

LETTER • OPEN ACCESS

## The environmental neighborhoods of cities and their spatial extent

To cite this article: M Llaguno-Munitxa and E Bou-Zeid 2020 *Environ. Res. Lett.* **15** 074034

View the [article online](#) for updates and enhancements.

### Recent citations

- [Humans in the city: Representing outdoor thermal comfort in urban canopy models](#)  
I. Pigliautile *et al*

# Environmental Research Letters



## LETTER

# The environmental neighborhoods of cities and their spatial extent

### OPEN ACCESS

RECEIVED  
23 November 2019

REVISED  
9 April 2020

ACCEPTED FOR PUBLICATION  
27 April 2020

PUBLISHED  
2 July 2020

Original content from this work may be used under the terms of the [Creative Commons Attribution 4.0 licence](https://creativecommons.org/licenses/by/4.0/).

Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.



M Llaguno-Munitxa  and E Bou-Zeid 

Department of Civil and Environmental Engineering, Princeton University, Princeton, NJ, United States of America

E-mail: [mllaguno@princeton.edu](mailto:mllaguno@princeton.edu)

**Keywords:** urban environment, air quality, urban planning, sensor networks, spatial heterogeneity

Supplementary material for this article is available [online](#)

## Abstract

We define the new concept of an environmental neighborhood as the surrounding area influencing the environmental quality at a given point in a city, and propose a novel methodology to measure its spatial extent. We compute the spatial correlation of air quality and urban parameters from high spatial resolution datasets for New York City, where the urban characteristics are averaged over variable urban footprint sizes, ranging from 25 m × 5 m to 5000 m × 5000 m. The scale at which these correlations peak indicates the extent of the neighboring area that influences pollutant concentrations deviations from the city-wide average. The results indicate that the scale of these environmental neighborhoods ranges from ~1000 m (for attributes such as road area or building footprint) down to ~200 m (for building use or green area). Selecting this optimal neighborhood scale is thus critical for identifying the urban fabric and activity attributes that have the largest influence on air quality; smaller footprints do not contain all the pertinent urban surface information while larger footprints contain irrelevant, potentially misleading information. The quantification of this scale of influence therefore enables more effective and localized policies and interventions to improve urban environmental quality and reduce urban health disparities. More broadly, the findings indicate that, in a wide range of environmental and ecological applications where surface heterogeneity is a primary driver, the scale of analysis is not an external parameter to be chosen, but rather an internal parameter dictated by the problem physics.

## 1. Introduction

Urban environmental quality has in the last decade become a global health concern. It has recently been reported at the World Health Organization (WHO) Global Conference on Air Pollution and Health that 'Air pollution—both ambient and household—is estimated to cause 7 million deaths per year' (WHO 2018a). Poor air quality impacts are exacerbated by the urban heat island effect, which itself is exacerbated by climate-change related heat waves that are increasing in frequency (Li and Bou-Zeid 2013, WHO 2018b; Li *et al* 2019). As reported by the Environmental Protection Agency (EPA), the increased daytime temperatures and reduced night-time cooling, coupled to the higher air pollution levels correlated with urban heat islands and anthropogenic emissions, affect human health by producing general discomfort, heat cramps and exhaustion, respiratory difficulties, heat strokes, and heat-related mortality (Dye 2008, EPA 2019, Goldman and Dominici 2019, Editorial

2019, Caplin *et al* 2019). Exposure to poor air quality is also linked to increased death risk from pathogenic respiratory diseases such as SARS (Cui *et al* 2003), and this has prompted recent media reports projecting these findings to the novel SARS-CoV-2 virus (The Washington Post 2020). Interventions to curb this multi-hazard risk have thus never been more crucial, but the science of such interventions remains more qualitative than quantitative.

The introduction of trees and vegetation, green roofs, cool roofs and pavements, as well as urban strategies to enhance urban mechanical ventilation, have proven effective in reducing the intensity of the urban heat island and air contaminant concentration levels (Grimmond and Oke 1999, Yin 2011, Santamouris 2013, Akbari 2016, Kroeger *et al* 2018, Llaguno-Munitxa and Bou-Zeid 2018). These measures are being adopted in various cities all around the world. Urban planning strategies to alleviate air pollution in cities through traffic restriction policies or Floor Area Ratio (FAR) regulations are also starting to

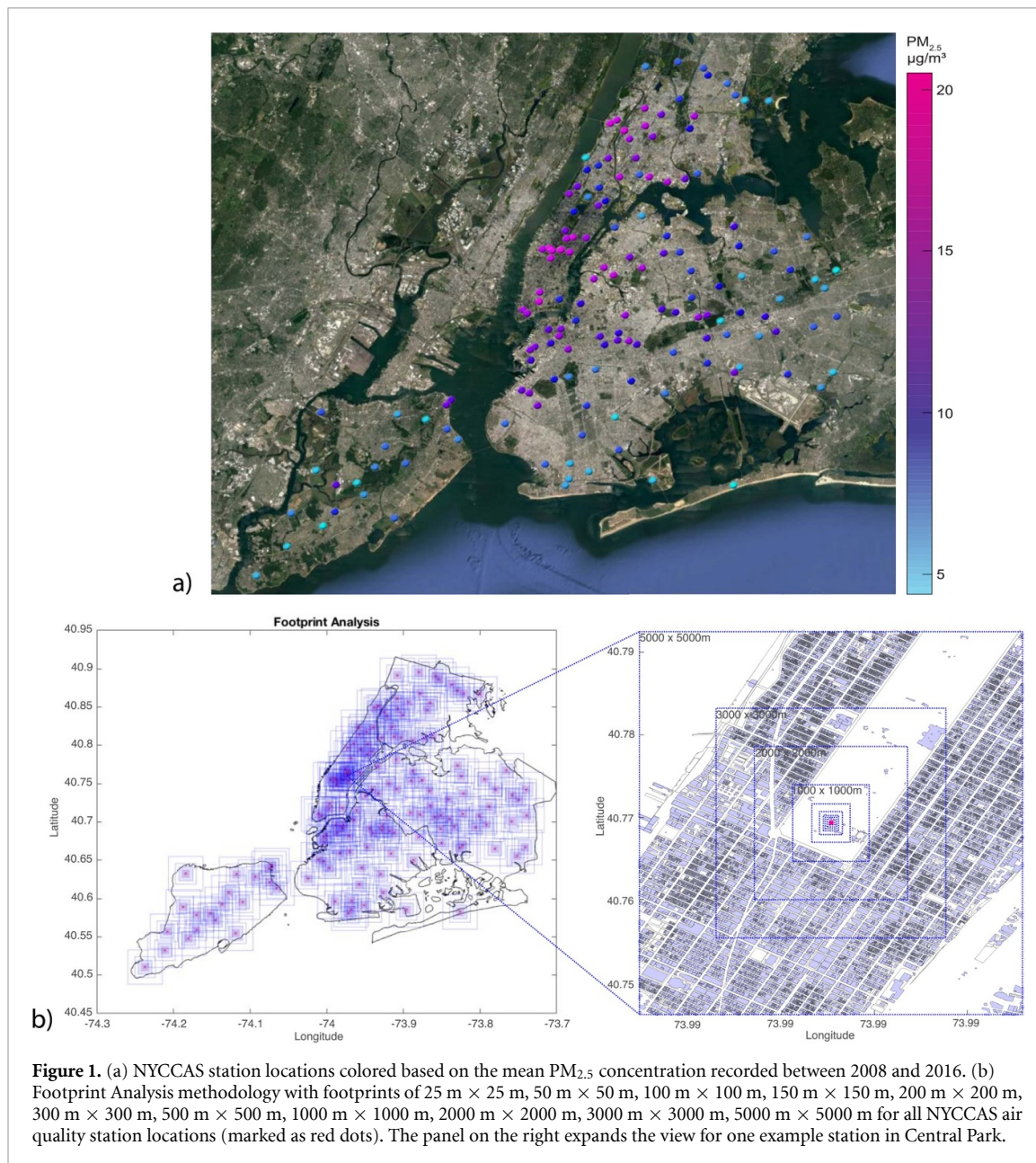
be considered (Editorial 2018, Chandra 2018). Concurrently, many cities are starting to invest in extensive weather and air quality sensing networks to understand their urban environmental quality gradients. The C40 Cities air quality network led by the cities of London in the UK and Bengaluru in India, for example, claims to be implementing the ‘most sophisticated air quality air monitoring system’ (C40 Cities 2018). The Array of Things (AoT) project in Chicago aims to deploy 500 stationary environmental sensing nodes in the city center to monitor urban floods, heat island and air quality, with the goal of informing strategic investment in urban policy and municipal socioeconomic initiatives (Chen *et al* 2016). New York City on the other hand is following an ambitious goal set by PlaNYC to become the city with the cleanest air of any major US city by 2030. As part of the PlaNYC’s air strategy, the city’s Department of Health and Mental Hygiene’s New York City Community Air Survey (NYCCAS) was launched to collect high spatial resolution air pollution data. 150 street level air pollution stations were installed from 2008 to 2010, reduced to 100 from 2010 to 2013 (NYCCAS 2018; Matte *et al* 2013).

