-012-0032-x.

Dueñas, C., Fernández, M.C., Cañete, S., Carretero, J., Liger, E., 2004. Analyses of ozone in urban and rural sites in Málaga (Spain). Chemosphere 56, 631–639. https://doi.org/10.1016/j.chemosphere.2004.04.013.

Gan, C., Yang, X., Liu, W., Zhu, Q., Jin, J., He, L., 2014. Propagation of computer virus both across the Internet and external computers: a complex-network approach. Commun. Nonlinear Sci. Numer. Simulat. 19, 2785–2792. https://doi.org/10.1016/j.cnsns.2013.12.026.

Guicherit, R., van Dop, H., 1977. Photochemical production of ozone in Western Europe (1971–1975) and its relation to meteorology. Atmos. Environ. 11, 145–155. https://doi.org/10.1016/0004-6981(77)90219-0.

He, H., 2017. Multifractal analysis of interactive patterns between meteorological factors and pollutants in urban and rural areas. Atmos. Environ. 149, 47–54. https://doi.org/10.1016/j.atmosenv.2016.11.004.

Joyce, K.E., Laurienti, P.J., Burdette, J.H., Hayasaka, S., 2010. A new measure of centrality for brain networks. PloS One 5. https://doi.org/10.1371/journal.pone.0012200.

Kampa, M., Castanas, E., 2008. Human health effects of air pollution. Environ. Pollut. 151, 362–367. https://doi.org/10.1016/j.envpol.2007.06.012.

Lacasa, L., Luque, B., Ballesteros, F., Luque, J., Nuño, J.C., 2008. From time series to complex networks: the visibility graph. Proc. Natl. Acad. Sci. Unit. States Am. 105, 4972–4975. https://doi.org/10.1073/pnas.0709247105.

Lacasa, L., Luque, B., Luque, J., Nuño, J.C., 2009. The visibility graph: a new method for estimating the Hurst exponent of fractional Brownian motion. Europhys. Lett. 86, 30001. https://doi.org/10.1209/0295-5075/86/30001.

Lacasa, L., Nicosia, V., Latora, V., 2015. Network structure of multivariate time series. Sci. Rep. 5, 15508. https://doi.org/10.1038/srep15508.

Lacasa, L., Toral, R., 2010. Description of stochastic and chaotic series using visibility graphs. Phys. Rev. E 82, 036120. https://doi.org/10.1103/PhysRevE.82.036120.

Liu, C., Zhan, X.-X., Zhang, Z.-K., Sun, G.-Q., Hui, P.M., 2015. Events determine spreading patterns: information transmission via internal and external influences on social networks. New J. Phys. 17, 113045. https://doi.org/10.1088/1367-2630/17/11/113045.

Loutridis, S.J., 2007. An algorithm for the characterization of time-series based on local regularity. Phys. Stat. Mech. Its Appl. 381, 383–398. https://doi.org/10.1016/j.physa.2007.03.012.

Lozano-Pérez, T., Wesley, M.A., 1979. An algorithm for planning collision-free paths among polyhedral obstacles. Commun. ACM 22, 560–570. https://doi.org/10.1145/359156.359164.

Mali, P., Manna, S.K., Mukhopadhyay, A., Haldar, P.K., Singh, G., 2018. Multifractal analysis of multiparticle emission data in the framework of visibility graph and

sandbox algorithm. Phys. Stat. Mech. Its Appl. 493, 253—266. https://doi.org/10.1016/j.physa.2017.10.015.

Miao, W., Huang, X., Song, Y., 2017. An economic assessment of the health effects and crop yield losses caused by air pollution in mainland China. J. Environ. Sci. 56, 102—113. https://doi.org/10.1016/j.jes.2016.08.024.

NAPCA, 1970. NAPCA publication No. AP 63. In: Air Quality Criteria for Photochemical Oxidants..

Newman, M.E.J., 2003. The structure and function of complex networks. SIAM Rev. 45, 167—256. https://doi.org/10.1137/S003614450342480.

Pavón-Domínguez, P., Jiménez-Hornero, F.J., Gutiérrez de Ravé, E., 2015. Joint multifractal analysis of the influence of temperature and nitrogen dioxide on tropospheric ozone. Stoch. Environ. Res. Risk Assess. 29, 1881—1889. https://doi.org/10.1007/s00477-014-0973-5.

Pierini, J.O., Lovallo, M., Telesca, L., 2012. Visibility graph analysis of wind speed records measured in central Argentina. Phys. Stat. Mech. Its Appl. 391, 5041—5048. https://doi.org/10.1016/j.physa.2012.05.049.

Qin, Y., 2004. Weekend/weekday differences of ozone, NOx, Co, VOCs, PM10 and the light scatter during ozone season in southern California. Atmos. Environ. 38, 3069—3087. https://doi.org/10.1016/j.atmosenv.2004.01.035.

Sannino, S., Stramaglia, S., Lacasa, L., Marinazzo, D., 2017. Visibility graphs for fMRI data: multiplex temporal graphs and their modulations across resting state networks. Netw. Neurosci. 208—221. https://doi.org/10.1101/106443.

Shang, P., Lu, Y., Kama, S., 2006. The application of Hölder exponent to traffic congestion warning. Phys. Stat. Mech. Its Appl. 370, 769—776. https://doi.org/10.1016/j.physa.2006.02.032.

Song, C., Havlin, S., Makse, H.A., 2006. Origins of fractality in the growth of complex networks. Nat. Phys. 2, 275—281. https://doi.org/10.1038/nphys266.

Soni, G., 2019. Signed Visibility Graphs of Time Series and Their Application to Brain Networks. University of British Columbia, Okanagan.

Stam, C.J., 2010. Characterization of anatomical and functional connectivity in the brain: a complex networks perspective. Int. J. Psychophysiol. 77, 186—194. https://doi.org/10.1016/j.ijpsycho.2010.06.024.

Turner, A., Doxa, M., O'Sullivan, D., Penn, A., 2001. From isovists to visibility graphs: a methodology for the analysis of architectural space. Environ. Plann. Plann. Des. 28, 103—121. https://doi.org/10.1068/b2684.

Yue, H., Yan, W., Ji, X., Zhang, Y., Li, G., Sang, N., 2018. Maternal exposure to NO2 enhances airway sensitivity to allergens in BALB/c mice through the JAK-STAT6 pathway. Chemosphere 200, 455—463. https://doi.org/10.1016/j.chemosphere.2018.02.116.

Zhou, C., Ding, L., Skibniewski, M.J., Luo, H., Jiang, S., 2017. Characterizing time series of near-miss accidents in metro construction via complex network theory. Saf. Sci. 98, 145—158. https://doi.org/10.1016/j.ssci.2017.06.012.

Zou, Y., Donner, R.V., Marwan, N., Donges, J.F., Kurths, J., 2019. Complex network approaches to nonlinear time series analysis. Phys. Rep. 787, 1—97. https://doi.org/10.1016/j.physrep.2018.10.005.

Publication 8
Multifractal Characterisation of Particulate Matter (PM$_{10}$) Time Series in the Caribbean Basin Using Visibility Graphs

Journal

Atmospheric Pollution Research (Ed. Elsevier)

Position / Quartile

(80/265) / Q2 (Environmental Sciences)

Impact factor

3.527 (2019)

Multifractal characterisation of particulate matter (PM_{10}) time series in the Caribbean basin using visibility graphs

Thomas Plocoste [a,b,*], Rafael Carmona-Cabezas [c], Francisco José Jiménez-Hornero [c], Eduardo Gutiérrez de Ravé [c], Rudy Calif [b]

[a] Department of Research in Geoscience, KaruSphère SASU, Abymes, 97139, Guadeloupe (F.W.I.), France
[b] Univ Antilles, LaRGE Laboratoire de Recherche en Géosciences et Energies (EA 4935), F-97100, Pointe-à-Pitre, France
[c] Complex Geometry, Patterns and Scaling in Natural and Human Phenomena (GEPENA) Research Group, University of Cordoba, Gregor Mendel Building (3rd floor), Campus Rabanales, 14071, Cordoba, Spain

ARTICLE INFO

ABSTRACT

A good knowledge of pollutant time series behavior is fundamental to elaborate strategies and construct tools to protect human health. In Caribbean area, air quality is frequently deteriorated by the transport of African dust. In the literature, it is well known that exposure to particulate matter with an aerodynamic diameter of 10 μm or less (PM_{10}) have many adverse health effects as respiratory and cardiovascular diseases. To our knowledge, no study has yet performed an analysis of PM_{10} time series using complex network framework. In this study, the so-called Visibility Graph (VG) method is used to describe PM_{10} dynamics in Guadeloupe archipelago with a database of 11 years. Firstly, the fractal nature of PM_{10} time series is highlighted using degree distribution for all data, low dust season (October to April) and high dust season (May to September). Thereafter, a profound description of PM_{10} time series dynamics is made using multifractal analysis through two approaches, i.e. Rényi and singularity spectra. Achieved results are consistent with PM_{10} behavior in the Caribbean basin. Both methods showed a higher multifractality degree during the low dust season. In addition, multifractal parameters exhibited that the low dust season has the higher recurrence and the lower uniformity degrees. Lastly, centrality measures (degree, closeness and betweenness) highlighted PM_{10} dynamics through the year with a decay of centrality values during the high dust season. To conclude, all these results clearly showed that VG is a robust tool to describe times series properties.

1. Introduction

Every year, millions of tons of mineral dust are transported from the African continent towards the Atlantic Ocean (Schütz, 1980; Kaufman et al., 2005; Koren et al., 2006; Van Der Does et al., 2016). According to the literature, one of the most intense source of dust in the world is the Bodélé depression of the northern Lake Chad Basin (Choobari et al., 2014; Euphrasie-Clotilde et al., 2020). When dust sources are activated, main quantities of emitted dust can be lifted up in high altitudes and transported in the Saharan Air Layer (SAL) towards the Caribbean basin (Cavalieri et al., 2010; Shao et al., 2011; Schepanski, 2018). The SAL is an hot and dry dust-laden layer confined between two inversion layers, i. e. at 1 and 5 *km* height (Carlson and Prospero, 1972; Prospero and

Carlson, 1972; Adams et al., 2012; Tsamalis et al., 2013). Because of intertropical convergence zone latitudinal movement through the year, a seasonal behavior is observed in dust cloud over the Atlantic Ocean (Nicholson, 2000), moving southward ($0 - 10°N$) in winter and northward ($10 - 20°N$) in summer (Moulin et al., 1997; Holz et al., 2004; Adams et al., 2012).

