CHARACTERISING THE TEMPORAL VARIATIONS
OF GROUND-LEVEL OZONE AND ITS RELATIONSHIP
WITH TRAFFIC-RELATED AIR POLLUTANTS IN THE UNITED
KINGDOM: A QUANTILE REGRESSION APPROACH

S. MUNIR, H. CHEN & K. ROPKINS
Institute for Transport Studies, University of Leeds, UK.

ABSTRACT
Ground-level ozone is a secondary air pollutant and is photochemically produced by solar radiation from the reaction of volatile organic compounds (VOCs) and nitrogen oxides (NOx). Ground-level ozone is considered a harmful pollutant due to its adverse impact on human health, agricultural crops and materials. The concerning factor is that in spite of decreasing trends in some other air pollutants (e.g. NOx), ozone concentrations are still increasing. This paper describes the temporal variations of ozone at four air quality monitoring sites (Harwell, Leeds, Marylebone and Strath Vaich) in the United Kingdom for the year of 2008. The association of ozone with some traffic-related air pollutants has been explored applying a quantile regression model (QRM). The traffic-related air pollutants considered as predictors for this study are hydrocarbons (HC), nitric oxides (NO), nitrogen dioxides (NO₂), carbon monoxides (CO) and particulate matter (PM₂.₅). QRM can handle the non-linearities in the relationship of ozone and its predictors and is applicable to non-normal air quality data distribution. The behaviour and interaction of ozone with its predictors vary at different regimes of ozone distributions, which remains hidden when applying an ordinary least square regression model. QRM explains significantly more variations in ozone concentrations (global goodness of fit $R^2 = 0.88$) as compared to ordinary least square regression (coefficient of determination $R^2 = 0.32$) and is therefore better suited for ozone data analysis and prediction.

Keywords: Air pollution, ground-level ozone, nitrogen oxides, ozone variations, quantile regression.

1 INTRODUCTION
Ozone is not emitted directly by combustion processes; it is rather formed photochemically from the sunlight-initiated oxidation of volatile organic compounds (VOCs) in the presence of nitrogen oxides (NOx) in the atmosphere and is therefore considered as a secondary air pollutant. Ozone concentrations are not only dependent on meteorological variables (e.g. solar radiation, temperature, relative humidity, wind speed and wind direction) but also on other air pollutants. Figure 1 shows
how ozone is involved in chemical reactions with various species in the atmosphere including NOx, hydrocarbons (HC), carbon monoxides and hydroxyl radical (OH). In addition to local photochemical ozone production, ozone rich air advection (regional ozone transportation) and stratospheric−tropospheric ozone exchange may contribute a significant amount to the local observed ozone concentrations. Nitric oxide (NO) titration and dry deposition of ozone are considered as the main sinks for ground-level ozone.

The concerning factor about ozone is that although the concentrations of other pollutants, including NOx, sulphur dioxides (SO2) and VOCs, have been declining for the last 20 years or so, ozone does not seem to suggest a declining trend. Background ozone concentrations have rather increased since about 1990 in the United Kingdom [5]. The increasing trends in ozone concentrations are most apparent at urban sites, but which to a less extent also influence the observations at the majority of rural locations [1]. The increasing trend in ozone concentrations can be attributed to long distance migration of ozone from across the North Atlantic [1] or to reduction in local-scale removal of ozone by direct reaction with fresh NO [6], a trend that is now widely attributed to the ongoing improvement in vehicle NOx emission regulations and associated progressive policy practices.

Ozone data are not normally distributed [9] and its relationship is non-linear with its predictors (NO, NO2, CO, PM2.5) [10]. Classical (parametric) statistics explicitly assume normality and linearity.
of the data; therefore, linear regression may not be applicable to ozone data analysis, otherwise it may result in biased estimation [11]. This paper studies the temporal variations of ozone and its association with traffic-related air pollutants. The study employs a quantile regression model (QRM) which can be applied to both normal and non-normal data distributions and is capable of addressing the non-linearities in the association of ozone and other air pollutants data.

2 METHODOLOGY

This study is based on mean hourly data of ozone (µg/m³), HC (µg/m³), NO (µg/m³), NO₂ (µg/m³), PM₂.₅ (µg/m³) and CO (mg/m³). Ozone data from four air quality monitoring sites have been analysed to establish ozone temporal variations (daily, weekly and seasonal trends) at these sites. The four sites are Harwell (rural) in the southeast of England, Leeds (urban centre) in West Yorkshire, Marylebone (roadside) in London and Strath Vaich (remote) in Highlands of Scotland. The paper concentrates on Marylebone roadside monitoring site to study the relationship between ozone and some of the traffic-related air pollutants using a QRM. These air quality monitoring sites are part of the UK Automatic Urban and Rural Network (AURN), which is the United Kingdom’s largest automatic monitoring network; for details of these monitoring sites, see Ref. [12].