While mobile microclimate and air quality sensing technologies have been in use since the first half of the 20th century (Schmidt 1927, Peppler 1929, Budel and Wolf 1933), recent advancements in IoT technologies are making them increasingly popular (Kuttler and Strassburger 1999, Van Poppel *et al* 2013, Levy *et al* 2014). They can very effectively complement the fixed networks surveyed above (Yang and Bou-Zeid 2019a). The use of mobile air quality monitoring networks is enabling the acquisition of high spatial resolution air quality data affordably in various cities such as New York, Zurich, Oklahoma, California and London (Maciejczyk *et al* 2004, Mueller *et al* 2016, Apte *et al* 2017, BBC News 2019). Thus, for the first time in history, through the data collected by dense fixed networks and mobile sensing technologies, we are starting to capture and understand the strong spatiotemporal urban air pollution and temperature gradients present in our cities. What is needed now is a framework for applying this knowledge to quantitatively guide the interventions and policies aimed at mitigating the multi-hazard environmental risk in cities.

The effect of local urban intervention strategies on air quality, such as the planting of trees or modifications of urban roughness and city block densities, as well as the effect of urban characteristics such as population or building geometries, have been widely researched in prior literature (Grimmond and Oke 1999, Beckett *et al* 2000, Hang *et al* 2012, Vos *et al* 2013, Gromke and Blocken 2015, Llaguno-Munitxa and Bou-Zeid 2018). Correlations between air quality datasets and urban surface characteristics and anthropogenic activities have also been studied by various researchers (Martilli 2014, Rodriguez *et al* 2016, Yang *et al* 2016). However, such studies have

mainly utilized idealized models or limited experimental datasets; their uncertainties hence remain high. This may preclude accurate quantification of the benefits and the design of finely-targeted interventions. A particularly important question concerns the required spatial extent of these interventions to appreciably improve environmental quality in neighborhoods with severe problems (Yang and Bou-Zeid 2019b). The expanding availability of extensive air quality datasets collected at high spatiotemporal resolution, or mobile air quality sensing technologies, now enables analyses that can answer such questions concerning complex urban air quality spatial distributions (Chen *et al* 2016; NYCCAS 2018, Matte *et al* 2013, Mueller *et al* 2016, Apte *et al* 2017, BBC News 2019; Muller *et al* 2013, Ching *et al* 2019, Hasenfrazz *et al* 2015, Munir *et al* 2019). It has been recently reported, for example, that concentrations of pollutants can vary by over five times within a single city block, and by over eight times at the city scale due to variability in local emissions and traffic (Apte *et al* 2017). If we combine detailed urban environmental and attribute datasets, can we distill the linkages of distinct urban characteristics and activity to local air quality? At what scale do these linkages manifest? How can this information guide the design of future cities to improve environmental quality?

This paper presents a novel approach to tackle these questions utilizing high spatiotemporal resolution air quality experimental data collected by NYC-CAS from 2008 to 2016, one of the most complete air quality datasets available today that includes data for Nitrogen Oxide (NO), Nitrogen Dioxide (NO<sub>2</sub>), Particulate Matter <2.5 μm (PM<sub>2.5</sub>), Ozone (O<sub>3</sub>), Elemental Carbon (EC), and Sulfur Dioxide (SO<sub>2</sub>). New York City also provides an extensive open data portal that enables us to relate the air pollutant concentration levels to urban characteristics. Stewart and Oke (2012) developed the Local Climate Zone (LCZ) concept to represent urban land patches with relatively uniform characteristics at the neighborhood scale that create spatially coherent thermal climates. Studies that followed which have made use of high spatial resolution environmental datasets to infer the influence of land use parameters have primarily focused on the study of the influence of urban attributes on local temperature, concluding that air temperature variability (e.g. Ziter *et al* 2019) and surface temperature variability (e.g. Buyantuyev and Wu 2009, Jenerette *et al* 2016) are linked to urban parameters at a very small, local scale. However, comparable analyses for urban air quality (Edussuriya *et al* 2011, Miskell *et al* 2015, Zhou *et al* 2018, Jung *et al* 2019) have struggled to converge on general conclusions on the links between urban attributes and air pollution, and have adopted arbitrary spatial scales in their analysis (ranging from a 200 m scale to the city scale). Therefore, the spatial extent of the influence of urban parameters and the links between urban



attributes and urban environmental conditions, particularly air quality, remain open research challenges.

## 2. Methods

In this paper, we propose a new approach to overcome these challenges, defining the novel concept of an environmental neighborhood as the area that exerts the most influence on environmental quality at a given point. We then demonstrate how the spatial extent of such a neighborhood can be measured based on a variable-footprint approach. We illustrate these ideas for the unique environmental contaminants and urban parameters available for New York City, but that are becoming commonly available in many other metropolises. For each NYCCAS air quality station location, urban parameters' metrics have been computed for distinct urban footprint sizes, ranging from  $25\text{ m} \times 25\text{ m}$  to  $5000\text{ m} \times 5000\text{ m}$ , allowing us

to understand the correlation between contaminant concentrations and neighboring urban parameters at multiple scales. The scale maximizing this correlation can then be defined as the area of influence of a given urban parameter on local air quality.

