

Spatial variability of forward modelled attenuated backscatter in clear-sky conditions over a megacity: implications for observation network design

Article

Published Version

Creative Commons: Attribution 4.0 (CC-BY)

Open Access

Warren, E. ORCID: <https://orcid.org/0000-0003-2757-7937>, Charlton-Perez, C., Lean, H. ORCID: <https://orcid.org/0000-0002-1274-4619>, Kotthaus, S. and Grimmond, S. ORCID: <https://orcid.org/0000-0002-3166-9415> (2022) Spatial variability of forward modelled attenuated backscatter in clear-sky conditions over a megacity: implications for observation network design. Quarterly Journal of the Royal Meteorological Society, 148 (744). ISSN 0035-9009 doi: 10.1002/qj.4253 Available at <https://centaur.reading.ac.uk/103328/>

It is advisable to refer to the publisher's version if you intend to cite from the work. See [Guidance on citing](#).

To link to this article DOI: <http://dx.doi.org/10.1002/qj.4253>

Publisher: Royal Meteorological Society

copyright holders. Terms and conditions for use of this material are defined in the [End User Agreement](#).

www.reading.ac.uk/centaur

CentAUR

Central Archive at the University of Reading

Reading's research outputs online

RESEARCH ARTICLE

Spatial variability of forward modelled attenuated backscatter in clear-sky conditions over a megacity: Implications for observation network design

Elliott Warren^{1,2}  | Cristina Charlton-Perez³  | Humphrey Lean³  | Simone Kotthaus⁴  | Sue Grimmond¹ 

¹Department of Meteorology, University of Reading, Reading, UK

²Met Office, Exeter, UK

³Met Office@Reading, Reading, UK

⁴Institut Pierre Simon Laplace (IPSL), CNRS, École Polytechnique, Institut Polytechnique de Paris, Palaiseau Cedex, France

Correspondence

E. Warren and S. Grimmond, Department of Meteorology, University of Reading, Reading, UK.

Email: elliott.warren@metoffice.gov.uk (E. W.); c.s.grimmond@reading.ac.uk (S. G.)

Funding information

European Cooperation in Science and Technology action PROBE, Grant/Award Number: CA18235; European Cooperation in Science and Technology action TOPROF, Grant/Award Number: ES1303; Newton Fund/Met Office CSSP China (SG), and European Research Council urbisphere, Grant/Award Number: 855005; EPSRC, Grant/Award Number: EP/P002331/1; Natural Environment Research Council, Grant/Award Numbers: NE/L008971/1, NE/H003231/1

Abstract

Sensors that measure the attenuated backscatter coefficient (e.g., automatic lidars and ceilometers [ALCs]) provide information on aerosols that can impact urban climate and human health. To design an observational network of ALC sensors for supporting data assimilation and to improve prediction of urban weather and air quality, a methodology is needed. In this study, spatio-temporal patterns of aerosol-attenuated backscatter coefficient are modelled using Met Office numerical weather prediction (NWP) models at two resolutions, 1.5 km (UKV) and 300 m (London Model [LM]), for 28 clear-sky days and nights. Initially, attenuated backscatter coefficient data are analysed using S-mode principal component analysis (PCA) with varimax rotation. Four to seven empirical orthogonal functions (EOFs) are produced for each model level, with common EOFs found across different heights (day and night) for both NWP models. EOFs relate strongly to orography, wind, and emissions source location, highlighting these as critical controls of attenuated backscatter coefficient spatial variability across the megacity. Urban–rural differences are largest when wind speeds are low and vertical boundary-layer dynamics can more effectively distribute near-surface aerosol emissions vertically. In several night-time EOFs, gravity-wave features are found for both NWP models. Increasing the horizontal resolution of native ancillaries (model input parameters) and improving the urban surface scheme in the LM may enhance the urban signal in the EOFs. PCA output, with agglomerative Ward cluster analysis (CA), minimises intra-group variance. The UKV and LM CA shape and size results are similar and strongly related to orography. PCA-CA is a simple, but adaptable methodology, allowing close alignment with observation network design goals. Here, CA is used with wind roses to suggest the optimised ALC deployment is one in the city to observe

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2022 Crown copyright. Quarterly Journal of the Royal Meteorological Society published by John Wiley & Sons Ltd on behalf of Royal Meteorological Society. This article is published with the permission of the Controller of HMSO and the Queen's Printer for Scotland.

the urban plume and others surrounding the city, with priority given to cluster size and frequency of upwind advection.

KEYWORDS

aerosols, lidar, network design, numerical weather prediction, urban

1 | INTRODUCTION

Automatic lidars and ceilometers (ALCs) measure the attenuated backscatter coefficient from atmospheric particles. Although many were originally used to measure cloud base height (Illingworth *et al.*, 2007), there are an increasingly wider variety of applications, including investigation of particle mass concentrations (Münkel and Roininen, 2010; You *et al.*, 2016), to measure boundary-layer height (Wang *et al.*, 2012; Peña *et al.*, 2013; Kotthaus and Grimmond, 2018b), and sea-breeze dynamics (Lemonsu *et al.*, 2006). As instrument sensitivity has increased (Kotthaus *et al.*, 2016), the utility of ALC-measured aerosol backscatter has grown (Illingworth *et al.*, 2013; Madonna *et al.*, 2014; Jin *et al.*, 2018; Cimini *et al.*, 2020), though many studies analyse relative variations or the signal-to-noise ratio of the measurements (Wiegner and Gasteiger, 2015). To relate modelled aerosol fields to the observed aerosol-attenuated backscatter coefficient directly, a forward operator (FO) is required.

An aerosol FO estimates the attenuated backscatter coefficient using aerosol variables as inputs. Typically, model aerosol-variable inputs are used from numerical weather prediction (NWP) or chemistry transport models to run an FO, so that the estimated attenuated backscatter coefficient can be evaluated against ALC observations or the latter can be assimilated (Benedetti and Dabas, 2016; Charlton-Perez *et al.*, 2016; Geisinger *et al.*, 2017; Chan *et al.*, 2018; Warren *et al.*, 2018). Several of these studies take advantage of ALC networks to obtain spatial aerosol information.

ALC networks cover spatial scales spanning cities (Kotthaus and Grimmond, 2018a), countries (Flentje *et al.*, 2010; Osborne *et al.*, 2018), and continents (Welton *et al.*, 2000; Illingworth *et al.*, 2007, 2015; Pappalardo *et al.*, 2014; Guerrero-Rascado *et al.*, 2016; Nishizawa *et al.*, 2016). However, the ideal positioning of an ALC within a network is uncertain and may depend on the location and spatial scale of dominant features, whether local, meso, or synoptic scale. Undesirable observational redundancy can exist when multiple ALCs are positioned such that they are capturing the same meteorological features and respond to the same emission sources, while

observations may miss (or only partially capture) features occurring elsewhere. Optimised instrument positioning maximises sampling efficiency and strengthens the observation products acquired across a network for a given equipment and operational cost.

Though practical constraints will impact final deployment locations (e.g., geopolitical boundaries or suitability of local infrastructure), an understanding of spatial variability and scale of meteorological features and their interaction with emission sources in the study area is critical. Most relevant processes are expressed in dynamics of the near-surface atmosphere; that is, the atmospheric boundary layer (ABL).

Network design for weather and hydrological instrumentation has been informed by the analysis of spatial or spatio-temporal variability in meteorological observations using (geo)-statistical techniques (e.g., Bastin *et al.*, 1982; Burn and Goulter, 1991; Bayat *et al.*, 2019), and regional climate models' climatological variability (St-Hilaire *et al.*, 2003). However, the resolution of the latter can be coarse compared with many meso-scale and local-scale processes driving aerosol distributions, and higher resolution models might better represent smaller scale processes that could be important.

The principal component analysis (PCA) technique can reduce a dataset to a series of orthogonal functions, or modes, that represent patterns of variability (Wilks, 2011). S-mode PCA (or empirical orthogonal function [EOF] analysis) is one variant that focuses on identifying spatial patterns in the different eigenvectors. It is used extensively to analyse meteorological variability, including rainfall (Smith and Phillips, 2013; Yu and Lin, 2015), wind (Álvarez-García *et al.*, 2020; Farjami and Hesari, 2020), and temperature (Li *et al.*, 2018). PCA has also been used to explore ABL variables, such as boundary-layer turbulence (Wilson, 1996; Lin *et al.*, 2008), urban heat island characteristics (Vicente-Serrano *et al.*, 2005; Qiao *et al.*, 2018), and air quality (Henry *et al.*, 1991; Chan and Mozurkewich, 2007; Fleming *et al.*, 2012; Rogula-Kozłowska *et al.*, 2015; Gupta *et al.*, 2018). Further in-depth reviews of different PCA approaches in meteorology and climatology can be found in Monahan *et al.* (2009), Schmidt *et al.* (2019), Wilks (2011), and

Zhang and Moore (2015). Some studies complement PCA with cluster analysis (CA) to explore the relations between modes or with other meteorological variables (Henry *et al.*, 1991; Beaver and Palazoglu, 2006) and to identify and group geographical regions according to similar target variable variability (Neal and Phillips, 2009; Smith and Phillips, 2013; Dogruparmak *et al.*, 2014). Identifying regions with co-varying meteorology or aerosol distribution based on PCA and CA can be exploited for network design by informing optimised placement of instrumentation to capture unique phenomena and maximise network utility.

In this study, our objectives are to: (a) identify the main spatial patterns of forward modelled attenuated backscatter coefficient β_m variability across a metropolitan area (London, UK), using NWP forecasts and PCA, (b) develop a method using PCA and CA to identify areas of similar β_m and inform ALC network design, and (c) understand the impact of horizontal resolution through the use of two NWP models.

2 | METHODS

2.1 | NWP data

The two models used are (a) the operational, convection-permitting UK regional model (UKV, 1.5 km; Tang *et al.*, 2013), and (b) the experimental London Model (LM, 300 m; Boutle *et al.*, 2016). These have been applied and evaluated for London, UK (Boutle *et al.*, 2016; Lapworth and Osborne, 2016; Warren *et al.*, 2018). Both are specific configurations of the Met Office Unified Model (Davies *et al.*, 2005). The urban surface energy balance scheme within the research LM at the time of this study was the JULES/Best one-tile scheme (Best, 1998; Best *et al.*, 2006; 2011) and within the UKV the JULES/MORUSES two-tile scheme (Porson *et al.*, 2010; Best *et al.*, 2011; Bohnenstengel *et al.*, 2011). The latter requires more detailed surface information and differentiates between urban canyons and roofs in a tiled scheme to better represent surface fluxes (Hertwig *et al.*, 2020). The aerosol emission ancillary is derived from the 1 km native resolution National Atmospheric Emissions Inventory dataset (Neal, 2019) to produce a dry mass mixing ratio of aerosol. The emission ancillary is based on a monthly climatology that is smoothly interpolated to a day scale using a running mean. Each day, a sinusoidal function is used to vary emissions across the day, which peak at 1200 UTC. The orography ancillary is from a digital terrain elevation data model with a native 100 m resolution (Boutle *et al.*, 2016). The nine land-use tiles are populated from the Institute of Terrestrial Ecology 25 m resolution dataset (Bunce *et al.*, 1990).

