scientific reports

Check for updates

OPEN

Investigating the spatiotemporal associations between meteorological conditions and air pollution in the federal state Baden-Württemberg (Germany)

Leona Hoffmann[1]✉, Lorenza Gilardi[2], Marie-Therese Schmitz[3], Thilo Erbertseder[2], Michael Bittner[2], Sabine Wüst[2], Matthias Schmid[3] & Jörn Rittweger[1,4]

When analyzing health data in relation to environmental stressors, it is crucial to identify which variables to include in the statistical model to exclude dependencies among the variables. Four meteorological parameters: temperature, ultraviolet radiation, precipitation, and vapor pressure and four outdoor air pollution parameters: ozone (O_3), nitrogen dioxide (NO_2), particulate matter ($PM_{2.5}$, PM_{10}) were studied on a daily basis for Baden-Württemberg (Germany). This federal state covers urban and rural compartments including mountainous and river areas. A temporal and spatial analysis of the internal relationships was performed among the variables using (a) cross-correlations, both on the grand ensemble of data as well as within subsets, and (b) the Local Indications of Spatial Association (LISA) method. Meteorological and air pollution variables were strongly correlated within and among themselves in time and space. We found a strong interaction between nitrogen dioxide and ozone, with correlation coefficients varying over time. The coefficients ranged from negative correlations in January (−0.84), April (−0.47), and October (−0.54) to a positive correlation in July (0.45). The cross-correlation plot showed a noticeable change in the correlation direction for O_3 and NO_2. Spatially, NO_2, $PM_{2.5}$, and PM_{10} concentrations were significantly higher in urban than rural regions. For O_3, this effect was reversed. A LISA analysis confirmed distinct hot and cold spots of environmental stressors. This work examined and quantified the spatio-temporal relationship between air pollution and meteorological conditions and recommended which variables to prioritize for future health impact analyses. The results found are in line with the underlying physico-chemical atmospheric processes. It also identified postal code areas with dominant environmental stressors for further studies.

Keywords Air pollution, Environmental stressors, Meteorological data, Cross-correlation, LISA

Abbreviations

BVOC	Biogenic volatile organic substances
BW	Baden-Württemberg
CAMS	Copernicus Atmosphere Monitoring Service
CCF	Cross-correlation function
COVID-19	Coronavirus SARS-CoV-2
CO_2	Carbon dioxide
C3S	Copernicus Climate Change Service
DLR	German Aerospace Center
ECMWF	European Centre for Medium-Range Weather Forecasts
EEA	European Environment Agency
LISA	Local indication of spatial association

[1]Institute of Aerospace Medicine, German Aerospace Center (DLR), Cologne, Germany. [2]German Remote Sensing Data Center, German Aerospace Center (DLR), Weßling, Germany. [3]Institute of Medical Biometry, Informatics and Epidemiology, University Hospital Bonn, Bonn, Germany. [4]Department of Pediatrics and Adolescent Medicine, University Hospital Cologne, Cologne, Germany. ✉email: leona.hoffmann@dlr.de

NO_2	Nitrogen dioxide
NO_x	Nitrogen oxide
O_3	Ozone
PCA	Principal component analysis
$PM_{2.5}$	Particulate matter with a diameter of 2.5 micrometers or smaller
PM_{10}	Particulate matter with a diameter of 10 micrometers or smaller
Prec	Precipitation
sd	Standard deviation
r	Correlation coefficient
Temp	Temperature
UV	ultraviolet radiation
VOC	Volatile organic compound
VP	Vapor pressure
WHO	World Health Organization

The human-made climate change threatens human health not only through extreme meteorological conditions, but also through polluted air that often accompanies it[1,2]. There are concrete plans for the reduction of air pollutant concentrations with guidelines from the World Health Organization (WHO)[3,4] and air quality directive of the European Parliament and the Council[5]. However, there is still a need to better understand outdoor air pollution and their internal relations to prevent potential misinterpretation of the outcomes. In the past, numerous studies were conducted on the effect of air pollution on human health[6–10]. All of them confirm that outdoor air pollution harms human health. Air pollution causes acute and chronic health effects and affects various systems, and organs[8].

Air pollutants are often released in conjunction, such as nitrogen dioxide (NO_2), carbon dioxide (CO_2) and particulate matter from combustion processes. The dispersion and deposition of air pollutants through meteorological factors is subject to variation given by emission sources, chemical transformations and average atmospheric lifetime. As a consequence there are spatiotemporal covariations between the various pollutants and meteorological variables.

One of the pollutants is particulate matter with a diameter up to 2.5 µm ($PM_{2.5}$) which is a variable that has been frequently studied in the literature and has often shown to have a strong negative effect on health. Short-term as well as long-term exposure to $PM_{2.5}$ has an impact on mortality and morbidity[6]. The evidence supports the possibility that both $PM_{2.5}$ and PM_{10} (particulate matter with a diameter up to 10 µm) are associated with increased mortality from all causes, cardiovascular disease, respiratory disease, and lung cancer[11,12]. In Europe in 2020, exposure to $PM_{2.5}$ concentrations above the WHO guideline level of 5 $\mu g/m^3$ resulted in 275,000 premature deaths[13]. Most individuals residing in Germany inhabit polluted regions[14]. Air pollution negatively impacts virus-transmitting infections such as influenza[15,16] and COVID-19[17–19]. Irrespective of which specific health outcome is considered, understanding the interdependencies between environmental variables and the selection of environmental stressors is a key aspect in epidemiological analyses. Previous studies that examined the relationship between meteorological and air pollution variables were conducted in Cairo (Egypt)[20], China[21], Rome (Italy)[22] and Stuttgart (Germany)[23], for example. Studies on the spatiotemporal variability of tropospheric ozone and nitrogen dioxide are available for Athens (Greece)[24], major cities in India[25] as well as in form of worldwide reviews[26].

In a study on the link between influenza and air pollution[15], a strong correlation was found between some of the environmental stressors considered, which included air pollutants and meteorological variables. Other studies[27,28] made use of the principal component analysis (PCA) method for dimension reduction[29]. However, PCA combines pollutants to create principal components. Understanding the individual coefficients can be difficult because they lack interpretability[30]. PCA is unsuitable for our analysis as we want to focus on the specific relationships between and among meteorological conditions and air pollution. Therefore, we opted for temporal and spatial analysis techniques that allow for a more conclusive interpretation of the environmental stressors and investigate the internal dependencies.

The aim of this analysis is to better characterize and identify the spatiotemporal relationships among environmental parameters. Our analysis takes into account multiple stressors and their spatial and temporal connections across the entire state of Baden-Württemberg. The scope is to reduce the number of variables needed in the epidemiological analysis and therefore simplify them and avoid biased results caused by correlated factors.

In the past, research has focused either on analyzing specific cities or studying one environmental factor affecting larger regions. This study is valuable because it considers multiple environmental stressors and covers at the same time the cross-sectional region of Baden-Württemberg (BW). Due to the combination of urban and rural areas including mountainous and river regions, Baden-Württemberg is particularly well suited for analyzing the spatial-temporal relationships between air quality and meteorological conditions. In addition, this paper provides decision support for the selection of environmental variables for future analyses. Understanding how different stressors are interconnected can offer valuable insights to aid in future health impact analyses and assist other researchers in related fields.

Through our spatial and temporal analysis, we have identified distinct differences and similarities in terms of spatiotemporal patterns in environmental stressors. Based on these findings, we suggest prioritizing certain variables for further investigation. Additionally, we categorized postal code areas into specific groups based on their environmental stressor pattern, providing a spatial delineation.

Methods

Study area

Baden-Württemberg is Germany's third most populated state, with an area of approximately 36 thousand km^2 and a population of approximately 11 million. Geographically, the state is located in southwestern Germany and includes urban and rural areas. The state capital Stuttgart is the largest city in BW with about 630 thousand inhabitants and is located in the center of the state. Other large cities are Mannheim (310 thousand inhabitants), Karlsruhe (308 thousand inhabitants), and Freiburg (231 thousand inhabitants). Especially in the south, southwest, and southeast of the state, there are many rural and mountain regions, including the Black Forest, the Lake Constance, the edge of the Alps, and the low mountain range Swabian Alb. Of all 1101 communities of BW, 586 have fewer than 5 thousand inhabitants[31].

We obtained a shapefile of the postal code areas in BW from the Esri Germany database[32] in order to handle the postal code areas.

The population density provides information about the number of inhabitants per km^2 and was calculated from the available data as follows:

$$\text{population density} = \frac{\text{number of inhabitants}}{\text{postal code area (km}^2\text{)}} \tag{1}$$

Based on a map of population density from the German Federal Institute for Population Research (Bundesinstitut für Bevölkerungsforschung)[33], we decided to introduce four population density categories. Figure 1 provides a graphical overview of the distribution of the population density categories for BW at postal code level.

Unsurprisingly, postal code areas near cities such as Stuttgart and Mannheim are assigned to category 4, that means densely populated areas. The areas of Black Forest, Swabian Alb, and northeastern of BW predominantly have a population density below 151 inhabitants per km^2.

Environmental variables

Eight environmental variables were considered in more detail. These were split into four meteorological parameters [temperature (*Temp*), precipitation (*Prec*), vapor pressure (*VP*) and UV radiation (*UV*)] and four outdoor air pollution parameters [ozone (O_3), nitrogen dioxide (NO_2), particulate matter ($PM_{2.5}$, PM_{10})]. Air pollution surface concentrations were retrieved from the European air quality reanalysis dataset provided by the Copernicus Atmosphere Monitoring Service (CAMS)[34]. The data source of the meteorological parameters was the ERA5 reanalysis dataset as provided by the Copernicus Climate Change Service (C3S) of the European Centre for Medium-Range Weather Forecasts (ECMWF)[35]. The datasets had a native spatial resolution of 0.1° × 0.1° and a temporal resolution of one hour. The variables were geographically aggregated to postal code areas as described in[36]. A detailed evaluation of uncertainties of the applied data can be found in[37–39]

Table 1 gives an overview of the environmental stressors. Daily mean values between 2010 and 2018 were available for each variable in the table, aggregated at the postal code level in BW and resulting in a total of 3,931,252 data. Next is a brief explanation of the environmental variables[40,41]. The variable *Temp* was measured two meters above the ground. *Prec*, expressed in mm/day, represented the depth of water if it were evenly

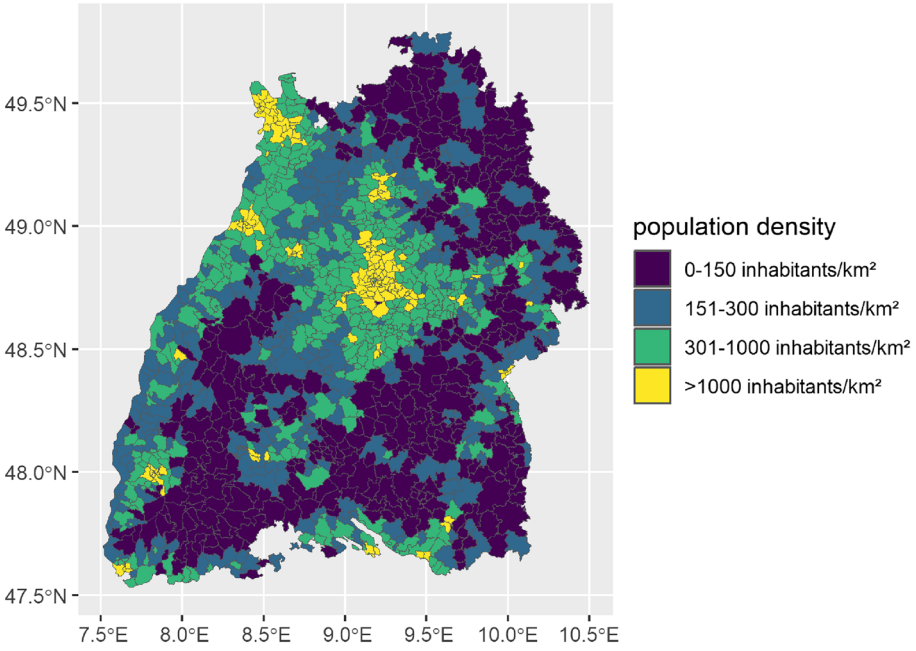

Figure 1. Postal code areas in BW categorized by population density.

