APPLICATION PAPER

Spatial analysis of tails of air pollution PDFs in Europe

Hankun He[1] (ORCID), Benjamin Schäfer[2] and Christian Beck[1]

[1]Centre for Complex Systems, School of Mathematical Sciences, Queen Mary University of London, London, UK
[2]Institute for Automation and Applied Informatics, Karlsruhe Institute of Technology, Karlsruhe, Germany
Corresponding author: Hankun He; Email: h.he@qmul.ac.uk

Abstract

Outdoor air pollution is estimated to cause a huge number of premature deaths worldwide. It catalyzes many diseases on a variety of time scales, and it has a detrimental effect on the environment. In light of these impacts, it is necessary to obtain a better understanding of the dynamics and statistics of measured air pollution concentrations, including temporal fluctuations of observed concentrations and spatial heterogeneities. Here, we present an extensive analysis for measured data from Europe. The observed probability density functions (PDFs) of air pollution concentrations depend very much on the spatial location and the pollutant substance. We analyze a large number of time series data from 3544 different European monitoring sites and show that the PDFs of nitric oxide (NO), nitrogen dioxide (NO_2), and particulate matter (PM_{10} and $PM_{2.5}$) concentrations generically exhibit heavy tails. These are asymptotically well approximated by q-exponential distributions with a given entropic index q and width parameter λ. We observe that the power-law parameter q and the width parameter λ vary widely for the different spatial locations. We present the results of our data analysis in the form of a map that shows which parameters q and λ are most relevant in a given region. A variety of interesting spatial patterns is observed that correlate to the properties of the geographical region. We also present results on typical time scales associated with the dynamical behavior.

Impact Statement

Our application paper offers a substantial research contribution at the interface of statistical physics, environmental sciences, geography, and data analytics. It addresses the complex dynamics of air pollution in Europe, analyzing, in particular, the tails of the observed PDFs which describe high-pollution events. Our work enables a better understanding of the time-varying dynamics of air pollution, which is essential for policy formulation and the construction of suitable stochastic models, as well as for analyzing the medical consequences of exposure to polluted air.

1. Introduction

Outdoor air pollution is estimated to cause 4.2 million premature deaths per year worldwide, according to estimates of the WHO, and is thus a major killer. It also causes many diseases on a variety of time scales World Health Organization (2020); Shah et al. (2015). There are highly detrimental effects on the environment, as air pollution affects vegetation and natural ecosystems and plays a role in climate change Ortiz et al. (2020). In all densely populated regions of the word, it represents a major environmental health risk Health Effects Institute (2019); World Health Organization et al. (2018); GBD 2016 Risk Factors Collaborators et al. (2017),

impacting human health, ecosystems, climate, infrastructure, and the economy Ortiz et al. (2020). There are many different types of air pollutants and they can interact in a complex way. Two air pollutants, namely particulate matter (*PM*) and nitrogen oxides (NO_x), pose a considerable threat to the health of citizens. In 2018, about 55000 and 417000 premature deaths in 41 European countries were attributed to NO_2 and $PM_{2.5}$, respectively Ortiz et al. (2020). In this contribution to Climate Informatics 2024, we focus on NO_x and *PM*, but our statistical data analytics methods can be applied to other substances as well.

The impact of air pollution on health does not only depend on the pollutant type but also on the type of surrounding area. The EU European Parliament (2011) uses environmental area types to classify air quality monitoring sites into traffic, industrial, background, urban, suburban, and rural, based on predominant emission sources and building density. From an air quality policy perspective, this allows for evaluating the effectiveness of measures targeting specific emissions sectors, assessing the impact of those pollutants which dominate the area surrounding a given monitoring station.

A lot of air pollution research concentrates on mean values, but having a thorough understanding of the time-varying statistics, i.e., of the entire probability density function (PDF) of air pollution concentrations is crucial for policymakers involved in defining thresholds or reducing overall exposure to air pollution by suitable mitigation measures. It is also crucial for the construction of suitable statistical physics-based models using stochastic methods. PDFs such as gamma, log-normal, and Weibull distributions. Hsin-Chung Lu (2003) has been widely used for fitting air pollutant concentration data. However, these distributions decay approximately like exponential functions at large values, while previous investigations have found heavy tails in air pollution statistics Williams et al. (2020); He et al., (2022), which are not well-described by the above distributions.

Previous papers Giri et al. (2023); Baldasano (2020); Environmental Research Group et al. (2020) have also explored the COVID-19 lockdown effects on air quality (for example, in megacities such as London and Delhi), focusing on comparing the PDFs or given moments of the PDFs before and during the lockdown. Superstatistical methods, originating from turbulence modeling Beck (2007) and applied to many fields Metzler (2020); Beck et al. (2020); Ourabah (2024), offer a powerful effective approach to describe the dynamics of air pollution. These types of models are based on the assumption of well-separated time scales Williams et al. (2020). Air pollutants such as NO_x have been described successfully using a superstatistical approach, taking into account nonequilibrium situations with fluctuating variance parameters Williams et al. (2020). However, the approach in that paper was verified for limited data sets only, chosen from the UK (London), and also only for a limited set of pollutants, namely *NO* and NO_2 only. For European data, recent progress was made by He et al., (2022), where it was shown that many measured air pollution time series data can be generically understood and modeled by superstatistical methods. In this conference proceedings, we follow up from that approach. We consider a very large set of measured time series data in Europe (3544 locations), extract the best-fitting parameters at each spatial location, and then plot a map of Europe where the best-fitting parameters are visualized at each point. The results are quite useful to know as they provide some information on what type of power law exponent is to be expected at a given location, thus describing the risk of high-pollution situations occurring at that location.