All ancillaries are rescaled to the respective NWP model resolution for implementation.

In this study, 28 clear-sky days are selected between April 1, 2018 and October 31, 2018 to avoid cloud-related processes (Table S1). Data from both models are used to force the aerosol lidar FO (aerFO; Warren *et al.*, 2018) to create three-dimensional hourly fields of attenuated backscatter β_m . Hourly aerFO calculations are conducted from midnight for 24 hr using the prior-day 2100 UTC forecast (i.e., 3 hr spin-up at the start of every model run are not used). The NWP data from both models are extracted for the London domain (Figure 1).

To account for diurnal variations in meteorological processes, such as the ABL height response to the surface energy balance, *day* (sunrise (SR) + 2 hr to sunset (SS) – 2 hr) and *night* (i.e., SS + 2 hr to SR – 2 hr) are analysed separately. From both NWP models, the 24 model levels from the bottom to 2,075 m above ground level (agl) are selected so as to include the ABL and (typically) the majority of the aerosol in the total atmospheric column (Seinfeld and Pandis, 2016). This ensures both the daytime boundary layer and night-time residual layer are always within the domain analysed. The heights of the model levels represent the midpoint of each layer. In London, the median boundary layer depth (1,704 m) and the typical daytime maximum mixed-layer height are below 2,075 m, as estimated from ALC observations within the London Urban Meteorological Observatory (Figure 1) (Kotthaus and Grimmond, 2018b). Manual quality control further ensures the ABL height does not exceed 2,075 m for any of the case-study days selected.

2.2 | Overview of the aerFO

Warren *et al.*'s (2020) aerFO version 2 is used to calculate β_m (at 905 nm). Briefly, the two critical variables are obtained from the NWP output: dry mass of total aerosol (m , kg·kg⁻¹) and relative humidity (RH). Model fields of air temperature T_{air} and pressure p_{air} are used with the specific water vapour mixing ratio q to compute the water vapour absorption. Both the UKV and LM provide m from the passive aerosol tracer (m_{MURK}) of the MURK visibility scheme, which considers the most common UK aerosol species (Neal, 2019). A constant lidar ratio of 43.1 sr is used. This lidar ratio, calculated specifically for urban areas, uses both in-situ urban aerosol observations from an urban background site (North Kensington) in the London Air Quality Network (DEFRA, 2018) and modelling (Warren *et al.*, 2020).

First, aerFO estimates physical properties of the aerosol in accumulation mode from m_{MURK} , including the dry mean volume radius and total number concentration.

This follows the MURK empirical parametrisation based on the climatological mean mass of dry aerosol, mean dry volume radius, and total number concentration (Clark *et al.*, 2008). The climatological (monthly) values and the geometric distribution for the accumulation mode are calculated from aerosol observations at an urban background site in London (NK; Figure 1; DEFRA, 2018).

Second, aerosol optical properties are calculated, including the extinction efficiencies for the dry aerosol particles and the effect of hygroscopic growth on particle extinction. For computational efficiency, precalculated look-up tables of dry extinction efficiency and hygroscopic growth effects are used. The estimation of the extinction coefficient includes the effect of water vapour. Third, the backscatter is calculated using a fixed lidar ratio and then converted into β_m by applying the transmission factor.

There are several sources of uncertainty in the spatial variation in β_m due to the NWP model aerosol composition being spatially invariant (in this application) that need to be considered. The lidar ratio, dry extinction efficiency, and extinction enhancement factor could vary spatially in reality, most notably when wind speeds are low and aerosol mixtures become more greatly influenced by local sources. Consequently, some spatial variability may be missed. However, the lidar ratio and extinction enhancement factor are highly sensitive to RH , and all three variables are highly sensitive to m_{MURK} (Warren *et al.*, 2020), which are spatially variable, and should still allow many spatial

patterns of β_m variability to be identifiable. For example, areas experiencing sea breezes also experience changes in RH , as well as aerosol composition, and the spatial variability in β_m may, therefore, be partially captured. To reduce the uncertainty in spatially varying aerosol proportions, speciated aerosol ancillaries can be used. However, speciated aerosol emission ancillaries were not available for the NWP models used in this study.

2.3 | S-mode PCA

To explore the nature of β_m variability in the LM domain (Figure 1), PCA is used to extract the most important spatial patterns from the original dataset and to identify when these spatial patterns are most prominent during the study period. As the distribution of β_m is positively skewed, it is not directly appropriate for PCA (Neal and Phillips, 2009). Therefore, a logarithmic transformation ($\log_{10}(\beta_m)$) is applied to reduce the skewness. For example, taking the logarithmic transformation of daytime β_m at 111.7 m agl reduced the skewness from 9.85 to 0.89.

Following Wilks (2011), S-mode PCA of $\log_{10}(\beta_m)$ is carried out for each model level separately. To begin, a vector is created containing the time series of $\log_{10}(\beta_m)$ from each grid cell for a model level ($x = x_1, \dots, x_n$, where n is the number of grid cells on the model level). x is then mean-centred elementwise (x'). PCA of x' will then identify positive and negative spatial patterns of variability

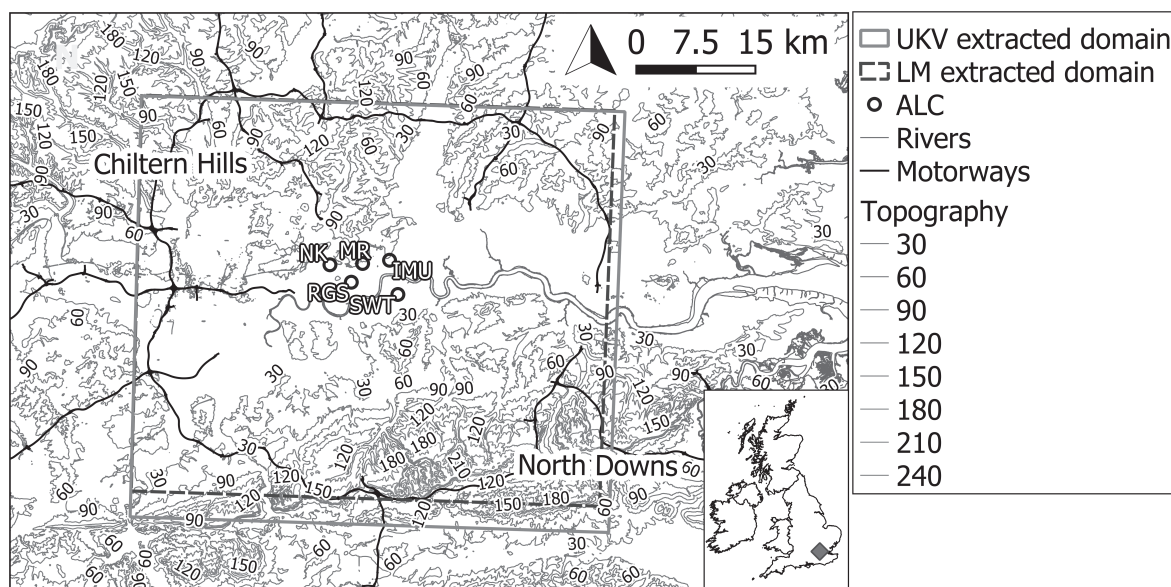


FIGURE 1 London Urban Meteorological Observatory automatic lidar and ceilometer (ALC) sites (circles with labels) within the “London” grid domains of the UKV (1.5 km) and London Model (LM; 300 m) extracted for the principal component analysis. Chiltern Hills and North Downs are highlighted as areas with significant orography. Contains Ordnance Survey data ©Crown copyright and database right (2019). Topography in metres above sea level [Colour figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com)]

with respect to the time average. The covariance matrix S of x' is then calculated, allowing PCA to emphasise identifying the largest covariances in x' as the main spatial patterns. Singular-value decomposition can then be carried out on S to produce unit-scaled eigenvectors (e_i , where $\|e_i\| = 1$) with paired eigenvalues λ_i , where $i = 1, \dots, n$, and where n is the number of original spatial variables x . The eigenvectors are then used to calculate a set of new, uncorrelated variables (the principal components [PCs]), that relate each e_i to each x_i (length of e_i is equal to the number of spatial points in x). A PC time series can be interpreted as a series of “scores”. High positive PC scores equate to the spatial pattern e_i being more relevant at a given time (length of a PC time series is equal to the length of the x time series), whereas more negative scores indicate the inverse of e_i is more relevant. Thus, each successive PC explains the maximum remaining variability in the original dataset.

By design, S-mode PCA produces orthogonal eigenvectors that represent statistical patterns of spatial variability. The orthogonality constraint of the method means that the first eigenvector captures the maximum variability in the data, and subsequent eigenvectors sequentially partition the remaining variability (Figure 2). However, these patterns may not be readily *physically* interpretable, because statistical patterns are not guaranteed to be matched to any single physical process. In fact, a single eigenvector could contain information related to multiple physical processes. Therefore, to ease physical interpretation,

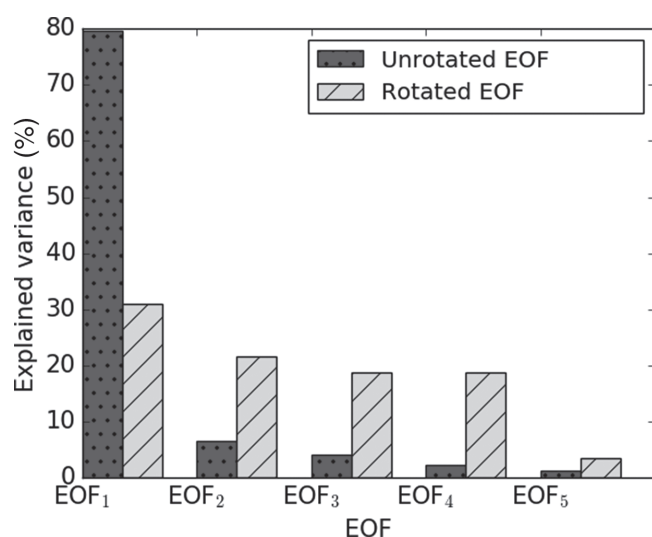


FIGURE 2 Explained variance of the first five unrotated empirical orthogonal functions (EOFs) (blue with dot hatching) and varimax rotated EOFs (red with line hatching), derived from UKV daytime forward-modelled attenuated backscatter coefficient β_m data at 111.7 m above ground level [Colour figure can be viewed at wileyonlinelibrary.com]

varimax rotation (Kaiser, 1958; Richman, 1986; Jolliffe and Cadima, 2016) is performed on a limited number of leading vectors in e . Varimax finds a new eigenvector rotation that maximises the sum of the variances of the squared loadings. Effectively, varimax redistributes the explained variability and identifies a new set of eigenvectors e_i , so that more unique spatial patterns are represented across fewer (or a single) eigenvector (Richman, 1986; Neal and Phillips, 2009; Figure 2). Here, only the e_i that explain more than 1% of the total variability in the original dataset are analysed. This threshold balances the need to retain as much of the original variability as possible while limiting the number of e_i that require physical interpretation. The selection criterion also limits the creation of multiplets (i.e., sets of e_i that effectively describe the same phenomena) (Wilks, 2011).