AURN uses Ultraviolet Analyser and Chemiluminescent Analyser for ozone and NOx measurements, respectively. CO concentrations are measured using Infrared Analyser; whereas PM₂.₅ concentrations are monitored by Tapered Element Oscillating Microbalance (TEOM), Beta Attenuation Monitor (BAM), Gravimetric Monitor or Filter Dynamics Measurement System (FDMS), which provides a continuous direct mass measurement of particulate matter [12]. These techniques represent the current state-of-the-art for automated monitoring networks and, with the exception of the automatic PM₁₀ and PM₂.₅ analysers, are the reference methods of measurement defined in the relevant EU Directives. The automated systems for particles are subject to correction for routine monitoring. The standard EU reference method for particulate measurement refers to three devices which might be used: (a) Low Volume System: the LVS-PM-10 sampler; (b) High Volume System: the HVS PM-10 sampler and (c) Super-High Volume System: the WRAC-PM10 sampler (Wide Range Aerosol Classifier) [13].

The mean hourly monitored data, after some initial screening process to exclude clearly faulty data, are uploaded as provisional data every hour. The data then go through a further two stages called data verification and data ratification before the data are marked as ‘Ratified’ data. All the data from AURN have a standard Quality Assurance (QA) and Quality Control (QC); for details of these procedures, see Ref. [12].

2.1 Model description

This study applies QRM proposed by Baur et al. [14] for ozone and air quality data analysis and has certain advantages over other methods. QRM can be used for both parametric and non-parametric regression methods, as this model does not depend on the single measure of the central tendency (mean or median) of the data distribution only; instead, it examines the entire distribution of the data and hence is robust to departures of the data from normality and skewed tails. QRM allows the covariates to have different impacts at different points of the data distribution and is, therefore, capable of handling the non-linearities in the association of dependent and independent variables.

The linear regression model (LRM) focuses on modelling the conditional mean of a response variable (in our case ozone) without addressing its full distribution, whereas the QRM accommodates analysis of the full distribution of the response variable. The QRM estimates the potential differential effect on various quantiles of the data distribution. Using ozone (O₃) as dependent
variable and HC, CO, NO, NO\textsubscript{2} and PM\textsubscript{2.5} as independent variables, LRM and QRM can be presented as below [11]:

\[
0_3 = \beta_0 + \beta_1 HC + \beta_2 CO + \beta_3 NO + \beta_4 NO_2 + \beta_5 PM_{2.5} + \varepsilon_1
\]

\[
0_3 = \beta_0^{(p)} + \beta_1^{(p)} HC + \beta_2^{(p)} CO + \beta_3^{(p)} NO + \beta_4^{(p)} NO_2 + \beta_5^{(p)} PM_{2.5} + \varepsilon_1^{(p)}
\]

where \(p\) shows the \(p\)th quantile and \(0 < p < 1\), \(\beta_0\) (constant) the intercept, \(\beta_1\) to \(\beta_5\) the slopes (gradients) of the independent variables and \(\varepsilon\) the error term. The error term in LRM is assumed to be independent of the value of the covariates (homoscedasticity). In contrast, QRM allows for the variance of the error term to vary (heteroscedasticity) and makes no assumptions about the variance structure. The constant \(\beta_0\) and the coefficients \(\beta_1\) to \(\beta_5\) are estimated for 99 different quantiles (\(p = 0.01, \ldots, 0.99\)) using each time the entire dataset. The 0.5th quantile represent the median, half of the data occur above the median and half below the median.

R package for statistical computing version 2.1.3.1 [15] and two additional packages ‘openair’ [16] and ‘quantreg’ [17] were used to perform the statistical analysis presented in this paper.

3 RESULTS AND DISCUSSION

Ozone distributions at Leeds, Strath Vaich, Harwell and Marylebone have been depicted in Fig. 2, which shows that statistically ozone distribution is not normally distributed at any of these sites. The distribution is different at various sites and is right skewed; see Ref. [9] for more details of ozone distribution at various monitoring sites and in various months.

Ozone temporal variations at roadsides, urban centres, rural and remote locations in the United Kingdom are briefly analysed in section 3.1; and the association of ozone with HC, NO, NO\textsubscript{2}, CO and PM\textsubscript{2.5} has been analysed with the help of QRM in section 3.2.