This paper is organized as follows. We first describe the data set considered and briefly comment on our method of finding the best-fitting parameters. Some relevant background on superstatitical methods and *q*-exponential functions will be given. We then present our main result on air pollution statistics in Europe in the form of a visual display of the map of Europe with the given power-law statistics parameters that we obtained in our detailed analysis. Finally, we discuss the observed patterns and spatial correlations.

2. Methods

2.1. The data set considered

We use air quality monitoring data from a large number of locations in Europe through the interface "Saqgetr", which is an R package available on the Comprehensive R Archive Network (CRAN) Grange (2019). Most of the data are openly Available at the European Commission's Airbase and air quality

e-reporting (AQER) repositories European Parliament, Council of the European Union (2008); European Commission (2011). We also import surface meteorological data from NOAA ISD Smith et al. (2011) via the "worldmet" Carslaw (2017) R package. Initially, we started with 9698 different locations and used measured data from January 2017 to December 2021, recorded at 1-hour intervals. However, with certain selection criteria applied on the quality of the measured data, this reduces to a set containing 3544 sites. Our selection criteria are described in detail in He (2024). A main criterium was that at least one year of measured data should be available. Our interest is in the behavior of the tails of the PDFs, as illustrated in Figure 1. These tails correspond to high pollution states and are most
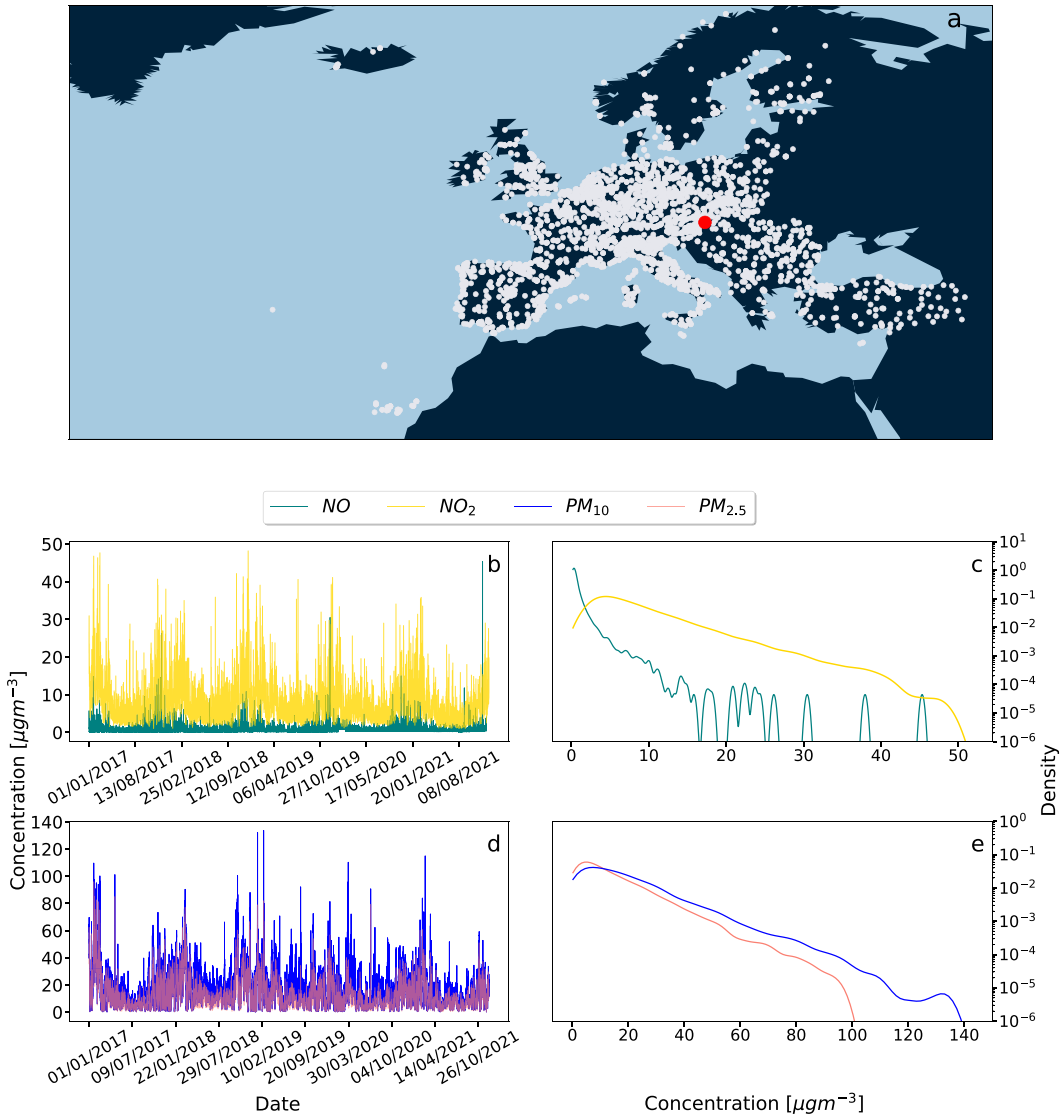

