**Atmospheric
Chemistry
and Physics**

Modelling trends in OH radical concentrations using generalized additive models

L. S. Jackson[1,*], N. Carslaw[1], D. C. Carslaw[2], and K. M. Emmerson[1,*]

[1]Environment Dept., University of York, York, YO10 5DD, UK
[2]Institute for Transport Studies, University of Leeds, Leeds, LS2 9JT, UK
[*]now at: School of Earth and Environment, University of Leeds, Leeds, LS2 9JT, UK

Abstract. During the TORCH campaign a zero dimensional box model based on the Master Chemical Mechanism was used to model concentrations of OH radicals. The model provided a close overall fit to measured concentrations but with some significant deviations. In this research, an approach was established for applying Generalized Additive Models (GAM) to atmospheric concentration data. Two GAM models were fitted to OH radical concentrations using TORCH data, the first using measured OH data and the second using MCM model results. GAM models with five smooth functions provided a close fit to the data with 78% of the deviance explained for measured OH and 83% for modelled OH. The GAM model for measured OH produced substantially better predictions of OH concentrations than the original MCM model results. The diurnal profile of OH concentration was reproduced and the predicted mean diurnal OH concentration was only 0.2% less than the measured concentration compared to 16.3% over-estimation by the MCM model. Photolysis reactions were identified as most important in explaining concentrations of OH. The GAM models combined both primary and secondary pollutants and also anthropogenic and biogenic species to explain changes in OH concentrations. Differences identified in the dependencies of modelled and measured OH concentrations, particularly for aromatic and biogenic species, may help to understand why the MCM model predictions sometimes disagree with measurements of atmospheric species.

1 Introduction

Advances in our knowledge of atmospheric chemistry are critical to developing effective policy measures for air quality; a major issue for human health, the growth of crops and natural vegetation. Environmental policies to address air quality issues have burgeoned worldwide since the 1950s. It is estimated that the European Community strategy on air pollution defined within the 6th Environmental Action Programme (European Parliament and Council, 2002) will reduce premature deaths in Europe by over 60 000 per year by 2020 (Bower et al., 2006).

The hydroxyl (OH) radical is one of the most important atmospheric chemical species. It participates in reactions with many longer lived chemical species including volatile organic compounds (VOCs), carbon monoxide, NO_X and ozone. OH radical reactions also lead to the production of other reactive species like the hydroperoxy (HO_2) and organic peroxy (RO_2) radicals (Fig. 1). OH plays a key role in photochemical reactions being both produced and destroyed in reaction cycles which contribute to the production of ozone in polluted atmospheres. Understanding the behaviour of the OH radical and its interactions with other chemical species is crucial to improving the accuracy of atmospheric models.

Recent advances in the understanding of OH radical chemistry have been developed from measurements of OH concentrations in different atmospheric environments, ranging from urban to extremely clean (e.g. see review by Heard and Pilling, 2003). In parallel, zero-dimensional box models constrained by observations of longer lived chemical species, have been employed to model the evolution of the atmospheric chemistry and provide insight into the chemical processing, for example see Emmerson et al. (2005, 2007).
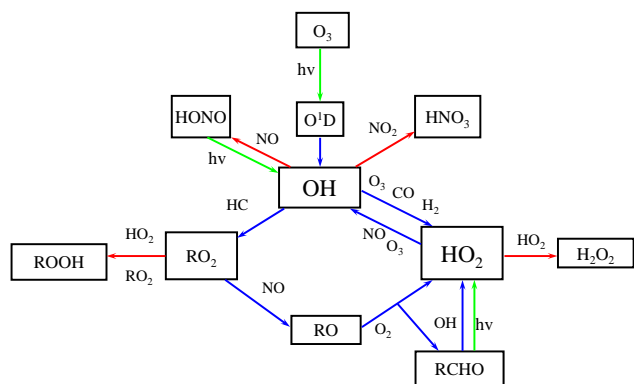
Fig. 1. Schematic of the key reactions of the OH radical in the atmosphere. The green lines show radical initiation, the red lines radical termination and the blue lines represent propagation reactions between radical species. RO_2 represents a generic peroxy radical, RO an oxy radical, ROOH an organic peroxide species and RCHO an aldehyde species.

The Tropospheric Organic Chemistry campaign (TORCH) took place during the summer of 2003 at a rural site about 25 miles north east of central London. The site at Writtle in Essex was surrounded by crop-based agriculture (sunflowers and grain). During the highly instrumented campaign, there were a large number of measurements made of long-lived and radical species as well as meteorological and aerosol parameters (Lee et al., 2006). The wealth of measurements allowed a zero dimensional box model based on a highly detailed chemical mechanism (MCM v3.1, Jenkin et al., 2003; Saunders et al., 2003) and constrained by longer-lived measurements, to be used to predict radical concentrations for comparison with measurements (Emmerson et al., 2007). On average, measured hydroxyl radical concentrations were over-predicted by 24%, those of the hydroperoxy (HO_2) over-predicted by 7%, and the sum of peroxy radicals (HO_2+RO_2) under-predicted by 22% (Emmerson et al., 2007). Although there was good overall agreement achieved between modelled and measured OH concentrations, there were also large differences on individual days (Emmerson et al., 2007).

There are a number of possible reasons for the observed discrepancies. There were gaps in data set owing to limitations on the number of atmospheric chemical species that could be measured. Potentially important omissions include biogenic hydrocarbons like the monoterpenes, larger (greater than C_{10}) hydrocarbons and nitrous acid (HONO). There are also experimental uncertainties. The spectroscopic method, Fluorescence Assay by Gas Expansion (FAGE) (Creasey et al., 2003; Heard and Pilling, 2003), was used to measure OH radical concentrations in the TORCH experiment. Uncertainty due to variation in FAGE measurement errors has

been estimated, at one standard deviation, to be 30% of the measured concentration (Smith et al., 2006). The measurement data used to constrain the model (for example NO_X, NMHC, O_3 etc.) also have associated uncertainties.

Areas of uncertainty also remain in the understanding of atmospheric chemistry and its representation in the MCM v3.1 mechanism. The MCM uses large numbers of rate coefficients, which in many cases have been estimated from analogous reactions rather than being evaluated directly from experimental results. There are almost certainly further unknown gaps and limitations in the current knowledge of OH radical chemistry and its representation in models. These experimental errors and areas of uncertainty will also contribute to discrepancies between measured OH radical concentrations and modelled concentrations.

An interesting observation reported in several papers is that the modelled and measured OH concentration dependencies on other species were found to differ. For instance, Emmerson et al. (2007) showed that the modelled HO_2:OH ratio had a stronger relationship with NO than measured during the TORCH campaign. Modelled HO_2:OH ratios also showed a stronger relationship with NO than measured in the "Pollution of the Urban Midlands Atmosphere" (PUMA) summer campaign of 1999 (Emmerson et al., 2005), in the "$PM_{2.5}$ Technology Assessment and Characterization" (PM-TAC) study (Ren et al., 2003) and during the "Berlin Ozone Experiment" (BERLIOZ) (Konrad et al., 2003). The stronger dependence on NO concentrations for model results compared to measurements could indicate that some of the rate coefficients tied up with NO_X concentrations in the models may be in error. This feature is a major motivation for the current research: by identifying differences between modelled and measured radical concentration dependencies on other species, it may be possible to begin to understand why models and measurements of atmospheric species sometimes disagree with each other.