As displayed in figure 1(b), the studied urban footprints cover ranges from the immediate neighboring area of a station, up to a big fraction of the city. Utilizing the  $3 \times 3\text{ ft}^2$  spatial resolution land use dataset available for New York City that provides a two dimensional characterization of the urban fabric, the building footprint area  $Building_a$ , road area  $Road_a$ , tree canopy plan area fraction  $Tree_{ca}$ , and the green grass/shrub area  $Green_{gsa}$  were computed for the different footprint sizes. Given that the traffic count was reported for specific locations within New York City, a traffic flow estimation model for the stations and footprints under study (details in the supplementary information available

**Table 1.** The linear Pearson correlations for the 1000 m × 1000 m urban footprint between all the studied urban parameters (note that the matrix is symmetric).

	$Tree_{ca}$	$Green_{gsa}$	$Building_{fa}$	$Road_a$	$Traffic_c$	$Building_{ta}$	$Height_{bm}$	$Non-Residential_{ta}$	$Residential_{ta}$
$Tree_{ca}$	—	0.29	-0.57	-0.59	-0.37	-0.34	-0.13	-0.32	-0.25
$Green_{gsa}$	0.29	—	-0.72	-0.54	-0.37	-0.45	-0.35	-0.35	-0.51
$Building_{fa}$	-0.57	-0.72	—	0.67	0.50	0.73	0.52	0.64	0.59
$Road_a$	-0.58	-0.54	0.67	—	0.44	0.44	0.26	0.37	0.40
$Traffic_c$	-0.37	-0.36	0.50	0.44	—	0.62	0.53	0.63	0.24
$Building_{ta}$	-0.34	-0.45	0.73	0.44	0.62	—	0.88	0.96	0.55
$Height_{bm}$	-0.13	-0.35	0.52	0.26	0.53	0.88	—	0.84	0.49
$Non-Residential_{ta}$	-0.32	-0.35	0.64	0.37	0.63	0.96	0.84	—	0.31
$Residential_{ta}$	-0.25	-0.51	0.59	0.40	0.24	0.55	0.49	0.31	—

online at [stacks.iop.org/ERL/15/074034/mmedia](https://stacks.iop.org/ERL/15/074034/mmedia)) was developed to produce the traffic count  $Traffic_c$  parameter we use in the analyses. From the building and tax lot information collected from the Primary Land Use Tax Lot Output (PLUTO) dataset, building use and area metrics were derived. Total building area (here total refers to the sum of the area for all floors)  $Building_{ta}$ , total residential building area  $Residential_{ta}$ , total non-residential area  $Non-Residential_{ta}$ , and mean building height metric  $Height_{bm}$  were computed for all NYCCAS air quality station locations (see figure 1(a) and figure S3), and for each of the studied footprints (see table S1 for further details).

### 3. Results and discussion

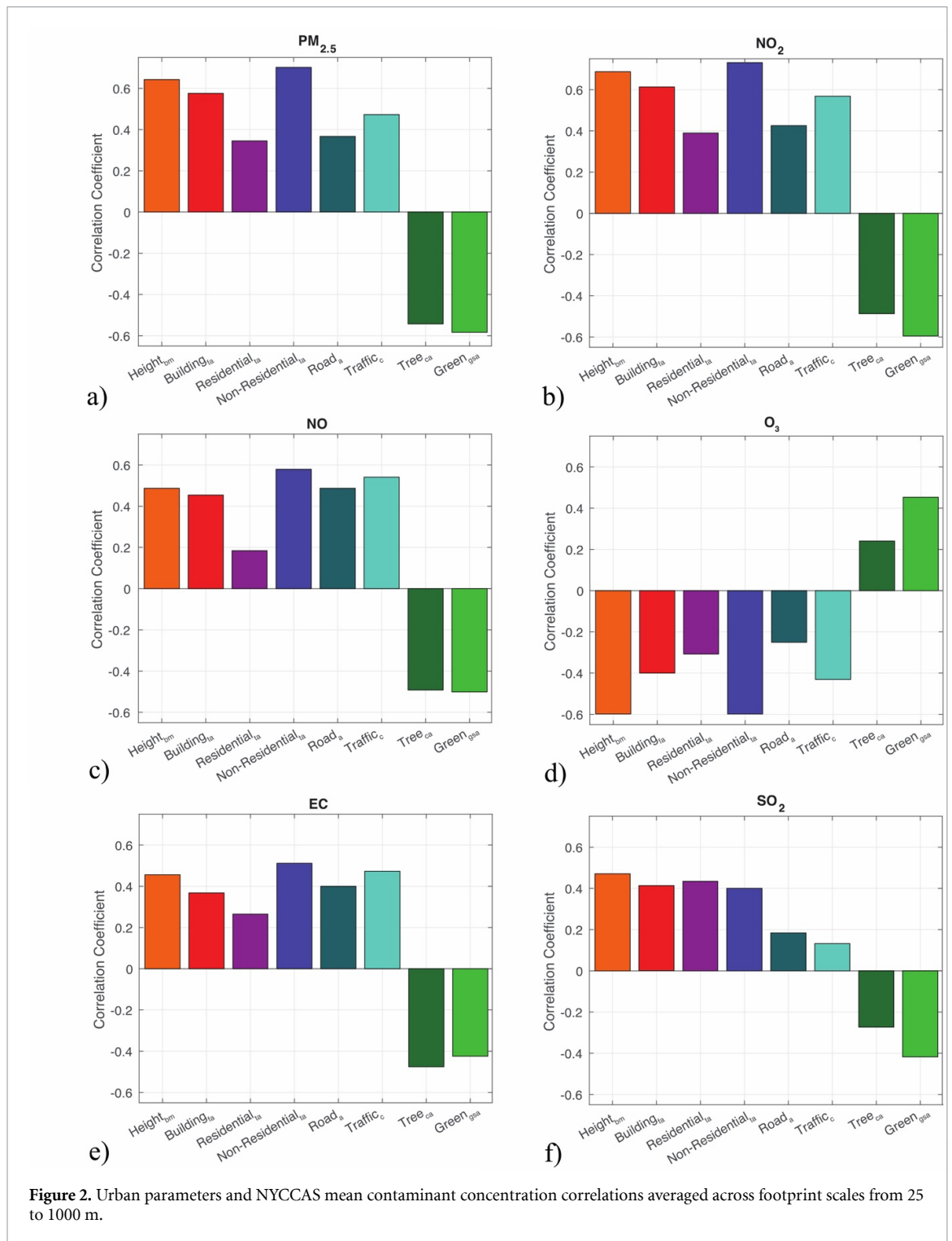
The linear Pearson correlation coefficients (Pearson 1895) of contaminant concentrations with each other, and of urban parameters with each other, were first computed to identify their linkages and inter-dependences. Amongst the six contaminants reported by NYCCAS, the highest correlations are observed between  $PM_{2.5}$  and  $NO_2$ , and between  $NO_2$  and  $NO$ , while  $O_3$  tends to be negatively correlated with the other pollutants (see table S2). Table 1 lists the Pearson correlation coefficients between the selected urban parameters for a 1000 m × 1000 m footprint as an illustration. The results of urban parameter correlations suggest that we can omit the  $Building_{ta}$  parameter as an independent input given the high correlation it has with the  $Height_{bm}$  parameter (see table 1). But all other parameters seem to be partially independent and need to be retained.