Figure 1. *(a) The white dots show the positions of all European measuring stations considered. A red dot singles out an example of a station at Illmitz, Austria. For such a given example location, the time series data of the four measured concentrations are shown in panels b and d. The corresponding tails of the PDFs (as shown in c, e) are then automatically analysed with a fitting program He (2024) that extracts the parameters (q, λ) for the best-fitting q-exponential. This is done for all 3544 sites.*

damaging to health. The measuring stations are divided into three categories: traffic, industrial, and background based on predominant emission sources; the surrounding areas are classified as urban, suburban, or rural based on the building density. Measuring station types are combined with area types to provide an overall station classification, and we analyze our data conditioned on this station classification.

2.2. Fitting with *q*-exponentials

In He et al., (2022), evidence was presented that generic air pollution PDFs often decay as a power law. To extract the power law exponent, it is best to apply a fitting approach based on *q*-exponential functions. These functions asymptotically approach a power law (with an exponent of $-1/(q-1)$ if $q > 1$) and also take into account the next-to-leading order terms in the asymptotics. The form of *q*-exponentials is motivated by generalized versions of statistical mechanics based on nonadditive entropies Tsallis (2009), as well as on superstatistical approaches Beck and Cohen (2003); Beck et al. (2005). The reason for the occurrence of *q*-exponentials is that exponentials with varying rate parameters, when integrated over the probability density of the rate parameter, in leading order lead to *q*-exponentials. The functional form of the *q*-exponential is given by

$$f_{q,\lambda}(x) = (2-q)\lambda[1-\lambda(1-q)x]^{\frac{1}{1-q}} \text{ for } 1-\lambda(1-q)x \geq 0, x > 0, \qquad (1)$$

where q is the entropic index Tsallis (2009); Hanel et al. (2011); Jizba and Korbel (2019), λ is a positive width parameter and x, in our case, denotes the air pollutant concentration. Eq. (1) contains the exponential distribution as a special case, namely that is obtained in the limit $q \to 1$, as the *q*-exponential function, defined as $e_q(x) = [1 + (1-q)x]^{\frac{1}{1-q}}$, converges to the exponential function in the limit $q \to 1$. For $q < 1$, $f_{q,\lambda}(x)$ is also well defined but in this case lives on a finite support and becomes exactly zero above a critical value x, since, by definition, $e_q(x) = 0$ for $1 - \lambda(1-q)x < 0$. In contrast, if $q > 1$, $1 - \lambda(1-q)x > 0$, then Eq. (1) exhibits power-law asymptotic behavior. Both cases are relevant for air pollution fits.

The occurrence of *q*-exponentials with $q > 1$ in PDFs of complex systems has been previously explained as originating from a superstatistical dynamics Beck and Cohen (2003); Beck, Cohen, and Swinney (2005); Ourabah (2024). In these types of models, one assumes a temporally fluctuating parameter $\hat{\lambda}$ for local exponential distributions $\sim e^{-\hat{\lambda}x}$. This means that in a sense the parameters of the system are random variables as well, but they change on a much larger time scale. These variations of $\hat{\lambda}$ take place on a long-time scale, which is much longer than the local relaxation time to equilibrium. Physically, one may think of this as describing changing weather conditions (e.g. changes in the wind pattern) at the measuring site. The marginal distribution, obtained by integration over all possible values of $\hat{\lambda}$ and describing the long-term behavior of the air pollution concentration dynamics, is then a *q*-exponential, with

$$q = \frac{\langle \hat{\lambda}^2 \rangle}{\langle \hat{\lambda} \rangle^2}. \qquad (2)$$

Here, $\langle \cdots \rangle$ denotes the expectation with respect to the PDF of $\hat{\lambda}$, see Beck and Cohen (2003) for more details. Strictly speaking, a *q*-exponential is only obtained exactly if $\hat{\lambda}$ is Γ-distributed, but the general idea of superstatistics is that a parameter q can be defined by Eq. (2) for more general distributions different than the Γ distribution as well.