Before varimax rotation is performed, the e_i are scaled by $\lambda_i^{1/2}$ ($\|e_i\| = \lambda_i^{1/2}$) to “load” the vectors. The individual elements of scaled eigenvectors ($e_{i,k}$, where $k = 1, \dots, n$) are hereafter referred to as loadings. This scaling effectively incorporates the information about the explained variability of e_i into the newly rotated eigenvectors (see table 12.3 in Wilks (2011)). Although it is possible for varimax rotation to maintain the orthogonality of e_i , the scaling used here means e_i are not orthogonal and may be correlated (Wilks, 2011). As the eigenvectors are created with respect to space—and to be consistent with the literature (e.g., Jolliffe and Cadima, 2016; Qiao *et al.*, 2018)—the retained, rotated, and loaded eigenvectors are hereafter referred to as EOFs.

Following varimax rotation, rotated PCs are calculated using regression (Field, 2009):

$$PC_i = x' (S^{-1} \cdot EOF_i) \quad (1)$$

where x' is the original, mean-centred dataset and $S^{-1} \cdot EOF_i$ act as PC coefficients for PC_i . As S is often an ill-conditioned matrix and unsuitable for inversion, to calculate S^{-1} we take the Moore–Penrose pseudo-inverse of S using singular-value decomposition (Strang, 1988):

$$S^{-1} = V \Sigma^{-1} U^T \quad (2)$$

Spearman correlation coefficients r are calculated between EOFs and PCs to identify EOF multiplets for interpretation.

To aid interpretation, the relations between EOFs and different meteorological variables are explored. For each EOF, x is subsampled twice in time: when the paired PC scores are (a) above the 90th percentile and (b) below the 10th percentile. These two data distributions are assumed to represent the meteorological conditions associated with an EOF or its inverse, respectively.

2.4 | Ward CA

CA is used with PCA output to find geographic regions of similar β_m variability, based on the spatial patterns captured. The geographic clusters that CA produces can then be interpreted to aid ALC network design. Agglomerative Ward minimum variance CA is chosen here, as this clustering method merges grid cells into groups based on their similarity (Ward, 1963; Wilks, 2011). Across the 24 model levels analysed, all the unrotated eigenvectors (uEOFs) that explain more than 1% of the total variability of their respective model level are selected and used collectively as CA inputs (i.e., 2,075 m model level and below). This will produce a single horizontal CA map. Although rotated EOFs are analysed to identify individual spatial patterns of variability (Section 3.1), uEOFs are used as input to the clustering as they contain the same statistical information on spatial variability but are statistically independent, and therefore better suited for statistical analysis.

For each k grid box, the unrotated eigenvector loadings from each uEOF ($e_{i,k}$) are combined into new vectors. CA is performed on these new vectors. The initial k separate groups (each with one variable member) are iteratively paired and merged to reduce the number of clusters until j specified clusters remain. The clusters are merged according to the minimum sum of squared distances between all cluster variable elements and cluster centroids (i.e., merging the two most similar clusters). The loadings represent the amount of original variability explained across x , both positive and negative, and are larger for uEOFs that explain more. The varied loading of uEOFs also provides a benefit to the CA. Vectors scaled with larger loadings have a larger range of values, which effectively gives those uEOFs larger weights in the CA (Kaufman and Rousseeuw, 2005). This means that cluster groups are weighted more towards important uEOFs that explained more of the original variability.

Central to the network design is the number of clusters to produce by the CA. This number should be set equal to the number of sensors to be deployed in the network. As Ward CA minimises the total variance within groups (Ward, 1963), one instrument per cluster is expected to optimise representation. In our analysis, EOFs are scaled relative to the total variability at each height (i.e., not weighted with respect to other heights). Therefore, all heights are considered equally important when clustered. Alternatively, elbow, silhouette, or gap statistic methods could be used with a dendrogram to help inform the optimum number of clusters to use if the number of sensors to be deployed is not already known.

3 | RESULTS

3.1 | Spatial variability of forward-modelled attenuated backscatter β_m

The PCA with varimax rotation for the daytime period produced between four and seven EOFs for each model level in both models that explained more than 1% of variability in the original dataset. In general, similar EOF patterns are identified across different heights. Examples of common daytime EOF patterns from the UKV (Figure 3) and the LM (Figure 4) are shown for the 111.7 m agl model level. At this height, five EOFs are needed to explain the β_m variance.

EOF patterns derived from both NWP models are strongly related to wind speed and direction. The highest loadings in each EOF (Figures 3a–e and 4a–e) are associated with areas downwind of aerosol emission sources (cf. Figures 3f and 4f) leading to an increase in aerosol and, consequently, β_m . As most aerosol emission sources are located near the centre of the domain (i.e., the most urbanised area), the highest EOF loadings are usually found near the domain edges. For example, the highest loadings in EOF₁ are in the northwest of the domain (Figure 3a) as EOF₁ is most prevalent under south–southeasterly winds (Figure 5a). The 28 case-study days used to derive the EOFs cover a wide range of wind directions and speeds (Figure 5), and therefore the importance of horizontal advection to β_m variability is highly likely at these NWP scales. A full day and night composite of wind speed and direction over all cases is in Figure S1.

A second factor determining the shape of the EOFs is topography. London, located in a river valley, is situated between the Chiltern Hills (north) and North Downs (south and southeast) (Figure 1), so its elevation is relatively low compared with its surroundings (>200 m above sea level; Figures 3a–e and 4a–e). The EOF₁, EOF₂, and EOF₄ spatial patterns occur when air is advected onto the hills in the northwest, southeast, and northeast, respectively. As the air is forced upward, part of the detected β_m variability is explained by the RH response to air temperature reduction. But the advected air also increases m_{MURK} locally, where background emissions are typically low. As $\exp(\beta_m)$ is proportional to RH and $\beta_m \propto m_{\text{MURK}}$, the advection increases the local variation of β_m . For example, if an air parcel with $m_{\text{MURK}} = 24 \mu\text{g}\cdot\text{kg}^{-1}$ is advected onto a hill with background concentration of $m_{\text{MURK}} = 18 \mu\text{g}\cdot\text{kg}^{-1}$, and raised adiabatically by 100 m such that RH increases from ~ 80 to 85% , β_m increases locally by $\sim 1.2 \times 10^{-6} \text{ m}^{-1}\cdot\text{sr}^{-1}$. This combined orographic effect and advection of aerosol emitted from the major sources (middle of the domain; Figure 3f) increases β_m in areas surrounding the city. The spatial variability of

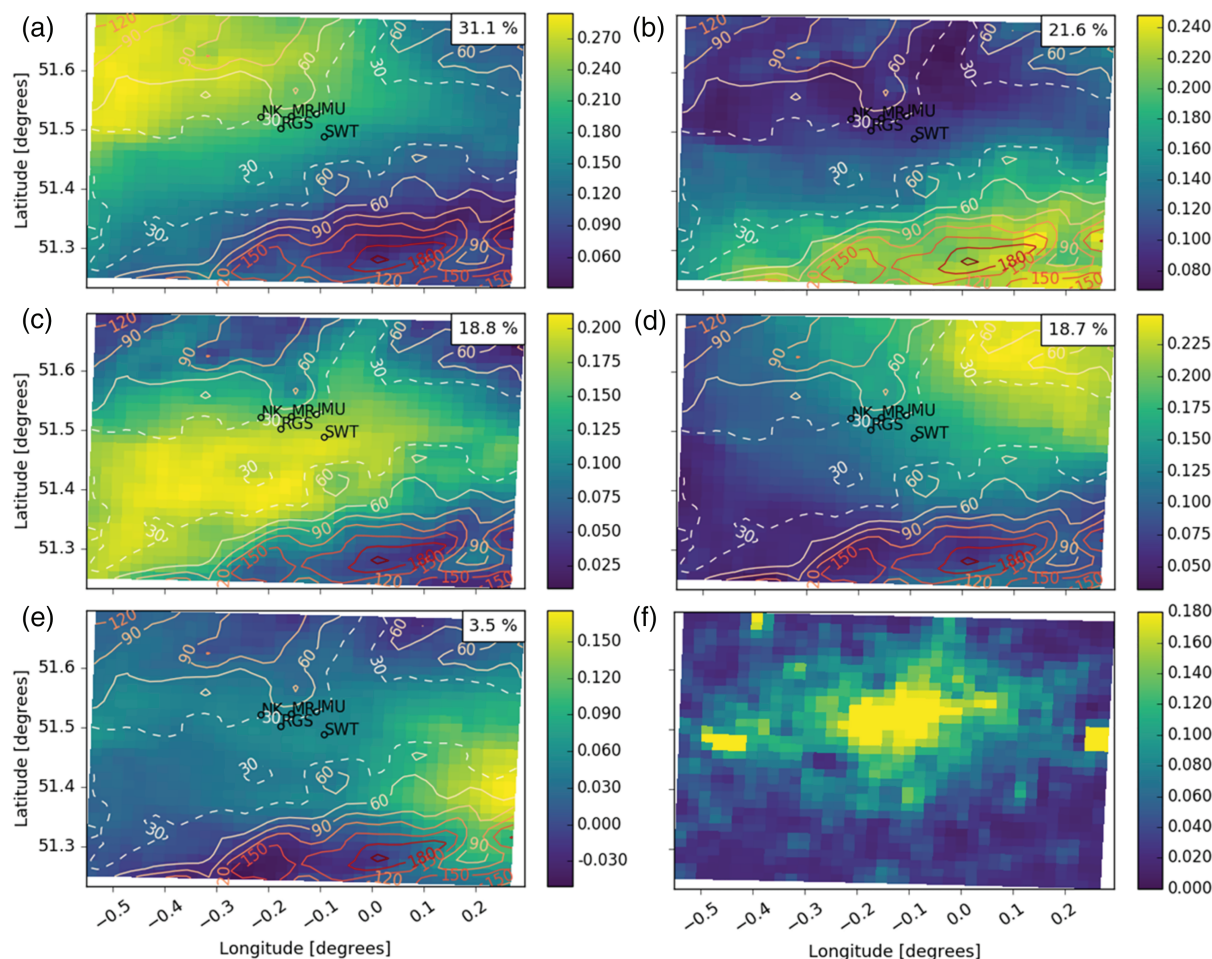


FIGURE 3 Examples of the main loaded, rotated empirical orthogonal function (EOF) patterns that typically occur across different model heights, derived from the UKV (1.5 km) during the daytime (colour) with topographic contours (lines, 30 m dashed). EOFs explain decreasing variability (%) of the original dataset (111.7 m above ground level): (a) EOF₁; (b) EOF₂; (c) EOF₃; (d) EOF₄; (e) EOF₅; (f) m_{MURK} background aerosol emissions ($\mu\text{g}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$) climatology for July. London Urban Meteorological Observatory automatic lidar and ceilometer network (Figure 1) shown as dots and labels in the domain centre [Colour figure can be viewed at wileyonlinelibrary.com]

β_m , m_{MURK} , and RH for EOF₁ from the UKV is shown in Figure S2.

