Environmental Science: Atmospheres

ROYAL SOCIETY OF CHEMISTRY

PAPER

View Article Online
View Journal | View Issue

Cite this: *Environ. Sci.: Atmos.*, 2022, **2**, 500

Measuring the impact of air quality related interventions†

Karl Ropkins, *[a] James E. Tate,[a] Anthony Walker [b] and Tony Clark[b]

As part of air quality management plans, administrative authorities commonly implement interventions, such as Low Emission Zones (LEZs) and Clean Air Zones (CAZs), to improve air quality. The associated benefits are often difficult to quantify due to the high variability in ambient time-series measurements and influence of contributions from meteorology, background and other emission sources. Break-point techniques have previously been used on their own to detect large changes, and in combination with deseasonalisation and deweathering methods to detect smaller changes. However, getting down to the detection limits needed to measure change at the levels predicted for most contemporary air quality interventions remains a challenge, as does the conversion of such higher-level analytical techniques into tools that are suitable for routine use by those tasked with the evaluation of interventions. Here, methods are presented that incorporate background subtraction to improve sensitivity and confidently quantify changes not readily detected in initial air quality time-series. Applied to air quality data collected in Leeds in the UK, the methods indicate a general reduction in the local NO_2 contribution across the studied period, 01 January 2015 to 31 January 2019, but also superimposed on that two discrete reductions: the first 2.4 μg m^{-3} (0.03 to −4.8 μg m^{-3}; 95% confidence) in late 2015, and a second of 3.6 μg m^{-3} (1.2–6.1 μg m^{-3}; 95% confidence), equivalent to a 12% (4% to 21%; 95% confidence) reduction in ambient air that coincides with the period when the local 2018 bus fleet was upgraded to cleaner Euro VI vehicles.

Received 16th September 2021
Accepted 4th April 2022

DOI: 10.1039/d1ea00073j

rsc.li/esatmospheres

Environmental significance

Break-point and change-segment detection methods can be used to independently detect and quantify change within time-series. However, the inherent variability in air quality data often hinders direct measurement of the impact of all but the largest change events. Here, deseasonalisation, deweathering and background subtraction are used to pre-process data, to improve sensitivity and detect change associated with an urban bus fleet upgrade not readily detected in ambient air. Such real-world evidence is much needed as part of efforts to evaluate the impacts of in-coming air quality initiatives, *e.g.* the Clean Air Zones in the UK, measure the impacts of discrete events, *e.g.* traffic network disruptions and forest fires, and to inform those developing next-generation environmental management policies.

1 Introduction

Despite decades of increasingly stringent vehicle fuel and emission control technology regulations, traffic-related air pollution (TRAP) is still a significant concern in urban areas in many countries.[1,2] Differences between laboratory type-approval testing and on-road vehicle performance, variable maintenance practices and deliberate tampering have all been cited as contributing to the limited environmental return on efforts to reduce TRAP.[3–5] In response, contemporary Governments are developing and implementing multi-part air quality improvement plans and strategies including a broad range of emission reduction and mitigation activities, many delivered at local scales by Regional and City Authorities. One example of this is the 2017 UK plan for tackling roadside NO_2 concentrations, which contains an overview of measures being undertaken to reduce NO_2 levels. The plan also defines the roles and responsibilities of stakeholders; from national government and devolved administrations, local authorities, businesses and manufacturers, to the Mayor of London and the public.[6,7]

Local air quality improvement activities include a wide range of interventions and actions to accelerate the renewal of vehicle fleets and reduction of on-road vehicle numbers and emission levels, including: vehicular access restrictions for vehicles deemed as excessive emitters *e.g.* in Low Emission, Ultra Low

[a]*Institute for Transport Studies, University of Leeds, Leeds, LS2 9JT, UK. E-mail: k.ropkins@its.leeds.ac.uk*

[b]*Joint Air Quality Unit, Department for Transport & Department for Environment, Food and Rural Affairs, Marsham Street, London, SW1P 4DF, UK*

† Electronic supplementary information (ESI) available: CRAN (stable) release version of associated R package at **https://CRAN.R-project.org/package=AQEval**; Project website and developer's code at **https://karlropkins.github.io/AQEval/**. See **https://doi.org/10.1039/d1ea00073j**.

Emission and Clean Air Zones (LEZs, ULEZs and CAZs), cleaner public transport services, alternative vehicle procurement and older vehicle retrofit and scrappage incentivisation schemes, infrastructure development (*e.g.* electric vehicle charging points and alternative fuelling stations), traffic flow management and calming activities, and the promotion of active travel.[8,9] Some large-scale interventions are reported to have demonstrable air quality benefits, *e.g.* reductions of 10–30% and 5–10% for PM_{10} and NO_2, respectively, have been reported following the introduction of some LEZs in Europe.[10–12] Integrated multi-action approaches have been reported to be even more effective, *e.g.* the combined air quality actions employed in China during the 2008 Beijing Olympic Games were reported to lower PM_{10} by 55% and NO_x by 47%, respectively.[13] The air quality impacts of most interventions are, however, more limited and less certain.[8,9,14]

Here, it is important to acknowledge both the complexity of such analyses, *e.g.* the inherent variability of air quality data,[15] the influence of meteorology,[16,17] the challenges of attributing impact to individual interventions that rarely occur in isolation,[18] and even the limitation of the analytical methods themselves.[14] It is also important to note that as air quality is improving (albeit slowly) in many countries with more progressive air quality management strategies,[15,19,20] any intervention-related improvements are becoming not just harder to earn but also more challenging to isolate and quantify. It is also perhaps all too easy to dismiss local interventions as local actions with only local relevance. They are key elements of city and regional plans being rolled out across many countries, and their effectiveness is critical to the delivery of national air quality strategies, internationally.[21–23] For example, UK Government is working with 61 local authorities across the country to tackle NO_2 exceedances, the allocated budget is £880 million and local interventions are central to all these activities.[7] Associated impact assessment is fundamental information, needed to support evidence-based policy making and those debating intervention performance, benchmarking, prioritisation and justification of investment.

Likewise, climate change is increasing the likelihood of atypical meteorological and environmental events, *e.g.* extreme weather events[24] and wildfires,[25] and with these the relative significance of related air pollution impacts. As a result, the identification, quantification and apportionment of air pollution associated with discrete changes is becoming an increasing important element of environmental research, and various statistical methods have been applied as part of this work.[26]

Break-point detection methods test for points in a series of observations that are better explained, with greater statistical significance, by an abrupt change within the monitored system rather than chance, noise or underlying trends.[27–30] They have been widely used in many commercial and research areas, including several air quality applications.[16,31–33] Various signal isolation methods have also been used as a data 'clean-up' step prior to air quality data analyses. Background subtraction or correction methods have perhaps been most widely used, and provide a measure of local contributions or 'increments'.[34–37] Classical trend deconvolution methods such as 'deseasonalisation' assume there are regular frequency cycles in time-series, *e.g.* hour-of-day, day-of-week, and week-of-year cycles, and that modelling and subtracting these frequency patterns from time-series provides a clearer measure of underlying trends.[38] Deweathering, sometimes also called 'weather normalisation' or 'meteorological detrending', extends this approach to the removal of variance associated with changes in meteorological conditions, such as wind speed and direction, air temperature and humidity.[16,39–41] Conditional extraction,[42] molecular tracers[43] and diagnostic ratios,[44,45] amongst other methods, have also all been used to isolate source-specific contributions. In the few cases where such signal isolation methods have been applied in combination with break-point methods, improved sensitivity[26] and/or easier trend visualisation[16] have been reported.

Methods extending the approach to break-segments, regions of change about a break-point, have been recently applied to the characterisation of major wide-scale air quality changes during the COVID-19 related UK lockdown.[46] There, NO, NO_2 and NO_x decreases of (on average) 32% to 50% were observed at roadside Automatic Urban and Rural Network (AURN) sites across the UK, and associated O_3 increases. That work also highlighted the extra insights to be gained if independent detection of the change-point is incorporated in such analysis, rather than explicitly assumed, as in conventional 'before and after' analyses. Here, extending on this work, a novel combination of local contribution isolation, employing a conservative implementation of deseasonalisation, deweathering and background subtraction, and break-point and change-segment analysis methods, is used to isolate, detect and then quantify change on smaller and more localised scales. The methods are applied to NO_2 air quality data from Leeds in the UK as part of an investigation of the potential environmental impact of a single traffic intervention. Leeds is one of the largest cities in the UK. It is 272 km NNW of London, has a population of 789 194 (third largest after London and Birmingham[47]), and the city centre, which is both heavily urbanised and densely trafficked, has been actively air quality managed by Leeds City Council (LCC) for over a decade. As part of local air quality actions, from Autumn 2016 the local bus operator introduced a rapid bus fleet overhaul, upgrading buses on selected routes to Euro VI. The investigation of the air quality impacts of interventions like this bus fleet upgrade, allows impact measurement methods to be developed and tested, and provides evidence on the potential effectiveness of bus-related interventions that is of relevance to both their use as standalone actions and as elements of larger air quality activities, *e.g.* the selection of vehicle-type-related strategies as part of LEZ and CAZ implementations.

2 Experimental

2.1. Data sources

Leeds City Council (LCC) provided 1 hour resolution NO_2 air quality monitoring data from three of their local monitoring stations (Headingley, Kirkstall Road and Temple Newsam) for the period 01 January 2015 to 31 January 2019 for this study. The locations of these and other monitoring sites used in this study are shown in Fig. 1, and further details are provided in Table 1.

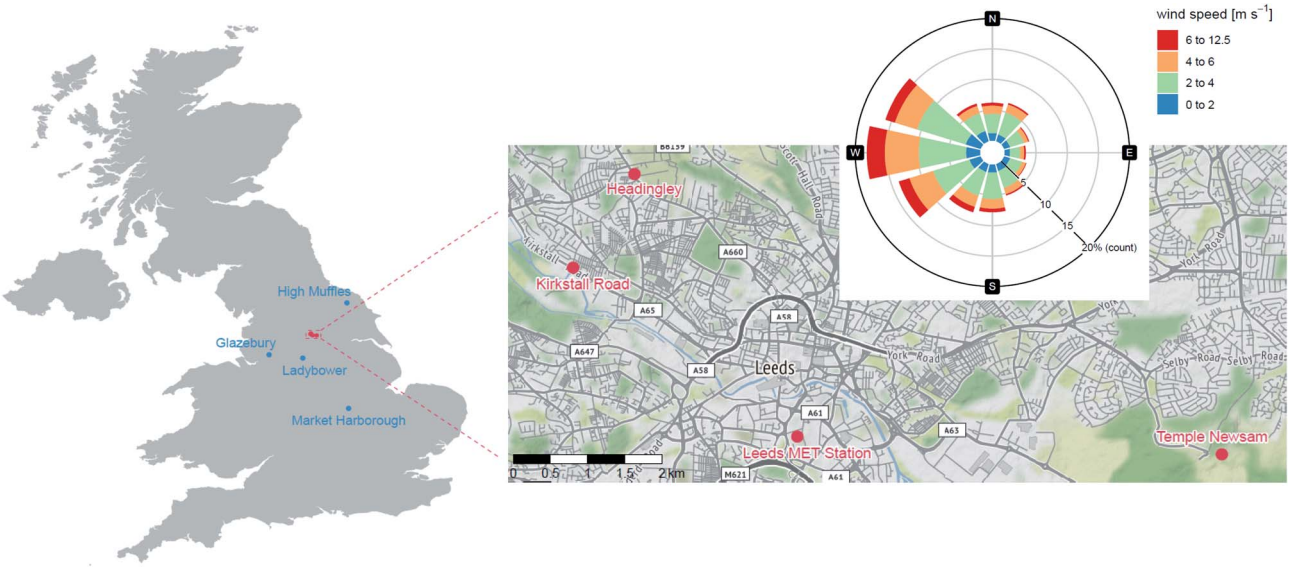

Fig. 1 Locations of monitoring stations: UK map (left) showing Automatic Urban and Rural Network (AURN) Rural Background sites used (in blue), and Leeds local map (right) showing Leeds City Council (LCC) monitoring stations (in red). Headingley is AURN affiliated, but shown in red here to indicate LCC as source of data used in this study. A wind rose is also included as an insert to indicate wind speeds and directions for the Leeds area during the study period, 01 January 2015 to 31 January 2019 (data source openair MET; see also Table 1) (map tiles produced by Stamen Design, under CC BY 3.0. Data under ODbL using R package OpenStreetMap[50]).