In Caribbean region, air quality is highly deteriorated by the seasonal transport of African dust (Prospero et al., 2014). Contrary to urban areas highly industrialized in Europe, the United States or China, Caribbean islands exhibit low emission of particulate matter linked to anthropogenic activity (Euphrasie-Clotilde et al., 2017; Plocoste et al., 2017, 2018). After the long-range transport from African coast to the Caribbean, i.e. ≈ 5–7 days for 4000 *km* (Velasco-Merino et al., 2018), dust

---

cloud is a mixture of fine and coarse particles (Petit et al., 2005; Van Der Does et al., 2016). The coarse particles (diameter > 70 μm) are expected to deposit within less than one day (Mahowald et al., 2014; Schepanski, 2018) while fine particles (diameter < 10 μm, PM_{10}) may travelled throughout the world (Mahowald et al., 2014). This study focused on PM_{10} which are predominant in dust cloud (Petit et al., 2005).

In the literature, it is well known that PM_{10} exposure has many adverse health effects as respiratory and cardiovascular diseases (Cadelis et al., 2014; Gurung et al., 2017; Zhang et al., 2017; Momtazan et al., 2019). By relating specific respiratory and cardiovascular emergency department admissions, Feng et al. (2019) study shows that elderly (age ≥ 65 years) and male subjects are more susceptible to specific respiratory diseases. In addition, PM_{10} exposure has also been associated with increased hospitalizations (Pun et al., 2014; Su et al., 2016). For these reasons, it is crucial to understand PM_{10} dynamics in order to elaborate strategies and construct tools to protect Caribbean population.

Over the past few decades, numerous studies have shown complex networks efficiency to treat many topics (Amaral and Ottino, 2004; Costa et al., 2007; Amancio et al., 2008; Stam, 2010; Gan et al., 2014; Zou et al., 2019), to cite a few. Usually, a complex network is used to characterize physical processes in nature and can be described as a graph showing non-trivial topological properties. During the last years, many methodologies have been developed in order to transform non-linear time series into complex networks (Zhang and Small, 2006; Lacasa et al., 2008; Zou et al., 2019), in order to take advantage of the many tools developed for analyzing the last ones. Thus, the so-called Visibility Graph (VG) algorithm introduced by Lacasa et al. (2008) is performed here. In several environmental time series (Donner and Donges, 2012; Pierini et al., 2012; Carmona-Cabezas et al., 2020a), this method has already shown that it perfectly inherits the main characteristics of the original time series (Lacasa et al., 2009; Lacasa and Toral, 2010).

In VG frame, two approaches have been frequently used to analyze time series properties, i.e. multifractal analysis and centrality measures (Donner and Donges, 2012; Carmona-Cabezas et al., 2019a, 2019b). Contrary to traditional descriptive statistics which are only based on a single scale analysis, multifractal method assesses raw data over a wider range of temporal scales and exponents order (Zeleke and Si, 2006). In addition, multifractaly is defined by self-similarity properties, i.e. they are scale independent (Mandelbrot, 1982). As regards centrality measures, several studies have shown that nodes distribution can be a useful way to describe time series dynamics (Donner and Donges, 2012; Carmona-Cabezas et al., 2020b). Here, the aim of this study is to perform a profound analysis of PM_{10} concentrations in Guadeloupe archipelago according to African dust seasonality. To our knowledge, no study has yet investigated PM_{10} dynamics in VG frame.

The remainder of this paper is organized as follows. Section 2 depicts the study area and the experimental data. Section 3 describes the theoretical framework. Section 4 presents the results and discusses them. Finally, Section 5 gives the conclusion and an outlook for future studies.

2. Study area and experimental data

Located in the middle of the Caribbean basin (see Fig. 1 (a)), Guadeloupe archipelago (16.25°N − 61.58°W, GPE) is a French overseas region of 390250 inhabitants which cover an area of ~1800 km^2 (Plocoste et al., 2019). Due to its geographical position, GPE experiences a tropical climate with uniformed temperature and continuous high humidity (Plocoste et al., 2014). Hourly PM_{10} measurements are released by GPE air quality network which is managed by Gwad'Air agency (http://www.gwadair.fr/). The Air Quality Stations (AQS) are principally at the center of the island where is located Guadeloupe's economic lung. A quarter of the population lives there and the topography is nearly flat (green space in Fig. 1(b)). In order to study the impact of large scale events on PM_{10} fluctuations, hourly data is converted into daily average values. From 2005 to 2017, only one PM_{10} sensor was available on this air quality network. It was successively placed in Pointe-à-Pitre (AQS1, 16.2422°N 61.5414°W) from 2005 to 2012 and Baie-Mahault (AQS2, 16.2561°N 61.5903°W) from 2015 to nowadays. Due to their geographical proximity (~8.1 km), both measurements are made under the same environmental conditions. Consequently, 11 years of PM_{10} data are available to conduct this study, i.e. 3849 points.

Fig. 2 shows a sequence of PM_{10} daily signal analyzed in this work illustrating huge fluctuations. The strong oscillations observed in the middle of each year are attributed to PM_{10} related to dust outbreaks coming from the African coast from May to September, i.e. the high dust season (Plocoste and Pavón-Domínguez, 2020). From October to April (low dust season), the PM_{10} concentrations are mainly linked to anthropogenic activity and marine aerosols (Euphrasie-Clotilde et al., 2020). In order to analyze the seasonal behavior of PM_{10} data, the signals are reconstructed for the low season (2253 points) and the high season (1596 points) over 11 years.
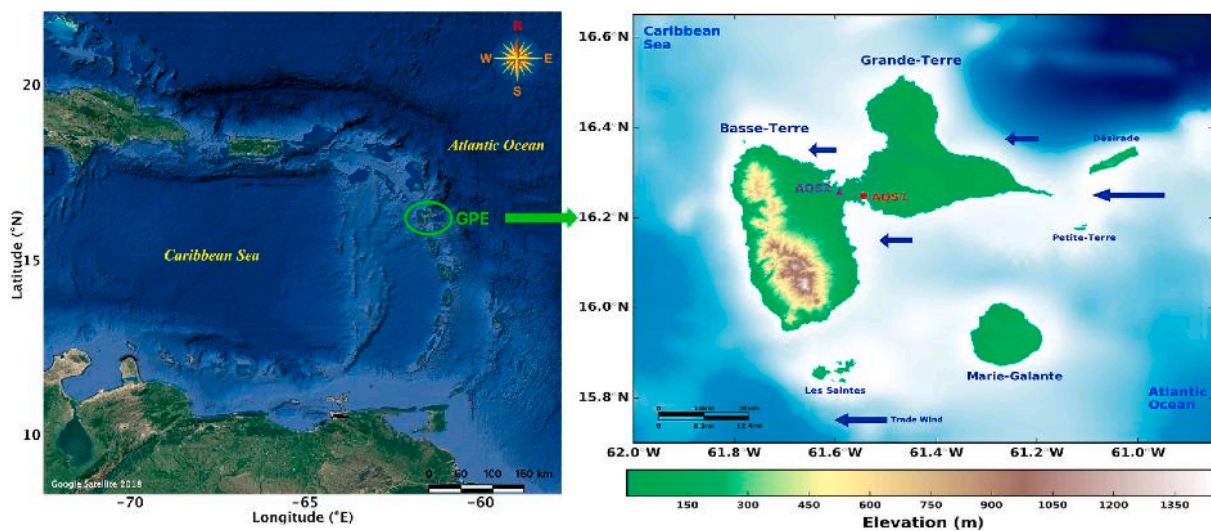

Fig. 1. (a) shows an overview of the Caribbean basin with Guadeloupe archipelago located in the middle of the West Indies arc (16.25°N, − 61.58°W; GPE in green). (b) is a zoom of Guadeloupe archipelago with the locations of the Air Quality stations at Pointe-à-Pitre (AQS1, in a red circle) and Baie-Mahault (AQS2, in a purple triangle). The blue arrows indicate Trade winds direction.
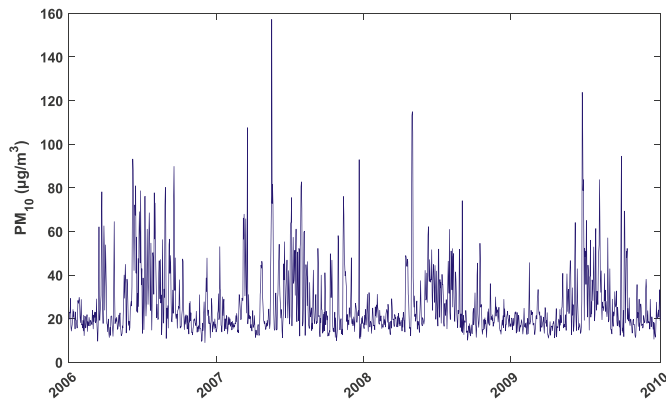
Fig. 2. A sequence of daily average data of PM_{10} concentrations between 2006 and 2009.

3. Theoretical framework

3.1. Visibility graphs

In graph theory, a graph is defined as a set of vertices, points or nodes connected to each other by lines that are usually called edges. Visibility Graph (VG) is a robust tool that transform time series into a graph. Firstly introduced by Lacasa et al. (2008), this new complex network has the benefit to inherit many properties of the original time series.

In order to obtain the VG of a time series, it is essential to establish a criterion to determine which points (or nodes) are linked to each other, i. e. have visibility properties. Consequently, let $y(t)$ be a function of time (time series), two arbitrary points (t_a, y_a) and (t_b, y_b) will have visibility, i.e. will be connected in the graph, if any given point (t_c, y_c) between them $(t_a < t_c < t_b)$ meets the following condition (Lacasa et al., 2008):

$$y_c < y_a + (y_b - y_a)\frac{t_c - t_a}{t_b - t_a} \tag{1}$$

After this condition for every pair of points in the series is checked, it is possible to determine which pairs have visibility or not. As an example, Fig. 3 illustrates the transformation of a time series into a complex network through VG frame.