The general distribution of loadings in the EOFs across the London domain are similar for the two NWP models, as the synoptic winds are the most important factor driving the spatial patterns. Whereas the LM EOFs have more spatial detail in the loading distribution around complex orography (Figure 3b cf. Figure 4b), the UKV topographic resolution is sufficient to generate orographic uplift and the associated peak in loadings.

EOF₁ and EOF₂ for both NWP models, across each of the model levels analysed, explain a large amount of the total variability (typically between ~25–40% for EOF₁ and ~20–40% for EOF₂). EOF₃ explains ~13–25% of the total variability in β_m and can be linked to two processes that increase β_m : relatively higher RH in the southwest, and high aerosol variability in the centre of the domain. This EOF often occurs with low winds speeds or northeasterlies

(Figure 5). In the UKV above 471.7 m, the two processes are instead represented by two separate EOF patterns, here labelled as EOF₃ sub-patterns, EOF_{UKV,3-1} (peak in loadings in the southwest of the domain driven by RH) and EOF_{UKV,3-2} (peak loadings in the centre of the domain driven by urban aerosol exchange) (Figure S4). In the LM, EOF_{LM,5} is most similar to EOF_{UKV,3-2} from the UKV, with respect to the loading spatial distribution.

EOF_{UKV,3-2} and EOF_{LM,5} likely portray rural–urban differences in vertical aerosol transport and changes in RH . PC time series corresponding to these EOFs peak with lower wind speeds, suggesting vertical aerosol transport is more important when wind speeds are low. Positive PC scores indicate higher β_m over the city, and negative scores indicate higher β_m in the surrounding rural areas. Below 111.7 m (not shown), soon after SR, the PC scores are high, which reflects the greater build-up of m_{MURK} overnight in urban areas compared with the rural surroundings.

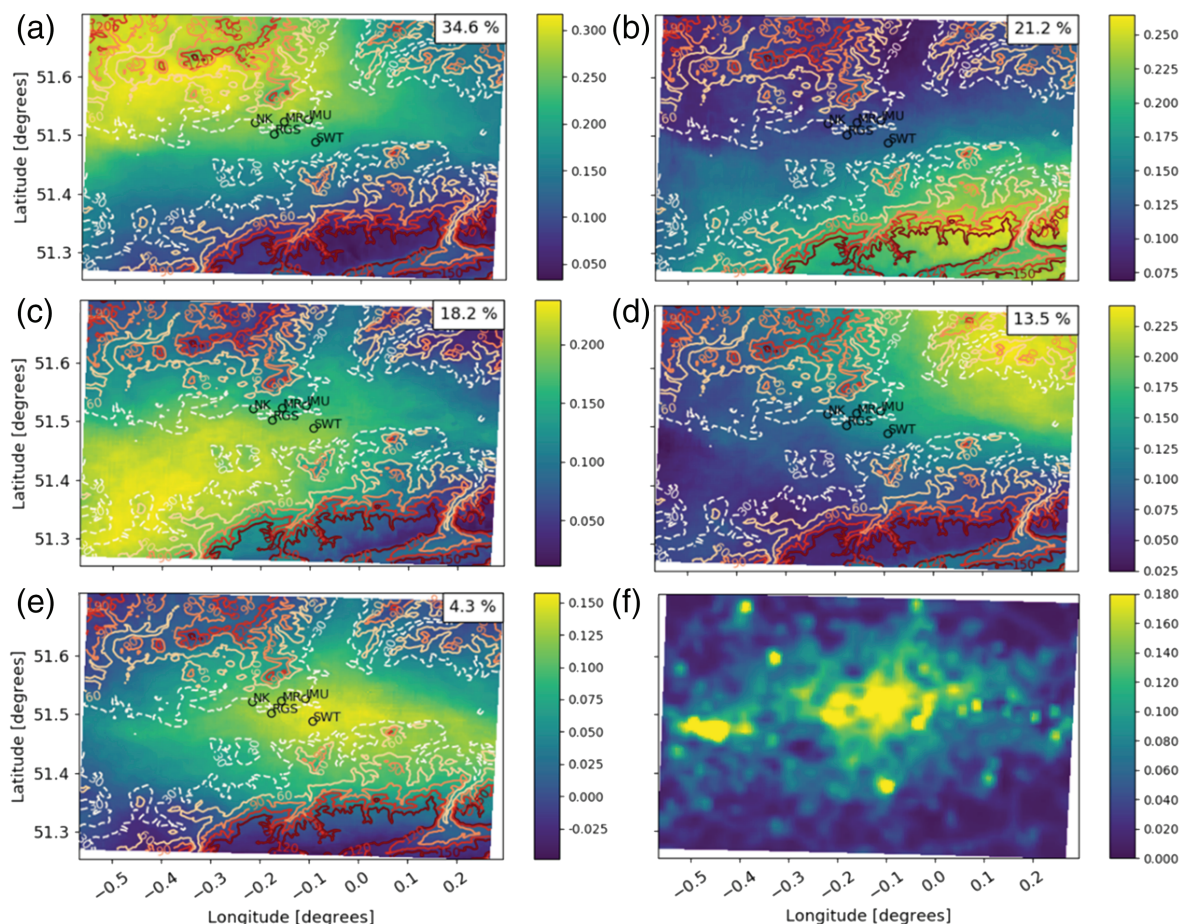


FIGURE 4 As Figure 3, but derived from the London Model (300 m) [Colour figure can be viewed at wileyonlinelibrary.com]

Greater buoyancy over built-up areas due to greater urban heat and roughness, in the absence of higher wind speeds, encourages greater upward transport of m_{MURK} and β_{m} in the domain centre. This vertical redistribution reduces the near-surface build-up of m_{MURK} , and consequently emphasises the urban–rural difference.

Above 111.7 m, EOF_{UKV,3–2} and EOF_{LM,5} PC scores typically have a diurnal pattern, and are low after SR, peak midday, and decrease before SS (Figure S5a). Furthermore, the scores also increase later after SR and decrease sooner before SS for level heights further from the surface (Figure S5b). This smaller range (later rise, lower fall) of PC scores at greater heights could reflect the extra time required for boundary-layer mixing processes to reach greater heights and to entrain the existing residual layer while redistributing relatively high near-surface concentrations of m_{MURK} further vertically. Again, this suggests the importance of m_{MURK} emissions and ABL dynamics in determining the variability of β_{m} .

EOF_{LM,5} tends to represent a greater proportion of total β_{m} variability than its UKV counterpart EOF_{UKV,3–2}, as urban effects are better resolved by the higher resolution

of both the urban characteristics and aerosol emission sources in the LM ancillaries. Aerosol and emission ancillaries with a native resolution of 1 km are coarsened for use in the UKV (to 1.5 km) and interpolated for the LM (to 300 m). Higher resolution LM ancillaries are likely to provide greater variability in β_{m} , particularly where emission sources are more heterogeneous.

For the nocturnal analysis, as with the daytime, the EOFs are derived from PCA with varimax rotation. Again, EOFs relate strongly to wind speed, wind direction, emission source location, and orography (not shown), with peak loadings in similar positions. However, there are some differences between the daytime and night-time EOFs.

Between model heights 325.0 and 955.0 m, several EOFs display wavelike patterns that are not present during the day; for example, EOF₂ at 417.7 m in both models (Figure 6), which generally occurs under northwesterly winds. Similar wave structures are found in EOFs associated with southerly winds (not shown). These patterns likely represent gravity waves produced as northwesterly flow passes over the Chiltern Hills (northwest London

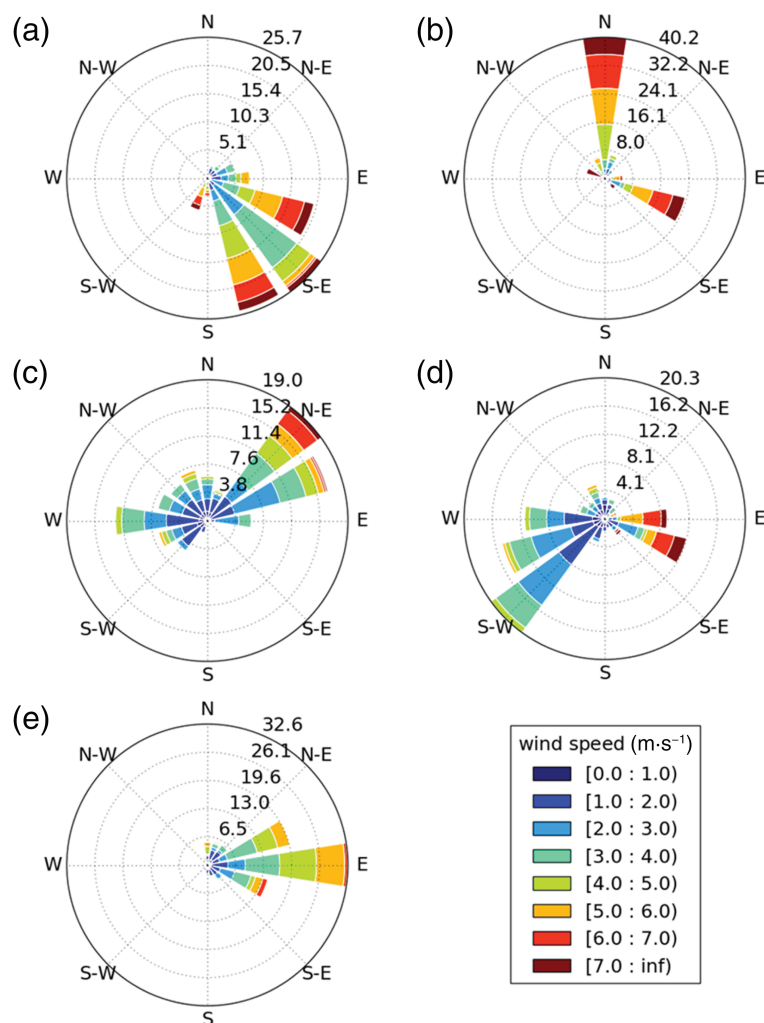


FIGURE 5 Paired daytime wind roses of UKV model wind speed ($\text{m}\cdot\text{s}^{-1}$) and direction (degrees) for each rotated empirical orthogonal function (EOF) at model height 111.7 m (Figure 3): (a) EOF₁; (b) EOF₂; (c) EOF₃; (d) EOF₄; (e) EOF₅. Wind speeds are temporally sampled when the scores for each PC are above the 10th percentile, and spatially sampled across the domain (wind characteristics most associated with each EOF). Bins extend from low inclusive and high exclusive (key). Radial axis (%) frequency of occurrence by wind direction bin. See Figure S1 for a composite of all wind speeds [Colour figure can be viewed at wileyonlinelibrary.com]