Headingley is operated as an affiliated site as part of Defra's national Automatic Urban and Rural Network (AURN; see also **https://uk-air.defra.gov.uk/networks/network-info?view=aurn**), while Kirkstall Road and Temple Newsam are operated by LCC as independent local monitoring stations. Headingley and Kirkstall Road are roadside sites on major roads (A660 and A65, respectively), both have similar road layouts, traffic junctions, signalised pedestrian crossings, and are subject to similar traffic flows/volumes. As the Euro VI bus fleet upgrade was only initially applied to buses on routes on the A660, the combination of Headingley (roadside, intervention) and Kirkstall Road (roadside, non-intervention) provides a classical 'test and control' combination. They exhibit similar air quality characteristics, as indicated by their NO_2 polar plots (Fig. 2), which are both dominated by similar SE (S to ESE) features: main maxima, 40–55 and 40–45 μg m^{-3}, respectively, at *ca.* 5 m s^{-1} suggesting highest contributions from nearby sources in that direction and secondary maxima at *ca.* 0 m s^{-1}

Table 1 Summary of data 1 hour resolution time-series and related monitoring station locations

Site	Site type[a]	Data type[b]	Data source[c]	Latitude	Longitude	Altitude (m)	Data capture (%)
Headingley[d]	Kerbside	CL NO_2	LCC	53.82035	−1.57669	85	99
Kirkstall road	Roadside	CL NO_2	LCC	53.80873	−1.58929	34	90
Temple Newsam	Background	CL NO_2	LCC	53.78557	−1.45607	67	86
Glazebury	Rural background	CL NO_2	AURN	53.46008	−2.47206	21	90
High Muffles	Rural background	CL NO_2	AURN	54.33494	−0.80855	267	90
Ladybower[e]	Rural background	CL NO_2	AURN	53.40337	−1.75201	420	95
Market Harborough	Rural background	CL NO_2	AURN	52.55444	−0.77222	145	95
Leeds MET[f]	Meteorological	WS, WD	LCC	53.78634	−1.54166	30	97
Leeds Bradford MET	Meteorological	WS, WD	NOAA	53.866	−1.66100	208	93
Openair MET[g]	Meteorological	WS, WD	WRF	—	—	—	98
Headingley A660 traffic	Traffic counter	Vehicle flow	LCC	53.80539	−1.57843	31	94
Kirkstall road A65 traffic	Traffic counter	Vehicle flow	LCC	53.81230	−1.55824	87	80

[a] Headingley and AURN site type assignments by Defra classification scheme, Kirkstall Road and Temple Newsam assignments by Local Authority. [b] CL NO_2 nitrogen dioxide measured using chemiluminescence analyser, WD and WS wind direction and speed, both measured by anemometer or model thereof. [c] LCC Leeds City Council, AURN Automatic Urban and Rural Network, NOAA National Oceanic and Atmospheric Administration Integrated Surface Database, WRF Ricardo Weather Research And Forecasting meteorological model. [d] Although Headingley data was obtained from LCC for this study, the station is AURN affiliated so data is also in the AURN archives. [e] All time-series were for full study period except Ladybower which was non-operational in 2015 (see also Table 2). [f] Although Leeds MET data capture was relatively high, some reported values were questionable, so openair MET data was used for main analysis reported here. [g] Downloaded from AURN archive using openair R Package, data 10 × 10 km resolution model output so area-specific rather than site-specific.

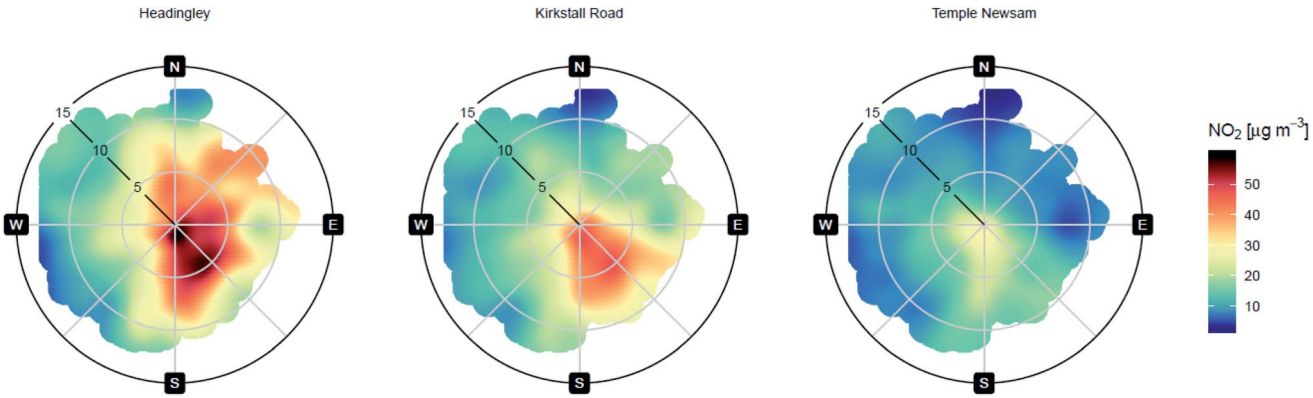

Fig. 2 Polar plots of NO_2 at Headingley, Kirkstall Road and Temple Newsam for the study period, 01 January 2015 to 31 January 2019. All plots generated using WRF meteorological data from Headingley AURN 'openair' dataset, see also Table 1 and discussion in Section 2.1.

consistent with local NO_2 accumulation during periods of air stagnation. The third site, Temple Newsam, is on rural land to the SE of Leeds, where NO_2 concentrations are typically much lower (compare *e.g.* polar plots in Fig. 2). Prevailing winds (indicated by the wind rose for the full study period, inset to Fig. 1) show that both roadside sites were often up-wind of Temple Newsam during the study period, making this a less than ideal candidate for use as an associated background. However, as Stedman and colleagues observe, ideal background sites are exceptional rare,[34] but with respect to both Headingley and Kirkstall Road, Temple Newsam meets the key criteria they recommend for a viable background, most importantly that is it is near, all evidence indicates that it is relatively unexposed to local pollution, and foreground/background concentration ratios are typically >1 (see *e.g.* Fig. 3).

In addition, 1 hour resolution NO_2 data was obtained for the nearest AURN rural background sites for the same period: Glazebury, High Muffles, Ladybower and Market Harborough. These were selected as the nearest sites (all within 200 km of Headingley) that were classified as 'Rural Backgrounds' and had NO_2 data for the study time period, and accessed using R package 'openair'.[48,49] Data captures were lower at the non-AURN-affiliated background site Temple Newsam, as it is arguably understandable that its upkeep might be of lower priority.

Meteorological data from three sources was used in this study: the Leeds Meteorological Station (Pottery Fields House, Kidacre Street), the Leeds Bradford Airport Meteorological Station, the nearest site submitting data to the NOAA Integrated Surface Database (**https://www.ncdc.noaa.gov/isd**), accessed using R package 'worldmet',[50] and modelled data generated by the Ricardo WRF model (**https://ee.ricardo.com/air-quality**) and supplied with AURN data when downloaded using 'openair' function importAURN. Data from the three sources were typically similar for near locations (*e.g.* $r \approx 0.77$–0.88 for wind speed). Here, Ricardo WRF meteorological data was selected for use in the main study for three reasons: (1) where the largest

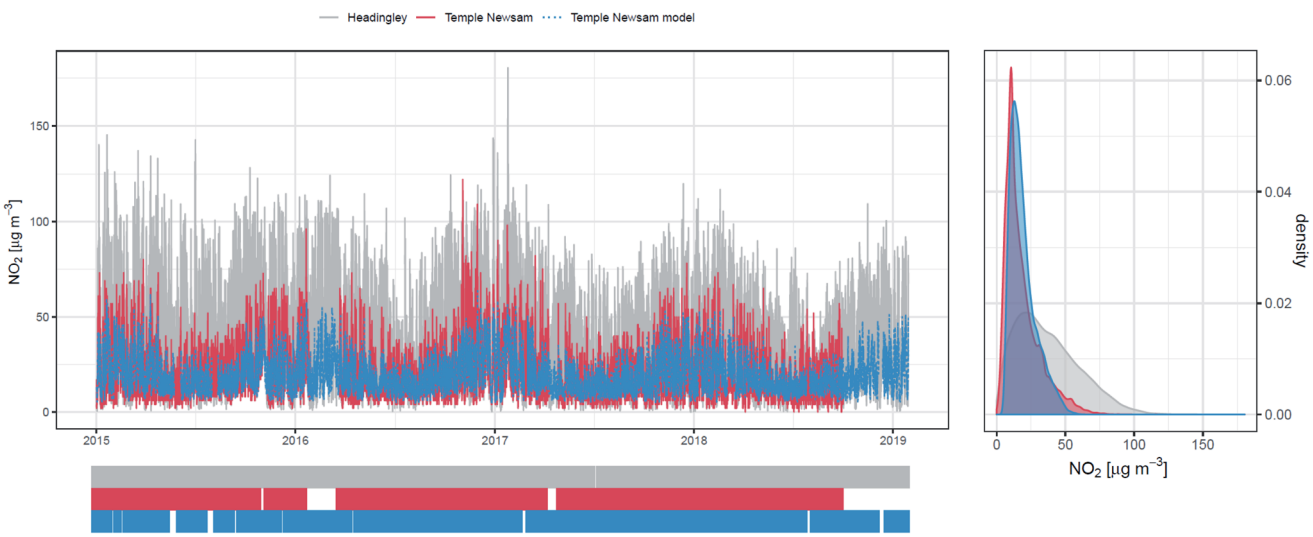

Fig. 3 NO_2 time-series, density plot (right) and rug plot (below) for the original Temple Newsam monitoring data-series (red) and the Temple Newsam model data-series (blue dashed) for the study period. Data from Headingley is also included (grey) to demonstrate the classic background/foreground relationship between data from Temple Newsam and Headingley.

differences were observed between the datasets, the Leeds Meteorological data was often most suspect (*e.g.* highly noisy or very different from neighbouring measurements). (2) The Ricardo WRF model is implemented at about 10 km^2 grid resolution, so meteorological data from the nearest AURN site could be assigned to any site in the Leeds area without dedicated meteorological time-series of its own, making the method readily transferrable to nearby non-AURN sites. And, (3) although break-point/segment analyses using either Ricardo WRF and NOAA meteorological data provided highly similar results, data capture levels were highest for Ricardo WRF meteorological data, an important consideration when using both signal isolation and change detection methods (see also Section 2.3 regarding Background Data handling and ESI† regarding selection of data and methods used for signal isolation). Meteorological data selection should, however, be considered on a case-by-case basis, because, *e.g.*, not all AURN datasets include as extensive WRF time-series as Leeds Headingley and at other sites, NOAA or other local meteorological data sources may provide a better proxy.