Once the independent urban parameters were selected, the retained ones were qualitatively coalesced into four categories to aid in the interpretation of the result. The first, *Pedestrian Level Emissions*, groups the urban parameters  $Road_a$  and the  $Traffic_c$  that inform us about the activity and emissions from near street level sources. The second category, *Green Infrastructure*, includes the urban parameters  $Tree_{ca}$  and  $Green_{gsa}$ , which can potentially act as air pollution sinks. The third category, *Building Density*, comprises the urban parameters  $Building_{ta}$ ,  $Building_{fa}$  and  $Height_{bm}$  that represent the ventilation

and sunlight ingress potential of urban street canyons. The final category, *Building Type*, includes the parameters  $Residential_{ta}$  and  $Non-Residential_{ta}$ , which reveal information related to emissions from household or tertiary use building facilities, both through pedestrian or rooftop level sources.

Now we can proceed to evaluate which urban parameters have the largest impact on air quality, and what is their spatial scale of influence. To that end, the urban parameters were spatially averaged (for urban footprints from 25 × 25 m to 1000 m × 1000 m); subsequently, three correlation coefficients (the linear Pearson correlation, the power law Pearson correlation and the non-parametric Spearman correlation) were computed for each footprint scale, and then averaged across these footprint scales (only for this analysis). The goal is to (i) confirm that the parametric and non-parametric correlation trends are consistent (which they were) and (ii) check whether the air quality and urban parameter relations are better described by a linear or a power law fit (further details included in the supplementary information and table S3). For most of the cases, a linear fit gave the highest correlation. However, some parameters such as the  $Tree_{ca}$  or  $Green_{gsa}$  show a higher correlation with the proposed Pearson ( $\ln(x)$ ,  $\ln(y)$ ) fit instead, where  $\ln$  is the natural logarithm of the variable. We henceforth use the correlation (linear or logarithmic) for each of the contaminant and urban parameter pairs that gave the best fit. These correlation coefficient results are depicted in the bar plots of figure 2.

For  $SO_2$ , the highest correlation is with  $Height_{bm}$  and  $Building_{ta}$ .  $SO_2$  behavior is distinct from the other pollutants since (i) it is mainly emitted from building heating boilers, which explains its strong dependence on *Building Density* parameters, (ii) it was thus only measured during the winter in NYCCAS, and (iii) a large drop of 68% in its concentrations was noted during the experimental period (NYCCAS 2018; Matte 2013) due to heating fuel regulations imposed by NYC, as well as the increasingly strict federal level regulations that have resulted in reductions in emissions of  $SO_2$ . For all other pollutants,  $Non-Residential_{ta}$ , which better combines the influence of traffic and urban density, was the best single predictor. Generally, the urban parameters associated



**Figure 2.** Urban parameters and NYCCAS mean contaminant concentration correlations averaged across footprint scales from 25 to 1000 m.

with building density and type have the highest correlations with concentrations, i.e. the most impact on air quality. Road and traffic urban parameters show similar trends, with strong correlations especially with NO and NO<sub>2</sub>. The parameters associated with buildings have the opposite effect to those associated with green cover, as expected. PM<sub>2.5</sub>, NO<sub>x</sub> and EC concentrations display very similar dependences on urban attributes. While these pollutants can also be transported from outside of the city, from sources such as vessels or power plants (NYCCAS 2018; Matte

2013) the main sources are local and mostly related to fuel combustion from vehicles and boilers.

O<sub>3</sub> concentrations, which were only measured during the summer months, on the other hand show unique trends compared to the remaining contaminants. An increase of O<sub>3</sub> is observed in areas with denser green canopies, suggesting that the formation of O<sub>3</sub> in the city may be partially related to local biogenic emissions, or transported from the urban boundary layer and troposphere as reported in prior literature (Kuttler and Strassburger 1999,

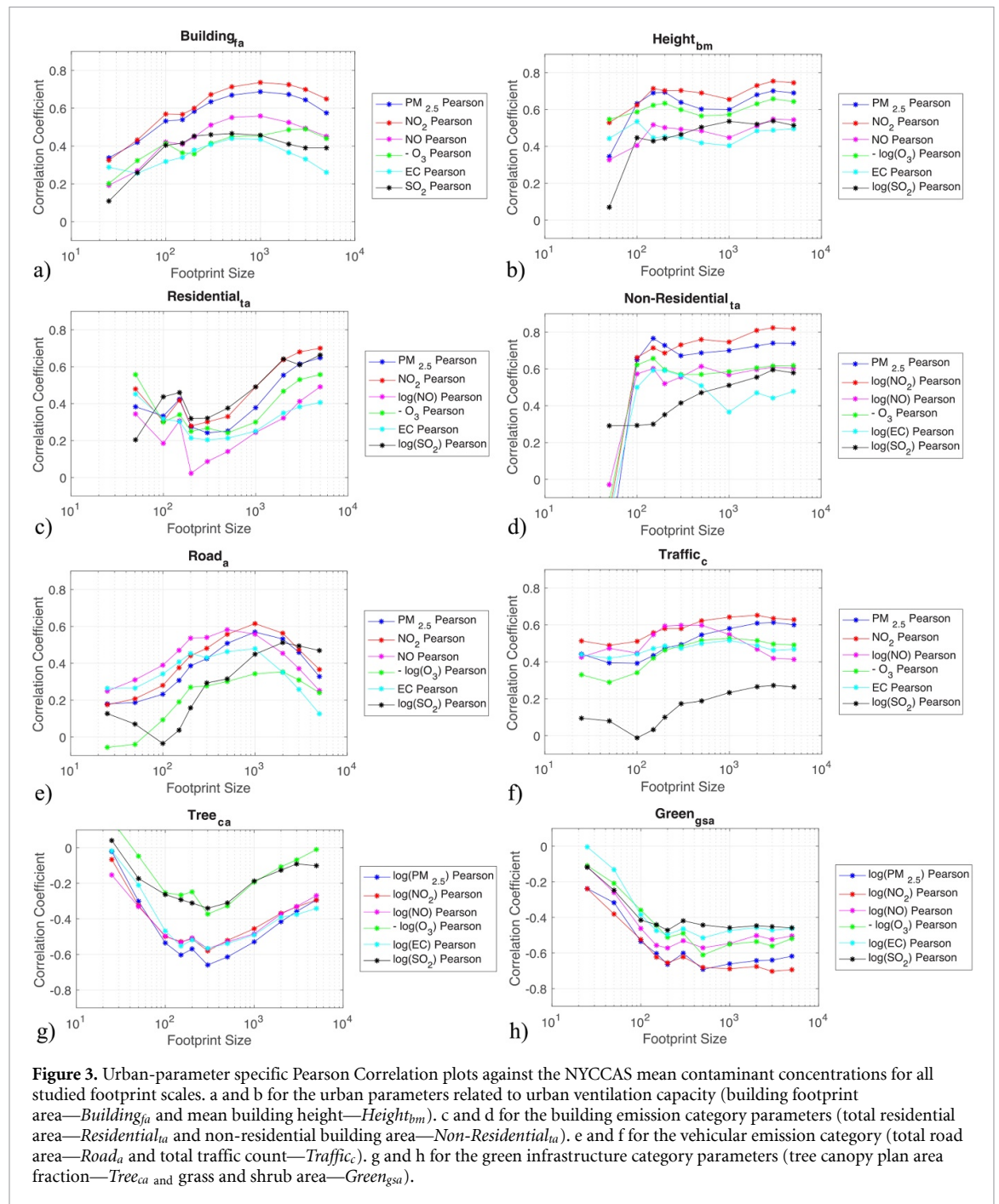
Sillman 1999, Oke *et al* 2017). As described by (Sillman 1999, Gu *et al* 2020). The larger and negative correlations with building density and traffic parameters suggest that the conventional main source of  $O_3$ , which is the reaction of combustion products, and  $NO_x$  in particular, with volatile organic compounds and sunlight, may not be active or effective in this dense urban area. In dense urban environments, especially in city centers with high buildings like New York, the penetration of sunlight into the street canyon is limited and thus the combustion emissions may not undergo a photoreaction. This trend is already visible in figure S2 where the  $O_3$  concentration is observed to be higher in less dense areas in boroughs outside of the denser Manhattan. Therefore, the reactants producing  $O_3$  might need to be lofted above the urban canopy, react and produce  $O_3$ , which is then re-entrained into the urban canopy layer from the urban boundary layer aloft. Another possible source of  $O_3$  is the transport from the troposphere, and since transported  $O_3$  is less likely to reach the pedestrian level in dense urban environments, that source would also explain the spatial variations observed in the data (Sillman *et al* 2003, Oke *et al* 2017). Future research can shed light on how the combination of these factors explains  $O_3$  variability, but what our paper aims to highlight is that the footprint extent needs to be correctly estimated to deduce the influence of urban parameters on air quality.