Variable	Parameter	Value
Temp (°C)	Mean (sd)	9.7 (7.7)
	Median [min, max]	9.8 [−18.9, 30.9]
Prec (mm/day)	Mean (sd)	3.4 (6.0)
	Median [min, max]	0.9 [0, 99.3]
VP (hPa)	Mean (sd)	10.0 (4.2)
	Median [min, max]	9.3 [1.0, 24.6]
UV (W)	Mean (sd)	15.4 (9.7)
	Median [min, max]	14.1 [0.4, 37.0]
O_3 ($\mu g/m^3$)	Mean (sd)	51.5 (23.0)
	Median [min, max]	52.4 [0.3, 149.0]
NO_2 ($\mu g/m^3$)	Mean (sd)	12.2 (7.3)
	Median [min, max]	10.4 [1.07, 66.4]
$PM_{2.5}$ ($\mu g/m^3$)	Mean (sd)	11.0 (6.7)
	Median [min, max]	9.5 [0.7, 72.3]
PM_{10} ($\mu g/m^3$)	Mean (sd)	14.9 (8.4)
	Median [min, max]	13.2 [0.9, 83.5]

Table 1. Overview of environmental stressors. Overview of meteorological data and air pollutants according to the statistical parameters mean, standard deviation (sd), median, minimum (min) and maximum (max) value. The variables cover all postal code areas in BW and are based on daily measurements from 2010 to 2018.

distributed over the area under consideration. The variable *VP* was a variable constructed from $2m$ dewpoint temperature, as expressed by the following empirical formula[42]:

$$e = 6.112 * \exp(\frac{17.67 * T_d}{T_d - 243.5}) \tag{2}$$

where e was the vapor pressure in hectopascal (hPa) and T_d was the dew point temperature in °C. For *UV*, the unit was converted from J/m^2 to W/m^2 by dividing the integration time in seconds, resulting in a mean value of 15.4 W/m^2 in the processed data. O_3 is a colorless and toxic gas in the atmosphere close to the ground (troposphere). O_3 is one of the main components of photochemical smog and is produced by complex photochemical processes during intense sunlight. NO_2 is a reactive nitrogen compound that is commonly released from the combustion of fuels in the transportation and industrial sectors. $PM_{2.5}$ and PM_{10} are not single pollutants but a mixture of many components such as sulfates, nitrates, ammonia, sodium chloride, black carbon, mineral dust, and water. Depending on the size of the particles, a distinction is made between PM_{10} and $PM_{2.5}$. Particulate matter is generated, in particular, by combustion processes in motor vehicles, power plants, small combustion plants, domestic heating as well as in metal and steel production.

Statistical methods

Pearson's correlation coefficient was used to determine the pairwise correlations between environmental stressors. The strength of correlation is classified as follows: $r > 0.9$ almost perfect, $0.7 < r \leq 0.9$ very large, $0.5 < r \leq 0.7$ large, $0.3 < r \leq 0.5$ moderate, $0.1 < r \leq 0.3$ small and $r < 0.1$ trivial[43]. From the pearson correlation coefficient emerges the concept of cross- and autocorrelations. In short, a cross-correlation examines the relationship between two or more parameters over time or in space, whereas an autocorrelation examines the relationship to itself (also possible in time and space). The concept of temporal cross-correlation was used to make more precise statements about the temporal internal dependencies of environmental stressors[22]. In a cross-correlation function (ccf), two time series, $x(t)$ and $y(t)$, are examined for correlations with a time offset[44,45]. The formula can be represented as follows:

$$\text{ccf} = \frac{\sum_{t=1}^{N-1}[(x(t) - \bar{x}) * (y(t - \text{lag}) - \bar{y})]}{\sqrt{\sum_{t=1}^{N-1}(x(t) - \bar{x})^2}\sqrt{\sum_{t=1}^{N-1}(y(t - \text{lag}) - \bar{y})^2}} \tag{3}$$

where $\bar{x}$ and $\bar{y}$ denote the mean over time of the corresponding series, respectively. The time series $x(t)$ is fixed, and $y(t \pm lag)$ has a time lag, which is possible in both directions, i.e., x leads y or x lags y. In a resulting correlation plot, horizontal lines represent the individual correlations of the two time series with the respective time lag.

A local indication of spatial association (LISA) model was used for the spatial analysis. The LISA statistic measures the degree of autocorrelation between a geographical location and its neighbors, identifying so-called hot and cold spots. For instance, hot spots refer to areas with significantly high values that are surrounded by postal code regions with high values. This method was developed by Luc Anselin[46] and, among others, applied in the context of air pollution[16,47]. Several statistics represent the measure of spatial autocorrelation[48,49]. Here, we used the local Moran statistic, whereby the statistic was applied to each environmental stressor individually. The local Moran's I statistic was given as follows:

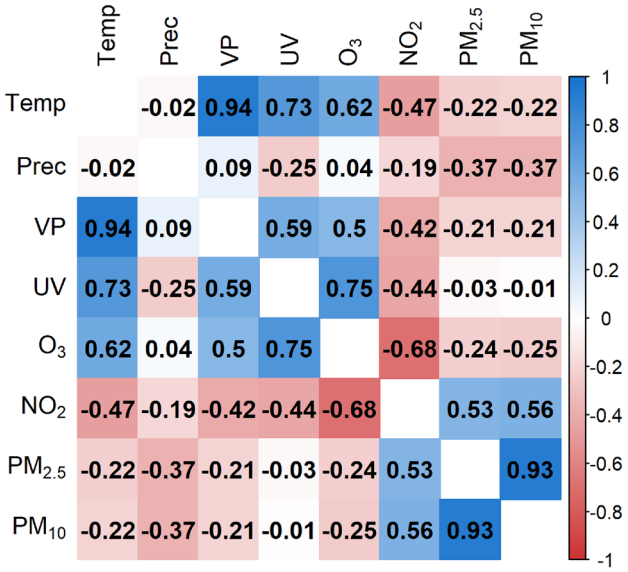

Figure 2. Pearson correlation matrix based on daily measurements from 2010 to 2018 across BW. The more intense the color, the stronger the correlation between the two variables, whereas the color blue indicated positive correlations and red negative correlations.

$$I_i = \frac{(x_i - \bar{X})}{S_i^2} \sum_{j=1, j\neq i}^{n} w_{i,j}(x_j - \bar{X}) \tag{4}$$

where x_i was the stressor concentration at location i, x_j was the concentration of spatial lag j (neighbors), and $\bar{X}$ was the global mean of the environmental stressor. The spatial weight between i and j was described by the matrix $w_{i,j}$, and the total number of observations was n. S_i^2 was a constant for all locations with:

$$S_i^2 = \frac{\sum_{j=1, j\neq i}^{n} (x_j - \bar{X})^2}{n-1} \tag{5}$$

Moran's test uses a null hypothesis of randomly dispersed data. All statistical analyses were conducted using R[50], and maps and figures 1 to 9 were generated in R version 4.3.0.

Results

A matrix of pairwise Pearson's correlation coefficients between the different environmental parameters was shown in Fig. 2.

The correlations represented associations across the entire observation period, that means over all days from 2010 to 2018 and total BW. As expected the temperature and vapor pressure (correlation coefficient (r) 0.94) and $PM_{2.5}$ and PM_{10} ($r = 0.93$) had a strong positive correlation. In addition, there were other relatively strong correlations. The correlation coefficient between *UV* radiation and O_3 was 0.75. This means that the stronger the radiation was, the higher the O_3 concentration. O_3 and NO_2 are negatively correlated ($r = -0.68$). If one air pollutant was low, the other was high.

Temporal analysis

Since this part focuses on temporal relationships, the spatial separation into postal code areas was neglected for this section and the values were averaged over time. It was investigated whether the pairwise Pearson correlations differ across months. We carefully examined correlation plots for all months and decided to include the first correlation plot of each quarter (January, April, July, and October) in the manuscript. Within the manuscript, we do not display all the months explicitly, as some of them represent transitions between the extreme figures shown. The correlation matrices for every month are available in the Supplementary information. The results for the four months January, April, July and October were presented in Fig. 3.

In January, there were many negative correlations. In April, the correlation matrix was primarily characterized by lower correlations. In July, stronger correlations dominate and in October, again, many weaker correlations were visible. The matrices of October and April were overall very similar. They seemed to be intermediate states between the more extreme correlations in January and July.