In order to investigate any such potential differences, Generalized Additive Models (GAMs) (Hastie and Tibshirani, 1990) have been employed. GAMs are an extension of generalized linear models (McCullagh and Nelder, 1989), and are a flexible statistical tool useful for fitting non-parametric relationships whilst retaining clarity of interpretation. The relationship between a response variable and selected predictor variables is expressed as the sum of a number of non-parametric predictor variable functions. Such models have proven useful for studying the complex non-linear relationships that exist between atmospheric chemical species. They have been applied to modelling nitrogen dioxide concentrations (Carslaw and Carslaw, 2007; Carslaw et al., 2007; Westmoreland et al., 2007), and those of benzene and 1,3-butadiene (Reiss, 2006). In these examples, the methodology has been employed to standardise data gathered under variable meteorological conditions adjusting for the non-linear effects involved. Such an adjustment facilitates investigation of the underlying trends in pollutant concentrations.

In this research, GAMs were used to construct statistical models for measured and modelled OH radical concentrations (response variables). Predictor variables were selected from measurements of meteorological parameters and concentrations of primary or secondary pollutants made during the TORCH campaign and used to constrain the chemical box model. This process permits identification of predictor variables which strongly influence the modelled or measured OH concentrations. In particular, it may be possible to infer sources of error in the MCM model, perhaps in the various rate coefficients that have been estimated. Further, it may be possible to reduce the number of input parameters needed to predict real atmospheric measurements. Finally, it may allow prediction of OH concentrations (and other species) where measurements are unavailable.

2 Methodology

2.1 Data

During the TORCH campaign, there were 59 model input constraints: Concentrations of 39 VOC species, NO, NO_2, CO, O_3, PAN, H_2O, 9 photolysis coefficients (of O_3 (to form $O(^1D)$), NO_2, H_2O_2, HCHO (to form HO_2), HCHO (to form H_2), HNO_3, HONO, acetaldehyde and acetone) and 5 physical parameters (temperature, density of air "M", aerosol surface loss rate of HO_2, measure of cloudiness and aerosol surface area). There were 1014 data points with coincident measurements of the relevant 59 parameters plus modelled and measured OH concentrations. Data were available for an initial period of 9 days and a second period of 13 days with the break in the middle due to instrument failure.

All input parameters were either averaged or interpolated to give 15-min input values as described in Emmerson et al. (2007). Further manipulation of these data points was necessary to remove outliers, which may unduly bias the GAM construction, leaving 933 points for further analysis. The mean values for key input constraints in the TORCH campaign are summarised in Table 1.

2.2 Generalized additive models

Separate GAMs were produced for modelled and measured OH concentrations. Non-parametric relationships between response and predictor variables were expressed in terms of smooth functions (smooths) developed using thin plate regression splines (Wood, 2006). The statistical software R (version 2.5.0 for Windows) was used for all calculations, with the integrated mgcv package (version 1.3–23) being used to produce the GAMs (R Development Core Team, 2007).

Figure 2 shows an example of a smooth function for a simple model relating [OH] with photolysis rate. The relation-
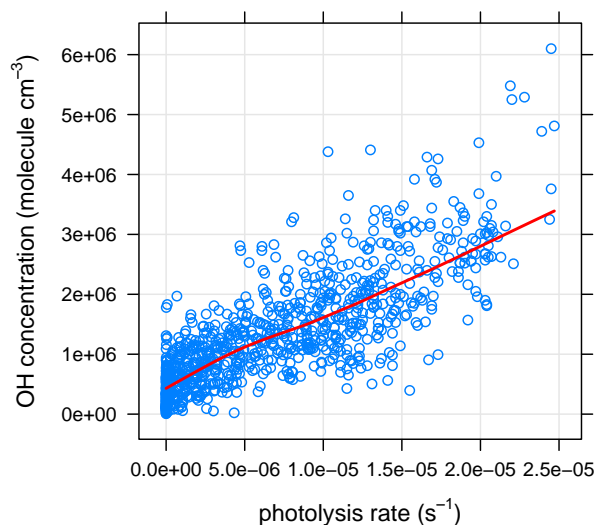
Fig. 2. Specimen smooth for OH concentration (molecule cm^{-3}), the response variable, plotted against the photolysis rate of O_3 (to form $O(^1D)$) (s^{-1}), the predictor variable. The original data is shown with blue circles and the smooth as a red line.

ship between the ith observation in the data, smooth function $s()$, constant a, and residual error ϵ_i is represented by:

$$[OH]_i = a + s(\text{photolysis rate}_i) + \epsilon_i \tag{R1}$$

For a model with n smooth functions (predictor variables) this relationship generalises to:

$$C_i = \sum_{j=1}^{n} s_j(x_i) + a + \epsilon_i \tag{R2}$$

The i-th concentration in the time series is C_i. $s_j(x_i)$ is the smooth for the j-th variable and gives the value of this smooth for the i-th observation. ϵ_i is the residual error for this observation and a is a constant.

An iterative process was used to select the predictor variables. New variables were added one at a time and the variable that maximised the level of deviance explained was retained:

Step 1 – The first variable was selected. Single variable GAMs were run for all variables and ranked in order of deviance explained; the proportion of the variance in OH data explained by a GAM model. The variable with the highest level of deviance explained was chosen as the first variable for the model.

Step 2 – The next variable was selected. Each of the remaining variables were added in turn to the one variable model from Step 1 and the deviance explained re-calculated. The additional variable that produced the highest level of deviance explained was then selected.

Step 3 – Confirm variable choice from Step 1. The variable selected in Step 1 was removed and each of the remaining variables

Table 1. Key input constraints for TORCH. (a) Mean value of all 933 data points. (b) Mean diurnal value calculated using the mean values of observations recorded at the same time of day.

Input constraint	Mean value	
	(a)	(b)
Modelled OH (molecule cm^{-3})	1.63×10^6	1.10×10^6
Measured OH (molecule cm^{-3})	1.32×10^6	9.46×10^5
CO (ppb)	184	191
NO (ppb)	3	3
NO$_2$ (ppb)	9	10
O$_3$ (ppb)	32	28
Ethane (ppb)	2	2
Ethene (ppt)	432	512
Acetaldehyde (ppb)	2	2
Methanol (ppb)	1	1
Acetone (ppb)	1	1
Isoprene (ppt)	115	77
Benzene (ppt)	124	147
Photolysis rate parameter (s^{-1})		
O$_3$ to O (^{1}D)	6.81×10^{-6}	4.03×10^{-6}
NO$_2$ to NO+O (^{3}P)	3.16×10^{-3}	1.95×10^{-3}
Mean Temperature (K)	293.5	291.5

were added in turn to the variable selected in Step 2. The variable that gave the highest level of deviance explained was selected. If it was a different variable to Step 1, the combination of two variables that gave the highest level of deviance explained was chosen.

Step 4 – Collinearity was tested. Collinearity occurs when two predictor variables have a near perfect linear relationship. Its presence in a regression model makes the contribution of each individual variable difficult to discern, introduces redundancy and it can cause results to be overly sensitive to changes in data. Collinearity was tested using the variance inflation factor (VIF) for each predictor variable:

$$\text{VIF} = \frac{1}{1 - R_j^2} \tag{R3}$$

where R_j^2 is the coefficient of determination from a linear regression of variable j in the model against the other variables (Freund and Wilson, 1998). A maximum value of five was accepted for the variance inflation factor (Montgomery and Peck, 1992). If this value was exceeded, the collinear variable making the least contribution in terms of deviance explained was removed from the model.