The same general observations are also supported by figure 3, but now the distinct correlations for the various studied footprints (25 m  $\times$  25 m, 50 m  $\times$  50 m, 100 m  $\times$  100 m, 150 m  $\times$  150 m, 200 m  $\times$  200 m, 300 m  $\times$  300 m, 500 m  $\times$  500 m, 1000 m  $\times$  1000 m, 2000 m  $\times$  2000 m, 3000 m  $\times$  3000 m, 5000 m  $\times$  5000 m) provide a much more complete picture, and allow us to derive the extents of the environmental neighborhoods. For the urban parameters that play an important role for urban ventilation such as  $Building_{fa}$ , the correlations peak at about 1000 m  $\times$  1000 m footprints, with a decrease at higher or smaller footprints, showing that the influence of building density is strongest within a few blocks. Thus, utilizing footprints that are too small or too big to understand the effect of building density on local air quality could yield misleading results or could underestimate the impact. The  $Height_{bm}$  parameter, on the other hand, plateaus beyond a  $\sim$ 150 m  $\times$  150 m footprint scale, revealing that the influence of the height of nearby buildings is the most significant, but that no penalty will be incurred if a larger footprint is used.

For the urban parameters associated with building emissions, that is the  $Residential_{ta}$  and  $Non-Residential_{ta}$  urban parameters, different trends are observed. For  $Residential_{ta}$ , high correlations are observed at very small footprints  $\sim$ 25 m  $\times$  25 m representative of the influence of nearby building

emissions on the local concentration. Correlations then drop at mid-range footprint sizes  $\sim$ 300 m  $\times$  300 m and increase again for 1000 m  $\times$  1000 m or larger footprints. This unique behavior may be related to the fact that residential household emissions take place on all floors and at roof levels, each of them with a distinct influence area. Residential emissions may therefore have a longer-range influence than those originating from commercial or urban industrial activities that generally take place in areas in closer proximity to the pedestrian level. The  $Non-Residential_{ta}$  urban parameter, on the other hand, shows a local peak at around  $\sim$ 150 m  $\times$  150 m footprints and plateaus for footprints  $>$ 200 m  $\times$  200 m. The  $Non-Residential_{ta}$  urban parameter includes total building floor areas for tertiary, industrial and logistical facilities and thus it is likely that such emissions occur mostly at the pedestrian level, or through high chimneys with a further reach. Also these tend to be areas with higher density of businesses and thus higher traffic (the correlation of  $Non-Residential_{ta}$  and  $Traffic_c$  is 0.63, table 1). Similarly to the  $Height_{bm}$  parameter examined before, linking the influence of  $Non-Residential_{ta}$  on local air quality at large footprints would not create a substantial penalty since the correlation plateaus. A conceptual diagram of characteristic emission sources in an urban street canyon in New York city is illustrated in figure 4.

The urban parameter categories associated with vehicular emissions, such as  $Road_a$ , show a peak with the highest correlation at around 1000 m  $\times$  1000 m footprints, with a steep decrease in the correlation for smaller and larger footprints. As with the  $Building_{fa}$  parameter,  $Road_a$  mostly affects local air pollution within a few block distance. The  $Traffic_c$  urban parameter on the other hand, shows different trends for the various pollutants, but except for  $O_3$ , it seems a plateau is reached somewhere between 300 and 1000 m footprint scales. It is important to note however that the present study does not get down to the street or building scales, which can present strong gradients specially for the  $Traffic_c$  parameter. This can be the case for example in proximity to heavy traffic intersections. Therefore, while the specific gradients in the main road arteries or street intersections of the city may not be accurately characterized, the presented footprint study provides a generalized correlation between the urban parameters and the averaged  $Traffic_c$  parameter at larger spatial, longer temporal, scales. Overall for the vehicular emission related parameters, the highest correlations are observed at 500 m  $\times$  500 m to 1000 m  $\times$  1000 m footprints, especially for the contaminants directly related to vehicular combustion  $PM_{2.5}$ ,  $NO_2$  and  $NO$ . The urban parameters related to green infrastructure,  $Tree_{ca}$  and  $Green_{gsa}$ , show the largest (negative) correlations at around 300 m  $\times$  300 m footprints, but while