For O_3, the sign of the correlation coefficient with *Prec* was not constant. The strength of the correlation differed in January ($r = 0.59$), April ($r = -0.13$), July ($r = -0.43$) and October ($r = 0.21$). The medium-large negative correlation between *UV* and *Prec* was almost constant over the months. *UV* and O_3 were slightly correlated

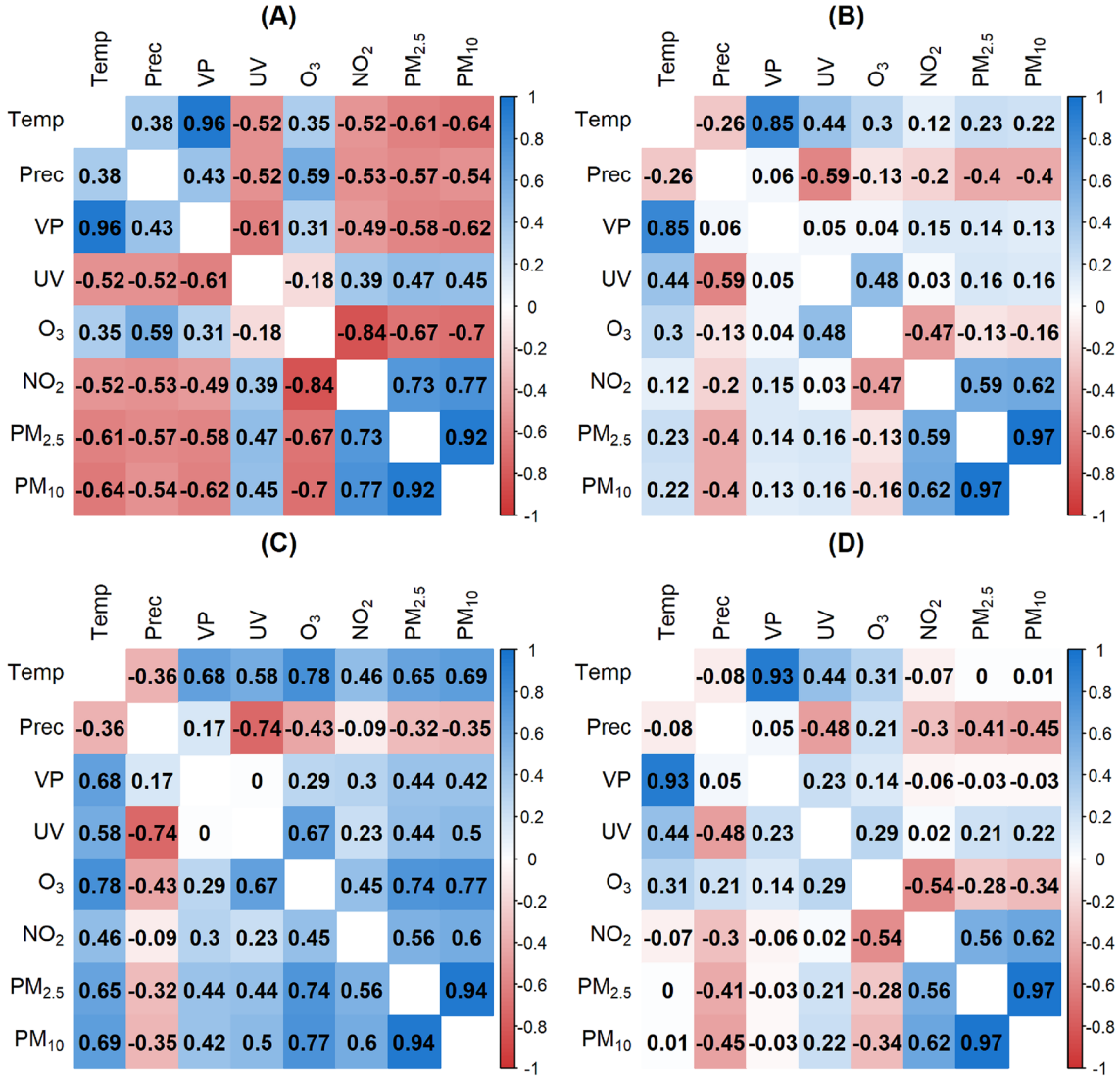

Figure 3. Pearson correlation matrix over months for January (**A**), April (**B**), July (**C**) and October (**D**).

in January ($r = -0.18$), moderately positively correlated in April ($r = 0.48$), strongly positively correlated in July ($r = 0.67$), and slightly correlated in October ($r = 0.29$). The variables O_3 and NO_2 were very strongly negatively correlated in January ($r = -0.84$), moderately negatively correlated in April ($r = -0.47$), moderately positively correlated in July ($r = 0.45$), and again negatively correlated in October ($r = -0.54$). The correlation between NO_2 and $PM_{2.5}$, as well as PM_{10}, was positively correlated over all months. In January, the correlation was stronger, with values around 0.75, than in the other months, with values around 0.6. To summarize, Fig. 3 showed numerous changes of sign over months. Specifically, 14 of the total 28 correlation coefficients showed a change in sign between January and July.

The correlation between NO_2 and O_3 was already prominent in Fig. 3. Therefore, the auto- and cross-correlation relationship between the O_3 and NO_2 variables was depicted in Fig. 4. Overall, the plots by year look very similar, so exemplary, the correlations of the environmental stressors in Fig. 4 were presented for 2018. All graphs had a low point at lag nine and a high point at lag 14. There were noticeable differences in the plots. Specifically, lag 32 in (A) showed a dip that was not present in (B) and (C). In Fig. 4D there was an annual periodicity with the highest positive Pearson correlation coefficient reaching around 0.4 when NO_2 comes before O_3 and approximately 0.3 when NO_2 follows O_3. In addition, there was a pattern that repeats about every 13 days in (C).

The output of the cross-correlation function between NO_2 and O_3 in Fig. 5 was essentially the same as the day-level correlation in Fig. 4. The figures for NO_2 and *Temp* as well as NO_2 and *UV* looked very similar to NO_2 and O_3. There was also a change in signs and another high point after about half a year in both directions. A small structure in Fig. 5 was seen between NO_2 and *Prec* in the correlation values shifted by months. That means a wave-like structure similar to NO_2 and O_3 can be observed. In contrast to NO_2 and O_3, the correlation coefficient starts in the positive range at lag 0, changes to negative with increasing lags, and returns to positive. However, these were relatively weak correlations. A shift made sense content-wise because rain washes the air clean of pollutants. NO_2 and $PM_{2.5}$ started at lag 0 with a relatively strong correlation value above 0.5. For positive lag values,

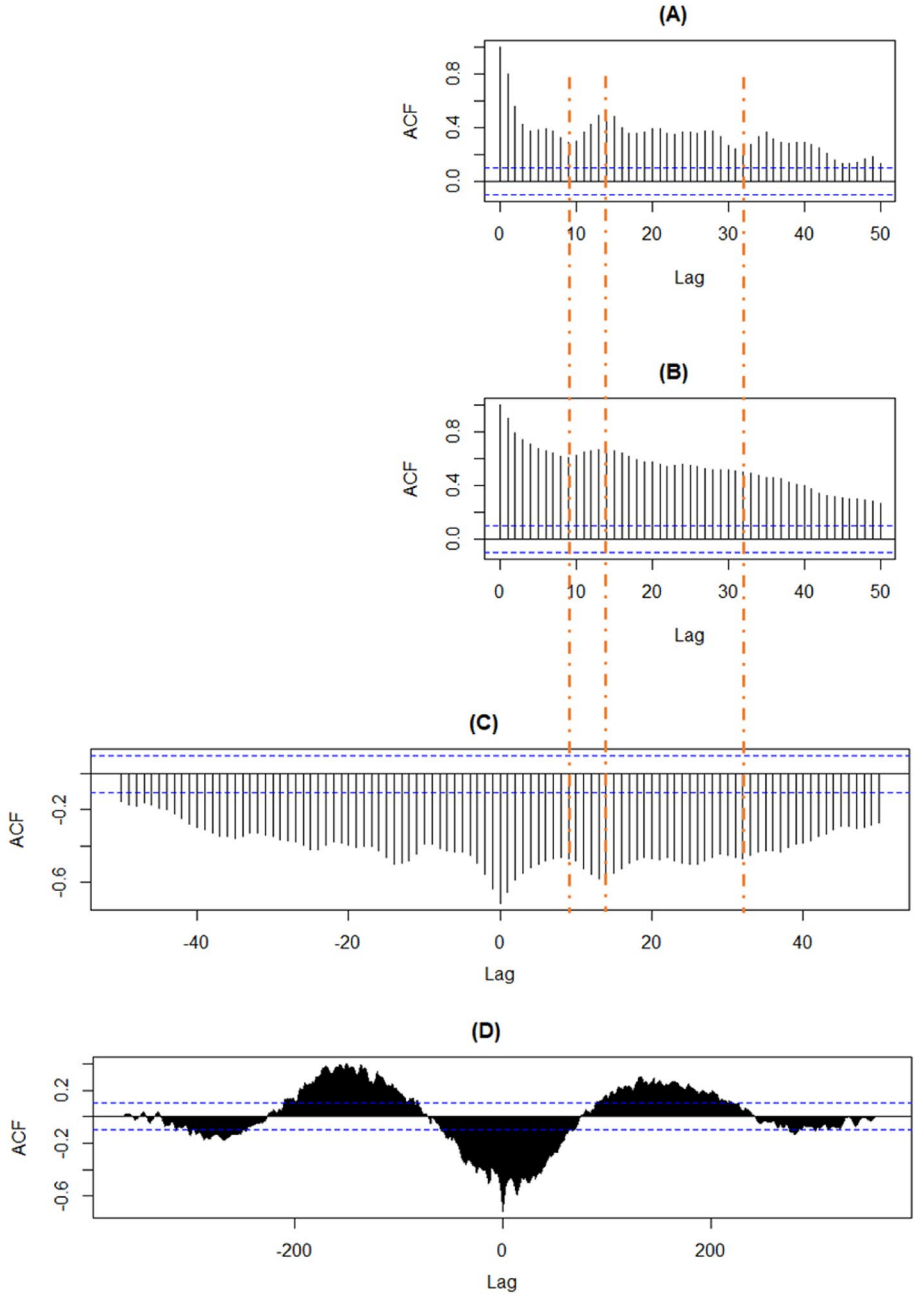

Figure 4. Autocorrelation function for NO_2 (**A**) and O_3 (**B**) of the year 2018. Cross-correlation function (ccf) for NO_2 and O_3 of the year 2018 with 50 days lag (**C**) and 400 days lag (**D**). The time series O_3 was fixed, and the time series NO_2 shifted by lags for the ccf function. Plots (**C**) and (**D**) showed the correlation between NO_2 at time $t \pm lag$ and O_3 at time t.

if NO_2 lags behind $PM_{2.5}$, there was a similar structure to the cross-correlation of NO_2 and *Prec* at low level. The ccf plot for PM_{10} and $PM_{2.5}$ gave a strong correlation at lag 0, followed by no clear structure at the monthly shift.

Spatial analysis

The aim of the the LISA analysis[46,47] was to identify the locations of clusters of LISA hot and cold spots of the environmental stressors throughout BW.

According to Fig. 6, regions with very high or low population density are most affected by air pollution variables such as PM_{10}, $PM_{2.5}$, NO_2 and O_3. The meteorological variables had various patterns. A more detailed representation of the spatial associations was given in Fig. 7 with a significance level set to 0.001. Using local cluster maps, the spatial associations between postal code areas were summarized into LISA hot and cold spots.