Step 5–Steps 2 to 4 were repeated. The process was repeated until an additional variable increased the deviance explained by less than one percent or a maximum of five variables was achieved. The limit of five variables was imposed to control the complexity of the resulting GAM and facilitate interpretation of the results. For both measured and modelled OH GAM models, an adequate fit to the data was achieved with five variables and adding a sixth variable yielded a minor improvement in the fit.

Step 6 – Robustness of the selected model was checked. Each variable was checked for a p-value significant at the 0.1% significance level. The sensitivity of results to changes in parameters in the smoothing process was also checked. The residuals were checked to confirm that their distribution was approximately a normal distribution with zero mean and that they exhibited no clear relationship with the predictor variables or fitted values. The Durbin-Watson test statistic was used to test the independence of GAM residuals (Chatfield, 1992). Autoregression was investigated using correlograms for measured OH concentrations, modelled OH concentrations and GAM residuals.

3 Results

To avoid confusion in terminology the following conventions have been used throughout: "modelled data" or "model results" refer to results from the MCM box models; "measured OH radical concentrations" refer to the concentrations measured during the TORCH campaign; the generalized additive models are referred to as GAMs and their results as GAM results; lastly "GAM$_{ME}$" and "GAM$_{MO}$" are GAMs for the TORCH measured and modelled OH concentrations, respectively.

3.1 TORCH Measured OH GAM$_{ME}$

The GAM produced for the TORCH measured OH data comprises a constant intercept and five smooths as shown in Table 2.

Table 2 shows the deviance explained by each variable: the photolysis rate for nitric acid was the most important predictor variable. The p-value associated with each of the smoothed terms provides a measure of the significance of the relationship for each predictor variable. All of the parameters were significant in explaining variation in measured OH concentrations at the 0.1% significance level. Collinearity was not a concern as the variance inflation factors were less than the prescribed limit of five for all variables. The value for the intercept was defined as the mean value of the measured OH radical concentrations, 1.32×10^6 molecule cm^{-3}. Including this intercept improved the fit of the models to the data and simplified interpretation of the smooth functions which represent variations from this mean value. When the constant intercept was removed from the model, the deviance explained reduced to just 24.7%.

As well as assessing the significance of the predictor variables through p-values, the relationship between the smoothed function of each predictor variable and the OH concentration was also explored (Fig. 3).
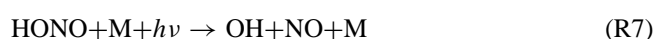
The photolysis reaction selected for GAM$_{ME}$ was

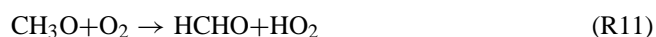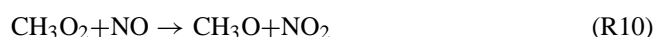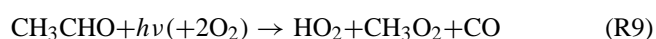$$HNO_3 + h\nu \rightarrow OH + NO_2 \tag{R4}$$

Figure 3a shows that the value of the smooth increases almost linearly with increasing photolysis rate for nitric acid. The increasing gradient at high photolysis rates should be

treated with caution owing to the limited amount of data as shown by the rug plot. The negative values of the smooth occur in periods when photolysis rates and OH radical concentrations were both low, for example at night-time.

Whilst the photolysis rate for reaction (R4) produces the highest level of deviance explained, all of the photolysis variables have similar explanatory power and all would have been suitable for including in the model. In terms of interpretation, therefore, emphasis is placed on the role of photolytic reactions in general which act as sources of OH radicals rather than the role of the specific nitric acid photolysis reaction. The linear relationship (Fig. 3a) reflects the formation of OH radicals through a number of reactions when photolysis is important. For instance, OH can also be formed through photolysis of O_3 and HONO:

$$O_3+h\nu \rightarrow O^1D+O_2 \text{ at wavelengths}<320\,nm \quad (R5)$$

$$O^1D+H_2O \rightarrow 2OH \quad (R6)$$

$$HONO+M+h\nu \rightarrow OH+NO+M \quad (R7)$$

and also indirectly as peroxy radicals are formed e.g. through the photolysis of carbonyl species (formaldehyde (R8) and acetaldehyde (R9)):

$$HCHO+h\nu(+O_2) \rightarrow CO+2HO_2 \quad (R8)$$

$$CH_3CHO+h\nu(+2O_2) \rightarrow HO_2+CH_3O_2+CO \quad (R9)$$

$$CH_3O_2+NO \rightarrow CH_3O+NO_2 \quad (R10)$$

$$CH_3O+O_2 \rightarrow HCHO+HO_2 \quad (R11)$$

$$HO_2+NO \rightarrow OH+NO_2 \quad (R12)$$

The large magnitude spanned by the y-axis in Fig. 3a also confirms that the photolysis variable is the most important in explaining the deviance in measured OH concentrations.

o-xylene is an aromatic hydrocarbon, a primary pollutant principally emitted by road vehicles through incomplete fuel combustion and fuel evaporation. During the day the degradation of o-xylene is typically initiated through reaction with OH radicals, with o-xylene having a lifetime with respect to OH of approximately 20 h based on the average diurnal OH radical concentration during TORCH.

As the o-xylene concentration increases (Fig. 3b), the OH concentration decreases, a relationship that would be expected with any primary emitted species (i.e. o-xylene has no secondary sources in the atmosphere). For o-xylene concentrations greater than about 8×10^8 molecule cm^{-3}, the relationship with OH is broadly flat. The start of the plateau region may indicate that the atmosphere switches from VOC dependence to NO$_X$ dependence, and so any further increases in o-xylene concentration beyond this point have little impact on OH concentrations. The change in shape beyond a concentration of 3×10^9 molecule cm^{-3} should not be

Table 2. TORCH measured OH (GAM$_{ME}$) results comprising the variables selected, deviance explained, p-values for each variable and variance inflation factors. Deviance explained is shown as the cumulative total for each variable and the preceding variables.

Variable name	Deviance explained (%)	p-value	VIF
Intercept	0.0	$<2\times10^{-16}$	n/a
Photolysis rate of HNO$_3$	67.6	$<2\times10^{-16}$	1.2
o-xylene	72.3	$<2\times10^{-16}$	1.2
Acetone	74.1	$<2\times10^{-16}$	1.2
Formaldehyde	75.7	2.15×10^{-13}	1.0
Water	77.9	3.11×10^{-13}	1.2

over-interpreted due to the paucity of data and wide confidence intervals.

There is also a fairly simple relationship with acetone (Fig. 3c). With increasing concentrations of acetone, the smooth initially increases in value indicating an associated increase in OH radical concentration. Beyond about 2.5×10^{10} molecule cm^{-3} the smooth is broadly flat indicating no further impact on OH radical concentrations. Changes in shape at high concentrations of acetone should be treated with caution due to the limited volume of data on which the smooth is based.