In our automatically applied fitting procedure for the 3544 measuring stations, we obtain the optimal fitting parameters via MLE, i.e. maximizing the likelihood under the assumption that the data originates from the assumed *q*-exponential distribution. One may ask whether other distributions, such as the previously studied lognormal distribution, Weibull distribution, or gamma distribution might yield equally good fits as the *q*-exponential distribution. To illustrate the goodness of fit, we compare all these functions with the measured histogram for *NO* data in Figure 2 for an example site (Barnsley Gawber in
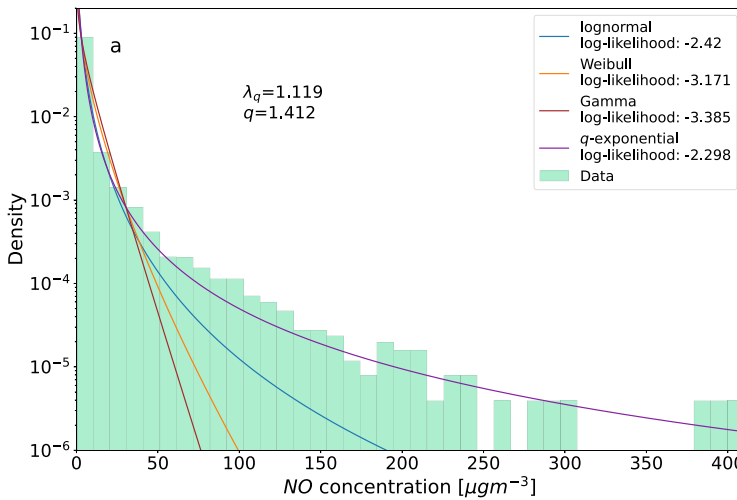
Figure 2. *Measured histogram of NO concentrations, recorded at the example site Barnsley Gawber, UK, together with the best log-normal (blue), Weibull (orange), Gamma (brown), and q-exponential (purple) fits, together with their respective log-likelihood values. Also displayed are the optimum values of the q and λ parameters for the q-exponential distribution. The q-exponential fits the measured data best, as indicated by the highest log-likelihood value. Similar plots can be produced for all 3544 sites.*

the UK). Clearly, the q-exponential fit yields the best fit of the tails of the distribution. It also yields the largest log-likelihood value compared to the other distributions. We systematically compared the goodness-of-fit using maximum likelihood estimation to determine the best fitting parameters and hence computed the likelihood for each fitting function. Overall, we find that q-exponentials generate the highest log-likelihoods for all four pollutants, when averaging over all sites. Hence, we apply q-exponentials as our main method for analyzing the tails of air pollution statistics. See also github He (2024) for full technical details.

2.3. Extracting the long superstatistical time scale T

Utilizing the q-exponential distribution for fitting the PDFs opens the option to identify a *long time scale* T, as discussed in the superstatistics approach Beck and Cohen (2003); Beck et al. (2005). In this view, we assume that on time scales shorter than T, we observe random samples from a simple distribution, which we assume to be exponential distributions here. Meanwhile, the overall complex time series, characterized frequently by heavy-tailed probability distributions, is seen as a superposition of several of these simpler, local distributions.

When considering shorter time slices, if the local time series approximately follows an exponential distribution, the kurtosis for a local snapshot should be $K_{\exp} = 9$. To determine T, we compute the local average kurtosis Beck et al. (2005) as follows:

$$\overline{\kappa}(\Delta t) = \frac{1}{t_{max} - \Delta t} \int_0^{t_{max} - \Delta t} dt_0 \frac{\langle (u - \overline{u})^4 \rangle_{t_0, \Delta t}}{\langle (u - \overline{u})^2 \rangle_{t_0, \Delta t}^2}, \tag{3}$$

where t_{max} is the length of the time series and $\langle \rangle_{t_0, \Delta t}$ denotes the expectation for the time slice of length Δt starting at t_0. The long-time scale T is then defined by the condition $\overline{\kappa}(T) = K_{exp}$. In this case, the average kurtosis of windows of length T has a kurtosis $K_{exp} = 9$. Once T is determined, the time series can be divided into multiple snapshots, each spanning a length of T. This provides a collection of slices with approximately local exponential distributions, each with a distinct decay parameter $\hat{\lambda}$ that fluctuates on the time scale T.

3. Results

Figure 3 shows the best-fitting parameters q and λ for the tails of the observed PDFs at all 3544 sites evaluated. As mentioned above, it is natural to expect q-exponential distributions as these simply arise from the agglomeration of many exponential distributions that have temporal fluctuations of the effective decay rate.

What is noticeable is the fact that one observes an immensely large range of values of the parameter λ for the best-fitting q-exponential as given in Eq. (1) for the various measuring stations. Note the logarithmic scale of the plots. The parameter λ can take on values as small as 10^{-2} up to values as large as 10^2, which spans four orders of magnitude. Typical q-values are in the range $0.8 - 1.4$, but there are subtle differences between the various substances, with NO reaching large q-values such as 1.6, and NO_2 reaching small q-values such as 0.6 in the scattering plots. Figure 3, panels a and b, indicate that roughly q grows linearly with $\log \lambda$ for NO_x. Traffic areas are clustered on the left, while urban/suburban background points are in the middle part, and rural background points are scattered widely at the right. The PDF decay rate increases as λ increases from highly polluted urban traffic sites to less polluted rural background areas. For $PM_{2.5}$ and PM_{10}, there is a slightly different statistical relationship between q values and λ, as can be seen in Figure 3 panels c and d. The urban points approach small λ values such as 0.01 for PM_{10}, and suburban/rural traffic, suburban/rural industrial, and rural background points cluster at the right-hand side with large λ. The observed range of λ values is smaller as compared to the case of NO_x, and the shape of possible values (q, λ) as displayed in the figure is more spherical. The stronger color mixing for PM may be due to the fact that by air movement there is transport of the pollutants, hence the distributions cannot be uniquely identified with the original environmental types where the PM-particles were produced.