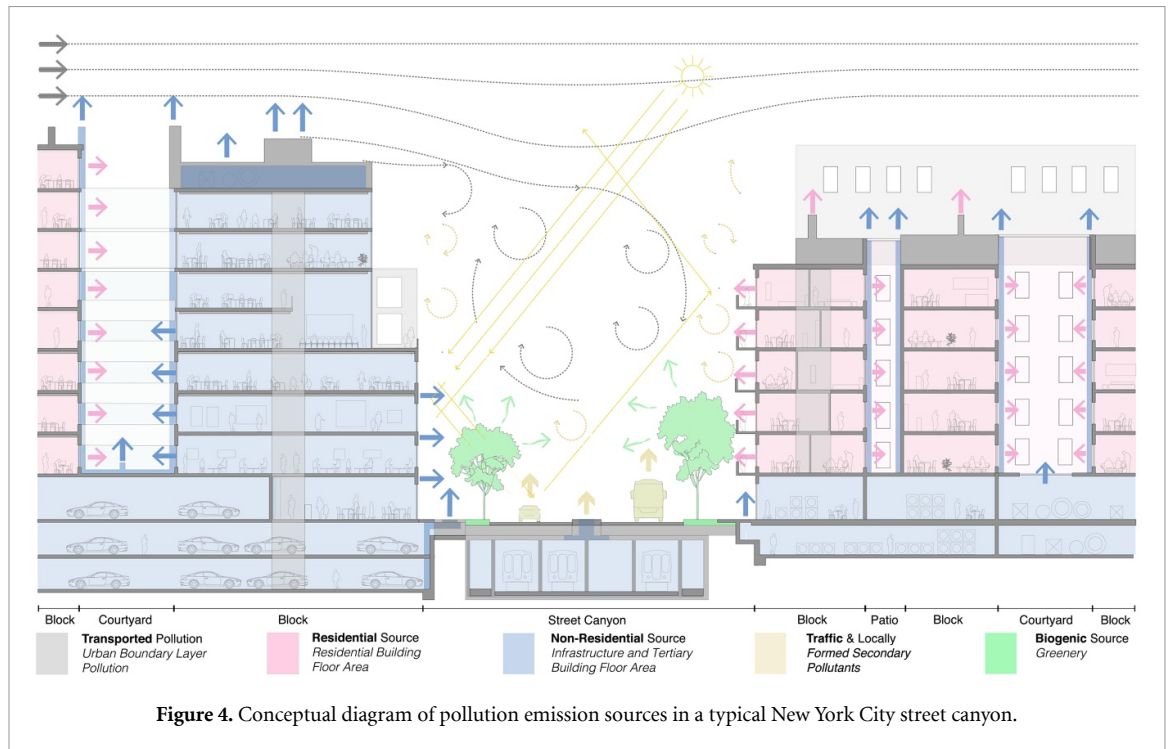


correlations drop for larger footprint with  $Tree_{ca}$ , they plateau with  $Green_{gsa}$ .

Of all the studied urban parameters, the highest correlations have been observed with building density and use-related (type) urban parameters, with highest correlations starting at  $\sim 150 \text{ m} \times 150 \text{ m}$ . The  $Non-Residential_{ta}$  parameter was the only one giving correlations exceeding 0.8, and was a very good predictor of  $NO_2$ ,  $PM_{2.5}$ , and  $NO$  concentration. More importantly, the accuracy penalties that would be associated with an inadequate choice of footprint scale, when aiming to understand the local air quality, are the largest for the urban parameters  $Building_{fa}$ ,  $Road_a$  and  $Tree_{ca}$ . Thus, it is critical that when studying the influence of such urban parameters

on the local air quality, the right footprint scale is analyzed.

The analysis presented in the paper has compiled the data collected from all NYCCAS stations together and thus it not only includes the stations located in the dense Manhattan street canyons, but also those located in parks and within different less dense boroughs of New York City, as shown in figures S2 & S4. It is then useful to evaluate whether including or excluding the stations located in parks or the stations located in different boroughs affects the trends observed in the correlations. To that end, we analyzed the results of the Pearson correlations between the NYCCAS  $PM_{2.5}$  and  $O_3$  data and the  $Building_{fa}$  and  $Tree_{ca}$  including and excluding the stations located in



parcs, and only including the data collected by the stations located in Manhattan (figure S4b). As demonstrated in figure S4b, the correlation trends remain consistent in all cases where parks are included and excluded from the computation, as well as when the district of Manhattan is studied in isolation or all stations belonging to different districts are studied together. This consistency underlines the robustness of our findings.

In order to check whether the analyzed air quality statistics are solely dictated by urban parameters or another analysis artifact, the trends of the urban parameters alone, for the different footprint scales, are displayed in figure S3. Comparing against figure 3, we note that (i) the correlation scaling trends are generally different from the urban parameter scaling trends and thus the former are not simply inherited from the spatial variability of the later, and (ii) there is a large variability in the urban parameters at a given scale (figure S3) and in air quality (figure S2) amongst the stations, which indicates that the signal in their correlation trends is large and supports the physical inferences and conclusions made in this paper.

#### 4. Conclusions

We investigate the correlation of urban air quality and physical attributes of the city at various urban footprint scales ranging from  $25\text{ m} \times 25\text{ m}$  to  $5000\text{ m} \times 5000\text{ m}$ . This allows us to define an environmental neighborhood as the surrounding area of influence the attributes of which are the most relevant for predicting air quality anomalies of a given location, and to measure its extent. The analyses are illustrated for the air quality dataset from NYCCAS

collected from 2008 to 2016 at 150 air quality stations. The urban parameters are compiled from the New York City land-use dataset at  $3\text{ ft} \times 3\text{ ft}$  resolution, the NYC Open data GIS dataset, the NYC PLUTO dataset for building and lot information, and the NYC traffic dataset for the year 2012–2013. While the spatial extent of the environmental neighborhoods might change for different cities, the main innovation in this study are the concept and methodology, which are generalizable.

Broadly, all combustion related parameters,  $\text{PM}_{2.5}$ ,  $\text{NO}_2$ ,  $\text{NO}$  and  $\text{SO}_2$  show consistent trends: the highest concentrations are in dense urban environments.  $\text{O}_3$  on the other hand follows a unique pattern, where the highest concentrations are observed in the less dense areas of the city. The results show that urban fabric parameters associated with building morphology and building use play the most critical role in determining the spatio-temporal gradients of air quality. The correlation coefficients associated with the mean building height ( $Height_{bm}$ ) and total non-residential areas ( $Non-Residential_{na}$ ) attain the highest magnitudes (negative for  $\text{O}_3$ ) at  $\approx 150\text{ m} \times 150\text{ m}$  footprints. On the other hand, the urban parameters Building Footprint Area ( $Building_{fa}$ ) or Road Area ( $Road_{ra}$ ) have a larger reach and affect the local concentrations at spatial extents  $\approx 1000\text{ m} \times 1000\text{ m}$ . Urban parameters associated with green infrastructure, such as tree canopy plan-area fraction ( $Tree_{ca}$ ) and grass/shrub ( $Green_{gsa}$ ), display negative correlations (except with  $\text{O}_3$ ) that peak at about  $\approx 300\text{ m} \times 300\text{ m}$ . A similar trend is observed with the total traffic count  $Traffic_c$ , which for most of the studied pollutants plateaus beyond  $\approx 300\text{ m} \times 300\text{ m}$ .

The environmental neighborhood scale deduced in our study pertains to Manhattan, but the same approach can be used to measure it in other cities. The extent of these neighborhoods is unlikely to be consistent, and the generalization of the findings might then appear challenging. Nevertheless, given the present analyses, it would seem plausible to hypothesize that the extent of environmental neighborhoods is closely linked to land use and land cover, and as such a suitable framework to use is that of the local climate zones reviewed in the introduction section (Stewart and Oke 2012, Muller *et al* 2013b). The LCZ classification can distill the commonalities and unique features between physical neighborhoods and cities, and we expect that the scale of environmental neighborhoods varies between different LCZs, but much less so among areas belonging to the same LCZ type (regardless of what city they are in).