3.1 Temporal variations of ozone

In this section, ozone data from four monitoring sites have been analysed to estimate ozone daily, weekly and seasonal cycles. The four monitoring sites have different characteristics; the aim here is to obtain a broad picture of ozone variations at roadsides, urban centres, rural and remote areas. Ozone concentrations seem to have been affected by both meteorological conditions and sources of fresh NO (as explained below). Figure 3 shows the time variations plot of ozone for 2008 at Leeds centre, Harwell, Marylebone, and Strath Vaich. Generally ozone concentrations are found to be higher at rural and remote sites (Harwell and Strath Vaich) and lower at urban and roadside monitoring sites (Leeds and Marylebone). The lowest ozone concentrations are exhibited by Marylebone monitoring site which is located approximately 1 m from the edge of Marylebone road (A50). This road has six lanes and has a flow of 80,000 vehicles per day [12]. Most probably titration of ozone by fresh NO emitted by road transport keeps ozone concentrations low at this site.

Ozone concentrations at all sites follow a typical 24-h cycle, that is, lower during night and early morning hours and higher during day times (Fig. 3, bottom-left). The highest concentrations are achieved during the afternoon from 1200 to 1400 h, most probably caused by photochemical ozone formation as a result of higher amount of UV radiation during these hours. Harwell and Strath Vaich sites have slightly different daily cycles than Marylebone and Leeds. At Marylebone and Leeds in addition to night times low, there is a further reduction in ozone concentrations in the morning (about 06:00 am), which cannot be observed at the other two sites. This affect is most probably caused by the early morning traffic which emits NO that quickly reacts with ozone and further reduces its concentrations.
On a weekly basis, generally weekend and Monday exhibit higher, whereas Wednesday and Thursday show relatively lower ozone concentrations (Fig. 3, bottom-right panel). For instance, mean ozone concentrations (µg/m³) at Marylebone site were 20 and 12 and at Harwell site 47 and 55 on Sunday (weekend) and Thursday (midweek), respectively. High level of ozone during weekend is a well-known phenomenon and is referred to as ‘ozone weekend effect’ (OWE)’ in the literature, for example, [18], and the references therein. Low levels of traffic during weekend is considered to be the main cause of OWE, as low levels of fresh NO reduce local ozone removal. Compared to other weekdays, the higher level of ozone on Monday is probably due to the carry over effect of ozone from weekend. At Strath Vaich monitoring site, the difference in ozone concentrations during different days is negligible, probably because the site is remote and is not affected much by road traffic.

Ozone concentrations show a clear seasonal effect. Ozone concentrations are higher during March, April and May and lower during autumn and winter months at all four sites (Fig. 3, bottom-middle). For instance, monthly mean ozone concentrations at Marylebone site were 35 and 7 and at
Leeds site were 79 and 17 for May and December, respectively. Cloudy and cold weather conditions during winter months result in very low UV radiation in the United Kingdom, which reduce photochemical ozone formation. Ozone titration by NO and low tropospheric–stratospheric exchange of ozone caused by the stagnant atmospheric conditions further reduces ground-level ozone concentrations, which explain low level of ozone in the winter months. In spring when UV radiation increases, it increases photochemical ozone formation and hence ozone level increases in the atmosphere. The point which needs further attention is that UV radiation are generally higher in June, July and August but Fig. 3 shows that after May ozone levels decrease gradually and in August reach a level which is normally observed in winter months. The reason is that for photochemical ozone formation in addition to UV radiation, the presence of VOCs is essential whose oxidation by OH in the presence of NOx leads to ozone formation. Data collected at Harwell and Marylebone monitoring sites show (Fig. 4) that HC accumulate in the atmosphere during winter months and reach the highest level in February. In March when the level of UV radiation increases, this consumes the HC through photochemical ozone formation and HC levels gradually decrease. In August in spite of relatively high level UV radiation, photochemical ozone formation is restricted by the limited availability of HC.

At Harwell and Marylebone about 30 HC are monitored (1,2,3-trimethylbenzene, 1,2,4-trimethylbenzene, 1,3,5-trimethylbenzene, 1,3-butadiene, 1-butene, 1-pentene, 2-methylpentane, 3-methylpentane, benzene, ethane, ethyl-benzene, ethene, ethyne, isoprene, propane, propene, toluene, cis-2-butene, cis-2-pentene, isobutane, isoctane, isopentane, mp-xylene, n-butane, n-heptane, n-hexane, n-octane, n-pentane, o-xylene, trans-2-butene, trans-2-pentene). Instead of analysing the data of all HC, their hourly concentrations are averaged to get one variable for each site. The HC levels at Marylebone are
much higher than at Harwell; therefore, normalised levels are applied to make the comparison easy. ‘Openair package’ simply divides the variable(s) by their mean values to calculate the normalised levels.