By performing visibility method, a $N \times N$ adjacency binary matrix is
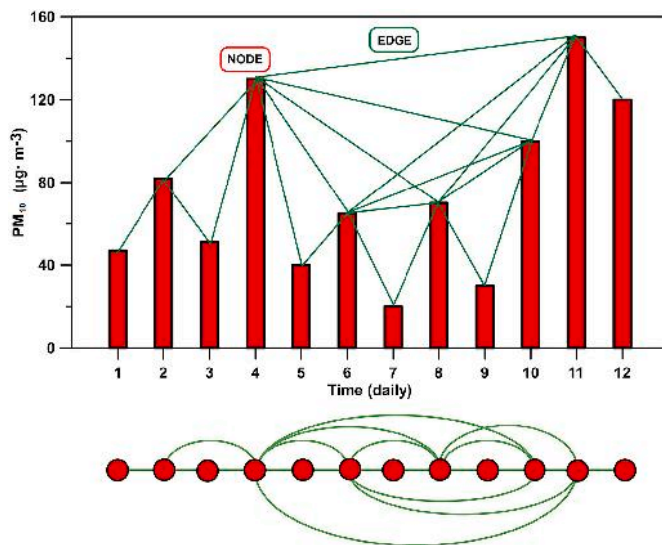

Fig. 3. Sample of PM_{10} time series transformed into a complex network through the visibility algorithm. The nodes of the graph are the data points (red bars), while the links among them (edges) are illustrated as solid lines. At the bottom, all the connections are illustrated in a more visual way.

achieved with N the total number of points in the studied time series. The information of the nodes is given by each row of the matrix. Thus, when $a_{ij} = 1$ the node i and j have visibility while $a_{ij} = 0$ means that no edge connects those two nodes, i.e. no visibility. In order to simplify the algorithm and reduce the computational required time, the resulting matrix exhibits several properties (Carmona-Cabezas et al., 2020b). Firstly, all the elements in the diagonal are zero ($a_{ij} = 0$). Indeed there is no visibility of an element with itself since there are no intermediate points to meets the criterion. Secondly, the condition of visibility is reciprocal, i.e. if a node sees another, this other sees the first one as well. This is translated into the adjacency matrix being symmetric: $a_{ij} = a_{ji}$. Thirdly, the elements surrounding the diagonal are always 1 ($a_{ij} = 1$ for $j = i \pm 1$) since each point always sees the closest previous and next node. When taking all of that into consideration, the visibility V matrix can be described as below (Carmona-Cabezas et al., 2020b):

$$V = \begin{pmatrix} 0 & 1 & \cdots & a_{1,N} \\ 1 & 0 & 1 & \vdots \\ \vdots & 1 & \ddots & 1 \\ a_{N,1} & \cdots & 1 & 0 \end{pmatrix} \tag{2}$$

The degree of a node (k_i) is the number nodes that have reciprocal visibility with the first one. As an example in Fig. 3, the degree of the first node is $k = 1$, for the second one $k = 3$, for the third one $k = 2$, etc. By taking into account the degree of each one of the nodes present in the VG, the degree distribution of the sample ($P(k)$) can be computed. In the literature, many studies have highlighted the efficiency of VG method to describe the nature of a time series (Lacasa et al., 2008; Mali et al., 2018; Carmona-Cabezas et al., 2019a). Consequently, we can firstly overview the behavior of PM_{10} time series by studying the degree distribution before performing more complex multifractal analysis. According to Lacasa et al. (2009), Lacasa and Toral (2010) and Mali et al. (2018), scale free VG (which arise from fractal time series) have degree distributions that can be fitted to power laws such as $P(k) \propto k^{-\gamma}$. This is due to the so-called effect of hub repulsion (Song et al., 2006).

3.2. Multifractal analysis

In order to investigate more precisely the dynamics of a time series, it is essential to analyze more scaling exponents, including through fractal and multifractal analysis (Pamuła and Grech, 2014). Multifractal approach is regarded as the inherent property of complex and composite systems because it gives the possibility of having different densities depending on the region of application (Mandelbrot, 1974). In the literature, there are two ways of depicting multifractality behavior: the generalized fractal dimension D_q (Tél et al., 1989; Block et al., 1990; Schreiber and Grussbach, 1991; Posadas et al., 2001) and the singularity spectrum $f(\alpha)$ (Chhabra et al., 1989; Bacry et al., 1993; Lyra and Tsallis, 1998; Caniego et al., 2005). Multifractal analysis has been widely carried out using the fixed-size algorithms (FSA) (Halsey et al., 1986; Mach et al., 1995). In this study, the sandbox algorithm is applied in complex network frame as Liu et al. (2015) and Mali et al. (2018).

3.2.1. Generalized fractal dimension

The first method introduced to study multifractal formalism are the generalized fractal or Rényi dimensions D_q (Harte, 2001). This latter depicts the scaling exponents of the qth moments of the system and can be defined by (Feder, 1988):

$$D_q = \frac{1}{q-1} \lim_{\delta \to 0} \frac{ln Z_q(\delta)}{ln \delta} \quad \forall q \neq 1 \tag{3}$$

$$D_1 = \lim_{\delta \to 0} \frac{\sum_{i=1}^{N_c(\delta)} \mu_i(\delta) \, ln\mu_i(\delta)}{ln \delta} \tag{4}$$

where q is the moment order, δ is the size of the used cells to cover the system, $Z_q(\delta)$ is the partition function, $N_c(\delta)$ is the number of cells with

length δ and $\mu_i(\delta)$ is the probability measurement of each cell. By taking the limit of D_q when $q \to 1$, D_1 expression is achieved. The partition function previously introduced, is defined as follows (Liu et al., 2015):

$$Z_q(\delta) = \sum_{i=1}^{N_c(\delta)} [\mu_i(\delta)]^q \tag{5}$$

From Equation (3), several parameters can be obtained (Carmona-Cabezas et al., 2019a): i) $D_{q=0}$ is the fractal dimension of the given system or box-counting dimension; ii) $D_{q=1}$ corresponds to the so-called information entropy; iii) $D_{q=2}$ is the correlation dimension. D_q limits for q values ranging between -∞ and +∞, depict the scaling properties of the regions where the regions are respectively more rarefied and concentrated. Usually, the strength of the multifractality can be quantified by $\Delta D_q = max\,D_q - min\,D_q$ (Yu et al., 2016).

3.2.2. Singularity spectrum

Another way to assess the multifractal properties of a time series is the singularity or multifractal spectrum. Traditionally, the Legendre transformation from mass exponents $\tau(q)$ is used to determine it (Muzy et al., 1993). Nevertheless, several authors as Chhabra et al. (1989) and Veneziano et al. (1995) pointed out several drawbacks and errors in this method due to the inclusion of spurious points and error amplification from the derivative. In addition, the Legendre transform does not allow independent measurements of the Rényi spectrum as $\tau(q) = (1 - q)D_q$. To overcome this drawbacks, Chhabra et al. (1989) introduced a method to determine the α-spectrum directly from the original time series. In order to compute the probabilities of the boxes of radius r, this method is based on the normalized measure $\beta_i(q)$ and μ_i from the original time series with (Chhabra et al., 1989):

$$\beta_i(q,r) = [P_i(r)]^q \Big/ \sum_j [P_j(r)]^q \tag{6}$$

where $P_i(r)$ are the different fractal measurements for each box of radius r, i.e. the number of nodes. Thus, from Equation (6), $f(\alpha)$ and α can be computed by using the following equations (Chhabra et al., 1989):

$$f(q) = \lim_{r \to 0} \frac{\sum_i \beta_i(q,r)\,log[\beta_i(q,r)]}{log\,r} \tag{7}$$

$$\alpha(q) = \lim_{r \to 0} \frac{\sum_i \beta_i(q,r)\,log[P_i(r)]}{log\,r} \tag{8}$$

In the literature, α is termed the Lipschitz-Hölder exponent (Posadas et al., 2001). Technically, those parameters are estimated using the slope of $\sum_i \beta_i(q,r)\,log[\beta_i(q,r)]$ over $log\,r$ for $f(\alpha)$; and $\sum_i \beta_i(q,r)\,log[P_i(r)]$ over $log\,r$ for $\alpha(q)$. This slope is characterized by means of a linear regression in the same range of radii where the other fractal measures are calculated (Carmona-Cabezas et al., 2019a). As the generalized fractal dimension, the strength of the multifractality can be estimated by the width of the spectrum with $W = \alpha_{max} - \alpha_{min}$ (Mali et al., 2018).

3.2.3. Sandbox algorithm

For the purpose of easily computing fractal dimensions of real data, Tél et al. (1989) introduced a sandbox algorithm (SBA) which is originated from the box-counting algorithm (Halsey et al., 1986). SBA was developed by Vicsek et al. (1990) and firstly applied to multifractal analysis of complex networks by Liu et al. (2015). Contrary to other box-counting FSA, SBA is able to correctly estimate the side corresponding to the negative values of q from both Rényi and singularity spectra. According to Yu et al. (2016), SBA is the most effective, feasible and accurate algorithm to investigate the multifractal behavior and compute the mass exponent of complex networks.

In more concrete terms, a number of randomly placed boxes are selected for each radius. These latter are always centred in a non-zero point of the system, i.e. a node. Consequently, the entire network is covered with those boxes by choosing a sufficiently high number of them (Carmona-Cabezas et al., 2019a). The probability measurement used to calculate each box (B) is defined by the following equation (Liu et al., 2015):

$$\mu(B) = \frac{M(B)}{M_0} \tag{9}$$

After computed that quantity for each sandbox of a given radius, the generalized fractal dimensions defined in Equation (3) of section 3.2.1 can be improved by the following formula (Liu et al., 2015):

$$D_q^{sb} = \frac{1}{q-1} \lim_{r \to 0} \frac{ln\langle \mu(B)^{(q-1)}\rangle}{ln\,r} \quad \forall q \neq 1 \tag{10}$$

For $q = 1$, SBA can also adjust to Equation (4) with (Mali et al., 2018):

$$D_1^{sb} = \lim_{r \to 0} \frac{\langle ln\mu(B)\rangle}{ln\,r} \tag{11}$$

The SBA procedure for multifractal analysis of complex networks is well described by Liu et al. (2015), Yu et al. (2016) and Carmona-Cabezas et al. (2019a). In this study, the following parameters are used to perform the SBA: i) the interval used for the radii goes from 1 to 30 ($r \in [1, 30]$) according to the distance matrix between the nodes; ii) the range of moments is between $q = -5$ and $q = +5$ with an increment step of 0.25.

3.3. Centrality measures

In VG frame, it is essential to characterize nodes importance of the complex networks. There are different measures to quantified the relevance of individual vertices for structural as well as dynamic characteristics of the underlying system (Donner and Donges, 2012). To achieve this, the centrality measures are frequently used. Firstly introduced by Freeman (1978) for social network study, it was later used in complex networks for many fields (Donner and Donges, 2012; Natapov et al., 2013; Carmona-Cabezas et al., 2020b). Here, we focused on three centrality measures: degree, closeness and betweenness.

3.3.1. Degree centrality

In complex network, the degree centrality captures how many destinations can be seen from each node within given geometrical conditions. In other words, the degree of a node (k_i) can be defined as the number nodes that have reciprocal visibility with the first one. It is defined by the following equation (Donner and Donges, 2012):

$$k_i = \sum_j a_{ij} \tag{12}$$

As a reminder, by taking Fig. 3 as an example, the degree of the three first node are respectively, $k_1 = 1$, $k_2 = 3$ and $k_3 = 2$. The higher the value of k_i, the greater the centrality is.