As already showed in the local significance map, most air pollution variables showed a similar structure. The following applies to NO_2, PM_{10}, and $PM_{2.5}$: In urban areas (Stuttgart + Mannheim region), many postal code

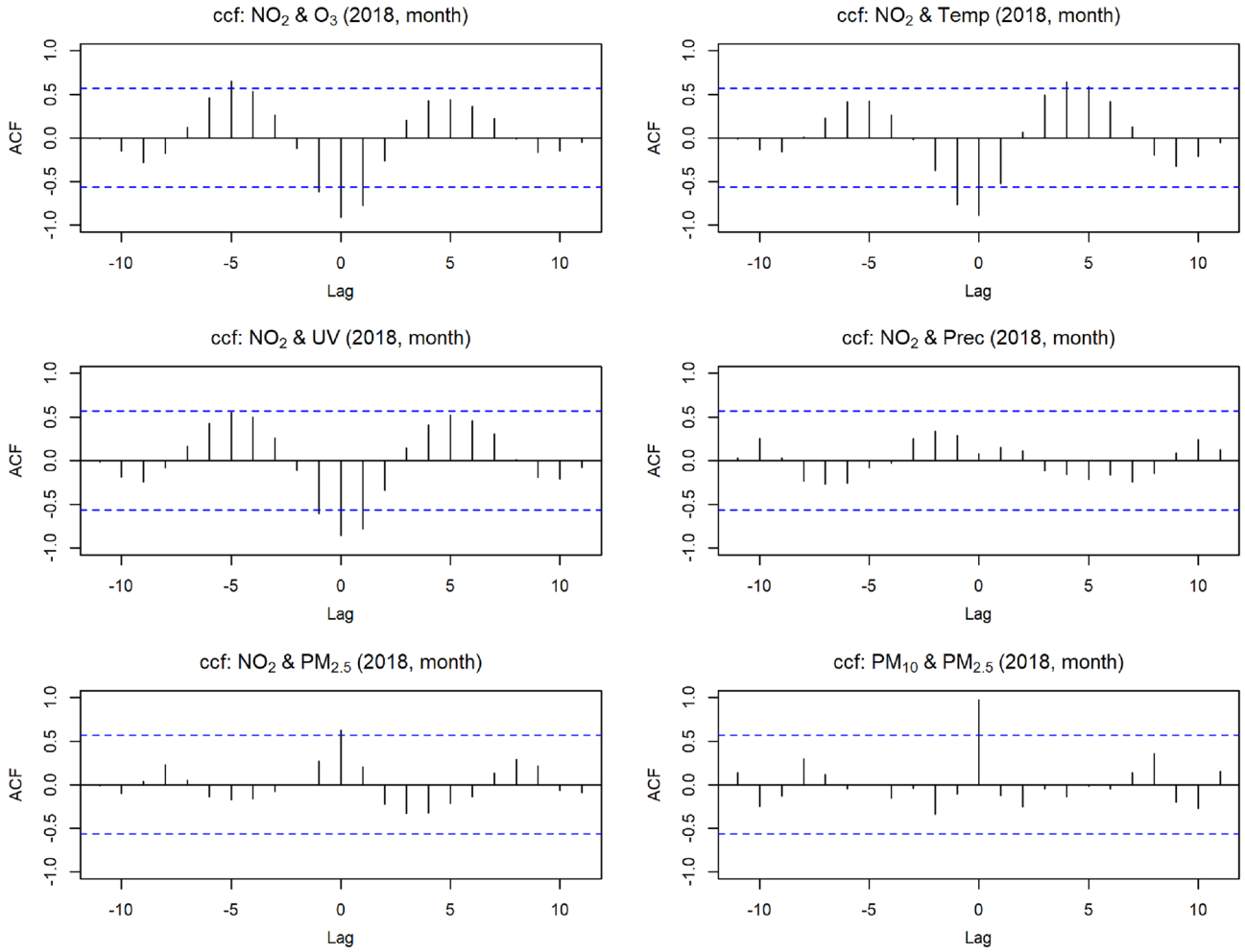

Figure 5. Different cross-correlation functions for the year 2018 (lag = month).

areas had neighbors with significantly similar high values. The cold spots geographically contain the Black Forest and parts of the Swabian Alb. For O_3, the distribution of LISA hot and cold spots was similar but reversed. Based on the meteorological variables, different postal code regions stand out, with *VP* and *Temp* showing similar trends. Accordingly, LISA hot spots were located at the western edge of BW. LISA cold spots were found along the Schwäbische Alb. A possible geographical connection could be the differences in altitude of the individual areas. For *UV*, LISA cold spots were located exclusively in the north, and LISA hot spots were exclusively in the south of the state. The Local Moran Map for *Prec* looked similar, i.e., LISA cold spots in the north and LISA hot spots in the south. However, no such clear and extreme assignment as for *UV* was recognizable. This could possibly be explained by the altitude, latitude and local air circulations. Furthermore, areas of the Black Forest showed LISA hot spot postal code areas. One could assume a connection between *Temp* and *UV*. However, this map showed no spatial relationship between LISA hot and cold spots for *UV* and *Temp*.

Pairwise correlations between stressors and cross-correlation plots were examined by the population density. Table 2 showed a summary of environmental stressors split by population density categories.

NO_2 showed the most remarkable change in concentration across the population density compared to the other environmental stressors with a mean value ranging from 9.6 $\mu g/m^3$ in rural to 16.5 $\mu g/m^3$ in urban areas. The more urban the area was, the higher the NO_2 concentration. For $PM_{2.5}$ and PM_{10}, a similar but less intense increase was observed. O_3 had a reverse effect ranging from 54.4 $\mu g/m^3$ to 47.2 $\mu g/m^3$. The environmental stressors *UV*, *VP*, *Temp*, and *Prec* were mainly constant over space.

Figure 8 showed the Pearson correlation matrix for all data separated by population density category. If we specifically compare the most rural and urban areas (categories 1 and 4), we noticed that NO_2 had different correlation values depending on the category. NO_2 correlated more negatively with *Temp* ($r = -0.67$), *VP* ($r = -0.6$), and *UV* ($r = -0.5$) in category 1 of population density than in category 4 with correlation coefficients of −0.55 (*Temp*), −0.48 (*VP*) and −0.45 (*UV*). The correlation between NO_2 and O_3, as well as *Prec*, remained stable mainly across the shifting categories of population density. Between NO_2 and $PM_{2.5}$ as well as PM_{10}, the positive correlation increased with rising population density from 0.51 and 0.53 (category 1) to 0.61 and 0.63 (category 4). All other correlations showed only minor variations across population density categories.

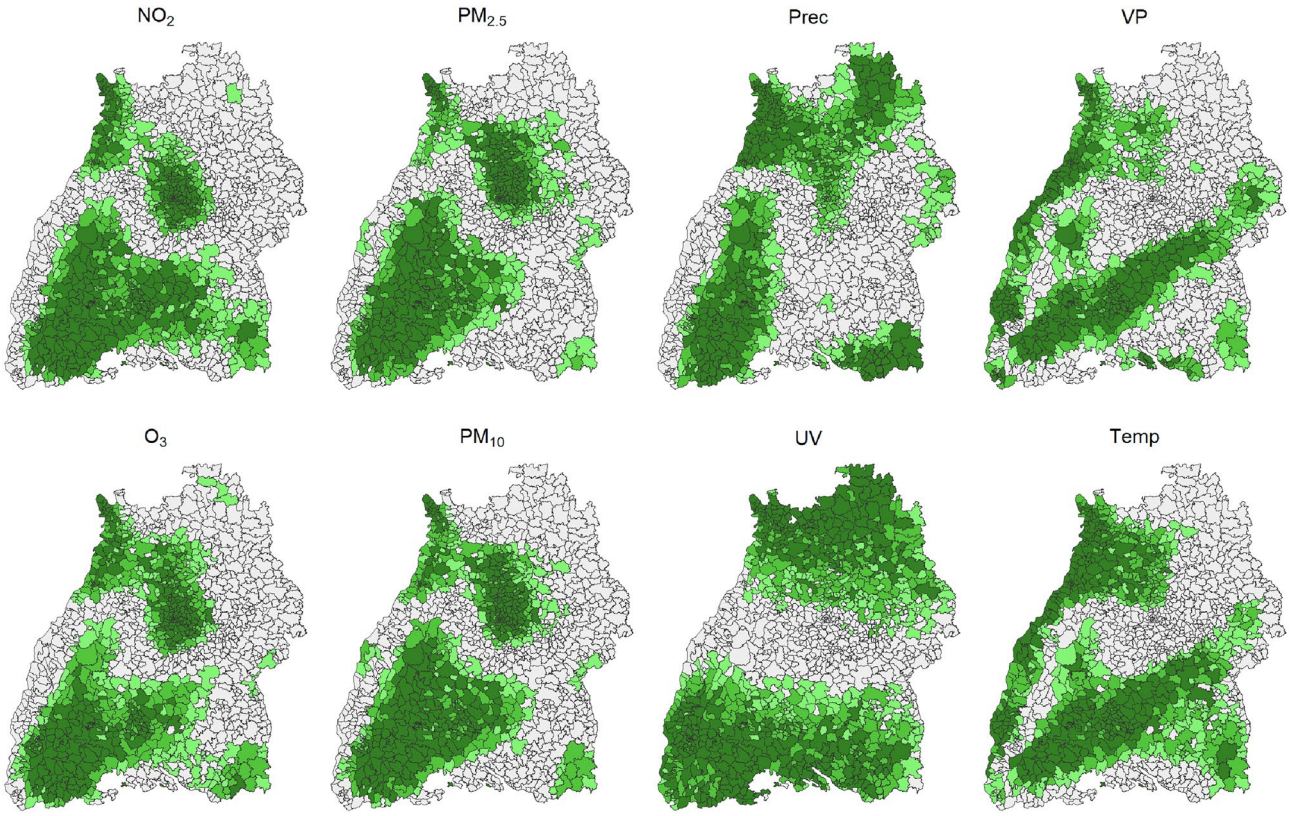

Figure 6. Local significance map of the individual environmental stressors in BW. The figure represented the results of the significance test of LISA analysis using Moran's I. Different shades of green indicated different thresholds: light green $p < 0.05$; medium green $p < 0.01$ and dark green $p < 0.001$.

The hot and cold spot resulted from the LISA analysis can be used to present a new level of spatial units in addition to the predefined categorical spatial units for population density in Table 2. When comparing the LISA hot and cold spots, it could be seen that they do not always corresponded to categories 1 through 4 of population densities. The parameters *Prec*, *UV*, *VP*, and *Temp*, in particular, differed from the distribution of population density in BW. Additionally, the LISA hot spots do not entirely matched the high population density in Table 2 for the NO_2, $PM_{2.5}$, PM_{10}, and O_3 parameters. Table 3 provided an overview of descriptive statistics for the new spatial units as obtained from the LISA analysis.

We introduced a new level of insight that was defined by the spatial variability of the stressors rather than relying on a predefined quantity such as population density. The strong discrepancy between the mean values was remarkable when looking at the hot and cold spots. These varied from 18.5 to 7.6 for NO_2, 59.7 to 45.1 for O_3, and 12.0 to 9.5 for $PM_{2.5}$.