Near continuous 15 minutes averaged data from Automatic Traffic Count (ATC) sites (two inductive loop per lane configuration) located on the Headingley Lane (A660) and Kirkstall Road (A65) arterials was accessed *via* the Drakewell C2-Cloud interface (**https://www.drakewell.com/c2-web**). The dates of the phases of the Euro VI bus fleet renewal were provided by LCC and the public transport operator First Bus (**https://www.firstgroup.com/leeds**). The month-by-month share of Euro V and VI double-decker Buses operating on the two arterials were provided by LCC and the public transport operator First Bus (**https://www.firstgroup.com/leeds**) on the 2nd May 2020.

2.2. Break-point detection and change-segment estimation

Change detection methods based on those of Zeileis *et al.*[28] for use in non-static systems, and Bai and Perron[27] for detecting multiple changes, as implemented in R package 'strucchange',[51] were used in this study. Here, a rolling window strategy is applied to test for changes in the linear regression properties of the investigated data-series. The associated hypothesis is that a change exists wherever the surrounding data is better explained by two discrete models rather than one general model, and significant changes, called break-points in the work of Zeileis and colleagues, were assigned on the basis of statistical significance. So, for example, for one change point in the types of data considered here eqn (1):

$$\begin{aligned}
&\text{Initial model:} && [NO_2]_{t=1:TW} = lm_1(trend_{t=1:TW})\\
&\text{Second model:} && [NO_2]_{t=2:TW+1} = lm_2(trend_{t=2:TW+1});\\
& && H_0\ \text{(no change)}:\ lm_2 = lm_1\\
&\text{Third model:} && [NO_2]_{t=3:TW+2} = lm_2(trend_{t=3:TW+2});\\
& && H_0(\text{no change}):\ lm_3 = lm_2\\
& && \text{And so on...}
\end{aligned} \quad (1)$$

where $[NO_2]$ is the concentration of NO_2, the species of interest, TW is the test window size, and $[NO_2]_{t=1:TW}$ is the first test window, or subset, TW width starting with the first NO_2 measurement, and subsequent test windows and models follow a rolling window sequence: $[NO_2]_{t=2:TW+1}$, $[NO_2]_{t=2:TW+1}$, *etc.* lm() represents a conventional linear model, so *e.g.* $[NO_2]_{t=1:TW}$ = slope × $[date/time]_{t=1:TW}$ + intercept, and H_0 represents the NULL hypothesis test which is applied to all sequential rolling window comparisons for the full time-series, T_{total}.

For all work reported here, the test window size, TW, was set to 10% of full data range, T_{total}, and F-Stat scores were used to test significance. Although the number of break-points searched was not limited during this step, window size and assignment strategy restricts the maximum number that can be detected to approximately (T_{total}/TW) − 2, or *ca.* 8 in this case.

All potential break-points identified based on F-Stat scores were tested and candidate break-points selected or discarded on the basis of Bayesian Information Criterion (BIC) in accordance with standard strucchange methods[20] and subsequent independent testing of BIC-selected break-points in the form eqn (2):

$$\begin{aligned}
[NO_2]_{t=1:T_{total}} = lm(&trend_{t=1:BP1} + trend_{t=BP1:BP2} + \ldots\\
&+ trend_{t=BPn:T_{total}})
\end{aligned} \quad (2)$$

where BP1, BP2, *etc*, are the identified breakpoints and model testing was on the basis of p-test significance.

Independent testing of BIC-selected break-points was applied in a stepwise-fashion. The initial model was accepted if it was statistically valid (or more specifically if all terms associated with individual break-points were all statistically significant, at $p < 0.05$). If not, all model combinations discarding one of the initial break-points were built, tested and compared, and the statistically valid model with the highest correlation (all break-points individually statistically significant, $p < 0.05$ and highest R) was accepted, or the process was repeated until a statistically valid model was obtained or all break-points were rejected. This step is included as an additional test of BIC-selected break-points, rather than an alternative to BIC based selected. Additional tests were investigated in light of concerns raised about BIC by the stucchange's authors,[28] and the current additional method was selected on the basis of performance in simulation testing. That said, at this stage, this is presented as an empirical solution and arguably more work may yet be required on break-point selection.

Elsewhere[31] it has been observed that such approaches test for instantaneous change, and that real environmental changes are more likely to happen gradually. For example, local residents and businesses would be expected to start purchasing compliant vehicles ahead of a LEZ/CAZ start date, and for purchases to increase as that deadline approached. To investigate evidence of more gradual change, break-point models were extended to test for more gradual changes using methods reported by Muggeo.[53–55] Here, change-segments (regions of change) were allowed around identified break-points. The break-point confidence intervals, calculated using the methods of Bai,[56] were applied as initial estimates of segment start and end points and then refined by iteratively testing neighbouring start and end points to provide estimates of break/change-segment time range and magnitude.

2.3. Background data handling

The value of background correction has been highlighted in multiple studies.[34,36] However, such approaches typically require paired data, and only generate valid local estimates of local increments for time intervals with valid measurements for both study site (or foreground) and background. As missing records in different data-series rarely overlap, corrected data losses tends to approach the sum of the losses from the individual time-series. So, data-series like Temple Newsam, which has a relatively low capture rate (86% compare with ≥90% at all other sites, Table 1), can significantly reduce time-series coverage and, therefore, representability of results. In addition, much of the missing data at Temple Newsam associated with long periods of continuous instrument down-time (see *e.g.* Fig. 3, Temple Newsam time-series, early 2016 and late 2018/early 2019), and wide gaps in data-series like these obviously hinder break-point detection and quantification in and about these time periods. So, a background model was used to improve coverage.

Here, a series of Generalised Additive Models (GAMs) were built for Temple Newsam in using R package 'mgcv'[57,58] and the AURN rural background data, starting with eqn (3):

$$[NO_{2,TN}] = s_1([NO_{2,BG1}]) + s_2([NO_{2,BG2}]) + s_3([NO_{2,BG3}]) + s_4([NO_{2,BG4}]) \quad (3)$$

where $[NO_{2,TN}]$ is the NO_2 time-series from Temple Newsam, $[NO_{2,BG1}]$ to $[NO_{2,BG4}]$ are the time-series from AURN sites, Glazebury, High Muffles, Ladybower and Market Harborough, and $s_1()$ to $s_4()$ are a GAM smoothing terms, (thin film) splines functions, used to fit inputs in the GAM model.

Model predictions were then generated for all valid cases (*i.e.* time intervals with measurements for all four AURN rural backgrounds). This process was then repeated for the best fitting three-input model and any new predictions added to the prediction time-series. This process was then repeated again with the next best fitting three input model and so on until predictions were generated for all cases with three valid inputs in the AURN datasets.

Fig. 3 and 4 provide time-series and polar plot comparisons of the original Temple Newsam monitoring data and model outputs. By comparison to the original Temple Newsam data, the model typically produces a smoothed estimate of local concentrations (lower maxima, high minima, correlation coefficient, r, 0.76). As work here focused on the detection of changes, the modelled outputs, rather than hole-filled original data were used in the subsequent analysis in case data source switching between measured and modelled data introduced artefact changes. As a result, small gaps can be seen in model outputs, *e.g.* mid and late 2015, indicating periods where NO_2 data was missing from 2/4 of the AURN time-series. This is considered an acceptable trade-off for the much larger data

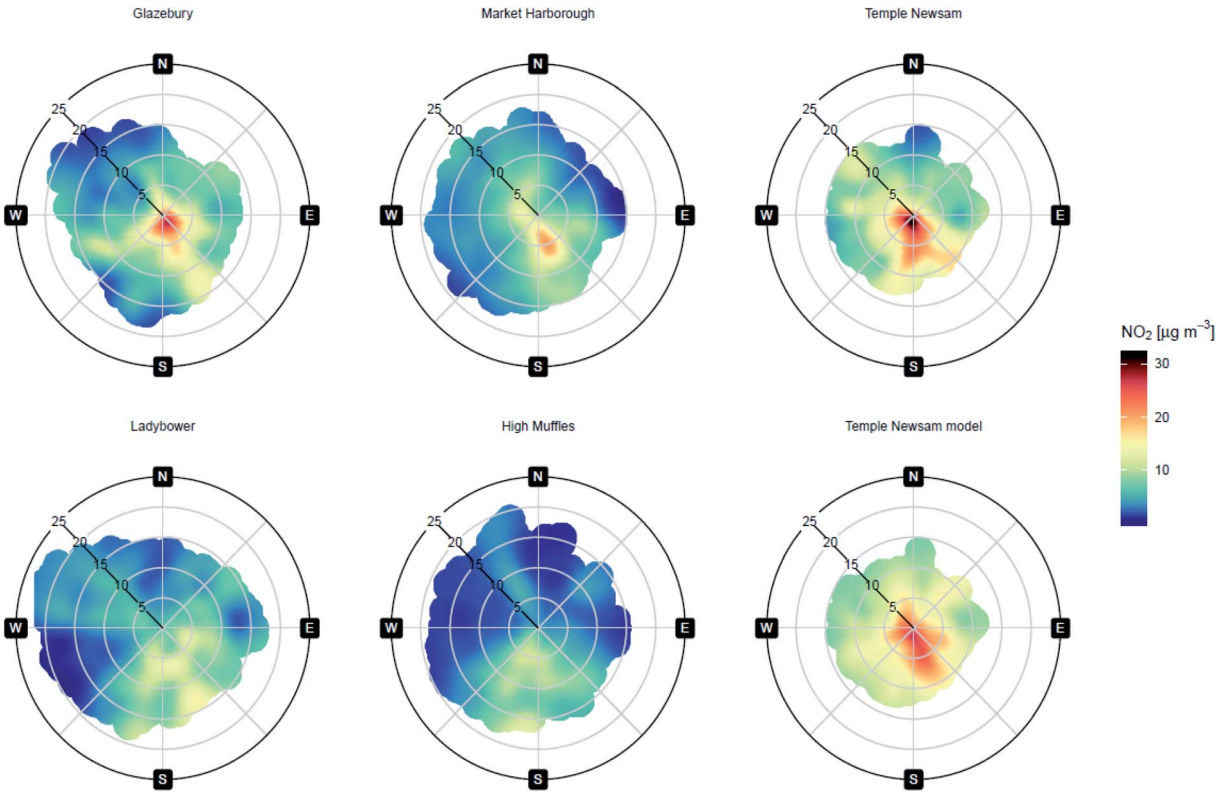

Fig. 4 NO_2 polar plots for original Temple Newsam monitoring data-series and the Temple Newsam model (right top and bottom, respectively) and the four AURN Rural Background monitoring sites used to build the Temple Newsam model. All plots generated using WRF meteorological data: for AURN sites own data; for Temple Newsam from Headingley AURN, all *via* 'openair', see also Table 1 and discussion in Section 2.1.

coverage gains, *e.g.* in early 2016 and late 2018/early 2019. One additional interesting feature worth noting here is that while the Temple Newsam model trends are most often similar to monitoring data trends, the model output does not include a high NO_2 period seen in Temple Newsam data in late 2016. As similarly elevated levels of NO_2 were not also seen at Headingley during this period, nor nearby AURN Rural Background sites either, it seems likely it is from a local pollution source near Temple Newsam rather than background levels more generally. The modelling process removed the high background NO_2 feature in late 2016. It is therefore proposed that this type of nearest-neighbour background modelling could also provide an option in cases where local background data was suspect or not available, and work is on-going to investigate this.