3.3.2. Closeness centrality

In order to define closeness centrality, it is important to introduce another main property of graph theory frame, the Shortest Path (SP). Contrary to degree centrality which takes into account the number of edges and the adjacency matrix properties, closeness centrality estimates how many steps are required to access every other node from a given node. More concretely, there is a different number of edges and paths between two distant nodes (i,j). However, there will be some of these paths where the number of edges will be minimum, i.e SP. For a node i, this quantity is defined by the inverse of the sum of distances from this node to others (Carmona-Cabezas, 2020b):

$$c_i = \frac{1}{\sum_{j=1}^{N} d_{i,j}} \tag{13}$$

with $d_{i,j}$ the (i,j) element of the distance matrix. Closeness centrality evaluates the relative importance of a node within the graph. The higher the value of c_i, the greater the nodes are integrated with SP from the others in the graph.

3.3.3. Betweenness centrality

As closeness centrality, betweenness centrality is also based on SP. It indicates how many times a node is passed through SP of other pairs of nodes. For a node i, betweenness centrality is defined as follows (Carmona-Cabezas et al., 2020b):

$$b_i = \sum_{\substack{j=1 \\ j \neq i}}^{N} \sum_{\substack{k=1 \\ k \neq i,j}}^{N} \frac{n_{jk}(i)}{2N_{jk}} \tag{14}$$

with $n_{jk}(i)$ the number of SP's from j to k that include i and N_{jk} the total number of SP's between j and k. Betweenness centrality estimated the influence a node has over the flow through the network. The higher the value of b_i, the greater the nodes weights are important in the graph.

It is important to underline that a node with a high degree centrality does not guarantee that it is well connected to all other nodes. In some cases, a node with few direct edges is more important since it can act as a bridge, i.e. without it a network may be broken into subgraphs (Natapov et al., 2013). Fig. 4 illustrates an example of these cases for 10 nodes based on Natapov et al. (2013) study for Integrative Visibility Graph. Here, node 1 shows one of the highest betweenness while node 3 has the highest degree ($k = 4$). Node 7 combine both previous centrality measures an add high closeness, meaning that is the most integrated node in the graph. Regarding node 2, it exhibits the smallest possible betweenness and degree ($k = 1$), however it shows high closeness. Consequently, it may be considered as well integrated.

4. Results and discussion

4.1. Degree distribution

Before performing an in-depth analysis of PM_{10} data, it is essential to firstly investigate the nature of this time series, i.e. periodic, random or fractal as an example. To achieve this, the degree distribution $P(k)$ of the VGs is analyzed. This quantity is defined as the number of nodes of a given degree divided by the total number of them in the VG. According to Carmona-Cabezas et al. (2019a), large degrees are related to nodes corresponding to highest concentrations of pollutants such as tropospheric ozone (so-called hub), they are much less likely to be repeat

within the network. Consequently, hubs which traditionally have high visibility should correspond to PM_{10} extreme values. In the Caribbean, PM_{10} extreme values are mainly attributed to African dust outbreak (Plocoste and Pavón-Domínguez, 2020). In other words, hubs likeliness will be less significant than nodes close to PM_{10} average. Further analysis for the relationship between hubs and empirical data will be carried out later on.

Fig. 5 illustrates the degree distribution obtained for all data, low dust season (October to April), and high dust season (May to September) over 11 years. For all cases, the tail region of the log-log plot of $P(k)$ can be fitted by a power law like $P(k) \propto k^{-\gamma}$. In VG analysis, the exponent γ is relevant because it is a scale free parameter for a large number of real networks (Mali et al., 2018; Carmona-Cabezas et al., 2019a). Furthermore, this exponent can be related to the Hurst exponent in some cases (Lacasa et al., 2009; Lacasa and Toral, 2010). The exponents computed from the slopes are estimated for $k \geq 10$ with $2.89 \leq \gamma \leq 3.17$. It can be noticed that γ value for the overall is centred between γ values for low and high dust seasons. All this results clearly show the fractal behavior of PM_{10} time series.

Classically, this approach is used to describe time series nature. However, for a profound description of PM_{10} time series dynamics, a multifractal analysis is required.

4.2. Multifractal analysis

4.2.1. Rényi spectrum

In order to study the multifractal properties of PM_{10} time series in VG frame with Rényi spectrum, the SBA procedure is firstly performed. Thus, the quantities $\frac{ln\langle\mu(B)^{(q-1)}\rangle}{q-1}$ for $q \neq 1$ and $<ln\mu(B)>$ for $q = 1$ are plotted against $ln\,r$. As the results of partition functions are widely presented in the literature (Liu et al., 2015; Yu et al., 2016; Mali et al., 2018; Carmona-Cabezas et al., 2019a), we just recall here the range where the linear regression is applied for all q values in order to build Rényi spectrum from Equations (10) and (11). From $q = -5$ to $q = +5$ with an increment step of 0.25, a linear regression was made for $0 \leq ln\,r \leq 2$. Fig. 6 illustrates the Rényi spectra obtained where D_q values are plotted against order number q. One can clearly observe the multifractal properties of PM_{10} time series in the Caribbean area with VG procedure because $D_0 > D_1 > D_2$ for all cases (see Table 1). This result validates those previously obtained by Plocoste et al. (2017) with the classical structure function analysis. Traditionally, the strength of the multifractality ΔD_q can be quantified by the difference between the maximum and the minimum values of D_q (Yu et al., 2016). Table 1 shows that ΔD_q value is higher for low dust season, followed by
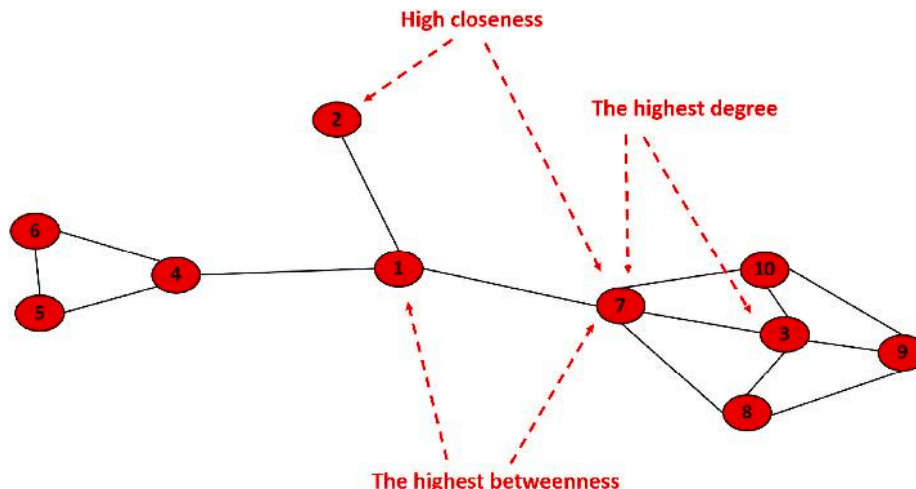

Fig. 4. An example of centrality measures for 10 nodes in Integrative Visibility Graph (Natapov et al., 2013).
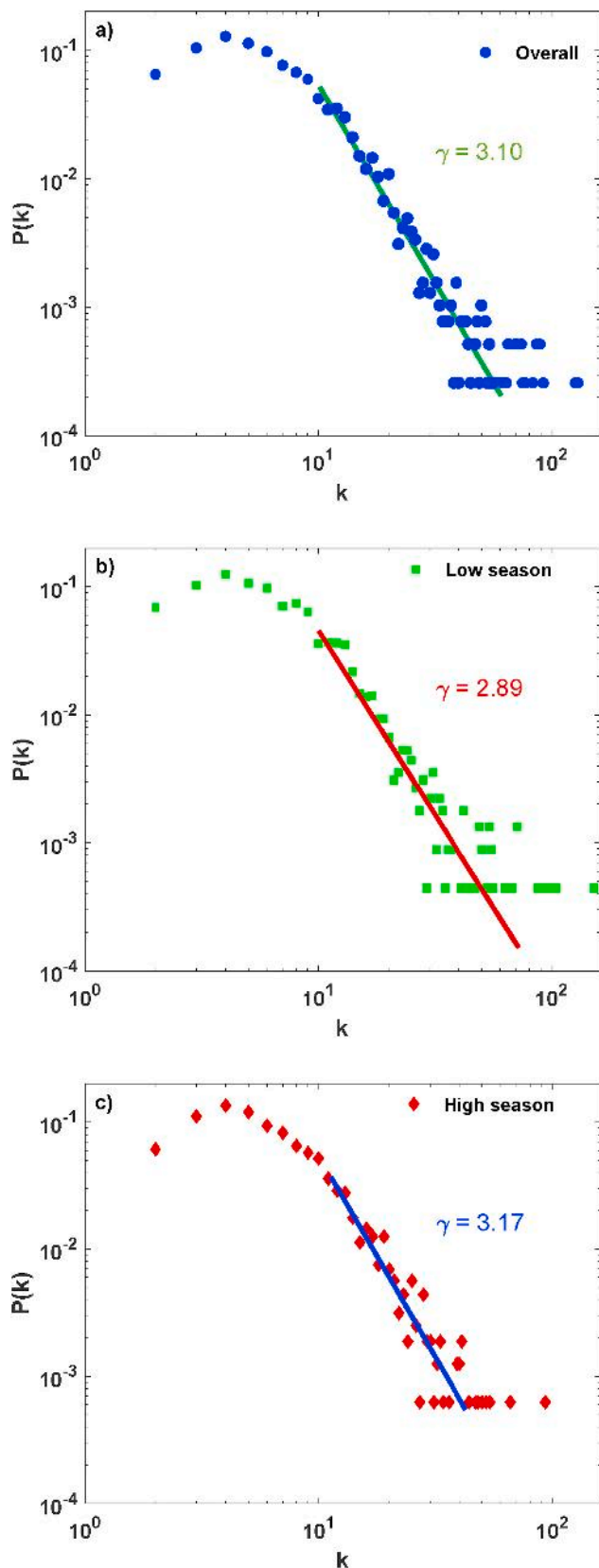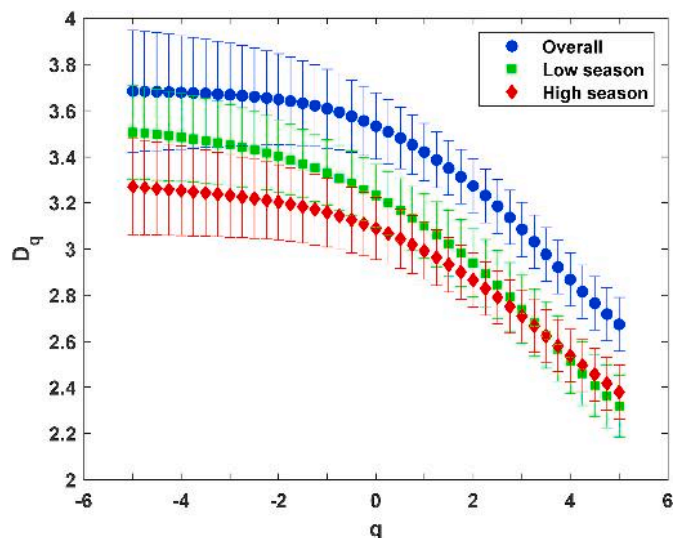
Fig. 6. Rényi dimensions for all data, low dust season (October to April) and high dust season (May to September). Standard deviations are illustrated by the whiskers.