The findings are consequential for urban planning and environmental quality management in many ways. First, they illustrate that selecting the optimal footprint size is critical for finding robust relations between air quality, and potentially other environmental quality parameters, and urban attributes at the neighborhood scale. Smaller footprints do not contain all the pertinent information, while larger footprints contain irrelevant, potentially misleading information; both might result in erroneous conclusions. In other words, the spatial extent is not an input to be chosen but is rather dictated by the urban physics. Second, spatial regression or machine learning models of air quality need to select the right environmental neighborhood scale to maximize model skill, and account for the full range of urban fabric and activity parameters. Finally, air pollutants and urban attributes do not all evolve similarly, underlining the complexity of airflow in cities, but general trends can be detected. The most consequential general trend is that for all urban characteristics the optimal footprint is ~200–1000 m, which implies that, like social neighborhoods, environmental neighborhoods in cities are quite limited and areas with worse air quality than the rest of the city can be improved with localized intervention measures to reduce environmental and health disparities.

## Acknowledgments

The study was supported by Princeton Engineering Innovation Funds and the campus-as-a-lab initiative, by Samsung Advanced Institute of Technology, and by the U.S. National Science Foundation under Grant 1664091 and through the UWIN Sustainability Research Network Cooperative Agreement 1444758. We would like to thank Sarah Johnson and the New York City Community Air Survey 2008–2016, the New York City Department of Health and Mental Hygiene, Queens College Center for the Biology of Natural Systems, and Zev Ross Spatial Analysis for collecting and

sharing the air quality data, and William Guthe from the Princeton University Geo spatial lab provided valuable help in the data formatting process.

## Competing interests

The authors declare that they have no competing interests related to the study performed and presented in this paper.

## Data availability statement

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

## ORCID iDs

M Llaguno-Munitxa  <https://orcid.org/0000-0002-5592-8901>

E Bou-Zeid  <https://orcid.org/0000-0002-6137-8109>

## References

- Akbari H, Cartalis C, Kolokotsa D, Muscio A, Pisello A L, Rossi F, Santamouris M, Synnef A, Wong N H and Zinzi M 2016 Local climate change and urban heat island mitigation techniques – the state of the art *J. Civil Eng. Manage.* **22-1** 1–16
- Apte J S, Messier K P, Gani S, Brauer M, Kirchstetter T W, Lunden M M, Marshall J D, Portier C J, Vermeulen R C H and Hamburg S P 2017 High-resolution air pollution mapping with google street view cars: exploiting big data *Environ. Sci. Technol.* **51** 6999–7008
- BBC News 2019 London air pollution: google street view cars to carry monitors ([www.bbc.com/news/uk-england-london-46878067](http://www.bbc.com/news/uk-england-london-46878067))
- Beckett K P, Freer-Smith P H and Taylor G 2000 Particulate pollution capture by urban trees: effect of species and windspeed *Glob. Chang. Biol.* **6** 995–1003
- Budel H and Wolf G 1933 Sonderklima der Stadte *Geog. Wochenschr.* **1** 25–31
- Buyantuev A and Wu J 2009 Urban heat islands and landscape heterogeneity: linking spatiotemporal variations in surface temperatures to land-cover and socioeconomic patterns *Landsc. Ecol.* **25-1** 17–33
- C40 Cities 2018 Air quality initiative ([www.c40.org/networks/air-quality](http://www.c40.org/networks/air-quality))
- Caplin A, Ghandehari M, Lim C, Glimcher P and Thurston G 2019 Advancing environmental exposure assessment science to benefit society *Nat. Commun.* **10** 1236
- Chandra S 2018 Delhi funds projects to cleanse its toxic air *Nat. News* (<https://doi.org/10.1038/d41586-018-03896-w>)
- Chen L J, Hsu W, Cheng M and Lee H C 2016 Demo: LASS: A Location-aware sensing system for participatory PM2.5 monitoring *MobiSys' Companion Proc. 14th Annual Int. Conf. on Mobile Systems, Applications, and Services Companion* pp 98–98
- Ching J *et al* 2019 Pathway using WUDAPT's digital synthetic city tool towards generating urban canopy parameters for multi-scale urban atmospheric modeling *Urban Climate* **28** 100459
- Cui Y, Zhang Z F, Froines J, Zhao J, Wang H, Yu S Z and Detels R 2003 Air pollution and case fatality of SARS in People's Republic of China: an ecologic study *Environ. Health* **2**
- Dye C 2008 Health and urban living *Science* **319-5864** 766–9
- Edussuriya P, Chan A and Ye A 2011 Urban morphology and air quality in dense residential environments in Hong Kong.