2.4. Local contribution isolation

Elsewhere, deseasonalisation, deweathering and background correction have all been used extensively in different combinations to isolate local contributions, although rarely as a prelude to change detection. Here, other contributions were initially predicted using a combined model (eqn (4)) and the normalised local contribution extracted as the mean-centred residuals (eqn (5)):

$$[NO_2] = s_1([NO_{2,TN}]) + s_2(wd, ws) + s_3(hd) + s_4(jd) \quad (4)$$

$$[NO_{2,local}] = ([NO_2] - [NO_{2,mod}]) + mean([NO_2]) \quad (5)$$

where $[NO_2]$ is the NO_2 concentration at the foreground (study) site, $[NO_{2,TN}]$ is the measure of NO_2 concentration at the background site, in this case the Temple Newsam model output, wd and ws are wind direction and speed, respectively, hd is hour-of-day, jd is day-of-year (or Julian date), and $s_1()$ to $s_4()$ are 'mgcv' GAM smoothing terms used to fit inputs to the GAM model of non-local contributions (eqn (4)), (thin film) splines for single input functions and (cubic regression) tensor for the two-input wd/ws term. $[NO_{2,mod}]$ is the prediction of $[NO_2]$ from the eqn (4) model, mean($[NO_2]$) is the mean local NO_2 concentration, and $[NO_{2,local}]$ is the estimated local NO_2 contribution calculated in eqn (5).

Here, non-local contributions are modelled in a single step, rather than sequential steps to minimise compounding of modelling errors, and a minimal-input design was adopted both as part of efforts focused on making these methods more widely accessible, and as a response to trends observed during simulation testing (see also ESI†).

In addition to GAMs, a number of other modelling strategies have been used in similar work, *e.g.* Boosted Regression Trees (BRTs)[36,58] and Random Forests (RFs).[16] One of the advantages commonly cited for many of these methods, GAMs, BRTs and RFs included, is that they are better able to handle nonlinearity and interactions than classical linear modelling approaches, and are not subject to collinearity. While this is true, it is important to recognise that the use of non-linear modelling does not prevent problems, rather it moves the potential modelling issue from collinearity to something more complex. In the case of GAMs, that fit curves, the issue becomes concurvity. Here, one of main reasons for using GAMs was that the R package 'mgcv' includes methods to test models for concurvity.[59] This potential issue is not widely acknowledged in the literature, and it is unclear how analogous issues can be confidently tested for when using modelling approaches that adopt more complex input response models.

The first GAM component of the model, $s_1([NO_{2,TN}])$, is the background contribution, in this case the Temple Newsam model output. The use of a GAM, rather than a direct subtraction as would be applied in *e.g.* local increment calculations, means that this is a non-linear estimate of background contribution at the study site rather than an explicit assumption that $[NO_{2,TN}]$ is an absolute measure of the background. Elsewhere, foreground NO_2 has been modelled as a function of foreground NO_x, and background ozone and NO_2 to provide a more sophisticated model of background and ambient NO_2 chemistry.[32] Here, the simpler background descriptor was adopted because it was more readily applicable and because exclusion of other inputs did not appear to significantly affect subsequent findings. It is, however, important to note that the focus in the current work was on break-point/segment performance rather than the statistical significance of the isolation model. The second GAM component, s_2(wd, ws), is the meteorological contribution. Here, wind speed and direction are modelled as a two-input (or surface) model to reflect the more complex wind speed and direction interactions. Elsewhere more meteorological descriptors have been included in such models, including modelling terms for measures such as air temperature, humidity and pressure,[17] but again, here, the more simplified description was adopted as a demonstration of what can be done even in cases where data is limited. The final two GAM components, s_3(dh) and s_4(jd), are included as seasonal terms. Here, these are not included as direct measures of any dependencies but rather as surrogates for contributions that have cyclic frequencies and are not captured by other model inputs. Comparisons of this and other models are provided in the ESI.†

2.5. Data resolution

When accounting for data variance, the significant diagnostic benefits of working with 1 hour data have been demonstrated in a number of studies.[36] So, here, all background modelling and local contribution isolation steps were carried out using data at the supplied 1 hour resolution. By comparison, testing suggests that when looking for break-points in several years of data, 1 hour resolution data is a significant additional computational burden that provides little benefit by comparison to *e.g.* 1 day resolution data. So, here, background modelling and local contribution isolation were carried out using the 1 hour resolution data and then time averaged to 1 day resolution for break-point testing and follow-on work as a practical compromise.

3 Results and discussion

The initial analysis of NO_2 data from Headingley, Kirkstall Road and Temple Newsam all indicated a likely downward trend in the measured air quality at all three sites (Table 2 and Fig. 5).

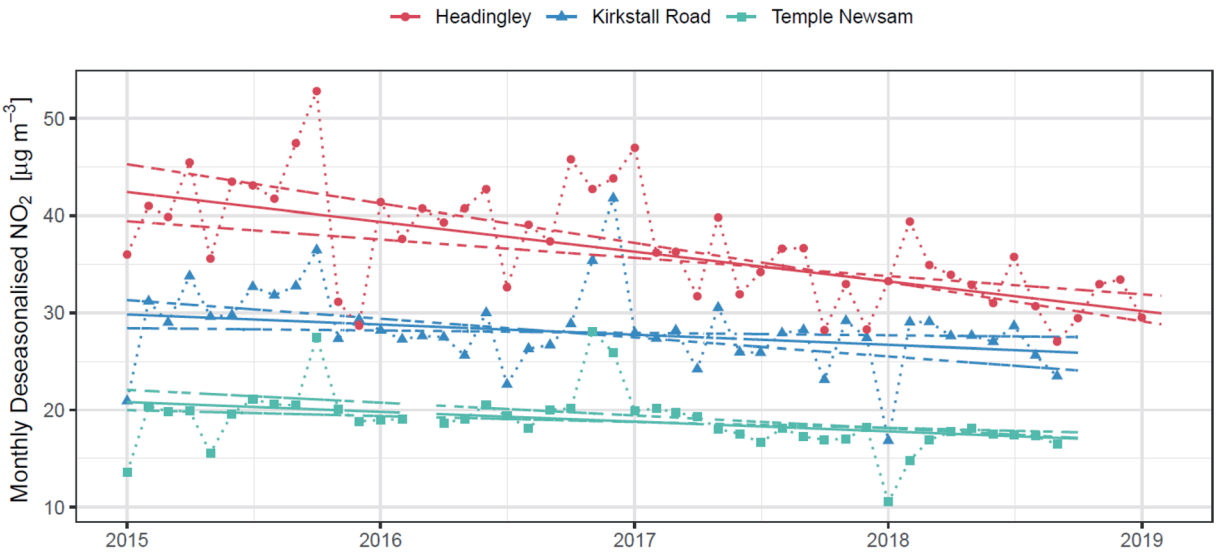

Fig. 5 General ambient NO_2 trends at Headingley (red lines and circles), Kirkstall Road (blue lines and triangles) and Temple Newsam (green lines and squares), estimated using deseasonalised month-average measurement and Theil-Sen method in openair[49] [all trends >95 confidence, $p < 0.05$, see also Table 2 and in ESI†].

Mean annual NO_2 typically decreases broadly year-on-year at all three sites. Relatively high mean NO_2 levels reported at all sites in 2019 reflect the early end of time-series used in this analysis, 31 January 2019, and should more strictly be regarded as one-month winter means. Overall trends measured using 'openair' R package Theil-Sen methods[49] were −3.06, −1.04 and −1.01 µg per m^3 per year (all statistically significant, all $p < 0.05$) at Headingley, Kirkstall Road and Temple Newsam, respectively. Theil-Sen uses the median of the slopes of all (different time) point-pairs in a data-series to provide a highly robust estimate of the underlying slope. Although arguably also possibly improving, trends at the four AURN Rural Backgrounds about Leeds were less distinct, suggesting that although NO_2 levels may be generally decreasing, benefits seen in Leeds, being more pronounced than benefits in the surroundings, are most likely benefits earned locally, *e.g.* from improvement in the local fleet or local air quality management activities in and about Leeds. This analysis, however, lacks the resolution to answer perhaps the most interesting question these findings prompt: is the higher of NO_2 reduction seen at Headingley the result of consistently higher rates of reduction at Headingley throughout the study period, perhaps reflecting a higher turn-over rate for the local vehicle fleet in the Headingley area, or are more abrupt step-changes, such as would be associated with an effective vehicle fleet intervention, contributing?

3.1. Standalone break-point testing

At the ambient NO_2 levels observed in Headingley, Kirkstall and Temple Newsam during the study (annual mean *ca.* 10–35 µg m^{-3}), seasonal trends are the dominant source of change, and in all three cases break-point methods identify highly similar and highly regular series of break-points when applied directly to ambient measurement data (Fig. 6). Here, the magnitude of changes associated with the observed seasonal break-points are of the order of 10 µg m^{-3} and above, and could easily obscure any non-seasonal break-point that were not of at least a similar magnitude. Others have previously discussed this limitation as part of the case for pre-treatment[16,32] but not attempted to characterise it further. Although not exhaustive, early direct testing of the break-point methods described here using 'raw' ambient concentration historical data (not the isolated local contribution) suggests a detection limit of about 35–50% of the local mean NO_2 level at the time of the intervention. The detection limit, which is at this stage an early estimate, is reported on a percentage basis rather than an absolute basis because sink as well as source elements cannot be ruled out for *e.g.* weather-related effects. Applying this standalone limit to Headingley NO_2 in 2018 indicates that an intervention would need to have delivered a change of at least 10–14 µg m^{-3} (29.55 × 0.35–0.5) to be detectable using break-point methods directly on ambient NO_2 data at the time.

3.2. Local contribution break-point/segment testing

Using the Temple Newsam model as a common background for both and applying the combined background, weather and seasonal corrections described in Section 2.4 to the ambient NO_2 data from Headingley and Kirkstall Road ($r = 0.80$ and 0.81, respectively) produces the normalised local contribution time-series shown in Fig. 7. Model input contributions for this step are shown as partial contribution plots in Fig. 8, and indicate that the background and combined wind speed/direction terms are the major contributors. As would be expected, by comparison to the original time-series (shown in Fig. 6, top and middle), the local contribution time-series are smoother, having had the dominant confounding variance removed. At this stage, linear regression of the full time-series indicates that average reductions are similar to those observed using Theil-Sen methods (Table 2) and deseasonalised