Acetone is formed in the atmosphere through primary emissions from various biogenic and anthropogenic sources and also through the oxidation of VOCs (Goldstein and Schade, 2000). For instance, the oxidation of propane and branched chain alkanes, branched chain alkenes and oxygenated species all contribute to acetone production in the atmosphere. Biogenic sources include α- and β-pinene and 2-methyl-3-buten-2-ol, which have acetone yields of 8–15% and 50% respectively (Goldstein and Schade, 2000).

Once emitted or formed, there are a number of ways that acetone can influence OH concentrations in the atmosphere. Firstly, it can react with OH through the following reaction:

$$OH+CH_3COCH_3+O_2 \rightarrow CH_3COCH_2O_2+H_2O \quad (R13)$$

The peroxy radical formed in R13 can undergo reaction with NO, other RO$_2$ or HO$_2$ to form a variety of products. However, under the TORCH conditions, reaction with NO is most likely (Emmerson et al., 2007):
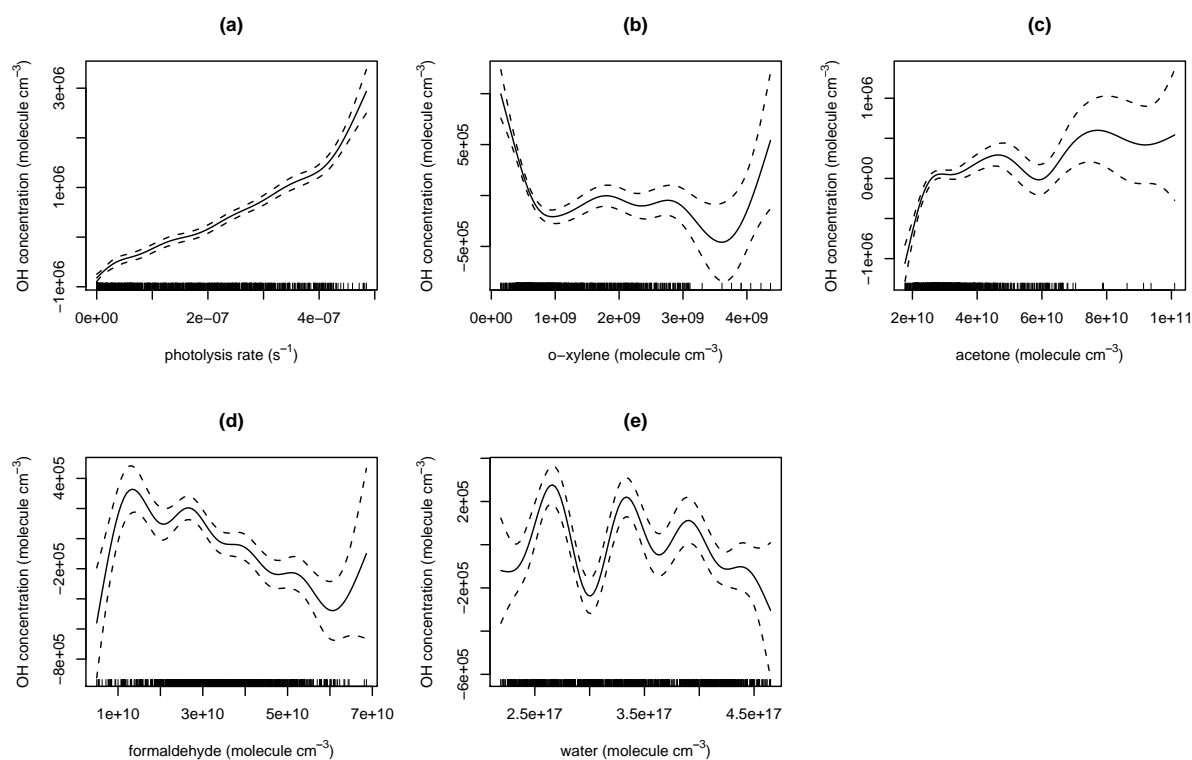
$$CH_3COCH_2O_2+NO \rightarrow CH_3COCH_2O+NO_2 \quad (R14)$$
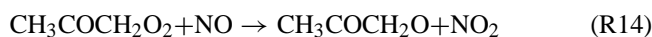
The CH$_3$COCH$_2$O radical decomposes very quickly to the peroxyacetyl (CH$_3$CO$_3$) radical and formaldehyde:

$$CH_3COCH_2O \rightarrow CH_3CO_3+HCHO \quad (R15)$$
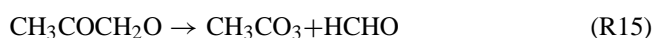
The HCHO feeds back to OH through R8 and R12. The CH$_3$CO$_3$ radical reacts with NO to form CH$_3$O$_2$ (R16), which itself feeds back to OH through R10–R12:

$$CH_3CO_3+NO+O_2 \rightarrow CH_3O_2+NO_2+CO_2 \quad (R16)$$
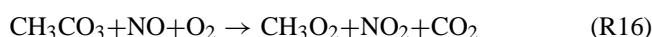
Fig. 3. Smooth functions of the predictor variables for measured OH **(a)** photolysis rate, **(b)** *o*-xylene, **(c)** acetone, **(d)** formaldehyde, **(e)** water. The concentrations of the predictor variables are shown along the x-axes and the value of the smooth function on the y-axis. The distribution of the data is indicated by the vertical marks along the x-axis (rug plot). Confidence intervals at 95% are shown by dotted lines.

Alternatively, the acetone can undergo photolysis to form CH_3O_2 and CH_3CO_3 radicals:

$$CH_3COCH_3 + h\nu \rightarrow CH_3CO_3 + CH_3O_2 \qquad (R17)$$
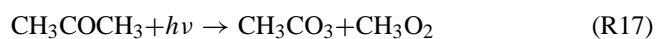
These radicals will then regenerate OH through R16, and R10–R12. The photolysis of acetone is therefore a radical and ultimately an OH source. The impact on OH of its reaction with acetone will depend on the acetone and NO concentrations. For the data used to construct the GAM, the destruction rate of acetone through photolysis is 5.0×10^3 molecule $cm^{-3} s^{-1}$ on average, compared with an OH destruction rate of 9.2×10^3 molecule $cm^{-3} s^{-1}$. Note that the data used to construct the GAM include both daytime and night-time data and so the photolysis source would be more important if just the daytime hours were considered. Warneck (2001) has calculated that, globally the photolysis sink of acetone is three times more important than that of OH.

However, both of these destruction rates become insignificant when considering the production rate of acetone through VOC oxidation. Although it is not possible to know for sure the production rate of acetone in the atmosphere during TORCH, particularly given that the biogenic species were not measured, a lower limit can be estimated from the relevant measured hydrocarbons that

go on to form acetone when they degrade. These are i-butene, i-pentane, propane, *o*-xylene, 2-methyl-2-butene, i-butane, 2-methylpentane, propanol and 2,2-dimethylbutane. The combined production rate from these 9 species is 1.2×10^5 molecule $cm^{-3} s^{-1}$, almost a factor of 10 more important than the combined destruction rates through reaction with OH and photolysis, despite this rate being a lower limit for reasons stated above. There is net production of acetone during TORCH, therefore, and the behaviour observed in Fig. 3c can be explained through OH and acetone being formed through the same degradation processes in the atmosphere.