"overall" and high dust season. These results are consistent with those previously found by Plocoste et al. (2020b) with descriptive statistics. Indeed, by computing the kurtosis parameter (K) which is a useful indicator of intermittency in pollution studies (Windsor and Toumi, 2001; Plocoste et al., 2018), they found $K_{Low}(33.0) > K_{Overall}(11.8) > K_{High}(5.9)$.

From Rényi spectrum, several parameters can be obtained. First of all, the capacity or box-counting dimension D_0 related to the fractal object, i.e. how many boxes are needed in order to have it covered. In VG frame, a high extension would mean a higher degree, since the maximum covering of the system would be a complete graph (K_n) whose degree are maximum. As it could be expected, D_0 values for low and high dust seasons are consistent with $\bar{k}$. Intuitively, one could make the assumption that D_0 value for the overall would be between D_0 values of low and high dust seasons as γ exponent previously found in section 4.1. Contrary to γ which is based on data weight, D_0 is related to time series dynamics. Overall period highlights the seasonal variation of PM_{10} data over 11 years while low and high dust seasons time series are reconstructed signals in order to take into account a specific period over 11 years. This may have an influence on fractal dimension behavior. In order to assess the relationship between the fractal object and the degree distribution, the values of PM_{10} concentration are plotted against the degree in VG. This procedure (v-k plot) is frequently used to illustrate if highest degrees (hubs) are related to the largest concentrations (Pierini et al., 2012). In Fig. 7, one can observed that highest degrees correspond basically to the largest concentrations, specially in the high season ($k > 40$), since the accumulation of points in the low season is more equally distributed along the degree. Nonetheless, in both cases one could make the assumption that high concentrations will be reflected by the hubs in the complex network. This behavior was previously observe by Carmona-Cabezas et al. (2019a) for tropospheric ozone time series in Cadiz, Spain. It is interesting to observe that v-k plot for overall is not just the sum of v-k plots for low and high dust seasons. As an example, 157.2 and 164.4 $\mu g/m^3$ respectively correspond to $k = 126$ and 129 in overall. On the other hand, $k = 93$ for 157.2 $\mu g/m^3$ in high season and $k = 152$ for 164.4 $\mu g/m^3$ in low season. For the same concentration values, the degree can differ according to time series layout. All these results clearly highlight the fact that VG method takes into account PM_{10} dynamics.

Thereafter, the entropy dimension D_1 is introduced. The latter is related to the uniformity of the system. In order to quantify uniformity

Fig. 5. Degree distribution of the visibility graph for (a) all data, (b) low dust season (October to April) and (c) high dust season (May to September) in log-log plot. As noted, the tail of the distribution of PM_{10} time series exhibits a scale-free behavior because it can be fitted by a power law in all cases.

Table 1

Characteristics values from VG frame for all data, low dust season (October to April) and high dust season (May to September).

Period	Degree distribution			Rényi parameters						Singularity parameters		
	γ	$\bar{k}$	σ_k	D_0	D_1	D_2	$D_0 - D_1$	$D_0 - D_2$	ΔD_q	α_{min}	α_{max}	W
Overall	3.10	65.5	37.10	3.5333	3.4209	3.2735	0.1124	0.2598	1.0106	2.663	3.761	1.098
Low season	2.89	77.0	43.73	3.2300	3.0986	2.9394	0.1314	0.2906	1.1899	2.293	3.616	1.323
High season	3.17	47.0	26.99	3.0878	2.9911	2.8648	0.0967	0.2230	0.8924	2.282	3.350	1.068

degree, the difference between D_0 and D_1 is computed. The greater $D_0 - D_1$, the less uniform it is. In VG frame, greater uniformity means less difference between the degrees of the sample, i.e. a decreased of standard deviation for the degree distribution. Table 1 shows that the most uniform period is the high dust season because it exhibits minimum values for $D_0 - D_1$ and σ_k. As illustrated in Fig. 2, one can observed in the middle of each year several peaks linked to African dust (Euphrasie-Clotilde et al., 2020). Overall, these peaks seem more homogeneous between May and September due to the continuous alternation between African easterly waves and dust outbreaks (Prospero and Carlson, 1981; Karyampudi et al., 1999) which generate a repetitive pattern in PM_{10} behavior. However, from October to April, some peaks with strong intensities can also be noted. These peaks can be caused by sporadic episodes of African dust transported by a particular circulation of air masses during Spring (Jury, 2017) or by extreme events as the eruption of Soufrière on Montserrat in February 2010 (Plocoste and Calif, 2019). This is the reason why they are called high and low seasons for African dust in the literature (Prospero et al., 2014; Velasco-Merino et al., 2018; Plocoste and Pavón-Domínguez, 2020) because depending on the synoptic conditions, dust outbreak can occur between October and April.

The last parameter estimated by Rényi spectrum is the correlation dimension D_2. This quantity is related to the recurrence in the series. In order to estimate recurrence degree, the difference between D_0 and D_2 is computed. The greater $D_0 - D_2$, the higher the probability of finding the same value in the time series. Table 1 indicates that during the low dust season, recurrence degree is higher because it exhibits the maximum value of $D_0 - D_2$. Based on $D_0 - D_1$ discussion, this result make sense because African dust is less frequent during that period. Indeed, between October and April, the main PM_{10} sources are marine aerosols and anthropogenic activity which composed the background atmosphere (Clergue et al., 2015; Rastelli et al., 2017). Euphrasie-Clotilde et al. (2017) study shows that this background atmosphere is roughly 20 $\mu g/m^3$. Due to Guadeloupe flat topography in the center of the island and its weak anthropogenic activity (Plocoste et al., 2014, 2018), the diurnal variation of PM_{10} concentrations is not significant as other megacities worldwide (Künzli et al., 2000; Mijić et al., 2009; Sansuddin et al., 2011; Xi et al., 2013; Ozel and Cakmakyapan, 2015; He et al., 2016; Momtazan et al., 2019). Thus, PM_{10} values highlight a more persistence behavior between October and April, i.e. more predictable.

4.2.2. Singularity spectrum

To investigate the multifractal properties of PM_{10} time series in VG frame through singularity spectrum, the Chhabra et al. (1989) procedure is firstly performed. Here, the quantities $\sum \beta_i(q,r) \ln[\beta_i(q,r)]$ and $\sum \beta_i(q,r) \ln[P_i(r)]$ have been plotted against $\ln r$ respectively for computed $f(\alpha)$ and α. In order to maintain the scale used for Rényi spectrum, the same range of linear regression has been chosen to construct the singularity spectrum, i.e. $0 \leq \ln r \leq 2$. The procedure to build singularity spectrum are widely described in the literature (Kelty-Stephen et al., 2013; Mali et al., 2018; Carmona-Cabezas et al., 2019a). In multifractal spectrum layout, the left side of $f(\alpha)$ spectrum is associated to $q > 0$ and it filters out the large fluctuations, whereas the right side of the spectrum is associated to $q < 0$ and it corresponds to small noise-like fluctuations. In VG theory, $q > 0$ is related to the highest values corresponding to the greatest degrees (so-called hubs) while $q < 0$ is related to the lowest values corresponding to the lowest degrees.

Fig. 8 shows the singularity spectra obtained where $f(\alpha)$ are plotted against α. From Rényi spectrum, we found that the low dust season time series has a higher multifractality degree. It is in accordance with this singularity spectra, as the width of the curves is also related to this multifractal behavior. This is because the wider the spectrum, the larger will be the range of fractal exponents required to describe the signal. This is translated into a richer structure in the series. Again, it seems that the high dust season is marked by a less multifractal dynamics compared to the low dust one, i.e. $W_{High} < W_{Low}$ (see Table 1).

Overall, one can notice that spectra are asymmetrical in shape with a left skewed distribution. This indicates dominance of large α values and therefore, a multifractal nature of large fluctuations. As it can be seen, the hubs are less likely and less homogeneous, as their α values differ more from each other. On the other hand, we observe that small noise-like fluctuations are more probable and homogeneous. By comparing both seasons, the left tail of the low season exhibits what could be expected, i.e. the large fluctuations are less probable ($f(\alpha)$ is lower in the left of low season). These results are consistent with what has been observed previously in the literature with high PM_{10} values due to African dust from May to September, i.e. the high dust season (Prospero et al., 2014; Velasco-Merino et al., 2018; Euphrasie-Clotilde et al., 2020; Plocoste et al., 2020b).

4.3. Centrality analysis

In VG theory, another way to retrieve information from a given complex network is to assess nodes importance through centrality measures. Here, the monthly dynamics of PM_{10} time series is investigated.