domain; Figure 3f) (Figure 6a), which is in agreement with earlier studies using UKV and LM data for the Greater London area (Lapworth and Osborne, 2016; 2019). Gravity waves can occur in statically stable conditions and have been found in EOFs derived from large-eddy simulations of the planetary boundary layer under weakly convective conditions (Wilson, 1996). Using lidar observations Gibert *et al.* (2011) found gravity waves to cause fluctuations in RH , which would translate into β_m variability captured by the EOFs.

The presence of waves is visible in the w -wind component when the PC scores for EOF₂ are >4 (e.g., Figure 6c,d). Under clear-sky conditions at night, the boundary layer is more likely to be stable, which can reinforce the presence of gravity waves (Wallace and Hobbs, 2006). The unstable daytime conditions limit their formation (Figures 3 and 4). However, the frequency and spatial prevalence of the gravity waves in the NWP model data may be too large compared with reality, leading to an overrepresentation in the EOFs. Both models have been found to overestimate atmospheric stability over the

urban area of London, which is caused by an underestimation of the anthropogenic heat emissions (Bohnenstengel *et al.*, 2014). Further, the simpler one-tile urban surface scheme used for the LM simulations also tends to underestimate London's sensible heat fluxes in the evening (Hertwig *et al.*, 2020). Thus, the gravity waves are likely a less important phenomenon than implied by the EOFs derived from the NWP model data.

3.2 | Application of spatial variability to network design

To find geographic regions with similar β_m variability, Ward CA (Ward, 1963; Wilks, 2011) is performed using the uEOFs for day/night, and for UKV/LM separately, with the cluster numbers (n) set to 5, 7 and 20. For example, Figure 7 shows the daytime UKV clustering of $n = 7$ clusters, ordered from the largest (1) to the smallest area (7).

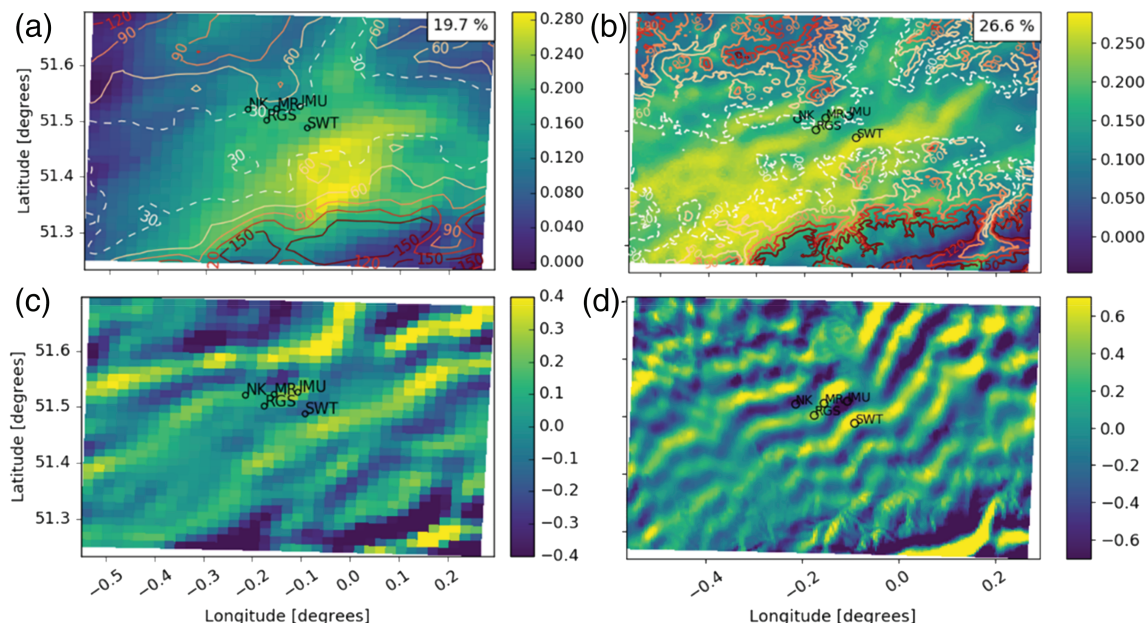
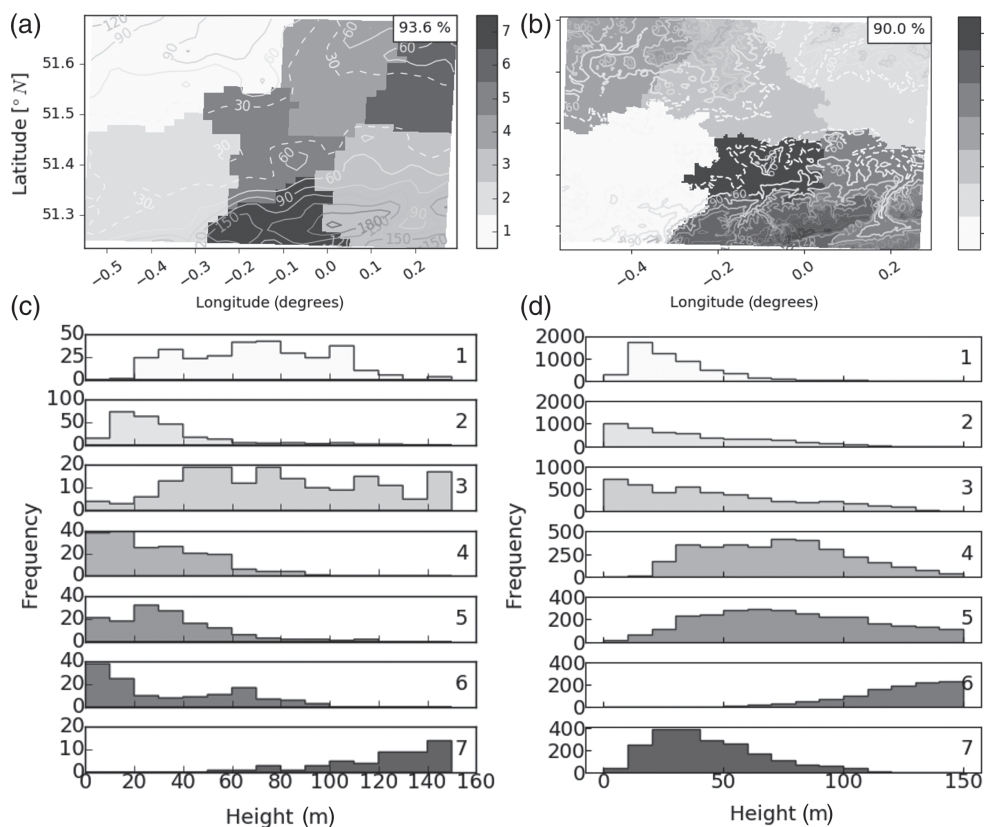


FIGURE 6 Nocturnal empirical orthogonal function EOF_2 at the 471.7 m model level derived separately from (a) UKV and (b) London Model. Horizontal cross-section of w -wind component when EOF_2 principal component scores are >4 on October 23, 2018, 2300 UTC (at 471 m) from (c) UKV and (d) London Model [Colour figure can be viewed at wileyonlinelibrary.com]

FIGURE 7 Ward cluster analysis ($n = 7$ clusters) of unrotated empirical orthogonal function (EOFs) from principal component analysis of daytime β_m . Colour patches numbered from the largest (C_1) to the smallest (C_7) derived from (a, c) UKV and (b, d) London Model; (a, b) location of cluster with topographic contours (lines, 30 m dashed), and (c, d) topographic variation in each cluster. Note frequency (y-axis) varies. All EOFs included explain $\geq 1\%$ of the variability in modelled attenuated backscatter coefficient [Colour figure can be viewed at wileyonlinelibrary.com]



The UKV and LM cluster maps are similar. Both have clusters located over the city centre and in the surroundings for different wind directions (Figure 7). The clusters are spatially confined, which is likely because

the covariance of eigenvector values between neighbouring grid cells is relatively high compared with grid cells that are further apart. The central LM cluster (7, $C_{LM,7}$) is smaller than its UKV counterpart ($C_{UKV,5}$), which could be

related to the higher loadings over central London compared with its surroundings in the first LM uEOF, and more nuanced skill possible with the smaller grid resolution. The clusters split into three elevation range types (Figure 7c,d): (a) wide range, in the northwest and south-east; (b) small range and low elevation, for most of the others; and (c) predominantly higher elevation, in the south.

Qualitatively, the boundaries between adjacent clusters follow the topographic contours for $n = 7$ (notably $C_{UKV,1}$, $C_{UKV,7}$ and $C_{LM,5}$, $C_{LM,6}$; Figure 7), which is the same for other cases ($n = 5$ or $n = 20$) analysed. When $n = 20$ (Figure S6), the smaller clusters align with more detailed topographic features (e.g., valley variations and hill tops, $C_{UKV,12}$ in the southwest). CA results are also similar between day and night, despite the wavelike patterns in night-time EOFs (Figure S7).