Figure 3d shows the profile of HCHO with OH. The tails of the smooth should be interpreted with caution due to the paucity of data and wide confidence intervals. Formaldehyde is both a primary pollutant (directly emitted from combustion processes for instance) and a secondary pollutant (from the degradation of hydrocarbons). There are two competing reactions to consider for HCHO destruction, photolysis (R8) and reaction with OH (R18):

$$HCHO + OH(+O_2) \rightarrow CO + HO_2 + H_2O \qquad (R18)$$

Under the conditions considered, the average rate of R8 was 2.9×10^5 molecule $cm^{-3} s^{-1}$, whilst that of R18 was 4.5×10^5 molecule $cm^{-3} s^{-1}$. Therefore, reaction with OH

is ~ 1.5 times more important than photolysis in terms of H CH O destruction. As the H CH O concentration increases, the rate of R18 will increase, which may be expected to suppress the OH concentration. However, R8 and R18 produce HO_2 radicals, which will subsequently reform OH through R12. These latter processes would suppress the observed decrease in OH concentrations. The competing and related processes for these species make specific interpretation of the GAM results difficult, particularly for secondary pollutants. For most of the profile where there are more data points, it appears that as the HCHO concentration increases that of OH decreases. The smooth, therefore, indicates that the destruction of formaldehyde through reaction with OH radicals is more important than formaldehyde regeneration through hydrocarbon oxidation.

Water plays a complex role in atmospheric chemistry affecting the rates of a wide range of reactions and is present in relatively large concentrations. The smooth reflects this complexity (Fig. 3e) with increases in water concentration associated with both increasing and decreasing OH radical concentrations. Owing to the formation of OH through R5 and R6, it may be expected that as the water concentration increases, so does that of OH. Indeed, at lower concentrations of H_2O, the OH does increase with H_2O. However, as the H_2O concentration increases past about 2.7×10^{17} molecule cm^{-3}, the OH concentration begins to decrease. Such variations probably indicate that there are physical processes occurring that are impacting the OH, perhaps to do with temperature effects, or aerosol formation, that are not immediately obvious. These variations are also much less significant in explaining variations in OH radical concentration than contributions from other smooth functions. Variations in the smooth function for photolysis rate, for example, are almost an order of magnitude greater than variations in the smooth for water.

3.2 TORCH modelled OH (GAM$_{MO}$)

The GAM produced for the TORCH modelled OH data comprises a constant intercept and five smooths as shown in Table 3.

The value of the intercept is 1.63×10^6 molecule cm^{-3}, the mean value of modelled OH. Removing the intercept from the GAM reduces the deviance explained from 83.1% to 27.6%. The observed relationship between the smoothed function of each predictor variable and the OH concentration is shown in Fig. 4. Collinearity was not a concern as the variance inflation factors were less than the prescribed limit of five for all variables.

The photolysis rate included in GAM$_{MO}$ represents the following reaction:

$$CH_3CHO + h\nu(+O_2) \rightarrow HO_2 + CH_3O_2 \qquad (R19)$$
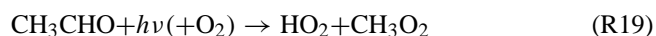
As for measured OH, the smooth increases with photolysis rate in a broadly linear manner (Fig. 4a). As with GAM$_{ME}$, the specific photolysis reaction is not particularly important

Table 3. TORCH modelled OH (GAM$_{MO}$) results comprising the variables selected, deviance explained, p-values for each variable and variance inflation factors. Deviance explained is shown as the cumulative total for each variable and the preceding variables.

| Variable name	Deviance explained (%)	p-value	VIF
Intercept	n/a	$<2\times10^{-16}$	n/a
Photolysis rate of CH_3CHO	68.3	$<2\times10^{-16}$	1.2
PAN	70.8	$<2\times10^{-16}$	1.4
Carbon monoxide	76.5	$<2\times10^{-16}$	2.8
Isoprene	80.6	$<2\times10^{-16}$	1.2
Ethanol	83.1	$<2\times10^{-16}$	3.3

as all photolysis variables had a similar level of explanatory power and similar smooths. The important point is that as photolysis rates increase, so do OH concentrations consistent with photolytic reactions acting as a source of OH radicals.

Peroxyacetyl nitrate (PAN) is a secondary pollutant produced from the reaction of the CH_3CO_3 radical with NO_2 (Baird and Cann, 2005):

$$CH_3CO_3 + NO_2 \rightleftharpoons CH_3CO_3NO_2 \qquad (R20)$$

CH_3CO_3 is produced through the degradation of many hydrocarbons in the atmosphere. The formation of PAN generally peaks in the late afternoon when concentrations of hydrocarbons and OH radicals are relatively high and concentrations of nitric oxide are relatively low (as NO also reacts with the CH_3CO_3 radical, R16). Figure 4b shows that PAN behaves as a typical secondary pollutant, increasing as the concentration of OH increases, showing that both of these species are formed through photochemical processes. The smooth flattens as the PAN concentrations increase beyond 1.0×10^{10} molecule cm^{-3}, although there are fewer data here and the smooth should be interpreted with caution.