Fig. 9 shows the monthly behavior of average degree ($\bar{k}$) and standard deviation (σ_k) from degree distribution over 11 years. Firstly, one can observe that the shape of both curves evolves simultaneously along the month. A trend emerges clearly with minimum values between May and September for both curves. In other words, during the high dust season a decay of hubs is observed. An opposite behavior was highlighted by Carmona-Cabezas et al. (2019b) for ozone time series study in Cadiz, Spain. In their cases, the degree dropped during the months where high concentrations (hubs) were less significant. This difference in behavior may be due to many factors. Firstly, the nature of the pollutant. Contrary to ozone which is a secondary pollutant formed by photochemical reactions involving nitrogen oxides, carbon monoxide and volatile organic compounds with solar energy (Blacet, 1952; Seguel et al., 2012); PM_{10} is a primary pollutant principally linked to anthropogenic activity, marine aerosols and African dust in insular context (Plocoste et al., 2017; Euphrasie-Clotilde et al., 2020). Secondly, the scale of sources. Traditionally, ozone is mainly generated by the reaction between local precursors and solar radiation. Consequently a strong diurnal behavior is observed in ozone concentrations (Carmona-Cabezas et al., 2019b). In Guadeloupe, the local emitters of PM_{10} exhibit low concentrations, i.e. $\approx 20 \ \mu g/m^3$ (Euphrasie-Clotilde et al., 2017). This is the reason why high values of PM_{10} are mainly attributed to African dust, i.e. large scale source (Plocoste et al., 2020a). The diurnal magnitude of PM_{10} values in Guadeloupe is weaker than megacities which have strong anthropogenic pollution (cars, factories etc). Thirdly, city configuration. In Guadeloupe, most of buildings are four stories high or less (Plocoste et al., 2014) contrary to megacities where high-rise
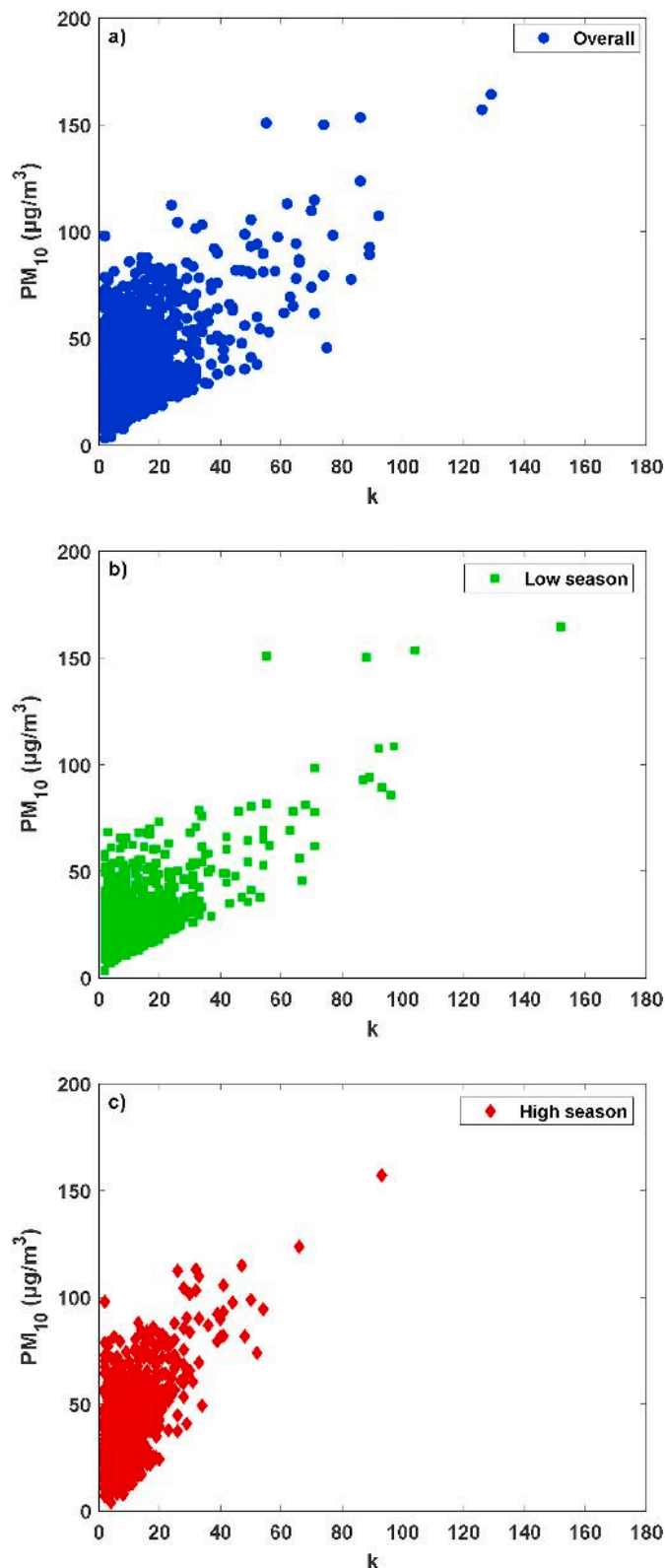
Fig. 7. Relationship between PM_{10} time series values and their degrees from VG frame for (a) all data, (b) low dust season (October to April) and (c) high dust season (May to September).

buildings may influence wind circulation and promote pollutants accumulation in the surface layer. Furthermore, the permanent trade winds may increase pollutants dispersion and decrease risk of stagnation of pollutants in the atmospheric boundary layer.
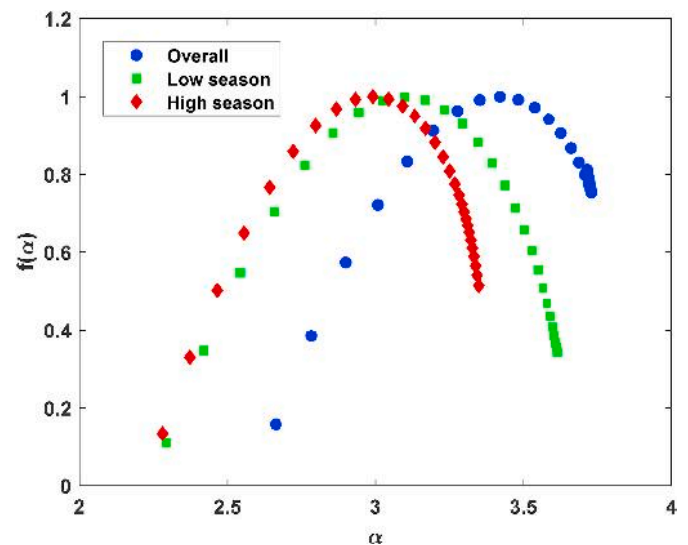

Fig. 8. Singularity spectra for all data, low dust season (October to April) and high dust season (May to September).
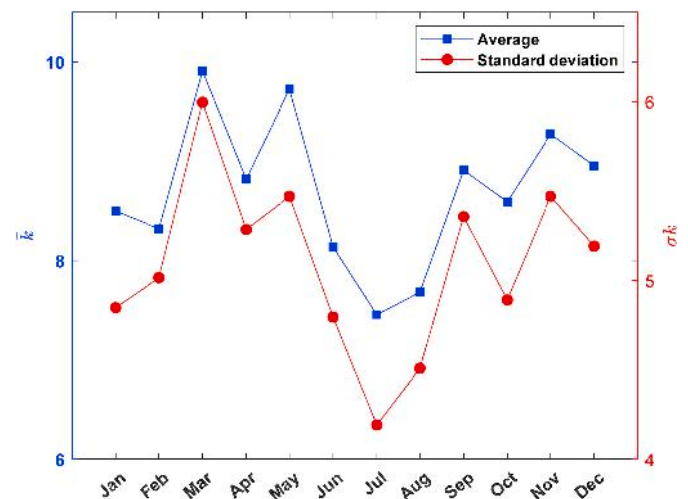

Fig. 9. Computed average degree and standard deviation from the degree distribution of each month over 11 years. Each monthly value is the average of the computed ones over 11 years.

Here, we also study PM_{10} dynamics according to both centrality measures related to SP, i.e. closeness and betweenness. For these both quantities, the average kurtosis seems more sensitive to PM_{10} dynamics than average, mediane and skewness. Fig. 10 depicts the monthly behavior of the average kurtosis values for both measures. Even if there are some fluctuations, a trend is clearly observed between May and September with a decrease of the average kurtosis values. The reasons previously mentioned to explain the decrease in degree centrality during the same period can also apply here.

All these results clearly highlight that centrality measures are able to describe PM_{10} time series dynamics. While a centred moment of order 2 is enough for degree centrality, a centred moment of order 4 is needed for centrality measures related to SP. This can be explained by the fact that degree centrality has an overview of the complex network while closeness and betweenness are more localized centrality measures, i.e. require a statistical parameter more sensitive to fluctuations.
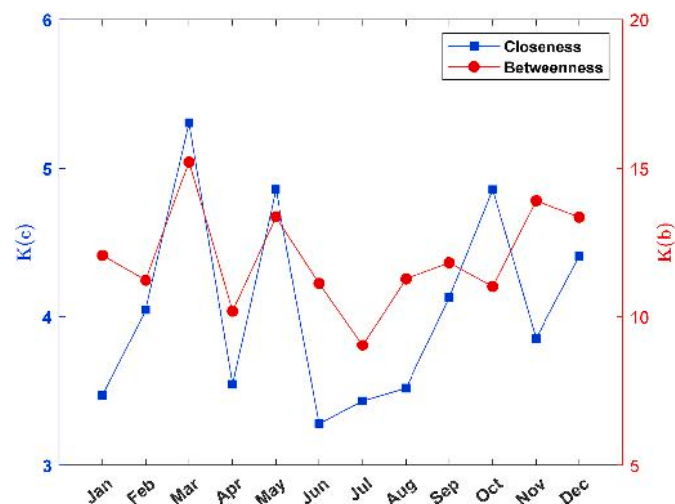
Fig. 10. Computed average kurtosis from closeness and betweenness of each month over 11 years. Each monthly value is the kurtosis average of the computed ones over 11 years.

5. Conclusion

In order to elaborate strategies and construct tools to predict dust events, it is fundamental to better understand particulate matter (PM_{10}) fluctuations. Here, the aim of this paper was to investigate PM_{10} dynamics in the Caribbean basin according to African dust seasonality. To achieve this, 11 years of daily PM_{10} database from Guadeloupe archipelago was analyzed using Visibility Graphs (VG).

Firstly, the degree distribution $P(k)$ showed the self-similar properties of PM_{10} data. Indeed, a scaling regime is found for all the studied periods highlighting the fractal nature of PM_{10} time series. The exponents γ computed from the slopes of $P(k)$ are estimated for $k \geq 10$ with $2.89 \leq \gamma \leq 3.17$.

Thereafter, a multifractal analysis is performed through 2 approaches, i.e. the generalized fractal dimension and the singularity spectrum. Both methods pointed out a higher multifractality degree in the low dust season showing that the inhomogeneity of the distribution of probability measured on the overall fractal structure increases faster with the order of moments from October to April. In addition, multifractal parameters exhibited that the low dust season has the higher recurrence and the lower uniformity degrees. Authors believe that the continuous alternation between African easterly waves and dust outbreaks during the high dust season tends to homogenize the multifractal characteristics of PM_{10} time series from May to September.

Lastly, the monthly behavior of PM_{10} time series is analyzed using the centrality measures, i.e. degree, closeness and betweenness. All the results showed that the centrality measures are sensitive to PM_{10} dynamics through the year with a decay of the centrality values during the high dust season. While the average is enough for degree centrality analysis, the kurtosis is required for the others quantities related to the shortest path.