So far we conditioned the observed distributions on the environmental type, such as high-traffic region, industrial region, or rural area. This leads to a pretty mixed pattern of states as displayed in Figure 4. However, it is useful to keep the spatial information where the actual measurements were done. Thus, the
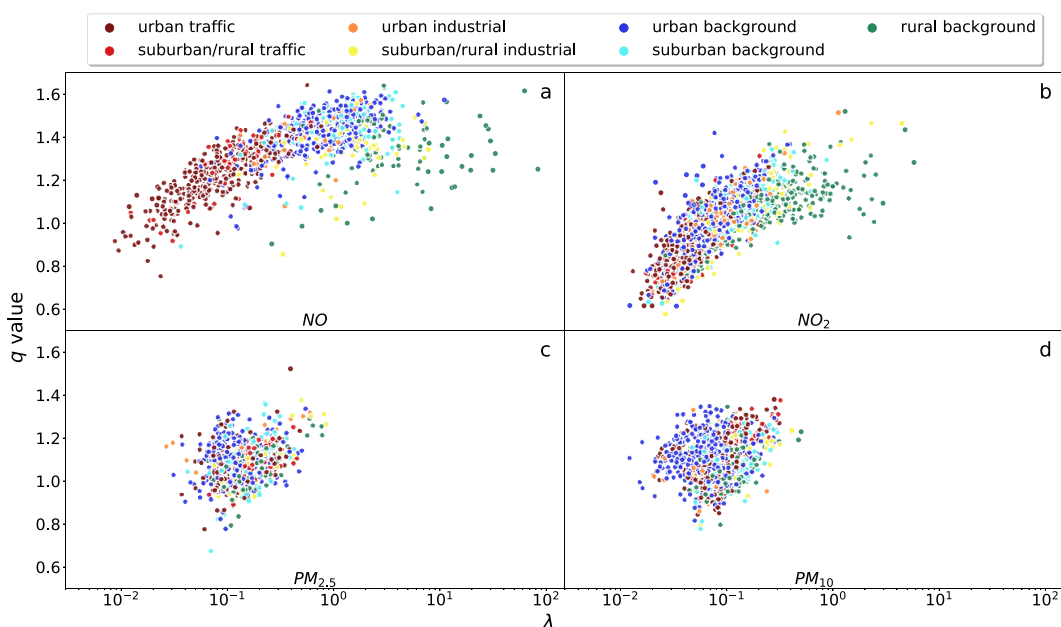