Carbon monoxide plays a significant role in OH radical chemistry. It can react with OH radicals to produce HO_2 radicals (Fig. 1), but is also produced as a by-product of reactions involving the OH radical. For example, the degradation of hydrocarbons by OH radicals can lead to the formation of formaldehyde which produces carbon monoxide on reaction with OH radicals (R18) or when broken down through photolysis (R8). Carbon monoxide is also a primary pollutant emitted from vehicle engines on incomplete combustion of hydrocarbon fuels. Such pollutants might be expected to display a mix of primary and secondary pollutant behaviour. The concentration of CO increased during each day due to the build-up of primary emissions and secondary production during the TORCH campaign. Low CO concentrations would be expected to be associated with low OH radical concentrations as both occur in the early morning. The shape of the smooth in Fig. 4c shows that there is a complex relationship between CO and OH for the reasons stated above.
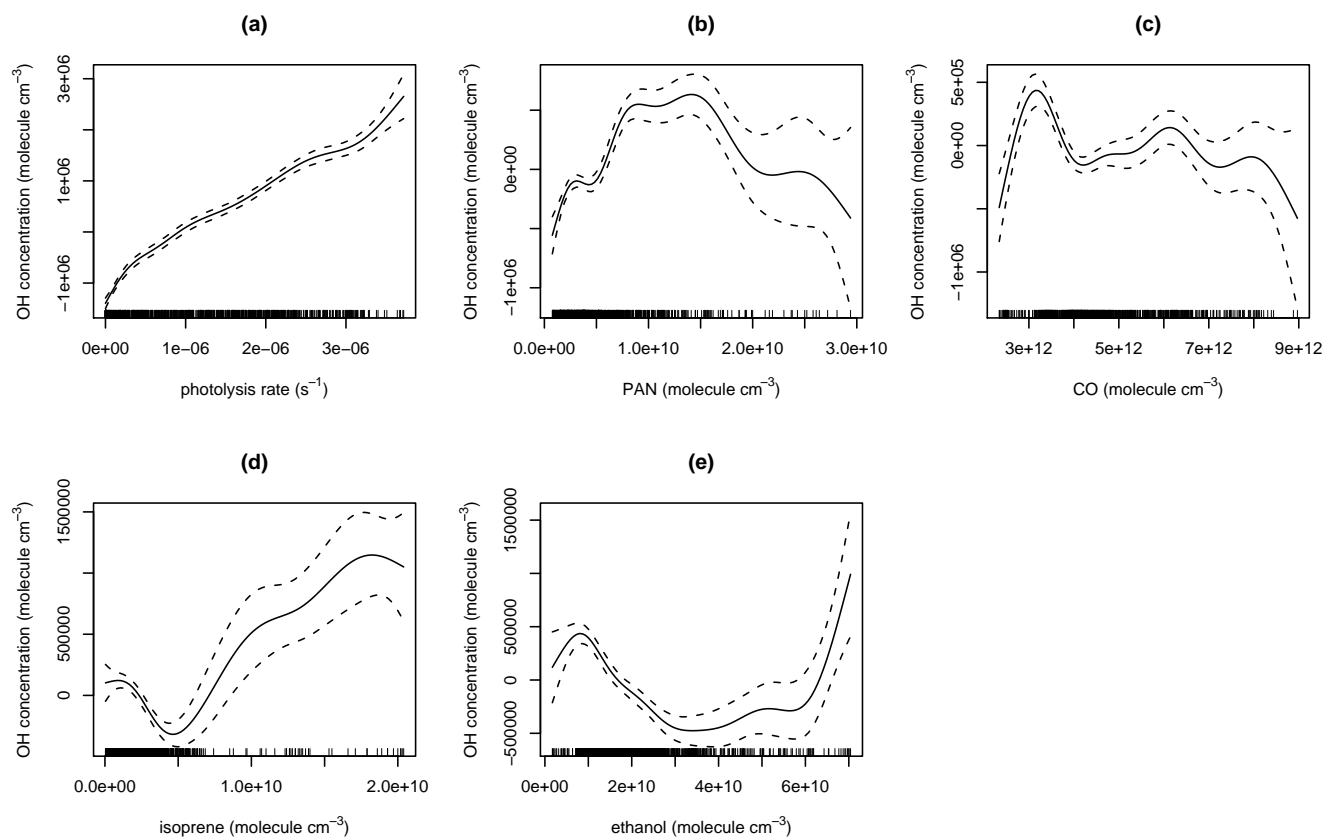
Fig. 4. Smooth functions of the predictor variables for modelled OH **(a)** photolysis rate, **(b)** peroxyacetyl nitrate (PAN), **(c)** carbon monoxide (CO), **(d)** isoprene, **(e)** ethanol. The concentrations of the predictor variables are shown along the x-axes and the value of the smooth function on the y-axis. The distribution of the data is indicated by the vertical marks along the x-axis (rug plot). Confidence intervals at 95% are shown by dotted lines.

Isoprene is a hydrocarbon whose rate of emission increases strongly with increases in temperature. It behaves as a primary pollutant with emissions from both plants and road traffic. Its fate in the environment is degradation through reaction with OH and NO_3 radicals, as well as through reaction with ozone. This behaviour is reflected in the smooth in Fig. 4d which shows reducing OH values with increasing isoprene concentrations where the majority of the data exist. The higher values for the smooth result mainly from late morning and early afternoon observations on a single day when OH radical concentrations were particularly high, probably due to high photolysis rates.

Ethanol is a primary pollutant; a volatile organic hydrocarbon emitted through solvent usage and other industrial processes. Its fate in the atmosphere is to react with OH radicals producing a range of secondary pollutant products such as acetaldehyde. The smooth (Fig. 4e) fits the shape of a primary pollutant with decreasing values associated with increasing concentration of ethanol as the ethanol acts as a sink for OH radicals.

3.3 Prediction of OH radical concentrations

GAM_{ME} and GAM_{MO} were validated by using them to predict OH radical concentrations. A bootstrapping approach was used which involved 1000 repeated samples from the data. Each sample, comprising 75% of the data, was used to calibrate the GAM model by re-calculation of the constant parameter and smooth functions. The calibrated GAM was then used to predict OH concentrations for the remaining 25% of the data. Figure 5 shows that the predicted values from GAM_{ME} and GAM_{MO} closely approximated the diurnal distribution of measured OH and modelled OH concentrations.

The GAM_{ME} model produced a superior fit to measured OH concentrations than the MCM model (Table 4). GAM_{ME} produced a practically unbiased estimate of the mean diurnal OH concentration whereas the MCM model over-estimated by 16.3%. The root mean squared error (RMSE) of GAM_{ME}, as a proportion of the mean diurnal concentration, was appreciably lower than the RMSE of the MCM model.

The accuracy of predictions produced by GAMs may be compromised by the presence of autocorrelation.
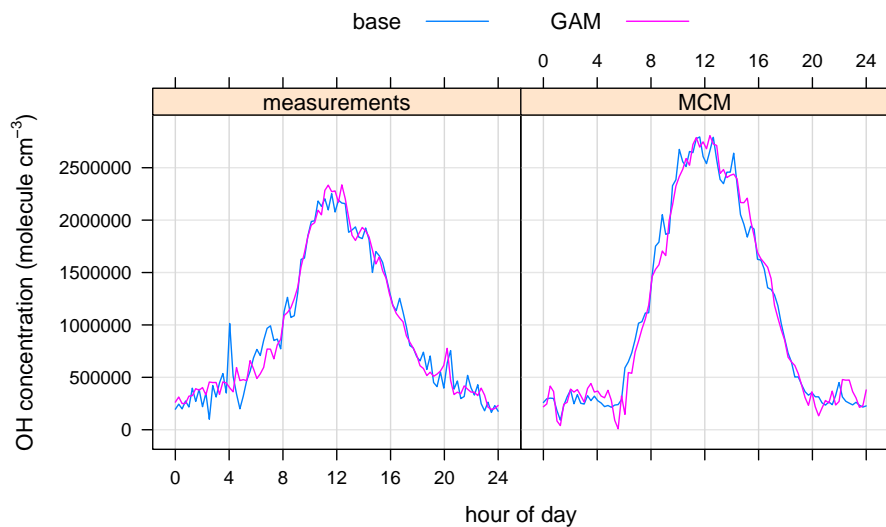
Fig. 5. Comparison of the measured (or modelled) OH concentration (shown as a blue line) against the predicted concentration (magenta line) produced using GAM_{ME} (or GAM_{MO}) calibrated by bootstrapping the data.

Examination of correlograms for measured OH on each day showed them to be autocorrelated and was confirmed by the Durbin-Watson test. The autocorrelation coefficient at lag 1 was 0.45 for the residuals of GAM_{ME} compared to 0.85 for the measured OH data. For the residuals of GAM_{MO} and modelled OH data, the autocorrelation coefficients at lag 1 were 0.54 and 0.93, respectively. Fitting auto-regression models (AR) to the measured OH data for each day and minimising the Akaike Information Criterion provided an estimate of the order of each AR process. The order was found to be 1 on 15 of the 22 days with values ranging from 0 to 8 on other days. The residuals for GAM_{ME} also exhibited autocorrelation with the estimated order of the AR processes varying between 0 and 8 for individual days. The GAM_{ME} model only eliminated autocorrelation in 6 of the 21 days on which it was originally present in the measured OH data.

4 Discussion

The shape of the smooths provides a general description of the role played by an individual variable. All the photolysis variables exhibited an approximate positive linear relationship with OH radical concentrations. The primary pollutants broadly displayed an inverse relationship between the concentration of the pollutant and the concentration of OH radicals. At low concentrations, the secondary pollutants generally showed a positive correlation between pollutant concentration and OH radical concentration. At medium to high concentrations the shape of the smooth generally became flat around the zero level. Whilst the shapes of the functions can provide an indication of the behaviour of a chemical, care

must be taken not to over-interpret the results especially for parts of the curves based on limited data.