To better understand the near-surface β_m variability (cf. variability in the full vertical domain as used in Figure 7, additional daytime cluster maps are produced, using EOFs from model height of (a) 5.0 m only, and (b) 5.0–111.7 m inclusively (Figure S8). Unlike the clusters in Figure 7, near-surface clusters have a greater east–west elongation and appear to be more tightly constrained by local near-surface emission sources, which also have an east–west elongation due to the presence of Heathrow Airport (west), the city centre (middle), and several emissions sources near the River Thames (east) (Figures 4f and 5f). Comparison of near-surface clusters with clusters in Figure 7 highlight that near-surface β_m variability is influenced more by local-scale emission sources, whereas clusters derived using the full sampled height domain are affected by larger-scale processes.

The CA maps can be used to inform the deployment of ALC instruments. As Ward clustering aims to minimise the variability in β_m within a cluster, one logical approach would be to locate a single ALC instrument in each cluster. Using the maps and wind roses together to assess aerosol advection patterns can help prioritise deployment. For example, larger clusters and those with frequent upwind advection would come first. A network of seven ceilometers deployed for data assimilation into the UKV could be distributed with one in the city centre, to observe the urban plume, and the other six in the surrounding hills (one in each cluster). Instruments in rural areas would monitor both advected aerosol emissions upwind from the city and the subsequent orographic lifting effect. If only two instruments are deployed, the CA maps derived in this study would suggest one in the city ($C_{UKV,5}$), and a second in the northwest cluster ($C_{UKV,1}$) given the relatively large cluster size and high frequency of southeasterlies. At night, the cluster shapes and areal extents can change. Therefore, where possible, the recommendation is

to use both day and night CA maps when selecting a site location.

4 | CONCLUSIONS

To assess what drives spatial variations in aerosol (and RH), within the boundary layer across a region in and around a megacity under cloud-free conditions, a method is developed to identify common spatial patterns of variability. The attenuated backscatter coefficient is modelled using the aerFO operator with input from two NWP models at different resolutions: the 1.5 km Met Office UKV and the 300 m research LM for a domain around Greater London, UK. PCA with varimax rotation is used for two periods (day, night) to create EOFs for each model level.

PCA results are combined with CA to identify areas of similar aerosol variability. The CA results can be used with wind roses to identify potential locations for ALC instruments to maximise uniqueness of observations across a network, with respect to observing features at spatial scales larger than the model resolution used, for the purpose of data assimilation and evaluation.

Common patterns in attenuated backscatter coefficient variability are identifiable across different model levels during both day and night, with similar results for both NWP models.

From the analysis of 28 days clear-sky case-study days, it is concluded that:

- Spatial variability in modelled attenuated backscatter coefficient β_m can be mostly explained by orography, transport of aerosols from source locations (mainly in city centre) and wind direction.
- The urban–rural difference in attenuated backscatter coefficient is most pronounced under low wind speeds and advection, allowing vertical boundary-layer dynamics to redistribute relatively higher urban near-surface aerosol concentrations over the city.
- Possible gravity waves influence the spatial variability of attenuated backscatter coefficient in the residual layer at night in both NWP models, through fluctuations in RH .
- Results with the coarser resolution NWP model (1.5 km) are comparable to the higher resolution (~ 300 m) NWP model, though the higher resolution NWP benefits from higher resolution orography.
- NWP models could potentially indicate more variability in the attenuated backscatter coefficient across a city if model inputs (i.e., ancillaries) have higher

resolution (e.g., aerosol emissions). For example, improved urban energy balance fluxes from anthropogenic heat emissions and heat storage could improve the representation of urban–rural contrasts.

- CA identifies distinct regions (clusters) of similar attenuated backscatter coefficient variability to inform instrument placement of a network for data assimilation into NWP models.
- CA results are similar between day and night, despite the wavelike patterns in night-time EOFs.
- When creating five or more clusters, the cluster shape and sizes relate to orography, aerosol emissions, and wind direction. One cluster is located in the domain centre (city), and others in surrounding rural areas with high orography.
- Near-surface clusters relate more to aerosol emissions due to greater local-scale influence, whereas clusters using the full vertical information can be influenced more by larger-scale influences.
- We recommend placing one instrument in the city to observe the urban plume and others in the rural surroundings, with priority given to the larger clusters and clusters with higher frequency of downwind aerosol advection from the city.
- Identification of distinct regions is mostly constrained by NWP model resolution and the spatial scales of features it can effectively resolve. To inform network design for ALCs to observe smaller spatial features, higher resolution NWP models are needed.

The PCA-CA technique is highly adaptable and could be modified or used in a wider variety of applications. The technique could be used on subsamples of NWP data to focus on better capturing information for particular meteorological situations or regions. For example, subsampling for above-average aerosol events, or the upper extent of the boundary layer where observations can be sparse in urban areas (Barlow, 2014). In addition, the PCA-CA technique could aid the spatial interpretation of verification statistics. The CA highlights spatial regions where the model grid cells covaried most similarly; therefore, the spatial applicability of verification statistics using ALC instruments located in identified cluster regions can be better understood. The PCA-CA technique could be applied to other meteorological variables, beyond aerosols (e.g., cloud base and boundary-layer heights), to inform network design that can target variability in those variables.

As two NWP models were used, the sensitivity of cluster maps to model biases was only partially explored.

Therefore, future work should consider generating cluster maps from more NWP models to better understand this sensitivity and ideally reduce the impact of model bias on instrument placement.

ACKNOWLEDGEMENTS

EW was funded by the NERC Scenario DTP and University of Reading (NE/L002566/1). We acknowledge the London Urban Meteorological Observatory-funded (NERC [NE/H003231/1, NE/L008971/1], EU7 211345, H2020 637519, KCL, UR) instruments, the vast number of people who help keep the equipment operating on a daily basis, and the institutions that provide sites. Funding support from EPSRC (EP/P002331/1), Newton Fund/Met Office CSSP China (SG), and ERC urbisphere (855005) are acknowledged. Mapped terrain, river, and road data were provided by Ordnance Survey. The following datasets are hosted on Zenodo: (1) ceilometer: <https://zenodo.org/record/4723564>; (2) UKV: <https://zenodo.org/record/4723576>; (3) LM: <https://zenodo.org/record/4723582>; (4) PCA results: <https://zenodo.org/record/4723806>. Python code used in analysis is available at <https://github.com/elliottwarren/improveNetworks>. We thank the European Cooperation in Science and Technology (COST) actions “TOPROF” (ES1303) and “PROBE” (CA18235) that fostered collaborations for this work. Thanks also to Anthony Illingworth and Helen Dacre for their helpful advice and comments.

AUTHOR CONTRIBUTIONS


Elliott Warren: Conceptualization; formal analysis; investigation; methodology; software; visualization; writing – original draft; writing – review and editing. **Cristina Charlton-Perez:** Conceptualization; investigation; supervision; writing – review and editing. **Humphrey Lean:** Conceptualization; investigation; supervision. **Simone Kotthaus:** Conceptualization; investigation; supervision. **Sue Grimmond:** Conceptualization; investigation; supervision; writing – review and editing.