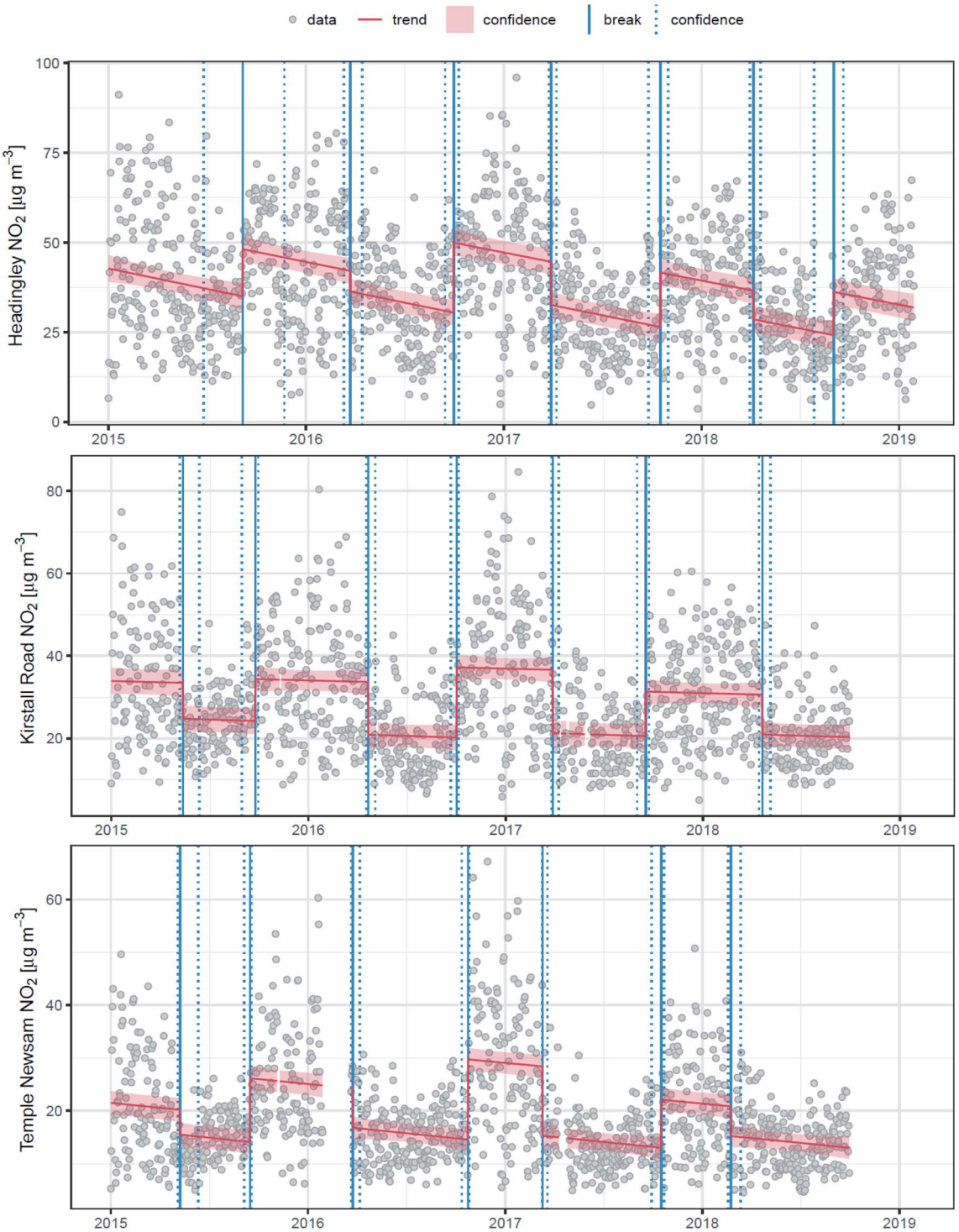

Fig. 6 Break-point detection in ambient NO_2 1 day resolution time-series for Headingley, Kirkstall Road and Temple Newsam (without local contribution isolation); data (grey), break-points, shown as trend breaks (blue) and associated confidence intervals (dashed blue) and broken trend (red).

data: −3.05 and −1.10 µg per m^3 per year for Headingley and Kirkstall Road, respectively.

Break-point tests identified two break-points in the local NO_2 at Headingley, the first on 30 October 2015 and the second on 07 April 2018. Change-segment modelling provided estimates of associated change periods of 25 September to 22 December 2015 (88 days) and 16 March to 27 April 2018 (42 days), respectively. The 2015 change was least distinct: −2.4 µg m^{-3} (0.03 to −4.8 µg m^{-3}; 95% confidence), equivalent to a change of −6.8% (−2.3 to 11%; 95% confidence) in ambient NO_2 at that

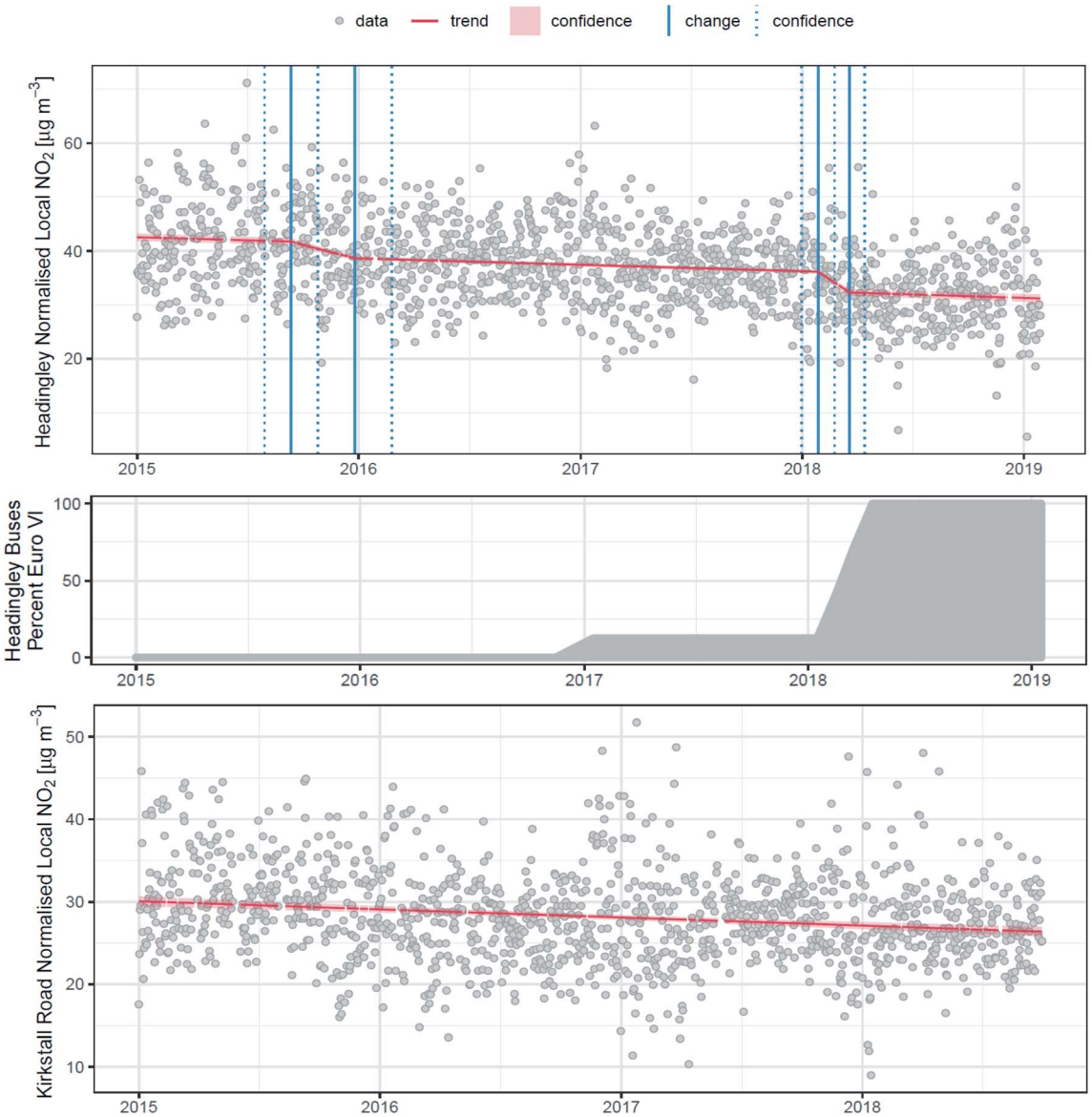

Fig. 7 Break-point detection and change-segment analysis of normalised local contributions for Headingley and Kirkstall Road time-series shown in Fig. 6 (top and bottom); data (grey), change-segments, with start and ends marked (blue) and associated confidence intervals (blue dashed) and segmented trends (red). The middle plot shows the percentage of Euro VI buses in the Euro V and VI fleet on routes through Headingley. By, comparison the local fleet on Kirkstall Road were 100% Euro V over the same timescales.

time. The second change was, by contrast, larger, more rapid and more confidently measured: −3.6 µg m^{-3} (−1.2 to −6.1; 95% confidence), equivalent to a change of −12% (−4 to −21%; 95% confidence) in ambient NO_2 at that time. By comparison, no break-points/segments were identified for local NO_2 at Kirkstall Road, which maintained a 100% older Euro V Bus fleet through this period. In the periods before, between and after the two changes, the gradients in local NO_2 at Headingley were all close to that observed at Kirkstall Road (−1.3, −1.5 and −1.3 µg per m^3 per year), supporting the interpretation that general trends at Headingley were the result of two discrete changes, one in late 2015 and the other early 2018, superimposed on a more general decrease, that was highly similar to that seen at Kirkstall Road.

3.3. Assignment of bus intervention-related change

Whilst it was known that the bus fleet was renewed on the A660 corridor in the 2017 to 2018 period when the analysis was run, the exact dates and proportions of Euro V and VI vehicles were only confirmed in May 2020, thereby making this a quasi-blind experiment.

As both the Headingley and Kirkstall Road sites are on urban arterial roads where flow capacities were unchanged throughout the study period and subject to fixed-time signalisation strategies that were also unchanged, overall traffic flows would be expected to be relatively constant. However, when quantifying the air quality impact of a traffic fleet upgrade, it is obviously important to confirm this because a change in local

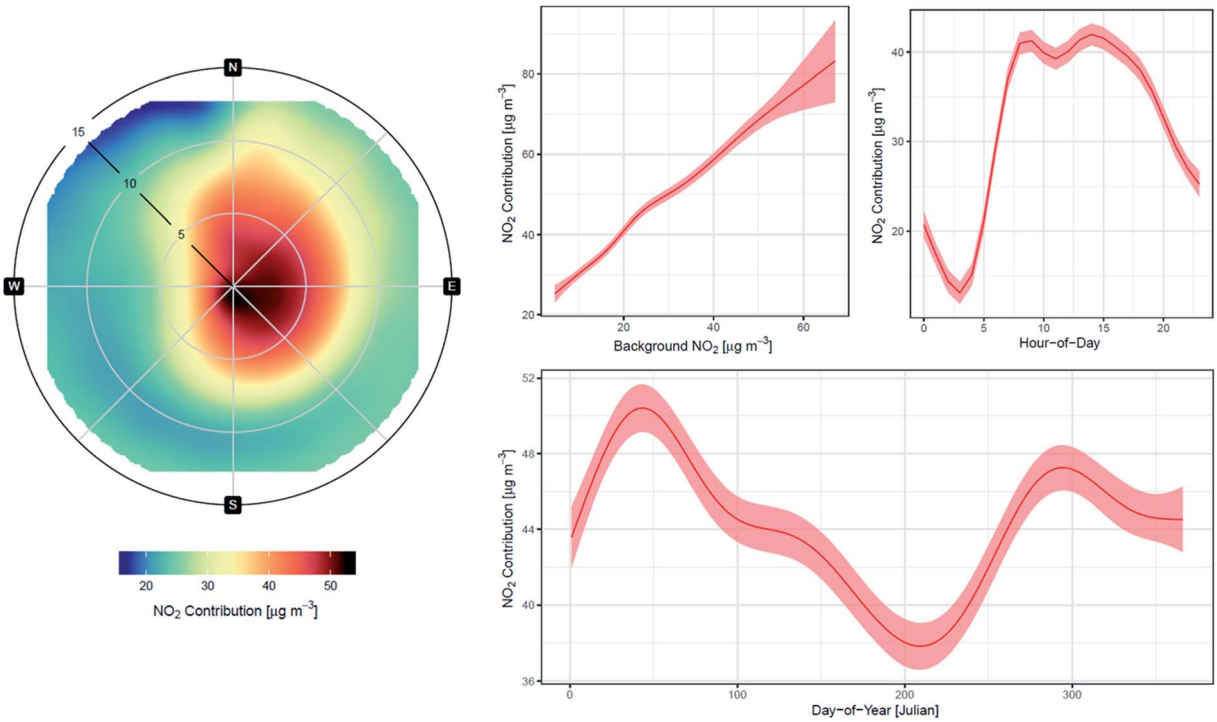

Fig. 8 Partial contribution plots for local contribution isolation (eqn (4) and (5)): left combined wind speed/direction term; middle top background NO_2 term; right top hour-of-day term; and, right bottom day-of-year term.

traffic flows, especially one that associates with the introduction of the upgrade, could make an indirect contribution to observed changes. Here, Theil-Sen analysis suggests small changes in traffic flows at both the Headingley and Kirkstall sites, 7.8 and −6.09 vehicles per hour per year, respectively, but neither were statistically significant ($p > 0.05$; Fig. 9 top), and prediction ranges are much larger (−9.13 to 23.55 and −18.01 to 4.12, respectively), indicating that overall traffic flows did not change significantly at either site during the timescales of this study. Similarly, break-point testing of the deseasonalised and deweathered traffic flow data identified no statistically significant break-points/segments in traffic flows at either site (Fig. 9 middle and bottom). These findings show that there were neither gradual nor abrupt changes in traffic flows at or about the times of the investigated bus fleet upgrade, supporting the assumption that changes seen at the time of the bus upgrade were due to the upgrade itself.