Figure 3. *Best-fitting parameters of q-exponentials at the various measuring stations. There is an increasing trend of q versus $\log \lambda$ for NO (a) and NO_2 (b), whereas a disk-shaped pattern is observed for $PM_{2.5}$ (c) and PM_{10} (d). The colors encode the area type where the measurements were done. There are predominant patches of a single color in the parameter space for NO and NO_2, where for $PM_{2.5}$ and PM_{10} the color pattern looks more mixed. Deep green points correspond to cleaner rural areas.*
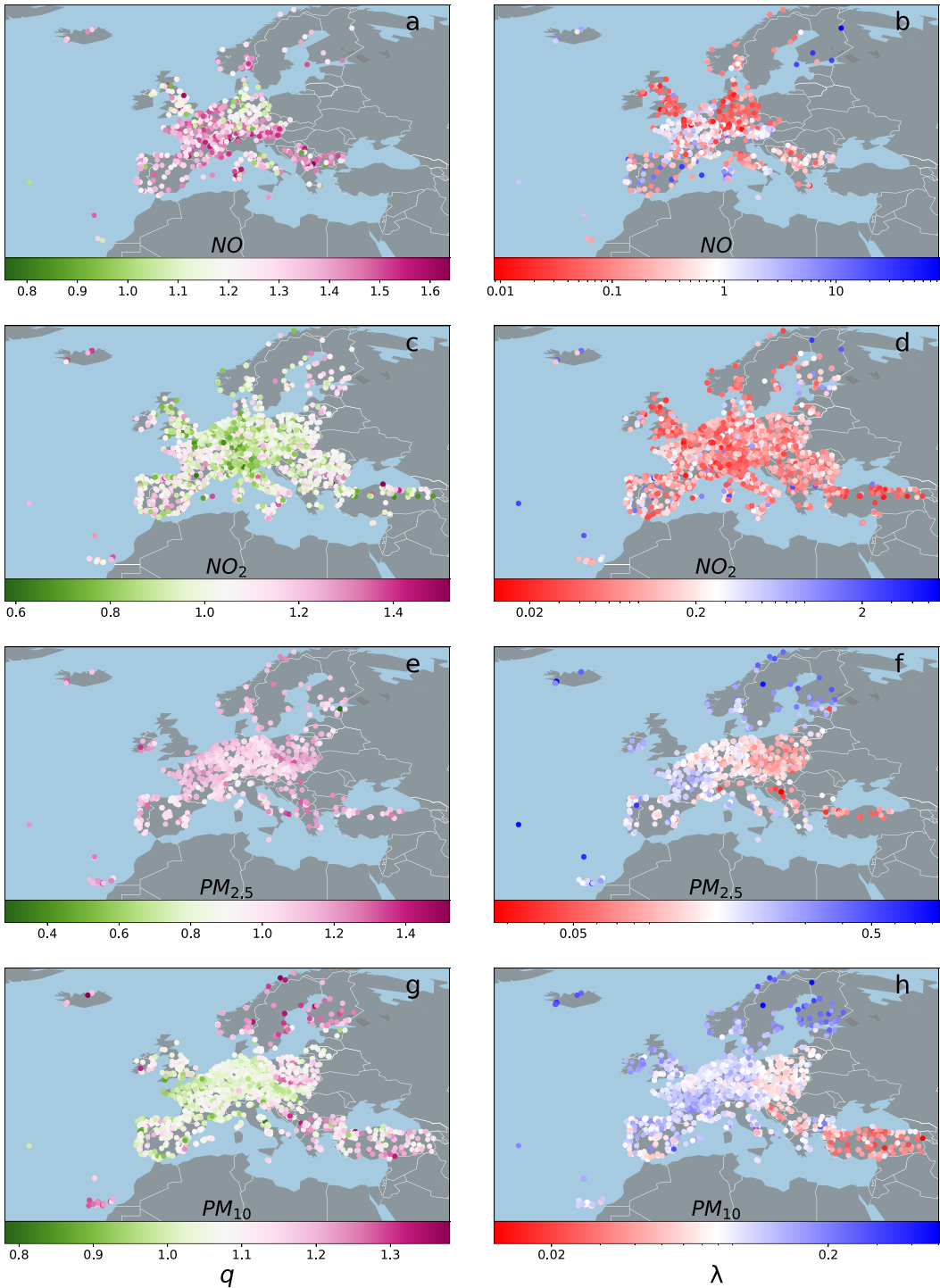
Figure 4. *Spatial distribution of best-fitting parameters* (q, λ) *characterizing the measured PDFs of* NO_x *and* PM_x *pollutants across Europe. The color codes are directly indicated in the individual figures. A large value of* q *indicates heavy tails in the distribution. A small value of* λ *indicates heavy average pollution. The pollutant characteristics across Europe is quite inhomogeneous.*
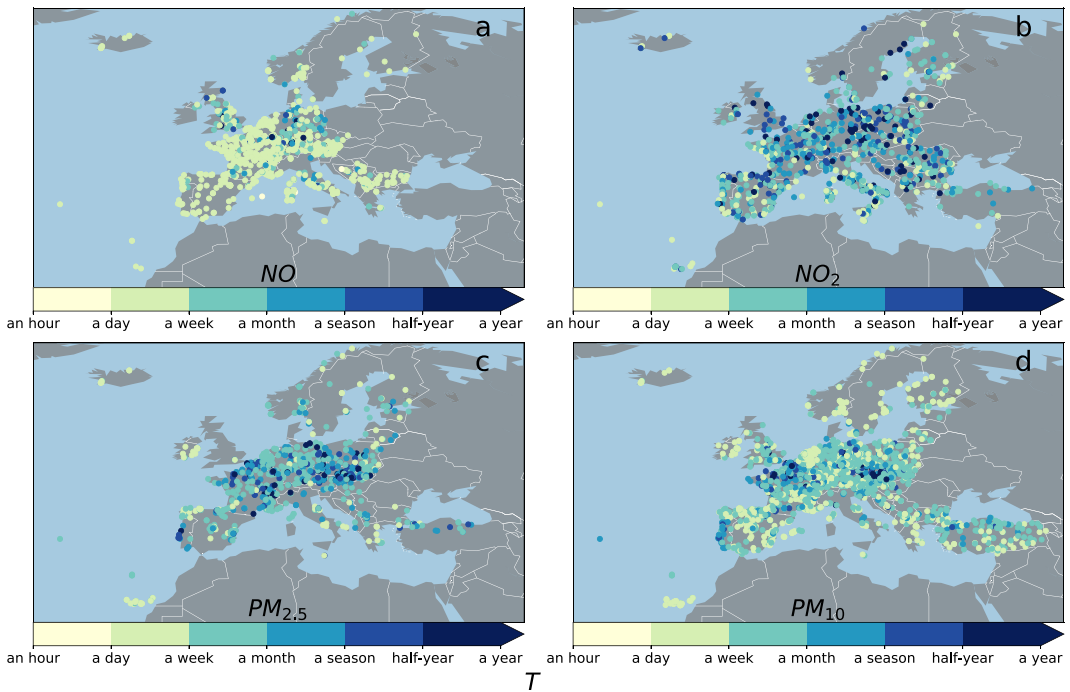
Figure 5. *The long time scales T describing the scale of typical changes of the temporal mean and variance of the measured time series for NO (a), NO$_2$ (b), PM$_{2.5}$ (c), and PM$_{10}$ (d). The color coding is explained at the bottom of each figure.*

idea in the following is to plot an air pollution map of Europe where the color encodes the value of the best-fitting fitting parameters (q, λ) relevant in that given region of Europe. This map is shown in Figure 5. The novel approach here is that we do not only consider average values of pollution concentrations as observed in a given region but encode information on the entire probability distribution. The parameter q describes the power-law tails of the probability densities, which decay with an exponent given by $-1/(q-1)$. Thus, q is related to the frequency of occurrence of extreme events describing high-pollution states.