Table 5 shows that both models have a photolysis variable which is the most important in terms of deviance explained (67–69 %). It is so important that a GAM model comprising only the photolysis variable produced predictions of measured OH concentration almost as good as the five variable GAM_{ME} model (see Table 4). The deviance explained by all photolysis variables was very high compared to other variables. There was also little difference between the variable selected for a GAM and the remaining photolysis variables. Photolysis is clearly important as it is a strong source for OH radicals both directly and indirectly. This is consistent with the strong correlation between measured OH concentrations and observed $J(O^1D)$ photolysis rates recorded at the Meteorological Observatory Hohenpeissenberg in Southern Germany (Rohrer and Berresheim, 2006). As well as having a photolysis variable, both models included a mixture of primary and secondary pollutants. Interestingly, the two GAMs contained no common species.

The location of the TORCH experiment was a semi-rural environment with significant atmospheric emissions from plants and trees. Such emissions would have been accelerated by the high temperatures experienced that summer, which may explain the presence of acetone and isoprene in these models, both of which can be emitted in part from biogenic sources. As with all of the variables, they do not solely represent the contribution of that named chemical but represent the role played by a range of species that exhibit broadly similar behaviour.

Table 4. Mean diurnal values and root mean squared error results for GAM and MCM models. GAM$_{ME}$ (photolysis) shows the results for GAM$_{ME}$ with only the photolysis variable and the constant term present.

TORCH	Mean diurnal value (molecule cm^{-3})	RMS error (molecule cm^{-3})
Measured concentration	9.46×10^5	n/a
MCM model	1.10×10^6	8.87×10^5
GAM$_{ME}$	9.44×10^5	1.39×10^5
GAM$_{MO}$	1.09×10^6	1.32×10^5
GAM$_{ME}$ (photolysis)	9.50×10^5	1.46×10^5

Table 5. GAM predictor variables classified by type.

GAM$_{ME}$	GAM$_{MO}$	Variable type
J (HNO$_3$)	J (CH$_3$CHO)	Photolysis rate
o-xylene	Isoprene	Primary pollutant
	Ethanol	Primary pollutant
Formaldehyde	Carbon monoxide	Primary/ secondary pollutant
Acetone	PAN	Primary /secondary pollutant
Water vapour		Meteorological data

So what can the results tell us? For the measurements, the results suggest that o-xylene, acetone, formaldehyde and water vapour are important, whilst for the model, isoprene, CO, PAN and ethanol affect the OH. Isoprene concentrations were remarkably high during the TORCH campaign, which explains the impact on the model results. However, its lack of impact on the measurements is surprising. This difference may reflect the fact that isoprene was just one of many biogenic species present during TORCH (unfortunately, no others were measured). It may only have represented a small proportion of the biogenic carbon present and as such, may only have had a small impact on the measured OH radicals. This conjecture is reinforced somewhat by the presence of acetone in GAM$_{ME}$, as acetone is a ubiquitous product of terpene degradation, and is a strong indicator for biogenic processing. The presence of acetone rather than isoprene in GAM$_{ME}$ could, also, indicate that the behaviour of isoprene is not representative of some other biogenic hydrocarbons.

Could efficient recycling of OH by isoprene, as observed by Lelieveld et al. (2008) over a tropical forest and not directly measurable using GAMs as they model net changes in concentrations, explain the omission of isoprene from GAM$_{ME}$ and possibly the flattening of the acetone smooth at higher concentrations? Modelled OH concentrations were not markedly different from measured OH at high levels of the pollutants NO or NO$_2$. For example, the ratio of modelled OH to measured OH increased from 123% for all data to 136% for OH associated with the highest quartile of NO concentrations. Higher measured OH concentrations would have been expected if there was significant recycling affecting OH measurements and not model calculations. Furthermore, the fit to data of GAM$_{ME}$ and GAM$_{MO}$ at high pollution levels was as robust as at low pollution levels. Again, this would not have been expected if OH recycling was strongly active at the high pollution levels. The deviance explained for GAM$_{ME}$ based on all data (77.9%) increased to 84.8% where NO concentrations were less than their median value and reduced to 75.3% where NO concentrations exceeded their median. The equivalent results for GAM$_{MO}$ were 86.1% and 88.0% respectively. The GAM ap-

proach is, however, limited by the quantity and quality of data available. The approach used would not have identified recycling of OH involving by-products which were not measured and which were also not correctly reproduced in the MCM model. Provided with additional data for an increased number of chemical species active in OH recycling the GAM approach would yield more conclusive results.

Another interesting difference is that the measurements show a dependence on o-xylene concentrations, whilst those of the model do not. There are known to be deficiencies with how the aromatic degradation schemes are represented in the MCM (Jenkin et al., 2003). The fact that the measured OH concentrations show a dependence on an aromatic species whilst the modelled OH values do not, may suggest that this area still needs work.

It is interesting that the difference in NO dependency for the modelled and measured HO$_2$/OH ratio, noted by Emmerson et al. (2007), was not prominent for OH in this study. Table 6 shows the deviance explained by GAM models using NO as the sole explanatory variable. Results for the log (HO$_2$/OH) ratio show a large discrepancy in the deviance explained between modelled and measured data. The shapes of the underlying smooth functions were also different; consistent with Emmerson et al. (2007). By contrast, neither modelled nor measured OH exhibited a strong statistical dependency on NO. Further, the root mean square of the difference between NO smooth functions for modelled and measured OH concentrations was found to be less than the equivalent result for the log (HO$_2$/OH) ratio. This indicates that the difference between modelled and measured OH dependencies on NO was not as marked as for (HO2/OH). For OH and HO2, there was also only a weak statistical association with NO concentrations for both modelled and measured data.

The fit of the GAM models to the data was good and the predictions of OH concentrations were an improvement over the MCM model. The model prediction approach using bootstrapping ensured that these predictions were based on data not used in the model construction reinforcing the credibility of the validated models. There remain limitations, however, with the results from the GAM models. Two sources of error were prominent in the model residuals. Firstly, extremely high OH concentrations were under-estimated by

Table 6. Deviance explained (%) by NO in single variable GAM models for OH, HO_2 and $\log(HO_2/OH)$.