Although month-by-month data on the share of Euro V and VI double-decker Buses operating on the two arterials was not of sufficient resolution for similar analysis, the percent Euro VI time profile for A660 Headingley bus routes included in Fig. 7 shows bus fleet composition changes during the study period. As part of a programme of on-going fleet improvement, the first batch of Euro VI Buses, 6 of 48, were introduced in December 2016, and a further 42 Euro VI Buses, taking the local fleet on the A660 to 100% Euro VI (all Wrightbus Streetdecks with Daimler OM934 diesel engines), were introduced January to April 2018. The first observation is that the second change

Table 2 General NO_2 concentration trends for the study period

	Mean NO_2 concentration (µg m^{-3})					
Site	2015	2016	2017	2018	2019[a]	Trend[b,c] (µg per m^3 per year)
Headingley	36.47	37.02	30.79	29.55	32.51	−3.06 (−1.88 to 4.04)***
Kirkstall road	27.0	23.0	23.0	21.0	—	−1.04 (−0.24 to −1.93)*
Temple Newsam	15.0	17.0	13.0	13.0	—	−1.01 (−0.61 to −1.32)**
Glazebury	11.14	11.21	9.01	10.19	21.54	−0.85 (−0.29 to −1.34)**
High Muffles	2.94	3.12	2.77	2.93	3.63	−0.09 (−0.33 to 0.16)
Ladybower	—	5.02	4.67	4.15	6.63	−0.44 (−1.11 to 0.04)
Market Harborough	6.79	7.25	6.58	5.72	10.73	−0.51 (−0.8 to −0.18)**

[a] Studied data range was 01 January 2015 to 31 January 2019, so mean for 2019 was for January only. [b] Trend was calculated using Theil-Sen method in openair with deseasonalisation[49] [***$p < 0.001$, **$p < 0.01$, *$p < 0.05$]. [c] Removing the 2019 data has but non-significant effect on these trends (about ± 15%) well within confidence intervals.

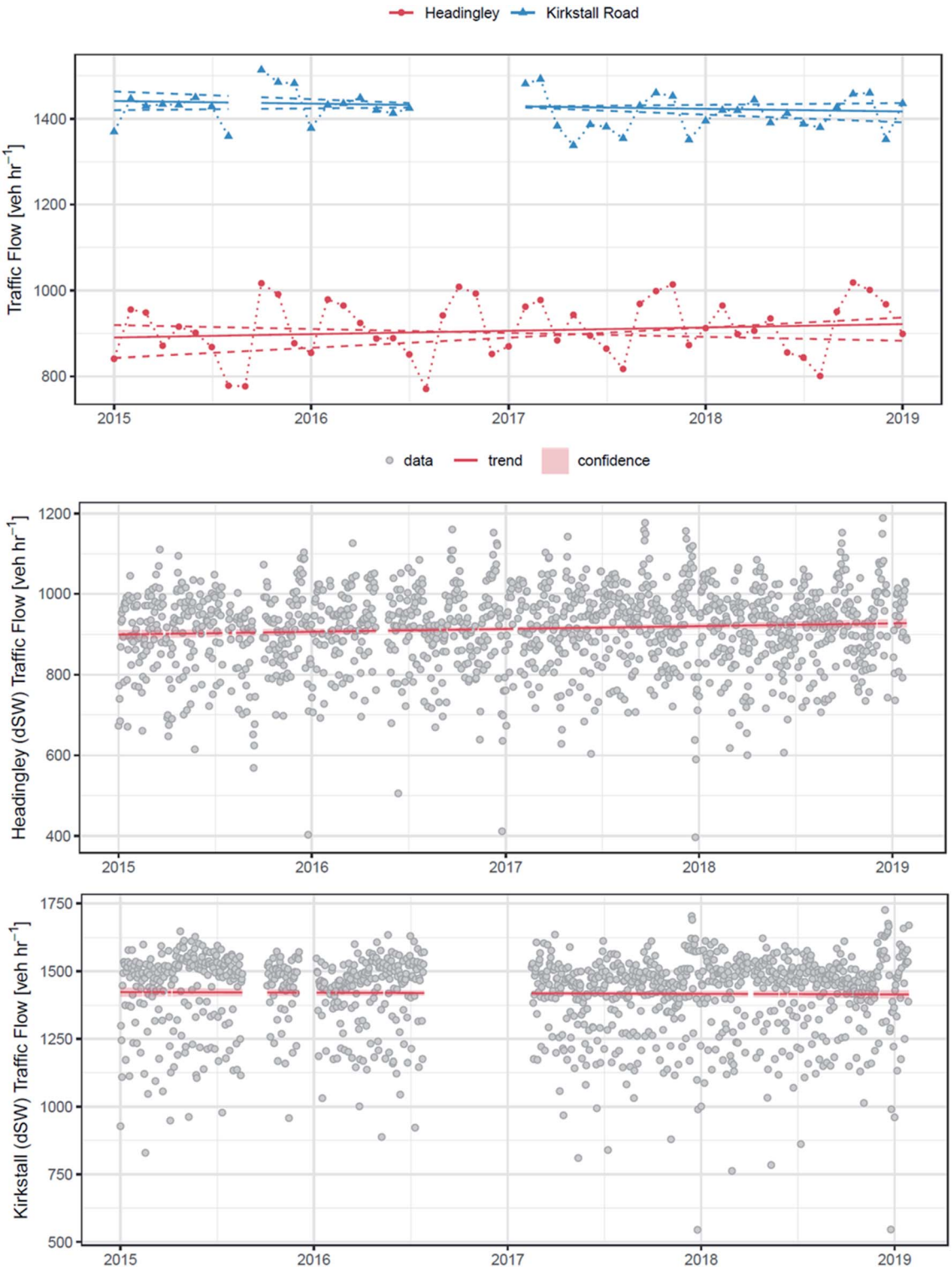

Fig. 9 Trend analysis of traffic flow measurements for Headingley and Kirkstall Road sites. Top, Thiel-Sen rate-of-change-in-traffic-flow analysis: Headingley 7.8 (range −9.13 to 23.55) vehicles per hour per year; and Kirkstall Road −6.09 (range −18.01 to 4.12) vehicles per hour per year neither statistically significant ($p > 0.05$). Middle and bottom, break-point/segment analysis of traffic data from Headingley and Kirkstall Road, respectively, method deweathering and deseasonalisation (dSW) with Leeds area meteorological data, (dSW) then break-point testing of residuals, which found no break-points for traffic counts for either site.

independently detected by change-point/segment methods coincides very closely with the time period of the main bus fleet upgrade. The combination of Selective Catalytic Reduction (SCR) and Diesel Particulate Filter (DPF) emission abatement used by Euro VI buses should, if correctly setup and maintained, deliver significant benefits by comparison to earlier technologies.

Transport for London (TfL), for example, reported that standard Euro VI London buses emitted an estimated 95% and 85% less NO_x and PM_{10}, respectively, by comparison to Euro V.[60] Concerns have however been raised about Euro VI emissions at lower speeds, when exhaust gases are cooler and SCRs tend to be less effective, but modelling based on emission inventories indicates that they should still deliver appreciable real-world air quality benefits.[61] Also, in areas on major bus routes where SCRs have been retrofitted to earlier bus fleets, air quality improvements of 9% and 14% have been reported for NO_x and NO_2, respectively.[62] So, the improvements estimated here for the period of the main intervention are of a magnitude that could realistically be attributed to the Euro VI bus fleet upgrade."

The earlier change-point/segment and earlier intervention are, however, not as readily associated. Simulation testing, described in further detail in ESI,† indicates that at the time of the intervention the methods used as reported here would have a higher than 70% likelihood of assigning the change to a point within *ca.* 2 months of actual date of occurrence (see *e.g.* Fig. S9–S12 and associated discussion in ESI†). So, the earlier break-point/segment, seen a year before the first upgrade, is highly unlikely to be an early prediction of this event. Similarly, the observed magnitude for the event (*ca.* 6.8%) is significantly larger than that would be anticipated for the first event if it were the result of an intervention one-quarter of the scale (6 *versus* 24 buses) of the second intervention. The anticipated magnitude, −3% (−12% × 6/24), is also almost half the 5% detection limit estimated for the method by simulation. It is therefore considered highly unlikely that the observed change could be the direct outcome of the first intervention.

Although not associated with method implementation or the investigated intervention, some further observations are made regarding the earlier 2015 event. Briefly summarising the preliminary analysis in Section 4 of the ESI,† although hindered by the completeness of the data, similar analyses of other nearby sites suggests that the 2015 event is more widely observed than the 2018 event, but urban rather than background/regional in nature. This is at about the time Euro 6 vehicle regulations were introduced in the UK and others[32] have reported similar changes that aligned with the introduction of earlier vehicle regulations. However, at this stage without further work and the analysis of more sites across the UK, a similar interpretation would be highly speculative in this case.

4 Main conclusions, discussion and future needs

Here, a novel combination of local contribution isolation, break-point and change-segment analysis was used to detect and quantify a discrete change-event in an environmental time-series. The methods were applied to ambient NO_2 air quality monitoring data from a Leeds City Council (LCC) operated and Automatic Urban and Rural Network (AURN) affiliated roadside air quality monitoring station in Headingley to demonstrate their use with typical data routinely collected by local authorities to quantify the impact of a contemporary traffic-related air quality intervention. In this case the studied intervention was the Leeds (UK) 2018 Euro VI upgrade of the high frequency local bus service fleet, but the methods could be used similarly to investigate the impacts of other interventions or disruptive events.

Direct analysis of the Headingley NO_2 time-series using break-point methods alone was hindered by seasonal, meteorological and background contributions to the area (see *e.g.* Fig. 6). Analysis suggests that these non-local contributes could hinder the detection of discrete changes of less than 35–50% of the local ambient mean concentration, which would be equivalent to a change of 10–14 μg m^{-3} at Headingley at the time of the interventions. As few air quality-related traffic interventions could realistically be expected to deliver such benefits, this obviously significantly limits the value of such break-point methods when applied to typical urban air quality data.