3.2 Output of quantile regression

The outputs of QRM have been depicted in Fig. 5, using ozone as a response variable and HC, NO, NO\textsubscript{2}, CO and PM\textsubscript{2.5} as explanatory variables measured at Marylebone road, 2008. The Barrodale and Roberts (BR) algorithm method for computing the fit has been adopted here. The ‘BR’ method has been described in details in Koenker and d’Orey [19] as an efficient technique for large datasets (e.g. up to several thousand observations). In Fig. 5 alongside quantile regression, the outputs of ordinary least square regression have also been visualised. In ordinary least square regression, only one regression coefficient represents the entire distribution of the explanatory variable (indicated by solid line along with its 95% confident interval), whereas in quantile regression, generally, several coefficients are given depending on the number of quantiles chosen. In Fig. 5 quantile regression coefficients have been given for 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9 and 0.99 quantiles (represented by dashed-dotted line with their 95% confident intervals). Quantile regression coefficients have been given at y axis and different quantile values (0.1 to 0.99) at x axis. In Fig. 5 the top left panel shows the intercepts of the model. The values of intercept (constant) are higher for higher quantiles and vice versa. For instance, the intercept value for 0.1 quantile is about 5, whereas it is about 75 for quantile 0.99.

The effect of HC on ozone concentrations have been presented in Fig. 5 (top-right). The effect of HC on ozone concentrations is significant as confident intervals of the quantile regression coefficients do not overlap with the zero line. Quantile regression coefficients are negative, which shows that HC have negative association with ground-level ozone. In other words, an increase in ozone concentrations causes a decrease in HC concentrations. HC levels are generally higher in winter months (Fig. 4) when ozone concentrations are lower and their concentrations start decreasing in
spring as photochemical ozone formation increases, which consumes HC. The strength of the relationship between ozone and HC increases gradually with the increase in ozone concentration and reaches the highest value at quantile 0.99, where the quantile regression coefficient value is ‘−2.02’. Spearman correlation between ozone and HC was also estimated and had a negative value, the highest correlation coefficient was estimated for the month of March (−0.80) and lowest for August (−0.34). Spearman correlation was applied because it can be used for both normal and non-normal distribution and is therefore best applicable to air quality and ozone data, which normally have non-normal distribution (Fig. 2).

The effect of CO on ozone concentrations is shown in Fig. 5 (middle-left). CO has a negative impact on ozone concentrations and the effect is significant at all quantiles. Like HC and NOx,
CO is considered as one of the precursors of ozone (as shown in Fig. 1) and therefore photochemical ozone formation will consume CO, which explains the negative correlation. The quantile regression coefficients of CO have the same pattern as that of HC. The strength of coefficients increases from quantile 0.1 to 0.99 gradually and range from −0.53 to −10.00. The correlation coefficients are much higher for CO than HC, NO, NO₂ and PM₂.₅, which may be due to the fact that CO is measured mg/m³, in contrast to the other pollutants which are expressed in µg/m³. The highest spearman correlation coefficient was shown by March (−0.81) and lowest for August (−0.37).

The effect of NO (Fig. 5, middle-right) on ozone concentration is negative and is significant at all quantiles except at quantile 0.99, where the confident intervals overlap with zero. The strength of coefficients slightly increases from quantile 0.1 to 0.5 and then stays almost constant. The effect is not significantly different from the mean effect at most of the quantiles. On the other hand, NO₂ (Fig. 5, bottom-left) has positive effect on ozone concentrations from quantile 0.1 to 0.6 and negative from quantile 0.7 to 0.99 of the ozone distribution.

Figure 5 (bottom-right) shows the effect of PM₂.₅ on ozone concentrations. Quantile regression coefficients are negative up to quantile 0.8 and positive at quantile 0.9 and 0.99. The effect is significant at all quantiles. Highest correlation coefficient was estimated for the month of March (−0.78) and lowest for the month of August (−0.46) using spearman correlation analysis.