The q values for NO_2 are smaller in Germany and the UK as compared to other countries, in fact below 1 (green colors), which indicates there are no heavy tails. Since the primary emission source of NO_x is fuel combustion, this pattern suggests that these regions cope better with extreme situations and have a decreased susceptibility to extreme NO_x events. The q values for PM pollutants exhibit a slightly different spatial pattern, with lower values (green) in the Western regions and higher values (red) in the East. The λ values in Central Europe and the UK are smaller compared to those in other European regions, indicating a higher average pollution, in particular for NO. As for the spatial distribution of λ values for PM pollutants, there is a slight eastward shift in the clustering of red dots, suggesting a higher pollution in Eastern Europe.

Note that the average concentration of pollutants is proportional to λ^{-1}. This means the smaller λ the bigger is the average air pollution concentration in that region. For this reason, we have color-coded smaller values of λ in red and bigger values in blue. Some of the Scandinavian countries appear to host a couple of blue dots, whereas in Turkey there are big concentrations of red dots, in particular for PM_{10}. For the q-values, describing extreme events, the tail in the distribution is more pronounced and the bigger q. Some Eastern European countries exhibit particularly heavy tails, as indicated by red colors in the q panels. On the other hand, if q is smaller than 1 then the distributions live on a finite support, and there are no heavy tails. This case is indicated by green colors in the q-plane. Some regions of Germany and Western Europe exhibit predominantly green colors.

Figure 5 displays the superstatistical time scale T as extracted for the four pollutants, indicating distinct temporal dynamics. NO overall has a majority of shorter time scales T, typically less than a week. NO_2, on the other hand, comprises higher values of time scales T, mostly ranging between a week to a year. This difference could be explained by the fact that NO is highly reactive and tends to quickly oxidize to form NO_2 in the presence of oxygen. This rapid transformation means that NO concentrations can change swiftly over short periods. The extracted time scales for $PM_{2.5}$ are of similar order of magnitude as for NO_2, whereas the extracted time scales for the PM_{10} dynamics appear to be smaller than for $PM_{2.5}$ on average.

4. Discussion

In this paper, we applied data analytics tools to better understand the probability densities of measured air pollution time series. In particular, we were interested in the tail behavior. The parameters q and λ, as obtained from the optimum fittings of data from 3544 measuring stations in Europe, exhibit interesting patterns in the (q, λ) plane. They also exhibit interesting patterns when associated with a map of Europe. There are significant differences between the NO, NO_2, and the PM statistics.

The shape of scattered (q, λ) values in the plane depends on the local characteristics, e.g. whether the measurements are done in a rural area or densely populated area. There is a typical range of parameter values for a given class of this type. We observe an immensely large range of values of the parameter λ for the various measuring stations.

We also observe a significant spatial heterogeneity in the best-fitting shape parameters of the PDFs. Eastern Europe shows a tendency for more severe PM pollution events and higher average concentration exposure, described by higher q values and lower λ values, respectively. Northern Europe shows a tendency for lower air pollutant concentration, described by higher λ values, but is more susceptible to extreme events, described by higher q values. Additionally, there are strong variations of PDFs at local level as well. These disparities highlight the importance of tailored strategies for air pollution mitigation strategies in different areas. Recall that q contains information about the tail, i.e. extreme events and λ about the scale and thereby the mean pollution level, i.e. we distinguish thereby between regions with low and high average pollution and simultaneously regions with many or few extreme events.

We also extracted the long superstatistical time scales T for the measured time series for each of the 3544 measuring sites and each of the four pollutants considered, thus incorporating dynamical information that amends the static information contained in the PDFs. This information was displayed in the form of a map of Europe, too, and shows a variety of time scales relevant at different spatial locations. The observed wide range of time scales, in particular for NO_2 and $PM_{2.5}$, can be attributed to a variety of factors, including but not limited to fluctuating local meteorological conditions and anthropogenic activities.

Future research, for each given location taking into account the local circumstances, should help to gain a clearer systematic understanding of the relationship between meteorological and anthropogenic factors and the relevant time scales and (q, λ)-values of air pollutants. A main result of our investigation is strong local heterogeneity, i.e. the temporal dynamics of air pollution can be quite different if one proceeds from one local measurement station to the next. Different local wind conditions are expected to play an important role in this context.

Further statistical analysis of more detailed data at local level could support policymakers to produce more precise rules and thresholds for individual types of environmental conditions and meteorological conditions, taking into account fluctuations, extreme events, and variations of concentrations on different time scales.

For our statistical analysis, the different types of air pollutants have different data availability in the various geographical regions. For this reason, in the air pollution maps that we have produced in this paper, some regions are less frequently covered by points than others. We hope that in the future more measurements at further locations will become available, given the importance of the air pollution problem. Our analysis could also be extended to other substances, such as sulfur oxide, carbon oxide,

and ozone, besides NO_x and PM. We stress again that the detailed description of the entire pollution PDF, including the exact behavior of the tails, seems critical here to better estimate the risks of very high pollution situations.

References

Baldasano JM (2020) COVID-19 lockdown effects on air quality by NO2 in the cities of Barcelona and Madrid (Spain). *Science of The Total Environment 741*, 140353. doi: 10.1016/j.scitotenv.2020.140353.