Previously, it has been reported that various time-series deconvolution procedures, *e.g.* background correction, dew-eathering and deseasonalisation,[16,32] can improve the sensitivity of break-points analysis. Here, using a relatively simply local contribution isolation method and relatively few inputs, all of which are readily available to many local authorities in the UK, methods are demonstrated that can be used to isolate local changes not readily detectable in ambient air measurement data-series (*cf e.g.* Fig. 6 and 7).

Using these methods, a discrete and significant NO_2 change was observed at Headingley, a decrease of 3.6 μg m^{-3} (1.2–6.1; 95% confidence), at the time of the major investigated intervention, equivalent to a 12% improvement in ambient NO_2 levels. This local change, not seen at the Kirkstall Road site, was superimposed on a less pronounced and similar general decrease seen at both sites across the study period, supporting the apportionment of both a smaller scale and more general improvement in air quality at the two sites and a more abrupt step-change in Headingley in early 2018. The break-point analysis was extended using change-segment prediction methods,[52–54] which provided an estimate of the change period of 16 March to 27 April 2018 (42 days).

As the urban arterials capacity was broadly stable through the study period, there was no statistically significant change in traffic demand and strong time alignment for the second major change-point/segment and the larger bus upgrade. This change is therefore attributed to the fleet upgrade to cleaner Euro VI powertrains.

Simulation studies indicate that, at the time of the intervention, the methods used as reported here would have a local contribution detection limit of *ca.* 5% for a decrease (<12% change observed) and a higher than 70% likelihood of assigning the change to a point within 2 months of actual date of occurrence (see *e.g.* Fig. S9–S12 and associated discussion in ESI†). We therefore conclude that, within the accuracy of the methods, this is also highly consistent with the observed change being the direct outcome of the intervention.

Further work with the simulation methods, included in ESI (Fig. S12–S19† and associated discussion), also provided useful insights into performance of the methods, *e.g.* predictive power at the start and end of time-series ranges and as detection limits

are approached, and method behaviour with near break-points, a situation that could be encountered if *e.g.* data properties such as the gradient are varied (Fig. S12† and associated discussion).

Looking forward it is important to note that many of the contemporary traffic interventions currently proposed by local authorities in the UK are of the order of 1–2% (ref. 7 and 63) and acknowledge that a detection limit of *ca.* 10% (5–15% depending on underlying trend and direction of change, see ESI†) is still a relatively large detection limit if these methods are to be used to robustly benchmark the performance of the full range of local air quality actions and interventions. But, here, simulation also provides useful insight regarding options for refinements, *e.g.* break-point input data variance has a significant effect of detection limit (see *e.g.* Fig. S12–S14 in ESI†). However, simulation tests using 1 hour to 1 week resolution data indicated an optimum combination of 1 hour resolution when applying the local contribution isolation and 1 day resolution when applying the break-point testing to the Leeds 01 January 2015 to 31 January 2019 time-series. In particular, when used in combination with 1 hour contribution isolation, break-point method performance did not improve significantly when used at resolutions lower than 1 day (*e.g.* 1 week or 1 month). This was because time averaging decreases both variance and mean. So, simply using lower time resolutions is not necessarily the answer.

Some signal isolation methods decrease variance without a pronounced effect on mean, *e.g.*, multiple-site time-series averaging, time-frequency filtering and 'ensemble of ensembles' methods (which build multiple models using different subsets of the input data and averaging the predictions of these). However, here again caution may be needed because simulation also suggests that there are likely to be trade-offs, and that perhaps sometimes some of the variance removed by more-aggressive or larger-scale normalisation strategies could actually be some of the variance needed to robustly detect and quantify small-scale change (see *e.g.* Fig. S13, S18 and S19 and associated discussion in ESI†). There was no need to apply further variance reduction here because the change of interest was detectable without more aggressive signal isolation. This is a 'conservative' strategy for studies applying combined signal isolation and break-point detection methods that it is arguably best both adopted and recommended, at least until any trade-offs between signal isolation and break-point detection are better understood. However, in other applications smaller changes will need to be detectable and such trade-offs do need to be further investigated as part of that process, along with options for the more strategic use of other data types to enhance signal isolation *e.g.* traffic data, vehicle fleet proportions.

By comparison to their use in other sectors, *e.g.* share trading in business, process line management in manufacturing and smart diagnostics in clinical practices where break-point methods are routinely applied in near-real time,[64] break-point methods have traditionally tended to be employed on time-series of years in environmental studies. While it is important to acknowledge that longer timescale implementations are obviously more easily justified, and that methods are unlikely to be completely transferable, more timely break-point methods were developed in other sectors because they were needed.[65] Similarly, local authorities need more timely assessments of the impacts of their traffic management activities and other interventions, so they can more confidently ensure they are delivering intended benefits, ideally at the earliest possible stage of the intervention process. So, there is also a need to investigate how much before and after data is actually needed to quantify a break-point, or more likely how predictive power changes with increasing monitoring time. In addition, there is also a need to consider methods for less certain measures of air quality, *e.g.* diffusion tubes and low cost-sensors, because in areas where conventional continuous analysers cannot be used, local authorities, and other interested parties, will undoubtedly be looking to use these. However, both case study findings and simulation studies reported here demonstrate that these approaches can already be used with confidence to measure the air quality impacts of larger traffic interventions using continuous analyser data.

Disclaimer

The views and opinions expressed herein by the authors are their own and do not necessarily reflect those of UK Government or any agency thereof.

Author contributions

All authors conceptualised the study, KR and JET collected data, KR, JET and AW contributed on software development and data analysis, all authors wrote, edited and reviewed paper.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

This work was funded by the Department for Environment, Food and Rural Affairs (Defra). The authors gratefully acknowledge contributions and input from colleagues at University of Leeds, Defra, IPSOS Mori, Leeds City Council and First Bus, and input from the Defra/Department for Transport Joint Air Quality Unit (JAQU) Technical Independent Review Panel (T-IRP) as part of internal review, and from editors and anonymous reviewers as part of the external peer-review process at Atmospheres. The authors also gratefully acknowledge the work of the R core team and their many collaborators in developing and maintaining the open-source statistical language R and associated packages (**https://www.r-project.org/**).

References

1 WHO (World Health Organization), *Ambient (outdoor) air quality and health: key facts*, **https://www.who.int/en/news-room/fact-sheets/detail/ambient-(outdoor)-air-quality-and-health**, 2018, accessed online 18/12/2019].

2 EEA (European Environmental Agency), *Air quality in Europe - 2019 report*. EEA Report No 10/2019, 2019, **https://www.eea.europa.eu/publications/air-quality-in-europe-2019**, as accessed 06/01/2020.

3 L. Ntziachristos, G. Papadimitriou, N. Ligterink and S. Hausberger, Implications of diesel emissions control failures to emission factors and road transport NOx evolution, *Atmos. Environ.*, 2016, **141**, 542–551, DOI: **10.1016/j.atmosenv.2016.07.036**.

4 N. Ligterink, *Real-world Vehicle Emissions*, International Transport Forum Discussion Paper 2017-06, 2017, DOI: **10.1787/2223439X**.

5 N. Hooftman, M. Messagie, J. Van Mierlo and T. Coosemans, A review of the European passenger car regulations–Real driving emissions vs. local air quality, *Renew. Sustain. Energy Rev.*, 2018, **86**, 1–21, DOI: **10.1016/j.rser.2018.01.012**.

6 Defra & DfT (Department for Environment, Food and Rural Affairs & Department for Transport), *Clean Air Zone Framework, Principles for setting up Clean Air Zones in England. Joint Air Quality Action Unit Report. OGL. 01 May 2017*, **https://www.gov.uk/government/publications/air-quality-clean-air-zone-framework-for-england**, as accessed 03/01/2020.

7 Defra & DfT (Department for Environment, Food and Rural Affairs & Department for Transport), *UK plan for tackling roadside nitrogen dioxide concentrations*. Detailed plan. OGL. July 2017, **https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/633270/air-quality-plan-detail.pdf**, as accessed 03/01/2020.

8 A. Y. Bigazzi and M. Rouleau, Can traffic management strategies improve urban air quality? A review of the evidence, *J. Transport Health*, 2017, **7**, 111–124, DOI: **10.1016/j.jth.2017.08.001**.

9 J. Burns, H. Boogaard, S. Polus, L. M. Pfadenhauer, A. C. Rohwer, A. M. van Erp, R. Turley and E. A. Rehfuess, Interventions to reduce ambient air pollution and their effects on health: An abridged Cochrane systematic review, *Environ. Int.*, 2020, **135**, p105400, DOI: **10.1016/j.envint.2019.105400**.

10 J. Cyrys, A. Peters, J. Soentgen and H. E. Wichmann, Low emission zones reduce PM_{10} mass concentrations and diesel soot in German cities, *J. Air Waste Manage. Assoc.*, 2014, **64**(4), p481–487, DOI: **10.1080/10962247.2013.868380**.

11 P. Panteliadis, M. Strak, G. Hoek, E. Weijers, S. van der Zee and M. Dijkema, Implementation of a low emission zone and evaluation of effects on air quality by long-term monitoring, *Atmos. Environ.*, 2014, **86**, 113–119, DOI: **10.1016/j.atmosenv.2013.12.035**.

12 F. M. Santos, Á. Gómez-Losada and J. C. Pires, Impact of the implementation of Lisbon low emission zone on air quality, *J. Hazard Mater.*, 2019, **365**, 632–641, DOI: **10.1016/j.jhazmat.2018.11.061**.

13 W. Wang, T. Primbs, S. Tao and S. L. M. Simonich, Atmospheric particulate matter pollution during the 2008 Beijing Olympics, *Environ. Sci. Technol.*, 2009, **43**(14), 5314–5320, DOI: **10.1021/es9007504**.

14 C. Holman, R. Harrison and X. Querol, Review of the efficacy of low emission zones to improve urban air quality in European cities, *Atmos. Environ.*, 2015, **111**, 161–169, DOI: **10.1016/j.atmosenv.2015.04.009**.

15 F. Kelly, H. R. Anderson, B. Armstrong, R. Atkinson, B. Barratt, S. Beevers, D. Derwent, D. Green, I. Mudway and P. Wilkinson, The impact of the congestion charging scheme on air quality in London. Part 1. Emissions modeling and analysis of air pollution measurements, *Res. Rep. Health Eff. Inst.*, 2011, (155), 5–71.

16 S. K. Grange and D. C. Carslaw, Using meteorological normalisation to detect interventions in air quality time series, *Sci. Total Environ.*, 2019, **653**, 578–588, DOI: **10.1016/j.scitotenv.2018.10.344**.

17 J. L. Pearce, J. Beringer, N. Nicholls, R. J. Hyndman and N. J. Tapper, Quantifying the influence of local meteorology on air quality using generalized additive models, *Atmos. Environ.*, 2011, **45**(6), 1328–1336, DOI: **10.1016/j.atmosenv.2010.11.051**.

18 A. M. Jones, R. M. Harrison, B. Barratt and G. Fuller, A large reduction in airborne particle number concentrations at the time of the introduction of "sulphur free" diesel and the London Low Emission Zone, *Atmos. Environ.*, 2012, **50**, 129–138, DOI: **10.1016/j.atmosenv.2011.12.050**.

19 A. Font and G. W. Fuller, Did policies to abate atmospheric emissions from traffic have a positive effect in London?, *Environ. Pollut.*, 2016, **218**, 463–474, DOI: **10.1016/j.envpol.2016.07.026**.

20 A. Font, L. Guiseppin, M. Blangiardo, V. Ghersi and G. W. Fuller, A tale of two cities: is air pollution improving in Paris and London?, *Environ. Pollut.*, 2019, **249**, 1–12, DOI: **10.1016/j.envpol.2019.01.040**.