ORCID

Elliott Warren  <https://orcid.org/0000-0003-2757-7937>

Cristina Charlton-Perez  <https://orcid.org/0000-0002-7805-5955>

Humphrey Lean  <https://orcid.org/0000-0002-1274-4619>

Simone Kotthaus  <https://orcid.org/0000-0002-4051-0705>

Sue Grimmond  <https://orcid.org/0000-0002-3166-9415>

REFERENCES

- Álvarez-García, F.J., Fresno-Schmolk, G., OrtizBevia, M.J., Cabos, W. and RuizdeElvira, A. (2020) Reduction of aggregate wind power variability using empirical orthogonal teleconnections: an application in the Iberian Peninsula. *Renewable Energy*, 159, 151–161. <https://doi.org/10.1016/j.renene.2020.05.153>.
- Barlow, J.F. (2014) Progress in observing and modelling the urban boundary layer. *Urban Climate*, 10, 216–240. <https://doi.org/10.1016/j.uclim.2014.03.011>.
- Bastin, G., Lorent, B., Duqué, C. and Gevers, M. (1982) Optimal estimation of the average areal rainfall over a basin and optimal selection of rain-gauge locations. *IFAC Proceedings Volumes*, 15(4), 1571–1576. [https://doi.org/10.1016/S1474-6670\(17\)63223-3](https://doi.org/10.1016/S1474-6670(17)63223-3).
- Bayat, B., Hosseini, K., Nasser, M. and Karami, H. (2019) Challenge of rainfall network design considering spatial versus spatiotemporal variations. *Journal of Hydrology*, 574(March), 990–1002. <https://doi.org/10.1016/j.jhydrol.2019.04.091>.
- Beaver, S. and Palazoglu, A. (2006) Cluster analysis of hourly wind measurements to reveal synoptic regimes affecting air quality. *Journal of Applied Meteorology and Climatology*, 45(12), 1710–1726. <https://doi.org/10.1175/JAM2437.1>.
- Benedetti, A. and Dabas, A. (2016) Assessment of the necessary developments to assimilate Aeolus and EarthCARE aerosol profile products. Technical Note 1 of ESA contract number 4000116106.
- Best, M.J. (1998) A model to predict surface temperatures. *Boundary-Layer Meteorology*, 88(2), 279–306. <https://doi.org/10.1023/A:1001151927113>.
- Best, M.J., Grimmond, C.S.B. and Villani, M.G. (2006) Evaluation of the urban tile in MOSES using surface energy balance observations. *Boundary-Layer Meteorology*, 118(3), 503–525. <https://doi.org/10.1007/s10546-005-9025-5>.
- Best, M.J., Pryor, M., Clark, D.B., Rooney, G.G., Essery, R.L.H., Ménard, C.B., Edwards, J.M., Hendry, M.A., Porson, A., Gedney, N., Mercado, L.M., Sitch, S., Blyth, E., Boucher, O., Cox, P.M., Grimmond, C.S.B. and Harding, R.J. (2011) The Joint UK land Environment Simulator (JULES), model description – Part 1: Energy and water fluxes. *Geoscientific Model Development*, 4(3), 677–699. <https://doi.org/10.5194/gmd-4-677-2011>.
- Bohnenstengel, S.I., Evans, S., Clark, P.A. and Belcher, S.E. (2011) Simulations of the London urban heat island. *Quarterly Journal of the Royal Meteorological Society*, 137(659), 1625–1640. <https://doi.org/10.1002/qj.855>.
- Bohnenstengel, S.I., Hamilton, I., Davies, M. and Belcher, S.E. (2014) Impact of anthropogenic heat emissions on London's temperatures. *Quarterly Journal of the Royal Meteorological Society*, 140(679), 687–698. <https://doi.org/10.1002/qj.2144>.
- Boutle, I.A., Finnenkoetter, A., Lock, A.P. and Wells, H. (2016) The London Model: Forecasting fog at 333 m resolution. *Quarterly Journal of the Royal Meteorological Society*, 142(694), 360–371. <https://doi.org/10.1002/qj.2656>.
- Bunce, R.G.H., Barr, C.J., Clarke, R.T., Howard, D.C. and Lane, A.M.J. (1990) ITE land classification of Great Britain 1990. NERC Environmental Information Data Centre <https://doi.org/10.5285/ab320e08-faf5-48e1-9ec9-77a213d2907f>.
- Burn, D.H. and Goulter, I.C. (1991) An approach to the rationalization of streamflow data collection networks. *Journal of Hydrology*, 122(1–4), 71–91. [https://doi.org/10.1016/0022-1694\(91\)90173-F](https://doi.org/10.1016/0022-1694(91)90173-F).
- Chan, K.L., Wiegner, M., Flentje, H., Mattis, I., Wagner, F., Gasteiger, J. and Geiß, A. (2018) Evaluation of ECMWF-IFS (version 41R1) operational model forecasts of aerosol transport by using ceilometer network measurements. *Geoscientific Model Development*, 11(9), 3807–3831. <https://doi.org/10.5194/gmd-11-3807-2018>.
- Chan, T.W. and Mozurkewich, M. (2007) Application of absolute principal component analysis to size distribution data: Identification of particle origins. *Atmospheric Chemistry and Physics*, 7(3), 887–897. <https://doi.org/10.5194/acp-7-887-2007>.
- Charlton-Perez, C., Simonin, D., Ballard, S., Hopkin, E., Illingworth, A., Kotthaus, S., Westbrook, C. and Grimmond, C. S. B. (2016) Suitability of ceilometer observations for DA, Data Assimilation And Ensembles Science Report 10. Met Office.
- Cimini, D., Haeffelin, M., Kotthaus, S., Löhnert, U., Martinet, P., O'Connor, E., Walden, C., Coen, M.C. and Preissler, J. (2020) Towards the profiling of the atmospheric boundary layer at European scale – Introducing the COST Action PROBE. *Bulletin of Atmospheric Science and Technology*, 1(1), 23–42. <https://doi.org/10.1007/s42865-020-00003-8>.
- Clark, P.A., Harcourt, S.A., Macpherson, B., Mathison, C.T., Cusack, S. and Naylor, M. (2008) Prediction of visibility and aerosol within the operational Met Office unified model. I: Model formulation and variational assimilation. *Quarterly Journal of the Royal Meteorological Society*, 134(636), 1801–1816. <https://doi.org/10.1002/qj.318>.
- Davies, T., Cullen, M.J.P., Malcolm, A.J., Mawson, M.H., Staniforth, A., White, A.A. and Wood, N. (2005) A new dynamical core for the Met Office's global and regional modelling of the atmosphere. *Quarterly Journal of the Royal Meteorological Society*, 131(608), 1759–1782. <https://doi.org/10.1256/qj.04.101>.
- DEFRA (2018) Site information for London N. Kensington (UKA00253). Available at: https://uk-air.defra.gov.uk/networks/site-info?site_id=KC1#startcontent [Accessed 8th October 2018].
- Dogruparmak, S.C., Keskin, G.A., Yaman, S. and Alkan, A. (2014) Using principal component analysis and fuzzy c-means clustering for the assessment of air quality monitoring. *Atmospheric Pollution Research*, 5(4), 656–663. <https://doi.org/10.5094/APR.2014.075>.
- Farjami, H. and Hesari, A.R.E. (2020) Assessment of sea surface wind field pattern over the Caspian Sea using EOF analysis. *Regional Studies in Marine Science*, 35, 101254. <https://doi.org/10.1016/j.rsma.2020.101254>.
- Field, A. (2009) *Discovering statistics using IBM SPSS Statistics*, 5th edition. London: Sage.
- Fleming, Z.L., Monks, P.S. and Manning, A.J. (2012) Review: Untangling the influence of air-mass history in interpreting observed atmospheric composition. *Atmospheric Research*, 104–105, 1–39. <https://doi.org/10.1016/j.atmosres.2011.09.009>.
- Flentje, H., Heese, B., Reichardt, J. and Thomas, W. (2010) Aerosol profiling using the ceilometer network of the German Meteorological Service. *Atmospheric Measurement Techniques Discussions*, 3(4), 3643–3673. <https://doi.org/10.5194/amtd-3-3643-2010>.
- Geisinger, A., Behrendt, A., Wulfmeyer, V., Strohbach, J., Förstner, J. and Potthast, R. (2017) Development and application of a backscatter lidar forward operator for quantitative validation of aerosol dispersion models and future data assimilation.

- Atmospheric Measurement Techniques*, 10(12), 4705–4726. <https://doi.org/10.5194/amt-10-4705-2017>.
- Gibert, F., Arnault, N., Cuesta, J., Plougonven, R. and Flamant, P.H. (2011) Internal gravity waves convectively forced in the atmospheric residual layer during the morning transition. *Quarterly Journal of the Royal Meteorological Society*, 137(659), 1610–1624. <https://doi.org/10.1002/qj.836>.
- Guerrero-Rascado, J.L., Landulfo, E., Antuña, J.C., de Melo Jorge Barbosa, H., Barja, B., Bastidas, Á.E., Bedoya, A.E., da Costa, R.F., Estevan, R., Forno, R., Gouveia, D.A., Jiménez, C., Larroza, E.G., da Silva Lopes, F.J., Montilla-Rosero, E., de Arruda Moreira, G., Nakaema, W.M., Nisperuza, D., Alegria, D., Múnera, M., Otero, L., Papandrea, S., Pallota, J.V., Pawelko, E., Quel, E.J., Ristori, P., Rodrigues, P.F., Salvador, J., Sánchez, M.F. and Silva, A. (2016) Latin American Lidar Network (LALINET) for aerosol research: Diagnosis on network instrumentation. *Journal of Atmospheric and Solar-Terrestrial Physics*, 138–139, 112–120. <https://doi.org/10.1016/j.jastp.2016.01.001>.
- Gupta, S., Gadi, R., Sharma, S.K. and Mandal, T.K. (2018) Characterization and source apportionment of organic compounds in PM₁₀ using PCA and PMF at a traffic hotspot of Delhi. *Sustainable Cities and Society*, 39(February), 52–67. <https://doi.org/10.1016/j.scs.2018.01.051>.
- Henry, R.C., Wang, Y. and Gebhart, K.A. (1991) The relationship between empirical orthogonal functions and sources of air pollution. *Atmospheric Environment. Part A. General Topics*, 25(2), 503–509. [https://doi.org/10.1016/0960-1686\(91\)90322-X](https://doi.org/10.1016/0960-1686(91)90322-X).
- Hertwig, D., Grimmond, S., Hendry, M.A., Saunders, B., Wang, Z., Jeoffrion, M., Vidale, P.L., McGuire, P.C., Bohnenstengel, S.I., Ward, H.C. and Kotthaus, S. (2020) Urban signals in high-resolution weather and climate simulations: Role of urban land-surface characterisation. *Theoretical and Applied Climatology*, 142(1–2), 701–728. <https://doi.org/10.1007/s00704-020-03294-1>.
- Illingworth, A.J., Cimini, D., Gaffard, C., Haefelin, M., Lehmann, V., Löhnert, U., O'Connor, E.J. and Ruffieux, D. (2015) Exploiting existing ground-based remote sensing networks to improve high-resolution weather forecasts. *Bulletin of the American Meteorological Society*, 96(12), 2107–2125. <https://doi.org/10.1175/BAMS-D-13-00283.1>.
- Illingworth, A.J., Hogan, R.J., O'Connor, E.J., Bouniol, D., Brooks, M.E., Delanoë, J., Donovan, D.P., Eastment, J.D., Gaussiat, N., Goddard, J.W.F., Haefelin, M., Klein Baltinik, H., Krasnov, O.A., Pelon, J., Piriou, J.M., Protat, A., Russchenberg, H.W.J., Seifert, A., Tompkins, A.M., van Zadelhoff, G.J., Vinit, F., Willen, U., Wilson, D.R. and Wrench, C.L. (2007) Cloudnet: Continuous evaluation of cloud profiles in seven operational models using ground-based observations. *Bulletin of the American Meteorological Society*, 88(6), 883–898. <https://doi.org/10.1175/BAMS-88-6-883>.
- Illingworth, A.J., Ruffieux, D., Cimini, D., Löhnert, U., Haefelin, M. and Lehmann, V. (2013) COST Action ES0702 European ground-based observations of essential variables for climate and operational meteorology final report.
- Jin, Y., Sugimoto, N., Shimizu, A., Nishizawa, T., Kai, K., Kawai, K., Yamazaki, A., Sakurai, M. and Wille, H. (2018) Evaluation of ceilometer attenuated backscattering coefficients for aerosol profile measurement. *Journal of Applied Remote Sensing*, 12(04), 1. <https://doi.org/10.1117/1.jrs.12.042604>.
- Jolliffe, I.T. and Cadima, J. (2016) Principal component analysis: A review and recent developments. *Philosophical Transactions of the Royal Society A*, 374(2065), 20150202. <https://doi.org/10.1098/rsta.2015.0202>.
- Kaiser, H.F. (1958) The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, 23(3), 187–200. <https://doi.org/10.1007/BF02289233>.
- Kaufman, L. and Rousseeuw, P. (2005) *Finding groups in data: An introduction to cluster analysis*. Hoboken, NJ: Wiley.
- Kotthaus, S. and Grimmond, C.S.B. (2018a) Atmospheric boundary-layer characteristics from ceilometer measurements. Part 1: A new method to track mixed layer height and classify clouds. *Quarterly Journal of the Royal Meteorological Society*, 144(714), 1525–1538. <https://doi.org/10.1002/qj.3299>.
- Kotthaus, S. and Grimmond, C.S.B. (2018b) Atmospheric boundary-layer characteristics from ceilometer measurements. Part 2: Application to London's urban boundary layer. *Quarterly Journal of the Royal Meteorological Society*, 144(714), 1511–1524. <https://doi.org/10.1002/qj.3298>.
- Kotthaus, S., O'Connor, E., Munkel, C., Charlton-Perez, C., Haefelin, M., Gabey, A.M. and Grimmond, C.S.B. (2016) Recommendations for processing atmospheric attenuated backscatter profiles from Vaisala CL31 ceilometers. *Atmospheric Measurement Techniques*, 9(8), 3769–3791. <https://doi.org/10.5194/amt-9-3769-2016>.
- Lapworth, A. and Osborne, S.R. (2016) Evidence for gravity wave drag in the boundary layer of a numerical forecast model: A comparison with observations. *Quarterly Journal of the Royal Meteorological Society*, 142(701), 3257–3264. <https://doi.org/10.1002/qj.2909>.
- Lapworth, A. and Osborne, S.R. (2019) Gravity-wave drag in the stable boundary layer over moderate terrain. *Boundary-Layer Meteorology*, 171(2), 175–189. <https://doi.org/10.1007/s10546-018-00422-3>.
- Lemonsu, A., Bastin, S., Masson, V. and Drobinski, P. (2006) Vertical structure of the urban boundary layer over Marseille under sea-breeze conditions. *Boundary-Layer Meteorology*, 118(3), 477–501. <https://doi.org/10.1007/s10546-005-7772-y>.
- Li, Q., Chen, P., Sun, L. and Ma, X. (2018) A global weighted mean temperature model based on empirical orthogonal function analysis. *Advances in Space Research*, 61(6), 1398–1411. <https://doi.org/10.1016/j.asr.2017.12.031>.
- Lin, C.-L., Xia, Q. and Calhoun, R. (2008) Retrieval of urban boundary layer structures from Doppler lidar data. Part II: Proper orthogonal decomposition. *Journal of the Atmospheric Sciences*, 65(1), 21–42. <https://doi.org/10.1175/2007JAS2329.1>.
- Madonna, F., Amato, F., Vande Hey, J. and Pappalardo, G. (2014) Ceilometer aerosol profiling vs. Raman lidar in the frame of INTERACT campaign of ACTRIS. *Atmospheric Measurement Techniques Discussions*, 7(12), 12407–12447. <https://doi.org/10.5194/amtd-7-12407-2014>.
- Monahan, A.H., Fyfe, J.C., Ambaum, M.H.P., Stephenson, D.B. and North, G.R. (2009) Empirical orthogonal functions: The medium is the message. *Journal of Climate*, 22(24), 6501–6514. <https://doi.org/10.1175/2009JCLI3062.1>.
- Munkel, C. and Roininen, R. (2010). Automatic monitoring of boundary layer structures with ceilometer. In: *TEC-2010*. https://library.wmo.int/pmb_ged/wmo-td_1546_en/P2_17_Muenkel_Germany.pdf