Beck C (2007) Statistics of three-dimensional lagrangian turbulence. *Physical Review Letters 98*(6), 064502. doi: 10.1103/PhysRevLett.98.064502.

Beck C, **Benedek G**, **Livadiotis G**, **Rapisarda A**, **Tirnakli U and Tsallis C** (2020) Nonextensive statistical mechanics, super-statistics and beyond: theory and applications in astrophysical and other complex systems. *The European Physical Journal Special Topics 229*(5), 707–709.

Beck C and Cohen EGD (2003) Superstatistics. *Physica A: Statistical Mechanics and its Applications 322*, 267–275.

Beck C, **Cohen EGD and Swinney HL** (2005) From time series to superstatistics. *Physical Review E 72*(5), 056133.

Carslaw DC (2017) *Worldmet: Import Surface Meteorological Data from NOAA Integrated Surface Database (ISD), GitHub Repository.* Available at https://github.com/davidcarslaw/worldmet.

Environmental Research Group and others (2020) The Effect of COVID-19 Lockdown Measures on Air Quality in London in 2020. Report (King's College London, 2020), Available at https://assets.ctfassets.net~\protect\global\let\T1\textellipsis.\kern\fontdimen3\font.\kern\fontdimen3\font.\kern\fontdimen3\font\T1\textelling

European Commission (2011) 2011/850/EU: Commission Implementing Decision of 12 December 2011 laying down rules for Directives 2004/107/EC and 2008/50/EC of the European Parliament and of the Council as regards the reciprocal exchange of information and reporting on ambient air quality (notified under document C(2011) 9068). *Official Journal of the European Union 54*, 86–106. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX

European Parliament (2011) Directive 2011/24/EU of the European Parliament and of the Council of 9 March 2011 on the application of patients' rights in cross-border healthcare. *Official Journal of the European Union 50*(88), 45–65.

European Parliament, Council of the European Union (2008) DIRECTIVE 2008/50/EC OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 21 May 2008 on ambient air quality and cleaner air for Europe, *Official Journal of the European Union.* Available at https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX

GBD 2016 Risk Factors Collaborators and others (2017) Global, regional, and national comparative risk assessment of 84 behavioural, environmental and occupational, and metabolic risks or clusters of risks, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016, *Lancet (London, England) 390*(10100), 1345.

Giri AJ, **Schäfer B**, **Verma R** et al. (2023) Lockdown effects on air quality in megacities during the first and second waves of COVID-19 Pandemic. *Journal of The Institution of Engineers (India): Series A 104* 155–165. doi: 10.1007/s40030-022-00702-9.

Grange SK (2019) saqgetr R package. Available at https://github.com/skgrange/saqgetr.

Hanel R, **Thurner S and Gell-Mann M**. (2011) Generalized entropies and the transformation group of superstatistics. *Proceedings of the National Academy of Sciences 108*(16), 6390–6394.

He H (2024) Code for Analyzing Spatial Heterogeneity of Air Pollution Statistics. Available at: https://github.com/hurst0415/Spatial-heterogeneity-of-air-pollution-statistics

He H, **Schäfer B and Beck C** (2022) Spatial heterogeneity of air pollution statistics in Europe. *Science Reports 12*, 12215. DOI: 10.1038/s41598-022-16109-2.

Health Effects Institute (2019) *State of Global Air 2019, Special Report*. Boston, MA: Health Effects Institute.

Jizba P and Korbel J (2019) Maximum entropy principle in statistical inference: case for Non-Shannonian Entropies. *Physical Review Letters 122*(12), 120601.

Lu H-C (2003) Comparisons of statistical characteristic of air pollutants in taiwan by frequency distribution. *Journal of the Air & Waste Management Association* 53(5) 608–616. doi: 10.1080/10473289.2003.10466194.

Metzler R (2020) Superstatistics and non-Gaussian diffusion. *The European Physical Journal Special Topics* 229(5), 711–728.

Ortiz AG, **Guerreiro C**, **Soares J** and **European Environment Agency** (2020) *Air Quality in Europe: 2020 Report*, EEA Report. Publications Office of the European Union. ISBN 9789294802927. Available at https://books.google.co.uk/books?id=swtez gEACAAJ.

Ourabah K (2024) Superstatistics from a dynamical perspective: entropy and relaxation. *Physical Review E 109*, 014127.

Shah ASV and **others** (2015) Short term exposure to air pollution and stroke: systematic review and meta-analysis. *BMJ 350*, h1295.

Smith A, **Lott N and Vose R** (2011) The integrated surface database: recent developments and partnerships. *Bulletin of the American Meteorological Society 92*(6), 704–708.

Tsallis C (2009) *Introduction to Nonextensive Statistical Mechanics: Approaching a Complex World*, Springer, 1(1), 2–1.

Williams G and Schäfer B and Beck C (2020) Superstatistical approach to air pollution statistics. *Physical Review Research 2*(1), 013019. doi: 10.1103/PhysRevResearch.2.013019.

World Health Organization (2020) How Air Pollution Is Destroying Our Health. Available at https://www.who.int/news-room/spotlight/how-air-pollution-is-destroying-our-health.

World Health Organization and others (2018) *WHO Global Ambient Air Quality Database (Update 2018)*. Geneva, Switzerland: World Health Organization.