21 M. de Fatima Andrade, P. Kumar, E. D. de Freitas, R. Y. Ynoue, J. Martins, L. D. Martins, T. Nogueira, P. Perez-Martinez, R. M. de Miranda, T. Albuquerque and F. L. T. Gonçalves, Air quality in the megacity of São Paulo: Evolution over the last 30 years and future perspectives, *Atmos. Environ.*, 2017, **159**, 66–82, DOI: **10.1016/j.atmosenv.2017.03.051**.

22 K. Ravindra, T. Singh, A. Biswal, V. Singh and S. Mor, Impact of COVID-19 lockdown on ambient air quality in megacities of India and implication for air pollution control strategies, *Environ. Sci. Pollut. Res.*, 2021, **28**(17), 21621–21632, DOI: **10.1007/s11356-020-11808-7**.

23 S. Gulia, S. S. Nagendra, M. Khare and I. Khanna, Urban air quality management-A review, *Atmos. Pollut. Res.*, 2015, **6**(2), 286–304, DOI: **10.1007/s40572-014-0019-7**.

24 A. Jentsch, J. Kreyling and C. Beierkuhnlein, A new generation of climate-change experiments: events, not trends, *Front. Ecol. Environ.*, 2007, **5**(7), 365–374, DOI: **10.1890/1540-9295(2007)5[365:ANGOCE]2.0.CO;2**.

25 G. Di Virgilio, J. P. Evans, S. A. Blake, M. Armstrong, A. J. Dowdy, J. Sharples and R. McRae, Climate change increases the potential for extreme wildfires, *Geophys. Res. Lett.*, 2019, **46**(14), 8517–8526, DOI: **10.1016/j.etap.2017.08.022**.

26 S. Aminikhanghahi and D. J. Cook, A survey of methods for time series change point detection, *Knowl. Inf. Syst.*, 2017, **51**(2), 339–367, DOI: **10.1007/s10115-016-0987-z**.

27 J. Bai and P. Perron, Computation and analysis of multiple structural change models, *J. Appl. Econom.*, 2003, **18**, 1–22, DOI: 10.1002/jae.659.
28 A. Zeileis, C. Kleiber, W. Krämer and K. Hornik, Testing and dating of structural changes in practice, *Comput. Stat. Data Anal.*, 2003, **44**(1–2), 109–123, DOI: 10.1016/S0167-9473(03)00030-6.
29 T. S. Lee, Change-point problems: bibliography and review, *J Stat Theory Pract.*, 2010, **4**(4), 643–662, DOI: 10.1080/15598608.2010.10412010.
30 A. Amiri and S. Allahyari, Change point estimation methods for control chart postsignal diagnostics: a literature review, *Qual. Reliab. Eng. Int.*, 2012, **28**(7), 673–685, DOI: 10.1002/qre.1266.
31 D. C. Carslaw, K. Ropkins and M. C. Bell, Change-point detection of gaseous and particulate traffic-related pollutants at a roadside location, *Environ. Sci. Technol.*, 2006, **40**(22), p6912–6918, DOI: 10.1021/es060543u.
32 D. C. Carslaw and N. Carslaw, Detecting and characterising small changes in urban nitrogen dioxide concentrations, *Atmos. Environ.*, 2007, **41**(22), p4723–4733, DOI: 10.1016/j.atmosenv.2007.03.034.
33 A. G. Barnett, Air pollution trends in four Australian cities 1996-2011. *Air Quality and Climate Change*, 2012, vol. 46, 4, p. 28.
34 J. Stedman, J. Abbott, P. Willis and J. Bower, *Review of Background Air Quality Data and Methods to Combine these with Process Contributions*, Environment Agency for England and Wales, Bristol, 2006, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/291521/scho1205bkbn-e-e.pdf, as accessed 03/01/2020.
35 S. Visser, J. G. Slowik, M. Furger, P. Zotter, N. Bukowiecki, R. Dressler, U. Flechsig, K. Appel, D. C. Green, A. H. Tremper and D. E. Young, Kerb and urban increment of highly time-resolved trace elements in PM_{10}, $PM_{2.5}$ and $PM_{1.0}$ winter aerosol in London during ClearfLo 2012, *Atmos. Chem. Phys.*, 2015, **15**(5), 2367–2386, DOI: 10.5194/acp-15-2367-2015.
36 A. Sayegh, J. E. Tate and K. Ropkins, Understanding how roadside concentrations of NOx are influenced by the background levels, traffic density, and meteorological conditions using Boosted Regression Trees, *Atmos. Environ.*, 2016, **127**, 163–175, DOI: 10.1016/j.atmosenv.2015.12.024.
37 X. Basagaña, M. Triguero-Mas, D. Agis, N. Pérez, C. Reche, A. Alastuey and X. Querol, Effect of public transport strikes on air pollution levels in Barcelona (Spain), *Sci. Total Environ.*, 2018, **610**, 1076–1082, DOI: 10.1016/j.scitotenv.2017.07.263.
38 M. Kendall and A. Stuart, *Adv. Theory Appl.*, 1983, **3**, 410–414.
39 J. Kuebler, H. van den Bergh and A. G. Russell, Long-term trends of primary and secondary pollutant concentrations in Switzerland and their response to emission controls and economic changes, *Atmos. Environ.*, 2001, **35**(8), 1351–1363, DOI: 10.1016/S1352-2310(00)00401-5.
40 L. R. Henneman, H. A. Holmes, J. A. Mulholland and A. G. Russell, Meteorological detrending of primary and secondary pollutant concentrations: Method application and evaluation using long-term (2000–2012) data in Atlanta, *Atmos. Environ.*, 2015, **19**, 201–210, DOI: 10.1016/j.atmosenv.2015.08.007.
41 H. L. Walker, M. R. Heal, C. F. Braban, S. Ritchie, C. Conolly, A. Sanocka, U. Dragosits and M. M. Twigg, Changing supersites: assessing the impact of the southern UK EMEP supersite relocation on measured atmospheric composition, *Environ. Res. Commun.*, 2019, **1**(4), 041001, DOI: 10.1088/2515-7620/ab1a6f.
42 A. R. Malby, J. D. Whyatt and R. J. Timmis, Conditional extraction of air-pollutant source signals from air-quality monitoring, *Atmos. Environ.*, 2013, **74**, 112–122, DOI: 10.1016/j.atmosenv.2013.03.028.
43 G. R. Cass, Organic molecular tracers for particulate air pollution sources, *Trac. Trends Anal. Chem.*, 1998, **17**(6), 356–366, DOI: 10.1016/S0165-9936(98)00040-5.
44 M. Tobiszewski and J. Namieśnik, PAH diagnostic ratios for the identification of pollution emission sources, *Environ. Pollut.*, 2012, **162**, 110–119, DOI: 10.1016/j.envpol.2011.10.025.
45 J. G. Watson, L. W. Antony Chen, J. C. Chow, P. Doraiswamy and D. H. Lowenthal, Source apportionment: findings from the US supersites program, *J. Air Waste Manage. Assoc.*, 2008, **58**(2), 265–288, DOI: 10.3155/1047-3289.58.2.265.
46 K. Ropkins and J. Tate, *Early Observations on the Impact of the COVID-19 Lockdown on Air Quality Trends across the UK*. Science of The Total Environment, 2020, p.142374, DOI: 10.1016/j.scitotenv.2020.142374.
47 ONS (Office for National Statistics), *Population Estimates for UK, England and Wales, Scotland and Northern Ireland, Mid-2018*, Office for National Statistics Report, OGL, 28 June 2019.
48 R Core Team, *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria, 2019, https://www.R-project.org/.
49 D. C. Carslaw and K. Ropkins, openair - an R package for air quality data analysis, *Environ. Model. Software*, 2012, **27–28**, 52–61, DOI: 10.1016/j.envsoft.2011.09.008.
50 I. Fellows, using the JMapViewer library by J.P. Stotz, *OpenStreetMap: Access to Open Street Map Raster Images*. 2019, R package version 0.3.4, https://CRAN.R-project.org/package=OpenStreetMap.
51 A. Zeileis, F. Leisch, K. Hornik and C. Kleiber, strucchange: An R Package for Testing for Structural Change in Linear Regression Models, *J. Stat. Software*, 2002, **7**(2), 1–38. https://www.jstatsoft.org/v07/i02/.
52 V. M. R. Muggeo, Estimating regression models with unknown break-points, *Stat. Med.*, 2003, **22**, 3055–3071, DOI: 10.1002/sim.1545.
53 V. M. R. Muggeo, segmented: an R Package to Fit Regression Models with Broken-Line Relationships, *R. News*, 2008, **8/1**, 20–25, https://cran.r-project.org/doc/Rnews/.
54 V. M. R. Muggeo, Interval estimation for the breakpoint in segmented regression: a smoothed score-based approach,

Aust. N. Z. J. Stat., 2017, **59**, 311–322, DOI: **10.1111/anzs.12200**.

55 J. Bai, Estimation of a change point in multiple regression models, *Rev. Econ. Stat.*, 1997, **79**(4), 551–563, DOI: **10.1162/003465397557132**.

56 S. N. Wood, Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models, *J. Roy. Stat. Soc. B*, 2011, **73**(1), 3–36, DOI: **10.1111/j.1467-9868.2010.00749.x**.

57 S. N. Wood, N. Pya and B. Saefken, Smoothing parameter and model selection for general smooth models (with discussion), *J. Am. Stat. Assoc.*, 2016, **111**, 1548–1575, DOI: **10.1080/01621459.2016.1180986**.

58 D. C. Carslaw and P. J. Taylor, Analysis of air pollution data at a mixed source location using boosted regression trees, *Atmos. Environ.*, 2009, **43**(22–23), 3563–3570, DOI: **10.1016/j.atmosenv.2009.04.001**.

59 S. N. Wood, *mgcv: Mixed GAM Computation Vehicle with Automatic Smoothness Estimation, R package version 1.8-31*. 2019, **https://CRAN.R-project.org/package=mgcv**.

60 TfL (Transport for London), *EURO VI Bus NOx abatement*, 2017, **https://content.tfl.gov.uk/pic-20170628-item14-euro-vi-bus-nox.pdf**.

61 T. Grigoratos, G. Fontaras, B. Giechaskiel and N. Zacharof, *Real World Emissions Performance of Heavy-Duty Euro VI Diesel Vehicles*, Atmospheric environment, 2019, vol. 201, pp. 348–359, DOI: **10.1016/j.atmosenv.2018.12.042**.

62 B. Barratt and D. C. Carslaw, *Impacts of the Bus Retrofit Programme on NO_2 Concentrations along Putney High Street*, Environ. Res. Group, Rep. Prep, London brgh. Wandsworth, 2014, **https://londonair.org.uk/London/reports/PHSSCRImpactsReport.pdf**, as accessed 06/03/2020.

63 DCC (Derby City Council), *Derby Local Air Quality Plan – Full Business Case for tackling roadside nitrogen dioxide exceedances*, 2019, DCC report, **https://www.derby.gov.uk/transport-and-streets/air-quality-in-derby/tackling-poor-air-quality/air-quality-business-case/**, as accessed 06/03/2020.

64 S. Liu, A. Wright and M. Hauskrecht, Change-point detection method for clinical decision support system rule monitoring, *Artif. Intell. Med.*, 2018, **91**, 49–56, DOI: **10.1016/j.artmed.2018.06.003**.

65 C. Y. Lu, G. Simon, S. B. Soumerai and M. Kulldorff, Counterpoint: Early Warning Systems are Imperfect, but Essential, *Med. Care*, 2018, **56**(5), 382–383, DOI: **10.1097/MLR.0000000000000896**.