To conclude, all these results clearly showed that VG is a robust tool to describe times series properties. Given the many possibilities offered by complex networks, a wide range of subjects can be dealt with. In order to quantify a possible hub repulsion phenomenon in PM_{10} time series, an analysis of edge repulsion force based on Zhang et al. (2013) work should be carried out in a future study.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

Adams, A.M., Prospero, J.M., Zhang, C., 2012. CALIPSO-derived three-dimensional structure of aerosol over the Atlantic Basin and adjacent continents. J. Clim. 25 (19), 6862–6879.

Amancio, D.R., Antiqueira, L., Pardo, T.A., COSTA L, da F., Oliveira Jr., O.N., Nunes, M. G., 2008. Complex networks analysis of manual and machine translations. Int. J. Mod. Phys. C 19, 583–598, 04.

Amaral, L.A., Ottino, J.M., 2004. Complex networks. The European Physical Journal B 38 (2), 147–162.

Bacry, E., Muzy, J.F., Arneodo, A., 1993. Singularity spectrum of fractal signals from wavelet analysis: exact results. J. Stat. Phys. 70 (3–4), 635–674.

Blacet, F., 1952. Photochemistry in the lower atmosphere. Ind. Eng. Chem. 44 (6), 1339–1342.

Block, A., Von Bloh, W., Schellnhuber, H., 1990. Efficient box-counting determination of generalized fractal dimensions. Phys. Rev. 42 (4), 1869.

Cadelis, G., Tourres, R., Molinie, J., 2014. Short-term effects of the particulate pollutants contained in saharan dust on the visits of children to the emergency department due to asthmatic conditions in Guadeloupe (French Archipelago of the Caribbean). PloS One 9, e91136.

Caniego, F., Espejo, R., Martın, M., San José, F., 2005. Multifractal scaling of soil spatial variability. Ecol. Model. 182 (3–4), 291–303.

Carlson, T.N., Prospero, J.M., 1972. The large-scale movement of Saharan air outbreaks over the northern equatorial Atlantic. J. Appl. Meteorol. 11 (2), 283–297.

Carmona-Cabezas, R., Ariza-Villaverde, A.B., Gutiérrez de Ravé, E., Jiménez-Hornero, F. J., 2019a. Visibility graphs of ground-level ozone time series: a multifractal analysis. Sci. Total Environ. 661, 138–147.

Carmona-Cabezas, R., Gómez-Gómez, J., Ariza-Villaverde, A.B., Gutiérrez de Ravé, E., Jiménez-Hornero, F.J., 2019. Can complex networks describe the urban and rural tropospheric O_3 dynamics? Chemosphere 230, 59–66.

Carmona-Cabezas, R., Gómez-Gómez, J., Ariza-Villaverde, A.B., Gutiérrez de Ravé, E., Jiménez-Hornero, F.J., 2020. Multiplex Visibility Graphs as a complementary tool for describing the relation between ground level O_3 and NO_2. Atmospheric Pollution Research 11 (1), 205–212.

Carmona-Cabezas, R., Gómez-Gómez, J., Gutiérrez de Ravé, E., Jiménez-Hornero, F.J., 2020b. Checking complex networks indicators in search of singular episodes of the photochemical smog. Chemosphere 241, 125085.

Cavalieri, O., Cairo, F., Fierli, F., Donfrancesco, G.D., Snels, M., Viterbini, M., Cardillo, F., Chatenet, B., Formenti, P., Marticorena, B., Rajot, J.L., 2010. Variability of aerosol vertical distribution in the Sahel. Atmos. Chem. Phys. 10 (24), 12005–12023.

Chhabra, A.B., Meneveau, C., Jensen, R.V., Sreenivasan, K., 1989. Direct determination of the f (α) singularity spectrum and its application to fully developed turbulence. Phys. Rev. 40 (9), 5284.

Choobari, O.A., Zawar-Reza, P., Sturman, A., 2014. The global distribution of mineral dust and its impacts on the climate system: a review. Atmos. Res. 138, 152–165.

Clergue, C., Dellinger, M., Buss, H., Gaillardet, J., Benedetti, M., Dessert, C., 2015. Influence of atmospheric deposits and secondary minerals on Li isotopes budget in a highly weathered catchment, Guadeloupe (Lesser Antilles). Chem. Geol. 414, 28–41.

Costa, L.d.f., Rodrigues, F.A., Travieso, G., Villas Boas, P.R., 2007. Characterization of complex networks: a survey of measurements. Adv. Phys. 56 (1), 167–242.

Donner, R.V., Donges, J.F., 2012. Visibility graph analysis of geophysical time series: potentials and possible pitfalls. Acta Geophys. 60 (3), 589–623.

Euphrasie-Clotilde, L., Molinié, J., Feuillard, T., Brute, F., 2017. The relationship between coastal West African dust level and Caribbean island dust. WIT Trans. Ecol. Environ. 211, 121–127.

Euphrasie-Clotilde, L., Plocoste, T., Feuillard, T., Velasco-Merino, C., Mateos, D., Toledano, C., Brute, F.N., Bassette, C., Gobinddass, M., 2020. Assessment of a new

detection threshold for PM10 concentrations linked to African dust events in the Caribbean Basin. Atmos. Environ. 224, 117354.

Feder, J., 1988. Fractals (Physics of Solids and Liquids), 1988th ed. Springer, Boston.

Feng, W., Li, H., Wang, S., Van Halm-Lutterodt, N., An, J., Liu, Y., Liu, M., Wang, X., Guo, X., 2019. Short-term PM10 and emergency department admissions for selective cardiovascular and respiratory diseases in Beijing, China. Sci. Total Environ. 657, 213–221.

Freeman, L.C., 1978. Centrality in social networks conceptual clarification. Soc. Network. 1 (3), 215–239.

Gan, C., Yang, X., Liu, W., Zhu, Q., Jin, J., He, L., 2014. Propagation of computer virus both across the Internet and external computers: a complex-network approach. Commun. Nonlinear Sci. Numer. Simulat. 19 (8), 2785–2792.

Gurung, A., Son, J.Y., Bell, M.L., 2017. Particulate matter and risk of hospital admission in the Kathmandu Valley, Nepal: a case-crossover study. Am. J. Epidemiol. 186 (5), 573–580.

Halsey, T.C., Jensen, M.H., Kadanoff, L.P., Procaccia, I., Shraiman, B.I., 1986. Fractal measures and their singularities: the characterization of strange sets. Phys. Rev. 33 (2), 1141.

Harte, D., 2001. Multifractals: Theory and Applications. CRC Press.

He, H.d., Pan, W., Lu, W.z., Xue, Y., Peng, G.h., 2016. Multifractal property and long-range cross-correlation behavior of particulate matters at urban traffic intersection in Shanghai. Stoch. Environ. Res. Risk Assess. 30 (5), 1515–1525.

Holz, C., Stuut, J.b.W., Henrich, R., 2004. Terrigenous sedimentation processes along the continental margin off NW Africa: implications from grain-size analysis of seabed sediments. Sedimentology 51 (5), 1145–1154.

Jury, M.R., 2017. Caribbean air chemistry and dispersion conditions. Atmosphere 8 (8), 151.

Karyampudi, V.M., Palm, S.P., Reagen, J.A., Fang, H., Grant, W.B., Hoff, R.M., Moulin, C., Pierce, H.F., Torres, O., Browell, E.V., Melfi, S.H., 1999. Validation of the Saharan dust plume conceptual model using lidar, Meteosat, and ECMWF data. Bull. Am. Meteorol. Soc. 80 (6), 1045–1076.

Kaufman, Y., Koren, I., Remer, L., Tanré, D., Ginoux, P., Fan, S., 2005. Dust transport and deposition observed from the terra-moderate resolution imaging spectroradiometer (MODIS) spacecraft over the Atlantic Ocean. J. Geophys. Res.: Atmosphere 110 (D10).

Kelty Stephen, D.G., Palatinus, K., Saltzman, E., Dixon, J.A., 2013. A tutorial on multifractality, cascades, and interactivity for empirical time series in ecological science. Ecol. Psychol. 25 (1), 1–62.

Koren, I., Kaufman, Y.J., Washington, R., Todd, M.C., Rudich, Y., Martins, J.V., Rosenfeld, D., 2006. The Bodélé depression: a single spot in the Sahara that provides most of the mineral dust to the Amazon forest. Environ. Res. Lett. 1 (1), 014005.

Künzli, N., Kaiser, R., Medina, S., Studnicka, M., Chanel, O., Filliger, P., Herry, M., Horak Jr., F., Puybonnieux-Texier, V., Quénel, P., Schneider, J., Seethaler, R., Vergnaud, J., Sommer, H., 2000. Public-health impact of outdoor and traffic-related air pollution: a European assessment. Lancet 356 (9232), 795–801.

Lacasa, L., Luque, B., Ballesteros, F., Luque, J., Nuno, J.C., 2008. From time series to complex networks: the visibility graph. Proc. Natl. Acad. Sci. Unit. States Am. 105 (13), 4972–4975.

Lacasa, L., Luque, B., Luque, J., Nuno, J.C., 2009. The visibility graph: a new method for estimating the Hurst exponent of fractional Brownian motion. EPL (Europhysics Letters) 86 (3), 30001.

Lacasa, L., Toral, R., 2010. Description of stochastic and chaotic series using visibility graphs. Phys. Rev. 82 (3), 036120.

Liu, J.L., Yu, Z.G., Anh, V., 2015. Determination of multifractal dimensions of complex networks by means of the sandbox algorithm. Chaos: An Interdisciplinary Journal of Nonlinear Science 25 (2), 023103.

Lyra, M., Tsallis, C., 1998. Nonextensivity and multifractality in low-dimensional dissipative systems. Phys. Rev. Lett. 80 (1), 53.

Mach, J., Mas, F., Sagués, F., 1995. Two representations in multifractal analysis. J. Phys. Math. Gen. 28 (19), 5607.

Mahowald, N., Albani, S., Kok, J.F., Engelstaeder, S., Scanza, R., Ward, D.S., Flanner, M. G., 2014. The size distribution of desert dust aerosols and its impact on the Earth system. Aeolian Research 15, 53–71.

Mali, P., Manna, S., Mukhopadhyay, A., Haldar, P., Singh, G., 2018. Multifractal analysis of multiparticle emission data in the framework of visibility graph and sandbox algorithm. Phys. Stat. Mech. Appl. 493, 253–266.

Mandelbrot, B., 1982. The Fractal Geometry of Nature. Freeman and Company, New York.