Figure 9 displayed the Pearson correlation matrices for the hot and cold spots of the LISA spatial units, similar to the population density categories.

The correlation directions in both matrices were identical. However, the correlation coefficients for NO_2 with *Temp*, *VP*, and *UV* were higher for the cold spots. In addition, the $PM_{2.5}$ and PM_{10} correlation values with all other stressors were overall slightly higher in the hot spots than in the cold spots.

We also generated cross-correlation plots for NO_2 and O_3 in 2018, categorized by both population density and hot and cold spots. However, the cross-correlations do not showed strong changes in the associations between environmental stressors.

Discussion

Meteorological and air pollution variables were strongly correlated between and among themselves, with specific seasonal and spatial features. For example, NO_2 and O_3 were strongly interdependent, and the Pearson correlation varied with time. In January, there was a negative correlation of −0.84, whereas in July, the correlation coefficient was 0.45. Figure 10 illustrated that NO_2 and O_3 correlated not only with each other but also with other environmental stressors. It is particularly intriguing to note the contrasting values of the two months, as their correlation directions often differed.

The cross-correlation showed a noticeable change in the correlation direction for O_3 and NO_2. Spatially, NO_2, $PM_{2.5}$, and PM_{10} concentrations were significantly higher in urban than in rural regions. For O_3, the effect is reversed. This result is also confirmed by LISA analysis, where distinct hot and cold spots of the different environmental stressors could be identified. In addition, the Pearson correlation coefficients suggest that PM_{10} variation was almost entirely explained by $PM_{2.5}$ and vapor pressure by temperature.

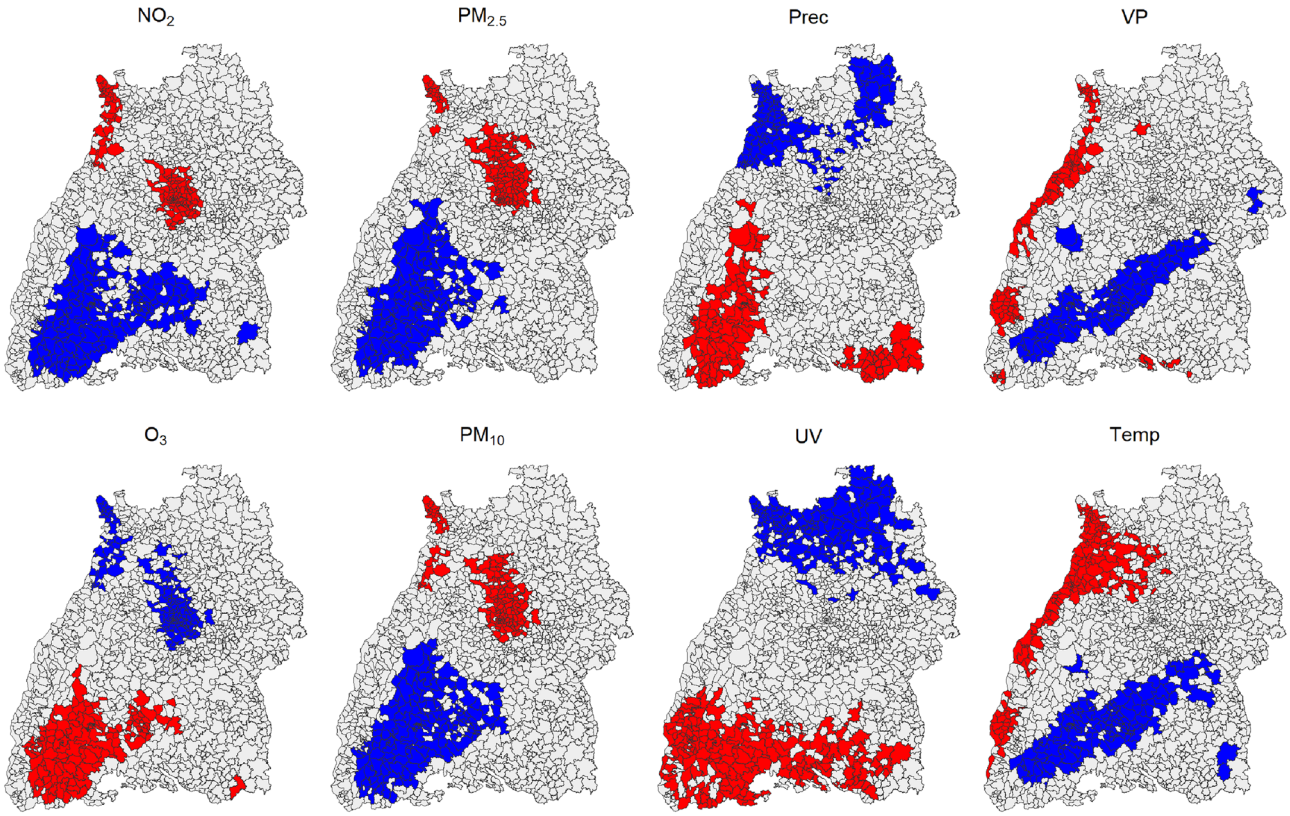

Figure 7. Local Cluster Map of the individual environmental stressors in BW. The significance level was set to $p \leq 0.001$. LISA hot spots were red, representing positive spatial autocorrelation with high values. LISA cold spots were blue, representing low values. BW had three postal code areas (78,266, 78,465, 78,479) that were isolated without neighboring postal code areas. We omitted these areas from any further consideration. The estimation was always relative to the mean, as constructed in formula (4). The combination of the red and blue areas results in the dark green areas in Fig. 6.

It is essential to emphasize that correlation does not necessarily imply causality. Correlation means that a change in one variable is related to another variable. However, correlation does not mean that one variable directly influences the other. Thus, there is not necessarily a cause-and-effect relationship. A third variable could affect the two variables that are supposed to be causally related.

Some atmospheric processes can explain the spatio-temporal variability of the correlation coefficients. From an atmospheric chemistry point of view NO_2 and O_3 need to be considered together since they are a function of each other which explains the opposite relationship[51]:

$$O_3 + NO \rightarrow O_2 + NO_2. \tag{6}$$

Their spatio-temporal variability is governed by superimposed emission-based and photochemistry-based regimes[52]. The predominantly anthropogenic nitrogen oxide (NO_x) emissions and the resulting NO_2 concentrations have a pronounced seasonal cycle, with higher values in winter than in summer[53]. This is due to the fact that in addition to the higher emissions also the lifetime of NO_2 is longer in winter (about 21 h) than in summer (about 6 h)[54,55]. Peak NO_2 concentrations in winter are resulting from superimposed atmospheric inversion conditions. The photochemically produced tropospheric O_3 exhibits higher levels in the summer months when there is more solar radiation. Nevertheless, the production depends strongly and non-linearly on precursors like NO_x and volatile organic compound (VOC) concentrations as well as meteorological conditions. However, photolysis of NO_2 is the primary chemical source of tropospheric ozone[51,56].

Consequently, the relationship between $O3$ and NO_2 is complex, influenced by a variety of factors and thus needs to be distinguished between rural and urban areas. Since NO_x are emitted from traffic, industrial processes, and other human activities the resulting NO_2 concentration are higher in urban areas and industrial agglomerations as can be inferred from Fig. 7. This is in line with the findings of[23] where the urban pollution island of the Stuttgart city region could be delineated from satellite.

Ozone in urban areas is primarily formed as a secondary pollutant through chemical reactions involving precursor pollutants, especially during sunny, warm weather conditions and can be transported to rural areas, affecting rural air quality. Paradoxically, locally high emissions of NO_x, such as from traffic, tend to favor ozone destruction in urban areas, resulting in NO_2 formation. As can be seen in Fig. 7 this results in ozone cold spots in urban areas. In rural areas, natural sources like vegetation (emitting biogenic volatile organic compounds) and soil contribute to ozone formation. O_3 levels tend to be higher in rural areas where there are fewer local emissions of NOx to destroy any O_3 that was photochemically produced. As can be inferred from Fig. 7 the Black Forest

Variable	Parameter	1	2	3	4
Temp (°C)	Mean (sd)	9.0 (7.7)	9.7 (7.7)	10.0 (7.7)	10.3 (7.6)
	Median [min, max]	9.2 [−18.5, 29.9]	9.9 [−18.9, 30.8]	10.2 [−18.6, 30.9]	10.4 [−18.2, 30.9]
Prec (mm/day)	Mean (sd)	3.6 (6.3)	3.4 (6.0)	3.3 (5.9)	3.1 (5.6)
	Median [min, max]	1.0 [0, 99.3]	0.9 [0, 98.1]	0,9 [0, 98.0]	0.8 [0, 85.4]
VP (hPa)	Mean (sd)	9.7 (4.2)	10.0 (4.2)	10.1 (4.3)	10.2 (4.3)
	median [min, max]	9.1 [1.0, 24.0]	9.4 [1.1, 24.6]	9.5 [1.1, 24.6]	9.6 [1.1, 24.3]
UV (W)	Mean (sd)	15.5 (9.6)	15.4 (9.7)	15.3 (9.7)	15.2 (9.7)
	Median [min, max]	14.1 [0.4, 37.0]	14.1 [0.4, 37.0]	14.0 [0.4, 37.0]	13.9 [0.4, 37.0]
O_3 ($\mu g/m^3$)	Mean (sd)	54.4 (22.3)	52.0 (22.8)	50.0 (23.2)	47.2 (23.5)
	Median [min, max]	55.1 [0.4, 149]	52.9 [0.6, 148]	51.0 [0.3, 149]	48.3 [0.5, 149]
NO_2 ($\mu g/m^3$)	Mean (sd)	9.6 (5.8)	11.4 (6.6)	13.5 (7.4)	16.5 (7.5)
	Median [min, max]	8.1 [1.7, 31.3]	9.7 [2.0, 38.3]	11.9 [2.5, 45.5]	14.8 [3.3, 53.0]
$PM_{2.5}$ ($\mu g/m^3$)	Mean (sd)	10.5 (6.1)	10.9 (6.3)	11.2 (6.5)	11.6 (8.7)
	Median [min, max]	8.0 [1.1, 66.4]	9.7 [1.1, 60.0]	11.7 [1.1, 66.0]	14.7 [1.4, 66.4]
PM_{10} ($\mu g/m^3$)	Mean (sd)	14.1 (7.9)	14.7 (8.3)	15.3 (8.7)	16.0 (9.1)
	Median [min, max]	12.6 [0.9, 79.6]	13.1 [1.0, 80.0]	13.6 [1.1, 83.5]	14.2 [1.3, 82.5]

Table 2. Overview of environmental stressors split by population density categories. Overview of environmental stressors split by population density category 1: 0–150 inhabitants/km^2, category 2: 151–300 inhabitants/km^2, category 3: 301–1000 inhabitants/km^2, category 4: > 1000 inhabitants/km^2 including information on standard deviation (sd), median, minimum (min) and maximum (max) value. The variables covered all postal code areas in BW and were based on daily data from 2010 to 2018.

mountain range depicts an O_3 hot spot in BW due to the high solar irradiation and the abundance of biogenic volatile organic substances *BVOC* ozone precursors. This is also substantiated in Table 2: the more inhabitants there are, the less O_3 and the more NO_2 occur.