- Neal, L. (2019) Air quality emissions data (Technical Report). Exeter: Met Office.
- Neal, R.A. and Phillips, I.D. (2009) Summer daily precipitation variability over the East Anglian region of Great Britain. *International Journal of Climatology*, 29(11), 1661–1679. <https://doi.org/10.1002/joc.1826>.
- Nishizawa, T., Sugimoto, N., Matsui, I., Shimizu, A., Higurashi, A. and Jin, Y. (2016) The Asian Dust and Aerosol Lidar Observation Network (AD-NET): Strategy and progress. *EPJ Web of Conferences*, 119(January), 19001. <https://doi.org/10.1051/epjconf/201611919001>.
- Osborne, M., Adam, M., Buxmann, J., Sugier, J., Marengo, F. and Haywood, J. (2018) Saharan dust and biomass burning aerosols during ex-hurricane Ophelia: Validation of the new UK lidar and sun-photometer network, (August).
- Pappalardo, G., Amodeo, A., Apituley, A., Comeron, A., Freudenthaler, V., Linné, H., Ansmann, A., Bösenberg, J., D'Amico, G., Mattis, I., Mona, L., Wandinger, U., Amiridis, V., Alados-Arboledas, L., Nicolae, D. and Wiegner, M. (2014) EARLINET: Towards an advanced sustainable European aerosol lidar network. *Atmospheric Measurement Techniques*, 7(8), 2389–2409. <https://doi.org/10.5194/amt-7-2389-2014>.
- Peña, A., Gryning, S.E. and Hahmann, A.N. (2013) Observations of the atmospheric boundary layer height under marine upstream flow conditions at a coastal site. *Journal of Geophysical Research Atmospheres*, 118(4), 1924–1940. <https://doi.org/10.1002/jgrd.50175>.
- Porson, A., Clark, P.A., Harman, I.N., Best, M.J. and Belcher, S.E. (2010) Implementation of a new urban energy budget scheme in the MetUM. Part I: Description and idealized simulations. *Quarterly Journal of the Royal Meteorological Society*, 136(651), 1514–1529. <https://doi.org/10.1002/qj.668>.
- Qiao, Z., Wu, C., Huang, N., Xu, X., Sun, Z. and Sun, X. (2018) Spatio-temporal structure of the urban thermal environment in Beijing based on an empirical orthogonal function. *Journal of Spatial Science*, 63(2), 297–310. <https://doi.org/10.1080/14498596.2018.1488632>.
- Richman, M.B. (1986) Rotation of principle components. *Journal of Climatology*, 6, 293–335.
- Rogula-Kozłowska, W., Majewski, G. and Czechowski, P.O. (2015) The size distribution and origin of elements bound to ambient particles: a case study of a Polish urban area. *Environmental Monitoring and Assessment*, 187(5), 240. <https://doi.org/10.1007/s10661-015-4450-5>.
- Schmidt, O.T., Mengaldo, G., Balsamo, G. and Wedi, N.P. (2019) Spectral empirical orthogonal function analysis of weather and climate data. *Monthly Weather Review*, 147(8), 2979–2995. <https://doi.org/10.1175/MWR-D-18-0337.1>.
- Seinfeld, J.H. and Pandis, S.N. (2016) *Atmospheric chemistry and physics: From air pollution to climate change*. Hoboken, NJ: Wiley.
- Smith, M.J. and Phillips, I.D. (2013) Winter daily precipitation variability over the East Anglian region of Great Britain and its relationship with river flow. *International Journal of Climatology*, 33(9), 2215–2231. <https://doi.org/10.1002/joc.3594>.
- St-Hilaire, A., Ouarda, T.B.M.J., Lachance, M., Bobée, B., Gaudet, J. and Gignac, C. (2003) Assessment of the impact of meteorological network density on the estimation of basin precipitation and runoff: A case study. *Hydrological Processes*, 17(18), 3561–3580. <https://doi.org/10.1002/hyp.1350>.
- Strang, G. (1988) *Linear algebra and its applications*, 3rd edition. San Diego, CA: Harcourt Brace Jovanovich.
- Tang, Y., Lean, H.W. and Bornemann, J. (2013) The benefits of the Met Office variable resolution NWP model for forecasting convection. *Meteorological Applications*, 20(4), 417–426. <https://doi.org/10.1002/met.1300>.
- Vicente-Serrano, S., Cuadrat-Prats, J. and Saz-Sánchez, M. (2005) Spatial patterns of the urban heat island in Zaragoza (Spain). *Climate Research*, 30, 61–69. <https://doi.org/10.3354/cr030061>.
- Wallace, M.J. and Hobbs, V.P. (2006) *Atmospheric science: An introductory survey*, 2nd edition. London: Academic Press.
- Wang, Z., Cao, X., Zhang, L., Notholt, J., Zhou, B., Liu, R. and Zhang, B. (2012) Lidar measurement of planetary boundary layer height and comparison with microwave profiling radiometer observation. *Atmospheric Measurement Techniques*, 5(8), 1965–1972. <https://doi.org/10.5194/amt-5-1965-2012>.
- Ward, H. (1963) Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, 58(301), 236–244.
- Warren, E., Charlton-Perez, C., Kotthaus, S., Lean, H., Ballard, S., Hopkin, E. and Grimmond, S. (2018) Evaluation of forward-modelled attenuated backscatter using an urban ceilometer network in London under clear-sky conditions. *Atmospheric Environment*, 191, 532–547. <https://doi.org/10.1016/j.atmosenv.2018.04.045>.
- Warren, E., Charlton-Perez, C., Kotthaus, S., Marengo, F., Ryder, C., Johnson, B., Green, D., Lean, H., Ballard, S. and Grimmond, S. (2020) Observed aerosol characteristics to improve forward-modelled attenuated backscatter in urban areas. *Atmospheric Environment*, 224, 117177. <https://doi.org/10.1016/j.atmosenv.2019.117177>.
- Welton, E.J., Campbell, J.R., Berkoff, T.A., Valencia, S., Spinhirne, J.D., Holben, B., Tsay, S., Goddard, N. and Flight, S. (2000) The NASA Micro-Pulse Lidar Network (MPLNET): Co-location of lidars with AERONET sunphotometers and related Earth science applications. Available at: <https://ams.confex.com/ams/pdfpapers/86086.pdf>.
- Wiegner, M. and Gasteiger, J. (2015) Correction of water vapor absorption for aerosol remote sensing with ceilometers. *Atmospheric Measurement Techniques*, 8(9), 3971–3984. <https://doi.org/10.5194/amt-8-3971-2015>.
- Wilks, D.S. (2011) *Statistical methods in the atmospheric sciences*, 3rd edition. Waltham, MA: Academic Press.
- Wilson, K.D. (1996) Empirical orthogonal function analysis of the weakly convective atmospheric boundary layer. Part I: Eddy structures. *Journal of the Atmospheric Sciences*, 53(6), 801–823. [https://doi.org/10.1175/1520-0469\(1996\)053<0801:EOFAOT>2.0.CO;2](https://doi.org/10.1175/1520-0469(1996)053<0801:EOFAOT>2.0.CO;2).
- You, W., Zang, Z., Zhang, L., Zhang, M., Pan, X. and Li, Y. (2016) A nonlinear model for estimating ground-level PM10 concentration in Xi'an using MODIS aerosol optical depth retrieval. *Atmospheric Research*, 168, 169–179. <https://doi.org/10.1016/j.atmosres.2015.09.008>.
- Yu, H.-L. and Lin, Y.-C. (2015) Analysis of space–time non-stationary patterns of rainfall–groundwater interactions by integrating empirical orthogonal function and cross wavelet transform methods. *Journal of Hydrology*, 525, 585–597. <https://doi.org/10.1016/j.jhydrol.2015.03.057>.

Zhang, Z. and Moore, J.C. (2015) Empirical orthogonal functions. In: Zhang, Z. and Moore, J.C.B.T.-M. (Eds.) *Mathematical and physical fundamentals of climate change*. Boston, MA: Elsevier, pp. 161–197. <https://doi.org/10.1016/B978-0-12-800066-3.00006-1>.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Warren, E., Charlton-Perez, C., Lean, H., Kotthaus, S. & Grimmond, S. (2022) Spatial variability of forward modelled attenuated backscatter in clear-sky conditions over a megacity: Implications for observation network design. *Quarterly Journal of the Royal Meteorological Society*, 148(744), 1168–1183. Available from: <https://doi.org/10.1002/qj.4253>