3.2.1 Goodness of fit for quantile regression
Coefficient of determination or goodness of fit (R²) is a measure used in statistical model analysis to assess how well a model explains and predicts future outcomes of the data. R² is the ratio of the explained variation to the total variation and is a measure that allows us to determine how certain one can be in making predictions from a certain model. The goodness of fit in ordinary least square is based on least squares criterion. R² values range from 0 to 1. Larger value of R² indicates a better model fit. In quantile regression, the goodness of fit is represented by R¹ (τ) and its values, like R², lies between 0 and 1 [20]. R² measures a global goodness of fit over the entire conditional distribution, whereas R¹ (τ) measures the local performance of model for a given quantile. Koenker and Machado [21] suggest measuring R¹ (τ) by comparing the sum of weighted distance for the model of interest with the sum in which only the intercept is used (for details, see Refs [20, 21]). R¹ (τ) and R² have different ways of calculating and have different natures, as the former is a local whereas the latter is a global measure of performance and therefore are not directly comparables. R¹ (τ) and R² that are estimated for 0.50 quantile (median of the data) and mean ozone, respectively, are different even if the models estimates of the ozone concentrations are similar. R¹ (τ) values for different quantiles have been shown in Fig. 6, which are relatively weaker as compared to global goodness of fit for the ordinary least square regression. The coefficient of determination for the linear model was 0.32.

To make the performance of QRM comparable with least square regression model, a global goodness of fit (denoted by R¹) can be estimated for QRM. To estimate R¹ for QRM, this study adopts the amalgamated QRM (AQRM) suggested by Baur et al. [14]. The first step in estimation of R¹ is to run QRM and determine quantile regression coefficients for 10 quantiles (0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 0.99) that were used in section 3.2, using ozone as dependent variable and NO, NO₂, HC, CO and PM₂.₅ as independent variables for the whole dataset. The test dataset (April 2008) was divided into 10 equal subsets according to the above quantile values of ozone data. Using quantile regression coefficients of each quantile, ozone was predicted for each subset and combined into one dataset ordering from 0.1 to 0.99. Finally predicted (estimated by the model) and observed (April 2008) ozone were compared for the test data (Fig. 7).
Figure 7 depicts predicted ozone versus observed ozone mixing ratios at Marylebone roadside monitoring site for April 2008. The scatter plot of observed ozone versus predicted ozone by QRM is shown in the left, whereas the scatter plot of observed ozone versus predicted ozone by ordinary least square (OLS) model is shown in the right–top panel of Fig. 7. Observed and predicted ozone depicted by lines rather than points to simplify the comparison are shown in Fig. 7 (bottom panel). Predicted ozone by QRM closely follows the observed ozone, particularly at extreme values where OLS fails to perform. QRM explains more of the ozone variations showing $R^1$-value of 0.88 in comparison to OLS which has $R^2$-value of 0.32. This indicates that QRM explain the ozone variation (88%) significantly better than OLS (32%). As shown in Fig. 2, the ozone concentrations at Marylebone roadside monitoring site have skewed distributions with the long right tails. As a result, OLS model underpredicts ozone concentrations in the right tail of the distribution.

4 CONCLUSION
This paper describes time variations of ground-level ozone concentrations (hourly mean data in µg/m³) at four air quality monitoring sites (Harwell, Leeds, Marylebone and Strath Vaich) for the year 2008. Daily, weekly and seasonal cycles of ozone concentrations have been visualized with the help of time variation plots. The temporal variations have been explained with the help of traffic-related air pollutants (HC, NO, NO₂, CO and PM₂.₅) data and meteorology. The paper focused on data from Marylebone site to investigate the relationship of ozone with traffic related air pollutants using a QRM. QRM is applicable to non-normally distributed air quality data and can effectively address the non-linearities in the association of ozone and its predictors. QRM studies the entire distributions of the data and unlike ordinary least square regression does not rely on the central value (mean, median) only. The behaviour and interaction of the independent variables with ozone change at different regimes of ozone concentrations, a characteristic that is normally obscured in the traditional
regression models. Comparing the performance of QRM and OLS, it is shown that QRM performs better (global goodness-of-fit $R^1 = 0.88$) than OLS (coefficient of determination $R^2 = 0.32$). Better correlation was observed between recorded and predicted ozone by QRM, particularly at extreme values as OLS underpredicts ozone concentrations in right skewed tail of ozone distribution. This study was conducted using data recorded during one year at four monitoring sites. Further work is required for spatial and temporal comparisons of ozone concentration and its new trends based on long-term data from more sites.

Figure 7: Comparison of predicted versus observed ozone concentration at Marylebone roadside site using AQRM $R^1 = 0.88$ and OLS model $R^2 = 0.32$ for April 2008.
ACKNOWLEDGEMENT
We gratefully acknowledge Economic and Social Research Council (ESRC) for providing funding for this study.

REFERENCES