The anticorrelation of O_3 and NO_2, for the reasons outlined above, can be confirmed for all temporal and spatial aggregation levels in the study: Fig. 2 (entire BW, entire period), Fig. 3 (entire BW; Jan, Apr, Oct), Fig. 8 (population density classes, entire period) and Fig. 9 (hot and cold spots; entire period) except for Fig. 3 (entire BW, Jul) where a positive correlation is reported. The negative correlations agree with in-situ measurements of O_3 and NO_2 for Munich, a similar city region to Stuttgart at almost the same latitude, with −0.58 and −0.64 for the period January to July 2019 and 2020, respectively[57]. The same authors find a positive correlation of O_3 and *Temp* of 0.67 and 0.49 for the time periods given and addresses also the interannual variability. This agrees with r = 0.62 in our study for the period 2010 to 2018. Further process-oriented regimes can be identified in the results of the study. Temperature-dominated effects can best be seen in the Pearson correlation matrix for July (Fig. 3C). In the first column the variables *VP*, *UV*, O_3, NO_2, $PM_{2.5}$ and PM_{10} are all positively correlated with *Temp*, partly because they are produced by photochemical processes and linked high solar irradiance (O_3, NO_2, *UV*) or increased by dry weather conditions ($PM_{2.5}$ and PM_{10})[1]. In the second column the effect of wet deposition and cleaning effect is evident by the negative correlation of all air pollutants with precipitation. *UV* shows the strongest negative correlation, due to the presence of clouds. Emission-dominated variables and effects can best be seen in winter (Fig. 3A). It can be inferred that low temperatures, low precipitation and low water pressure favor high concentrations of NO_2, $PM_{2.5}$ and PM_{10}, partly due to increased heating, longer photochemical lifetimes and accumulation under inversion conditions or low windspeeds[1,55]. More previous studies have discussed the effect of meteorological conditions on the concentration of atmospheric pollutants and meteorological variables such as wind direction, wind speed and precipitation that have a constraining effect on atmospheric pollutant concentrations, but not a simple linear relationship[20,21,58]. This study is limited to daily data. For further studies, hourly observations could be considered as applied in[10]. The different stressors usually interact differently during the day and at night since e.g. anthropogenic emissions exhibit a pronounced daily cycle and photochemical reactions are confined to sunlit conditions[26]. Another potential extension of the current analysis is to expand the study area to encompass all of Europe. The study uses air pollution and meteorological data that represents background conditions and mesoscale variability. As such air pollution from point sources or along roads cannot be resolved. However, such data is not yet available to our knowledge for the entire BW and the time period under investigation. A possible addition to the variables considered could be wind speed, wind direction and boundary layer height since these parameters have shown a large impact on the variability of particulate matter[59,60]. Furthermore, boundary layer height and O_3 showed the strongest positive correlation among all the analyzed variables in[57]. The Pearson correlation assumes a linear relationship between two continuous variables. Linearity was deemed sufficient for our initial analysis of the internal dependencies of environmental stressors, although there are other correlation coefficients like Spearman correlation that deal with nonlinear associations. We recommend using nonlinear statistical methods such as generalized additive models with splines for advanced studies of air pollution and health factors.

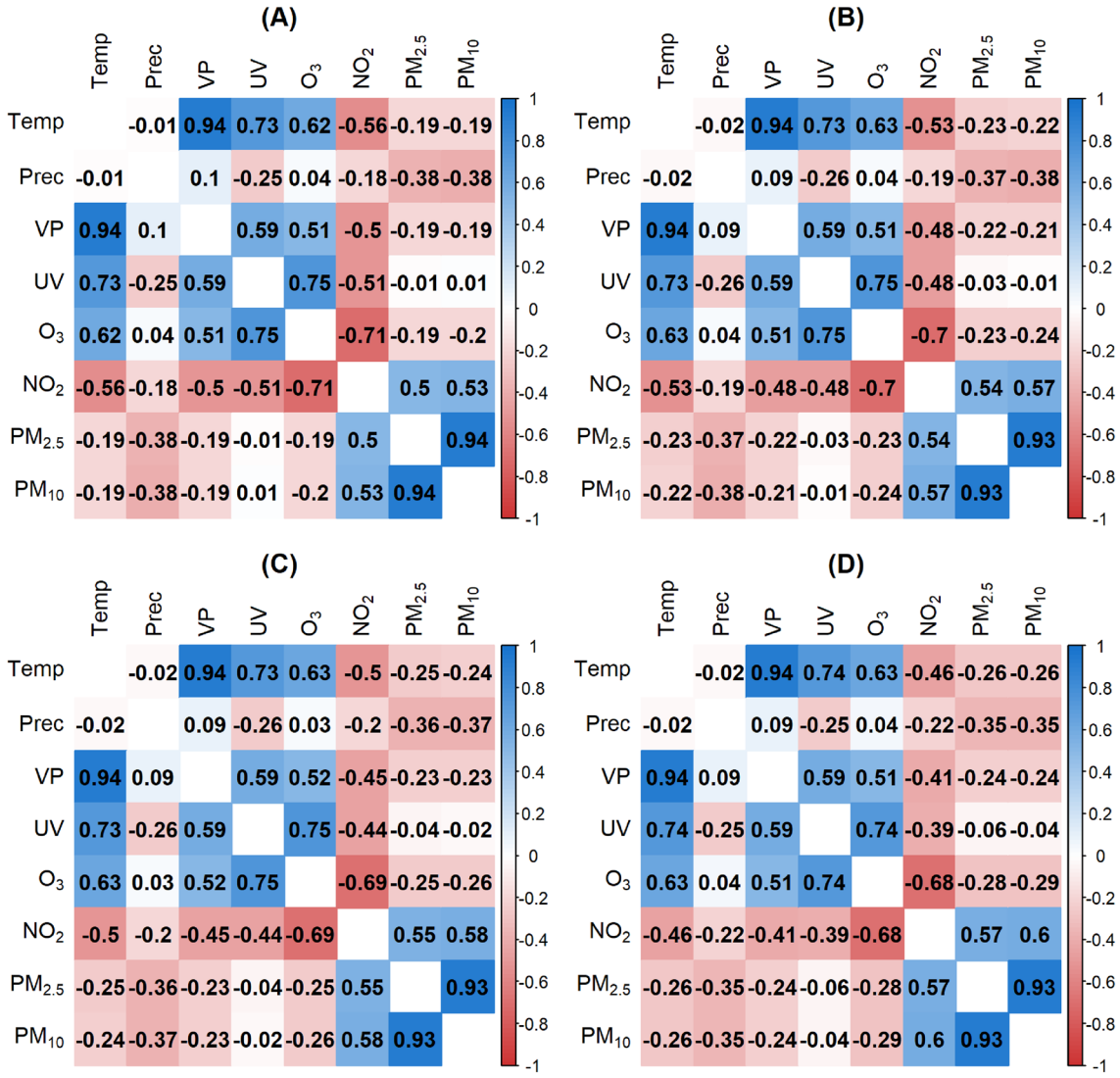

Figure 8. Pearson correlation matrices based on daily data aggregated by population density category 1 (**A**): 0–150 inhabitants/km^2, category 2 (**B**): 151–300 inhabitants/km^2, category 3 (**C**): 301–1000 inhabitants/km^2, category 4 (**D**): > 1000 inhabitants/km^2.

Variable	Parameter	LISA hot spots	LISA cold spots	Isolated	Non significant
NO_2 (µg/m^3)	Mean (sd)	18.5 (8.7)	7.6 (4.6)	11.2 (6.5)	12.0 (6.9)
	Median [min, max]	16.7 [2.5, 66.4]	6.24 [1.1, 42.9]	9.2 [1.7, 40.3]	10.3 [1.1, 66.4]
O_3 (µg/m^3)	Mean (sd)	59.7 (21.8)	45.1 (23.2)	53.7 (23.5))	51.5 (22.8)
	Median [min, max]	59.5 [1.1, 149]	46.4 [0.5, 136]	55.2 [3.3, 125]	52.4 [0.3, 149]
$PM_{2.5}$ (µg/m^3)	Mean (sd)	12.0 (7.1)	9.5 (5.9)	11.3 (6.9)	11.0 (6.7)
	Median [min, max]	10.5 [1.2, 63.1]	8.3 [0.8,50.4]	9.8 [1.1, 57.3]	9.6 [0.7, 72.3]

Table 3. Overview of environmental stressors split by LISA spatial units. Overview of NO_2, O_3 and $PM_{2.5}$ split by LISA spatial units hot spot, cold spot, isolated and other non significant postal code areas including information on standard deviation (sd), median, minimum (min) and maximum (max) value. The variables covered all postal code areas in BW and included all days from 2010 to 2018. Note that the LISA spatial units differed based on the environmental stressor, as illustrated in Fig. 7. As a result, the quantity of values in each category varied.

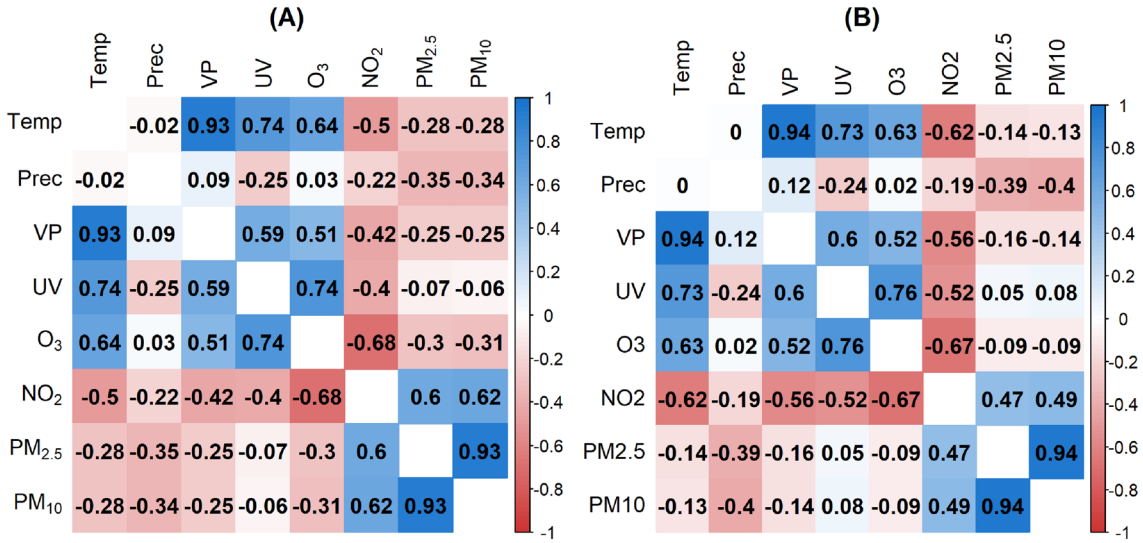

Figure 9. Pearson correlation matrices based on daily measurements aggregated by postal code areas in hot (**A**) and cold (**B**) spots as obtained from the LISA analysis.