Mandelbrot, B.B., 1974. Intermittent turbulence in self-similar cascades: divergence of high moments and dimension of the carrier. J. Fluid Mech. 62 (2), 331–358.

Mijić, Z., Tasić, M., Rajšić, S., Novaković, V., 2009. The statistical characters of PM10 in Belgrade area. Atmos. Res. 92 (4), 420–426.

Momtazan, M., Geravandi, S., Rastegarimehr, B., Valipour, A., Ranjbarzadeh, A., Yari, A. R., Dobaradaran, S., Bostan, H., Farhadi, M., Darabi, F., Khaniabadi, Y.O., Mohammadi, M.J., 2019. An investigation of particulate matter and relevant cardiovascular risks in Abadan and Khorramshahr in 2014–2016. Toxin Rev. 38 (4), 290–297.

Moulin, C., Lambert, C.E., Dulac, F., Dayan, U., 1997. Control of atmospheric export of dust from north africa by the north atlantic oscillation. Nature 387 (6634), 691–694.

Muzy, J.F., Bacry, E., Arneodo, A., 1993. Multifractal formalism for fractal signals: the structure-function approach versus the wavelet-transform modulus-maxima method. Phys. Rev. 47 (2), 875.

Natapov, A., Czamanski, D., Fisher-Gewirtzman, D., 2013. Can visibility predict location? Visibility graph of food and drink facilities in the city. Surv. Rev. 45 (333), 462–471.

Nicholson, S.E., 2000. The nature of rainfall variability over Africa on time scales of decades to millenia. Global Planet. Change 26 (1–3), 137–158.

Ozel, G., Cakmakyapan, S., 2015. A new approach to the prediction of PM10 concentrations in Central Anatolia Region, Turkey. Atmospheric Pollution Research 6 (5), 735–741.

Pamuła, G., Grech, D., 2014. Influence of the maximal fluctuation moment order q on multifractal records normalized by finite-size effects. EPL (Europhysics Letters) 105 (5), 50004.

Petit, R., Legrand, M., Jankowiak, I., Molinié, J., Asselin de Beauville, C., Marion, G., Mansot, J., 2005. Transport of saharan dust over the caribbean islands: study of an event. J. Geophys. Res.: Atmosphere 110, D18S09.

Pierini, J.O., Lovallo, M., Telesca, L., 2012. Visibility graph analysis of wind speed records measured in central Argentina. Phys. Stat. Mech. Appl. 391 (20), 5041–5048.

Plocoste, T., Calif, R., 2019. Spectral observations of PM10 fluctuations in the hilbert space. Functional Calculus. IntechOpen, pp. 1–13.

Plocoste, T., Calif, R., Euphrasie-Clotilde, L., Brute, F.N., 2020a. Investigation of local correlations between particulate matter (PM10) and air temperature in the Caribbean basin using Ensemble Empirical Mode Decomposition. Atmos. Pollut. Res. https://www.sciencedirect.com/science/article/abs/pii/S1309104220301902.

Plocoste, T., Calif, R., Euphrasie-Clotilde, L., Brute, F.N., 2020b. The statistical behavior of PM10 events over guadeloupean archipelago: stationarity, modelling and extreme events. Atmos. Res. 241, 104956.

Plocoste, T., Calif, R., Jacoby-Koaly, S., 2017. Temporal multiscaling characteristics of particulate matter PM$_{10}$ and ground-level ozone O$_3$ concentrations in Caribbean region. Atmos. Environ. 169, 22–35.

Plocoste, T., Calif, R., Jacoby-Koaly, S., 2019. Multi-scale time dependent correlation between synchronous measurements of ground-level ozone and meteorological parameters in the Caribbean Basin. Atmos. Environ. 211, 234–246.

Plocoste, T., Dorville, J.F., Monjoly, S., Jacoby-Koaly, S., André, M., 2018. Assessment of nitrogen oxides and ground-level ozone behavior in a dense air quality station network: case study in the lesser antilles arc. J. Air Waste Manag. Assoc. 68 (12), 1278–1300.

Plocoste, T., Jacoby-Koaly, S., Molinié, J., Petit, R., 2014. Evidence of the effect of an urban heat island on air quality near a landfill. Urban Climate 10, 745–757.

Plocoste, T., Pavón-Domínguez, P., 2020. Temporal scaling study of particulate matter (PM10) and solar radiation influences on air temperature in the Caribbean basin using a 3D joint multifractal analysis. Atmos. Environ. 222, 117119.

Posadas, A.N., Giménez, D., Bittelli, M., Vaz, C.M., Flury, M., 2001. Multifractal characterization of soil particle-size distributions. Soil Sci. Soc. Am. J. 65 (5), 1361–1367.

Prospero, J.M., Carlson, T.N., 1972. Vertical and areal distribution of Saharan dust over the western equatorial North Atlantic Ocean. J. Geophys. Res. 77 (27), 5255–5265.

Prospero, J.M., Carlson, T.N., 1981. Saharan air outbreaks over the tropical North Atlantic. Weather and Weather Maps. Springer, pp. 677–691.

Prospero, J.M., Collard, F.X., Molinié, J., Jeannot, A., 2014. Characterizing the annual cycle of African dust transport to the Caribbean Basin and South America and its impact on the environment and air quality. Global Biogeochem. Cycles 28, 757–773.

Pun, V.C., Yu, I.T.s., Ho, K.f., Qiu, H., Sun, Z., Tian, L., 2014. Differential effects of source-specific particulate matter on emergency hospitalizations for ischemic heart disease in Hong Kong. Environ. Health Perspect. 122 (4), 391–396.

Rastelli, E., Corinaldesi, C., Dell'Anno, A., Martire, M.L., Greco, S., Facchini, M.C., Rinaldi, M., O'Dowd, C., Ceburnis, D., Danovaro, R., 2017. Transfer of labile organic matter and microbes from the ocean surface to the marine aerosol: an experimental approach. Sci. Rep. 7 (1), 11475.

Sansuddin, N., Ramli, N.A., Yahaya, A.S., Yusof, N.F.F.M., Ghazali, N.A., Al Madhoun, W. A., 2011. Statistical analysis of PM10 concentrations at different locations in Malaysia. Environ. Monit. Assess. 180 (1–4), 573–588.

Schepanski, K., 2018. Transport of mineral dust and its impact on climate. Geosciences 8 (5), 151.

Schreiber, M., Grussbach, H., 1991. Multifractal wave functions at the Anderson transition. Phys. Rev. Lett. 67 (5), 607.

Schütz, L., 1980. Long range transport of desert dust with special emphasis on the sahara. Ann. N. Y. Acad. Sci. 338 (1), 515–532.

Seguel, R.J., Morales, S.R.G., Leiva, G.M.A., 2012. Ozone weekend effect in Santiago, Chile. Environ. Pollut. 162, 72–79.

Shao, Y., Wyrwoll, K.H., Chappell, A., Huang, J., Lin, Z., McTainsh, G.H., Mikami, M., Tanaka, T.Y., Wang, X., Yoon, S., 2011. Dust cycle: an emerging core theme in Earth system science. Aeolian Research 2 (4), 181–204.

Song, C., Havlin, S., Makse, H.A., 2006. Origins of fractality in the growth of complex networks. Nat. Phys. 2 (4), 275–281.

Stam, C., 2010. Characterization of anatomical and functional connectivity in the brain: a complex networks perspective. Int. J. Psychophysiol. 77 (3), 186–194.

Su, C., Breitner, S., Schneider, A., Liu, L., Franck, U., Peters, A., Pan, X., 2016. Short-term effects of fine particulate air pollution on cardiovascular hospital emergency room visits: a time-series study in Beijing, China. Int. Arch. Occup. Environ. Health 89 (4), 641–657.

Tél, T., Fülöp, Á., Vicsek, T., 1989. Determination of fractal dimensions for geometrical multifractals. Phys. Stat. Mech. Appl. 159 (2), 155–166.

Tsamalis, C., Chédin, A., Pelon, J., Capelle, V., 2013. The seasonal vertical distribution of the Saharan Air Layer and its modulation by the wind. Atmos. Chem. Phys. 13 (22), 11235–11257.

Van Der Does, M., Korte, L.F., Munday, C.I., Brummer, G.J.A., Stuut, J.B.W., 2016. Particle size traces modern Saharan dust transport and deposition across the equatorial North Atlantic. Atmos. Chem. Phys. 16 (21).

Velasco-Merino, C., Mateos, D., Toledano, C., Prospero, J.M., Molinie, J., Euphrasie-Clotilde, L., González, R., Cachorro, V.E., Calle, A., Frutos, A.M.d., 2018. Impact of long-range transport over the Atlantic Ocean on Saharan dust optical and

microphysical properties based on AERONET data. Atmos. Chem. Phys. 18 (13), 9411–9424.

Veneziano, D., Moglen, G.E., Bras, R.L., 1995. Multifractal analysis: pitfalls of standard procedures and alternatives. Phys. Rev. 52 (2), 1387.

Vicsek, T., Family, F., Meakin, P., 1990. Multifractal geometry of diffusion-limited aggregates. EPL (Europhysics Letters) 12 (3), 217.

Windsor, H., Toumi, R., 2001. Scaling and persistence of UK pollution. Atmos. Environ. 35 (27), 4545–4556.

Xi, W., Chen, R.J., Chen, B.H., Kan, H.D., 2013. Application of statistical distribution of PM10 concentration in air quality management in 5 representative cities of China. Biomed. Environ. Sci. 26 (8), 638–646.

Yu, Z.G., Zhang, H., Huang, D.W., Lin, Y., Anh, V., 2016. Multifractality and Laplace spectrum of horizontal visibility graphs constructed from fractional Brownian motions. J. Stat. Mech. Theor. Exp. 2016 (3), 033206.

Zeleke, T.B., Si, B.C., 2006. Characterizing scale-dependent spatial relationships between soil properties using multifractal techniques. Geoderma 134 (3–4), 440–452.

Zhang, H., Lan, X., Wei, D., Mahadevan, S., Deng, Y., 2013. Self-similarity in complex networks: from the view of the hub repulsion. Mod. Phys. Lett. B 27 (28), 1350201.

Zhang, J., Liu, Y., Cui, L.l., Liu, S.q., Yin, X.x., Li, H.c., 2017. Ambient air pollution, smog episodes and mortality in Jinan, China. Sci. Rep. 7 (1), 1–8.

Zhang, J., Small, M., 2006. Complex network from pseudoperiodic time series: topology versus dynamics. Phys. Rev. Lett. 96 (23), 238701.

Zou, Y., Donner, R.V., Marwan, N., Donges, J.F., Kurths, J., 2019. Complex network approaches to nonlinear time series analysis. Phys. Rep. 787, 1–97.