- Part I: district-level analysis *Atmos. Environ.* **45** 4789–803
- EPA 2019 Heat Island Impacts ([www.epa.gov/heat-islands/heat-island-impacts](http://www.epa.gov/heat-islands/heat-island-impacts))
- Editorial 2018 Filthy air is a global disgrace *Nature* **561** 285
- Editorial 2019 Stop denying the risks of air pollution *Nature* **568** 433
- Goldman G T and Dominici F 2019 Don't abandon evidence and process on air pollution policy *Science* **363-6434** 1398–400
- Grimmond C S B and Oke T R 1999 Aerodynamic properties of urban areas derived from analysis of surface form *J. Appl. Meteorol.* **38** 1262–92
- Gromke C and Blocken B 2015 Influence of avenue-trees on air quality at the urban neighborhood scale. Part II: traffic pollutant concentrations at pedestrian level *Environ. Pollut.* **196** 176–84
- Gu Y, Li K, Xu J, Liao H and Zhou G 2020 Observed dependence of surface ozone on increasing temperature in Shanghai, China *Atmos. Environ.* **221** 117108
- Hang J, Li Y, Sandberg M, Buccolieri R and Di Sabatino S 2012 The influence of building height variability on pollutant dispersion and pedestrian ventilation in idealized high-rise urban areas *Build Environ.* **56** 346–60
- Hasenfratz D, Saukh O, Walser C, Hueglin C, Fierz M, Arn T, Beutel J and Thiele L 2015 Deriving high-resolution urban air pollution maps using mobile sensor nodes *Pervasive Mob. Comput.* **16** 268–85
- Jenerette G D, Harlan S L, Buyantuev A, Stefanov W L, Declet-Barreto J, Ruddell B L, Myint S W, Kaplan S and Li X 2016 Micro-scale urban surface temperatures are related to land-cover features and residential heat related health impacts in Phoenix, AZ USA *Landscape Ecol.* **31-4** 745–60
- Jung M C, Park J and Kim S 2019 Spatial relationships between urban structures and air pollution in Korea *Sustain* **11**
- Kroeger T, Mcdonald R I, Boucher T, Zhang Z and Want L 2018 Where the people are: current Trends and future potential targeted investments in urban trees for PM<sub>10</sub> and temperature mitigation in 27 U.S. Cities *Landscape Urban Plan.* **177** 227–40
- Kuttler W and Strassburger A 1999 Air quality measurements in urban green areas - a case study *Atmos. Environ.* **33** 4101–8
- Levy I, Mihele C, Lu G, Narayan J and Brook J R 2014 Evaluating multipollutant exposure and urban air quality: pollutant interrelationships, neighborhood variability, and nitrogen dioxide as a proxy pollutant *Environ. Health Perspect.* **122** 65–72
- Li D and Bou-Zeid E 2013 Synergistic interactions between urban heat islands and heat waves: the impact in cities is larger than the sum of its parts *J. Appl. Meteor.- Climatol.* **52** 2051–64
- Li D, Liao W, Rigden A J, Liu X, Wang D, Malyshev S and Shevliakova E 2019 Urban heat island: aerodynamics or imperviousness? *Sci. Adv.* **5** eaau4299
- Llaguno-Munitxa M and Bou-Zeid E 2018 Shaping buildings to promote street ventilation: a large-eddy simulation study *Urban Climate* **26** 76–94
- Maciejczyk P B, Offenbergh J H, Clemente J, Blaustein M, Thurston G D and Chi C L 2004 Ambient pollutant concentrations measured by a mobile laboratory in South Bronx, NY *Atmos. Environ.* **38** 5283–94
- Martilli A 2014 An idealized study of city structure, urban climate, energy consumption, and air quality *Urban Climate* **10** 430–46
- Matte T D, Ross Z, Kheirbek I, Eisl H, Johnson S, Gorczynski J E, Kass D, Markowitz S, Pezeshki G and Clougherty J E 2013 Monitoring intraurban spatial patterns of multiple combustion air pollutants in New York City: design and implementation *J Expo Sci Env Epidemiol* **23** 223–31
- Miskell G, Salmond J, Longley I and Dirks D K 2015 A novel approach in quantifying the effect of urban design features on local-scale air pollution in central urban areas *Environ. Sci. Technol.* **49-15** 9004–11
- Mueller M D, Hasenfratz D, Saukh O, Fierz M and Hueglin C 2016 Statistical modelling of particle number concentration in Zurich at high spatio-temporal resolution utilizing data from a mobile sensor network *Atmos. Environ.* **126** 171–81
- Muller C L, Chapman L, Grimmond C S B, Young D T and Cai X 2013a Sensors and the city: a review of urban meteorological networks *Int. J. Climatol. RMets* **33** 1585–600
- Muller C L, Chapman L, Grimmond C S B, Young D T and Cai X M 2013b Toward a standardized metadata protocol for urban meteorological networks *Bull. Am. Meteorol. Soc.* **94** 1161–85
- Munir S, Mayfield M, Coca D and Jubb S A 2019 Structuring an integrated air quality monitoring network in large urban areas – discussing the purpose, criteria and deployment strategy *Atmos. Environ.* **2** 100027
- NYCCAS 2018 *The New York City Community Air Survey: Neighborhood Air Quality 2008–2016* ([www1.nyc.gov/?TeX=>assets/doh/downloads/pdf/environmental/comm-air-survey-08-16.pdf](http://www1.nyc.gov/?TeX=>assets/doh/downloads/pdf/environmental/comm-air-survey-08-16.pdf))
- Oke T R, Mills G, Christen A and Voogt J A 2017 *Urban Climates* (Cambridge: Cambridge University Press) p 546
- Pearson K 1895 Notes on regression and inheritance in the case of two parents *Proc. Royal. Soc. Lond.* **58** 240–2
- Peppeler A 1929 Die Temperaturverhältnisse von Karlsruhe an heissen Sommertagen. Deutsche Meteorologisches
- Rodriguez M C, Dupont-Courtade L and Oueslati W 2016 Air pollution and urban structure linkages: evidence from European cities *Renewable Sustainable Energy Rev.* **53** 1–9
- Santamouris M 2013 Using cool pavements as a mitigation strategy to fight urban heat island-A review of the actual developments *Renewable Sustainable Energy Rev.* **26** 224–40
- Schmidt W 1927 Die Verteilung der Minimumtemperaturen in derl Frostnacht des 12. 5.19.27 im Gemeindegebiet von Wien. In Fortschritte der Landwirtschaft
- Sillman S 1999 The relation between ozone, NOx and hydrocarbons in urban and polluted rural environments *Atmos. Env.* **33** 1821–45
- Sillman S 2003 Tropospheric ozone and photochemical smog *Treatise on Geochemistry, Environmental Geochemistry*, ed B Sherwood Lollar, 9407–31
- Stewart I D and Oke T R 2012 Local Climate Zones for urban temperature studies *Bull. Am. Meteorol. Soc.* **93** 1879–900
- The Washington Post 2020 The coronavirus is deadly enough. But some experts suspect bad air makes it worse ([www.washingtonpost.com/climate-environment/2020/03/15/smoking-air-pollution-coronavirus/](http://www.washingtonpost.com/climate-environment/2020/03/15/smoking-air-pollution-coronavirus/))
- Van Poppel M, Peters J and Bleux N 2013 Methodology for setup and data processing of mobile air quality measurements to assess the spatial variability of concentrations in urban environments *Environ. Pollut.* **183** 224–33
- Vos P E, Maiheu B, Vankerkom J and Janssen S 2013 Improving local air quality in cities: to tree or not to tree? *Environ. Pollut.* **183** 113–22
- WHO 2018a First WHO global conference on air pollution and health summary report ([www.who.int/phe/news/clean-air-for-health/en/](http://www.who.int/phe/news/clean-air-for-health/en/))
- WHO 2018b Climate Risks ([www.who.int/sustainable-development/cities/health-risks/climate-risks/en/](http://www.who.int/sustainable-development/cities/health-risks/climate-risks/en/))
- Yang J and Bou-Zeid E 2019a Designing sensor networks to resolve spatio-temporal urban temperature variations: fixed, mobile or hybrid? *Environ. Res. Lett.* **14** 074022
- Yang J and Bou-Zeid E 2019b Scale dependence of the benefits and efficiency of green and cool roofs *Landscape Urban Plan* **185** 127–40
- Yang L, Niyogi D, Chen F, Ni G, Tewari M, Tian F and Aliaga D 2016 Contrasting impacts of urban forms on the future thermal environment: example of Beijing metropolitan area *Environ. Res. Lett.* **11** 034018
- Yin S, Shen Z, Zhou P, Zou X, Che S and Wang W 2011 Quantifying air pollution attenuation within urban parks: an experimental approach in Shanghai, China *Environ. Pollut.* **159-8-9** 2155–63

Zhou C, Li S and Wang S 2018 Examining the impacts of urban form on air pollution in developing countries: a case study of China's megacities *Int. J. Environ. Res. Public Health* [15](#) 1565

Ziter C D, Pedersen E J, Kucharik C J and Turner M G 2019 Scale-dependent interactions between tree canopy cover and impervious surfaces reduce daytime urban heat during summer *PNAS* [116-15](#) 7575–80