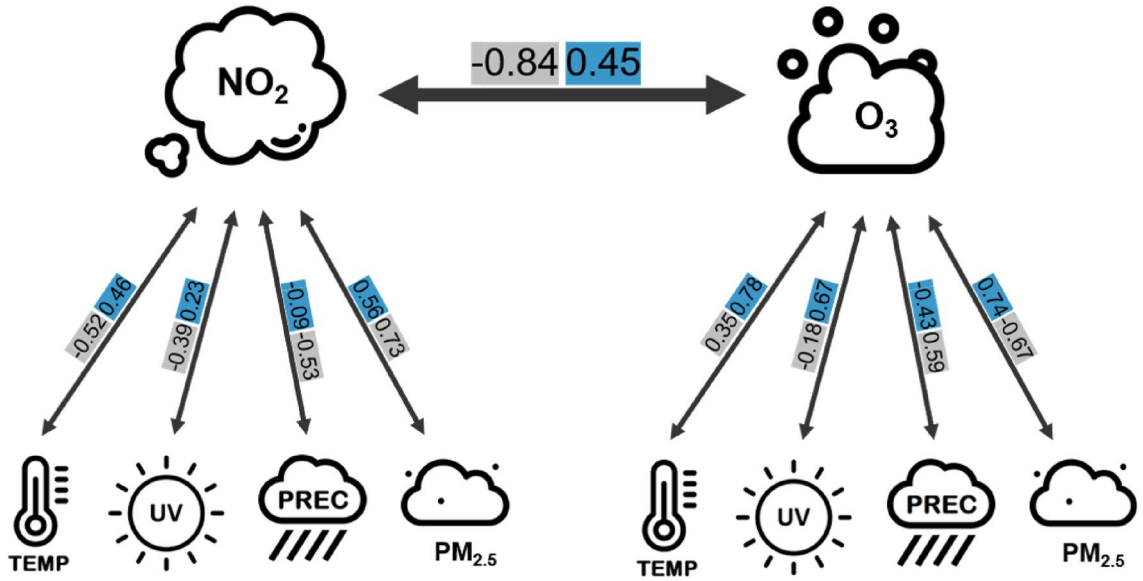

Figure 10. Illustration of the correlations between NO_2 and O_3 in January (gray) and July (blue). The correlation values originated from Fig. 3. This diagram had been designed using images from flaticon.com.

Conclusion

Selecting the appropriate variables for a statistical model can be a challenging task. This paper offers decision-making assistance for upcoming analyses describing the health effects of environmental stressors. Including a single environmental variable in the model may result in information loss, while including too many variables may lead to correlations and biases. Finding the right balance is important. The optimal choice of variables relies on the specific research question and the given data. However, this paper provides recommendations regarding the variable selection that can be considered. In this work, it turned out to be sufficient to consider $PM_{2.5}$. PM_{10} has larger particles but almost identical temporal and spatial characteristics. The only possible deviation would be for Saharan dust[61,62]. The variables *VP* and *Temp* show strong similarities by design so that future investigations can be limited to the temperature.

The opposite relationship between NO_2 and O_3 was confirmed both temporally and spatially. NO_2 is more often observed in metropolitan areas and O_3 in rural areas. How can this knowledge be addressed in a future model describing the health effects of environmental stressors? For future analyses, we propose incorporating interaction terms to effectively illustrate the relationships between the two variables, NO_2 and O_3, and their impact on the dependent variable. Considering only one environmental stressor in a future model may lead to loss of information and confusion in interpreting. Moreover, including both variables in the model without

an interaction term is not advisable, as high Pearson correlation coefficients may cause bias. Based on this, we recommend using an interaction term between NO_2 and O_3.

It is important to note that different stressors have different health effects. Although the previous analysis suggests that NO_2 and O_3 are opposite, they have different impacts on human health[40]. This fact makes the choice of model and the interpretation of the relationships more complex and must be considered in future analyses.

When applying the LISA model, we found substantial spatial differences in some variables (e.g. $PM_{2.5}$), but not in others (*UV*) between urban and rural areas. This indicates that this spatial variation can be statistically exploited for epidemiological studies. Notably, a large fraction of postal code regions show lack of coherence with their neighbors, as can seen from the high proportion of uncolored areas in Fig. 7. In addition, we also identified clear patterns of LISA hot and cold spots, particularly in urban areas, mountainous regions Schwarzwald, and Schwäbische Alb. All of these identified patterns show positive autocorrelations, and no negative autocorrelation was observed.

We spatially categorized the state of Baden-Württemberg in two ways: first, by population density (Fig. 2 and table 2) and second, by LISA hot and cold spots (Fig. 7 and table 3). The two categorizations matched spatially well for some air pollution variables (e.g. $PM_{2.5}$ and NO_2) and less well for some meteorological variables (e.g. *UV*).

To conclude, it will be straightforward to implement the principle findings of our study, namely (a) the temporal coherence of stressor patterns, and (b) the spatial clustering into statistical models for the epidemiological study of stressor effects on human health, e.g. by affording the necessary spatial categorical variables and the opportune interaction terms into the statistical models.

Data availability

The data that support the findings of this study are available on request from the corresponding author LH.

Received: 2 November 2023; Accepted: 7 March 2024
Published online: 12 March 2024

References

1. Jacob, D. J. & Winner, D. A. Effect of climate change on air quality. *Atmos. Environ.*https://doi.org/10.1016/j.atmosenv.2008.09.051 *(2009).*
2. Doherty, R. M., Heal, M. R. & O'Connor, F. M. Climate change impacts on human health over Europe through its effect on air quality. *Environ. Health.* **16**(1), 33–44. https://doi.org/10.1186/s12940-017-0325-2 (2017).
3. World Health Organization (WHO). *WHO Global Air Quality Guidelines: Particulate Matter (PM2.5 and PM10), Ozone, Nitrogen Dioxide, Sulfur Dioxide and Carbon Monoxide*. (World Health Organization, 2021). https://www.who.int/publications/i/item/9789240034228.
4. World Health Organization (WHO). *Air Quality Guidelines: Global Update 2005: Particulate Matter, Ozone, Nitrogen Dioxide, and Sulfur Dioxide*. (World Health Organization, 2006). WHO/SDE/PHE/OEH/06.02. https://www.who.int/publications/i/item/WHO-SDE-PHE-OEH-06.02.
5. The European Parliament and the Council of the European Union. 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. *J. Eur. Union.* (2008).
6. World Health Organization (WHO). *Review of Evidence on Health Aspects of Air Pollution: REVIHAAP Project: Technical Report*. (World Health Organization, Regional Office for Europe, 2021). WHO/EURO: 2013-4101-43860-61757. https://apps.who.int/iris/handle/10665/341712.
7. Manisalidis, I., Stavropoulou, E., Stavropoulos, A. & Bezirtzoglou, E. Environmental and health impacts of air pollution: A review. *Front. Public Health*https://doi.org/10.3389/fpubh.2020.00014 *(2020).*
8. Kampa, M. & Castanas, E. Human health effects of air pollution. *Environ. Pollut.* **151**(2), 362–367. https://doi.org/10.1016/j.envpol.2007.06.012 (2008).
9. Schwartz, J. Air-pollution and daily mortality: A review and meta analysis. *Environ. Res.* **64**(1), 36–52. https://doi.org/10.1006/enrs.1994.1005 (1994).
10. Gilardi, L., Marconcini, M., Metz-Marconcini, A., Esch, T. & Erbertseder, T. Long-term exposure and health risk assessment from air pollution: Impact of regional scale mobility. *Int. J. Health Geogr.* **22**(1), 11. https://doi.org/10.1186/s12942-023-00333-8 (2023).
11. Chen, J. & Hoek, G. Long-term exposure to PM and all-cause and cause-specific mortality: A systematic review and meta-analysis. *Environ. Int.*https://doi.org/10.1016/j.envint.2020.105974 *(2020).*
12. Burnett, R. *et al.* Global estimates of mortality associated with long-term exposure to outdoor fine particulate matter. *Proc. Natl. Acad. Sci.* **115**(38), 9592–9597. https://doi.org/10.1073/pnas.1803222115 (2018).
13. European Environment Agency (EEA). *Health Impacts of Air Pollution in Europe, 2022 [Web Page]*. European Environment Agency (EEA). https://www.eea.europa.eu/publications/air-quality-in-europe-2022/health-impacts-of-air-pollution.
14. Liu, X. *et al.* Assessment of German population exposure levels to PM10 based on multiple spatial-temporal data. *Environ. Sci. Pollut. Res.* **27**, 6637–6648. https://doi.org/10.1007/s11356-019-07071-0 (2020).
15. Rittweger, J. *et al.* Temperature and particulate matter as environmental factors associated with seasonality of influenza incidence: An approach using Earth observation-based modeling in a health insurance cohort study from Baden-Wurttemberg (Germany). *Environ. Health.*https://doi.org/10.1186/s12940-022-00927-y *(2022).*
16. Zhang, Y., Wang, S. J., Feng, Z. X. & Song, Y. Influenza incidence and air pollution: Findings from a four-year surveillance study of prefecture-level cities in China. *Front. Public Health.*https://doi.org/10.3389/fpubh.2022.1071229 *(2022).*
17. Villeneuve, P. J. & Goldberg, M. S. Methodological considerations for epidemiological studies of air pollution and the SARS and COVID-19 coronavirus outbreaks. *Environ. Health Perspect.*https://doi.org/10.1289/Ehp7411 *(2020).*
18. Katoto, P. D. M. C. *et al.* Acute and chronic exposure to air pollution in relation with incidence, prevalence, severity and mortality of COVID-19: A rapid systematic review. *Environ. Health.* https://doi.org/10.1186/s12940-021-00714-1 (2021).
19. Marques, M. & Domingo, J. L. Positive association between outdoor air pollution and the incidence and severity of COVID-19: A review of the recent scientific evidences. *Environ. Res.*https://doi.org/10.1016/j.envres.2021.111930 *(2022).*
20. Elminir, H. K. Dependence of urban air pollutants on meteorology. *Sci. Total Environ.* **350**(1–3), 225–237. https://doi.org/10.1016/j.scitotenv.2005.01.043 (2005).
21. Liu, Y. S., Zhou, Y. & Lu, J. X. Exploring the relationship between air pollution and meteorological conditions in China under environmental governance. *Sci. Rep.* https://doi.org/10.1038/s41598-020-71338-7 (2020).
22. Battista, G. & Vollaro, R. D. Correlation between air pollution and weather data in urban areas: Assessment of the city of Rome (Italy) as spatially and temporally independent regarding pollutants. *Atmos. Environ.* **165**, 240–247. https://doi.org/10.1016/j.atmosenv.2017.06.050 (2017).

23. Samad, A. *et al.* Meteorological and air quality measurements in a city region with complex terrain: Influence of meteorological phenomena on urban climate. *Meteorol. Z.*https://doi.org/10.1127/metz/2023/1124 *(2023).*
24. Varotsos, C., Christodoulakis, J., Tzanis, C. & Cracknell, A. P. Signature of tropospheric ozone and nitrogen dioxide from space: A case study for Athens, Greece. *Atmos. Environ.* **89**, 721–730. https://doi.org/10.1016/j.atmosenv.2014.02.059 (2014).
25. Venkitaswamy, S. & Bhaskar, V. Relationship between ozone with nitrogen dioxide and climatic impacts over major cities in India. *Sustain. Environ. Res.* **25**(6), 295–304 (2015).
26. Nguyen, D. H. *et al.* Tropospheric ozone and NOx: A review of worldwide variation and meteorological influences. *Environ. Technol. Innov.* https://doi.org/10.1016/j.eti.2022.102809 (2022).
27. Souza, J. B. *et al.* Generalized additive models with principal component analysis: An application to time series of respiratory disease and air pollution data. *J. R. Stat. Soc. C* **67**(2), 453–480. https://doi.org/10.1111/rssc.12239 (2018).
28. Statheropoulos, M., Vassiliadis, N. & Pappa, A. Principal component and canonical correlation analysis for examining air pollution and meteorological. *Atmos. Environ.* **32**(6), 1087–1095. https://doi.org/10.1016/S1352-2310(97)00377-4 (1998).
29. Stafoggia, M., Breitner, S., Hampel, R. & Basagana, X. Statistical approaches to address multi-pollutant mixtures and multiple exposures: The state of the science. *Curr. Environ. Health Rep.* **4**, 481–490. https://doi.org/10.1007/s40572-017-0162-z (2017).
30. Sun, Z. C. *et al.* Statistical strategies for constructing health risk models with multiple pollutants and their interactions: Possible choices and comparisons. *Environ. Health.* **12**(1), 1–19. https://doi.org/10.1186/1476-069X-12-85 (2013).
31. Höffner, J. Geografie. Staatsministerium Baden-Württemberg. https://www.baden-wuerttemberg.de/de/unser-land/land-und-leute/geografie.
32. Esri, D. Postleitzahlgebiete in Deutschland. Esri Deutschland Content. https://opendata-esri-de.opendata.arcgis.com/datasets/5b203df4357844c8a6715d7d411a8341_0.
33. Klüsener, S. Bevölkerungsdichte in Deutschland (Kreisebene, 2020). Bundesinstitut für Bevölkerungsdichte. https://www.bib.bund.de/DE/Fakten/Fakt/B77-Bevoelkerungsdichte-Kreise.html.
34. European Centre for Medium-Range Weather Forecasts (ECMWF). CAMS European air quality reanalyses. European Centre for Medium-Range Weather Forecasts (ECMWF). https://ads.atmosphere.copernicus.eu/cdsapp#!/dataset/cams-europe-air-quality-reanalyses?tab=overview.
35. Muñoz Sabater, J. ERA5-Land hourly data from 1950 to present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS). https://doi.org/10.24381/cds.e2161bac.
36. Gilardi, L. NO2, O3, PM10 and PM2.5 concentrations - Daily geographical aggregates at ZIP-code level from CAMS European Air Quality Re-analyses. Zenodo. https://doi.org/10.5281/zenodo.8325533 (2023).
37. Lavers, D. A., Simmons, A., Vamborg, F. & Rodwell, M. J. An evaluation of ERA5 precipitation for climate monitoring. *Q. J. R. Meteorol. Soc.* **148**(748), 3152–3165. https://doi.org/10.1002/qj.4351 (2022).
38. Marécal, V. *et al.* A regional air quality forecasting system over Europe: The MACC-II daily ensemble production. *Geosci. Model Dev.* **8**(9), 2777–2813. https://doi.org/10.5194/gmd-8-2777-2015 (2015).
39. Akritidis, D. *et al.* A complex aerosol transport event over Europe during the 2017 Storm Ophelia in CAMS forecast systems: Analysis and evaluation. *Atmos. Chem. Phys.* **20**(21), 13557–13578. https://doi.org/10.5194/acp-20-13557-2020 (2020).
40. World Health Organization (WHO). Ambient (outdoor) air pollution. World Health Organization (WHO). https://www.who.int/news-room/fact-sheets/detail/ambient-(outdoor)-air-quality-and-health.
41. Minkos, A. *et al*. Luftqualität 2015: Vorläufige Auswertung. Umweltbundesamt; 2016. https://www.umweltbundesamt.de/publikationen/luftqualitaet-2015.
42. Bolton, D. The computation of equivalent potential temperature. *Mon. Weather Rev.* **108**(7), 1046–1053. https://doi.org/10.1175/1520-0493(1980)108<1046:TCOEPT>2.0.CO;2 (1980).
43. Hinkle, D. E., Wiersma, W. & Jurs, S. G. *Applied Statistics for the Behavioral Sciences* (Houghton Mifflin, 2003).
44. Bourke, P. *Cross Correlation, Autocorrelation, 2d Pattern Identification*, vol. 2019 (1996). http://paulbourke.net/miscellaneous/correlate/.
45. Brockwell, P. J. & Davis, R. A. *Introduction to Time Series and Forecasting* (Springer, 2016).
46. Anselin, L. Local indicators of spatial association: Lisa. *Geogr. Anal.* **27**(2), 93–115. https://doi.org/10.1111/j.1538-4632.1995.tb00338.x (1995).
47. Müller, I., Erbertseder, T. & Taubenbock, H. Tropospheric NO_2: Explorative analyses of spatial variability and impact factors. *Remote Sens. Environ.*https://doi.org/10.1016/j.rse.2021.112839 *(2022).*
48. Getis, A. & Ord, J. K. The analysis of spatial association by use of distance statistics. *Geogr. Anal.* **24**(3), 189–206. https://doi.org/10.1111/j.1538-4632.1992.tb00261.x (1992).
49. Anselin, L. A local indicator of multivariate spatial association: Extending Geary's c. *Geogr. Anal.* **51**(2), 133–150. https://doi.org/10.1111/gean.12164 (2019).
50. R Core Team. *R: A Language and Environment for Statistical Computing*. https://www.R-project.org/.
51. Seinfeld, J. H. & Pandis, S. N. *Atmospheric Chemistry and Physics: From Air Pollution to Climate Change* (Wiley, 2016).
52. Han, S. *et al.* Analysis of the relationship between O_3, NO and NO_2 in Tianjin, China. *Aerosol Air Qual. Res.* **11**(2), 128–139. https://doi.org/10.4209/aaqr.2010.07.0055 (2011).
53. Beirle, S., Boersma, K. F., Platt, U., Lawrence, M. G. & Wagner, T. Megacity emissions and lifetimes of nitrogen oxides probed from space. *Science.* **333**(6050), 1737–1739. https://doi.org/10.1126/science.1207824 (2011).
54. Pommier, M. Estimations of NOx emissions, NO_2 lifetime and their temporal variation over three British urbanised regions in 2019 using TROPOMI NO_2 observations. *Environ. Sci. Atmos.* **3**(2), 408–421. https://doi.org/10.1039/D2EA00086E (2023).
55. Shah, V. *et al.* Effect of changing NOX lifetime on the seasonality and long-term trends of satellite-observed tropospheric NO_2 columns over China. *Atmos. Chem. Phys.* **20**(3), 1483–1495. https://doi.org/10.5194/acp-20-1483-2020 (2020).
56. Kleinman, L. I. *et al.* Dependence of ozone production on NO and hydrocarbons in the troposphere. *Geophys. Res. Lett.* **24**(18), 2299–2302. https://doi.org/10.1029/97GL02279 (1997).
57. Balamurugan, V. *et al.* Tropospheric NO_2 and O_3 response to COVID-19 lockdown restrictions at the national and urban scales in Germany. *J. Geophys. Res. Atmos.*https://doi.org/10.1029/2021JD035440 *(2021).*
58. Gong, W., Reich, B. J. & Chang, H. H. Multivariate spatial prediction of air pollutant concentrations with INLA. *Environ. Res. Commun.* **3**(10), 101002. https://doi.org/10.1088/2515-7620/ac2f92 (2021).
59. Handschuh, J., Erbertseder, T., Schaap, M. & Baier, F. Estimating PM2.5 surface concentrations from AOD: A combination of SLSTR and MODIS. *Remote Sens. Appl. Soc. Environ.* **26**, 100716. https://doi.org/10.1016/j.rsase.2022.100716 (2022).
60. Handschuh, J., Erbertseder, T. & Baier, F. Systematic evaluation of four satellite AOD datasets for estimating PM2.5 using a random forest approach. *Remote Sens.* **15**(8), 2064. https://doi.org/10.3390/rs15082064 (2023).
61. Klüser, L., Kleiber, P., Holzer-Popp, T. & Grassian, V. H. Desert dust observation from space: Application of measured mineral component infrared extinction spectra. *Atmos. Environ.* **54**, 419–427. https://doi.org/10.1016/j.atmosenv.2012.02.0 (2012).
62. Capraz, O. & Deniz, A. Particulate matter (PM10 and PM2.5) concentrations during a Saharan dust episode in Istanbul. *Air Qual. Atmos. Health.* **14**, 109–116. https://doi.org/10.1007/s11869-020-00917-4 (2021).

Acknowledgements
We kindly acknowledge the Copernicus Climate Change Service (C3S) and the Copernicus Atmosphere Monitoring Service (CAMS) for providing the environmental datasets.

Author contributions
LG provided the re-elaborated and pre-processed environmental data. JR, LH and MTS conceived and designed the analysis. LH performed the analysis and wrote the paper. TE wrote parts of the discussion. MTS and MS verified the analytical methods. TE, LG, SW and MB helped interpret the analysis results within physico-chemical atmospheric contexts. All authors reviewed and approved the manuscript.

Funding
Open Access funding enabled and organized by Projekt DEAL. This research has been supported by the Deutsche Forschungsgesellschaft (DFG) funded project "Influence of air quality on the expected burden on the health care system in the event of pandemics" (Project Number 458531714) and the German Aerospace Center (DLR) funded project "Environmental stressors and health costs".

Competing interests
The authors declare no competing interests.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-024-56513-4.

Correspondence and requests for materials should be addressed to L.H.

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

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