GAM model	Modelled	Measured
OH$\sim$s (NO)	19.2	8.7
$\log (HO_2/OH)\sim$s (NO)	72.5	18.1
$HO_2\sim$s (NO)	21.1	1.5

Fig. 6. Measured OH concentration plotted against the residuals for GAM_{ME} (blue circles) and compared to the zero residual level (dashed line).
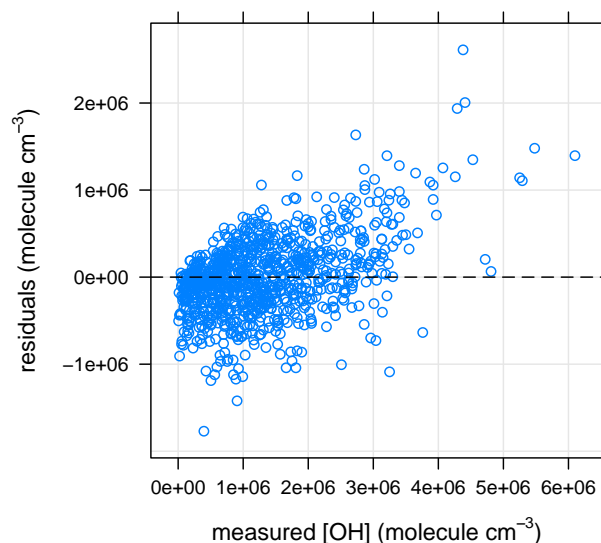
both GAM_{ME} and GAM_{MO}. Figure 6 shows that all observations greater than 4×10^6 molecule cm^{-3} were underestimated by GAM_{ME}. The discrepancy is substantial with the mean value of these observations being under-estimated by 27%. Observations at very high and also very low concentrations were not as accurately estimated as observations close to the mean level. The GAM models, therefore, appear to provide a less robust description of the atmospheric chemistry at these extremes.

Secondly, residuals for data points immediately following a gap in the time series were higher than average. Replacing the single constant term in the GAM models with separate constants for each day helped address this problem and improved the fit of the GAM_{ME} model from 77.9% of deviance explained to 94.1%. A more complete 24 h record of measurements over a greater number of days may resolve this issue without recourse to increasing the number of constant parameters.

Predictions produced by the GAM models, as well as being biased by errors associated with extremely high OH concentrations and by observations that immediately follow a gap in the measurement time series, are also prejudiced by autocorrelation in the data. This is not surprising. TORCH time series data were averaged or interpolated to give measurements at intervals of 15 min. OH radical concentrations and the concentrations of many atmospheric gases also follow a systematic diurnal pattern.

Autocorrelation could have been addressed by incorporating an explicitly defined autocorrelation process into the GAM models. The improvement in predictive capability was weighed against the practical difficulty of modelling the numerous autocorrelation processes present. Furthermore, even without explicitly defining the autocorrelation processes the GAM models explained some of the autocorrelation. For example, the autocorrelation at lag 1 was reduced by almost 50% by both GAM_{ME} and GAM_{MO}. The predictions achieved in Fig. 5 were judged adequate for this research so no adjustment was made. Further, the presence of autocorrelation had no effect on use of GAM_{ME} and GAM_{MO} for interpreting relationships and trends in the data.

The proportion of variation in OH data which can be explained by the GAM_{ME} and GAM_{MO} models is limited to less than 100% by the precision of experimental measurements. Imprecision in the measurement of OH concentrations introduces variability that affects the fit of GAM_{ME}

to measured OH concentrations but not the fit of GAM_{MO}. The deviance explained by GAM_{ME} is, therefore, lower than that explained by GAM_{MO}. Adjustment of the deviance explained requires an accurate estimate of the precision of OH measurements which is not available for the TORCH data. The estimate reported by Smith et al. (2006) from the NAMBLEX experiment (20% at an OH concentration of 3×10^6 molecule cm^{-3}) would limit the maximum deviance explained to 60% for GAM_{ME}, clearly too low given the 77.9% achieved. Uncertainties in measurements of predictor variables also introduced noise into the data used for GAM_{ME} and GAM_{MO} models and also data used to constrain the MCM box model. These measurement errors may affect the relative importance of some of the less influential predictor variables in GAM models but not the qualitative interpretation of the shapes of the smooth functions which is the primary focus of this research.

5 Conclusions

The GAM methodology successfully produced models of measured and modelled OH radical concentrations for the TORCH experiment. GAM_{ME}, the model for measured OH, explained 77.9% of the variation in the data and GAM_{MO}, the model for modelled OH, explained 83.1%. When used to predict OH concentrations the GAM_{ME} model produced better results than the MCM model. GAM_{ME} accurately predicted the diurnal profile of OH concentrations and the predicted mean diurnal concentration from GAM_{ME} underestimated the measured mean by only 0.2% compared to 16.3% over-estimation by the MCM model. The RMSE of

predicted values was reduced to 14.7% of the mean diurnal concentration by GAM$_{ME}$ from 93.4% achieved by the MCM model.

The main weaknesses in the fit of the GAMs were, firstly, a consistent under prediction of very high OH observations, and secondly, relatively large residual errors for observations immediately following a gap in the measurement time-series. The second point may be related to autocorrelation in the time-series data which was not allowed for in the GAM models. Whilst the predictions achieved using GAM$_{ME}$ and GAM$_{MO}$ were adequate for this research, autocorrelation may be sufficiently influential in other data to require explicit modelling within GAM models before they can be used for prediction purposes.

Weaknesses in the fit of GAM$_{ME}$ and GAM$_{MO}$ and the presence of autocorrelation in the data did not affect use of these models for interpretation of trends and relationships in the atmospheric chemistry. Both models identified the key role played by photolysis reactions in the generation of OH radicals with photolysis variables explaining between 67% and 69% of the variation in OH concentrations.

Both models also included a combination of primary and secondary pollutants. Due to the statistical nature of the GAM models, chemical parameters selected for GAM$_{ME}$ and GAM$_{MO}$ were not interpreted as possessing unique properties but properties representative of a range of species. Care was also required not to over-interpret the shapes of smooth functions for parts of the curves where there were limited data. In general, primary pollutants were found to act as sinks for OH radicals with high concentrations of the pollutant associated with low concentrations of the OH radical. Secondary pollutants were generally found to have a positive correlation with OH radicals at low concentrations. At medium to high concentrations, OH radicals were broadly insensitive to changes in concentration of the secondary pollutants.

Besides the photolysis variables, GAM$_{ME}$ suggested that o-xylene, acetone, formaldehyde and water vapour were influential for OH concentrations and GAM$_{MO}$ suggested that isoprene, CO, PAN and ethanol were influential. The differences in these dependencies of modelled and measured OH concentrations supplements knowledge of differences for NO$_X$ already noted by Emmerson et al. (2007). Of particular interest were: the inclusion of o-xylene in GAM$_{ME}$ and not in GAM$_{MO}$ when there are known to be uncertainties in the degradation schemes of aromatic species (Jenkin et al., 2003); and, the inclusion of acetone in GAM$_{ME}$ apparently as an indicator of biogenic processing whilst it was "replaced" by isoprene in GAM$_{MO}$.

The close fit to the data achieved by the GAM models and successful prediction of the diurnal profile of OH concentration supports the use of GAM models as a supplement to MCM modelling. Measurement of more chemical species, especially organic and biogenic species and by-products of OH recycling, with measurements over much longer periods of time would enhance the ability of GAMs to provide insight into the underlying atmospheric chemistry and help address the problems encountered with observations that immediately follow gaps in the data.

GAM models have the potential to be applied more widely in modelling atmospheric chemistry. They are particularly suited to identifying trends in historic data, filling-in gaps in measured data and supporting interpretation of the chemistry. They can be used to forecast future concentrations once the models have been calibrated for a specific location and once secular trends and autoregression have been addressed in the modelling.

References

Baird, C. and Cann, M.: Environmental Chemistry, Freeman, New York, USA, 3rd edn., 154 pp., 2005.

Bower, J., Lampert, J., Broughton, G., Stedman, J., Pye, S., Targa, J., Kent, A., and Grice, S.: Air pollution in the UK: 2005, available at: http://www.airquality.co.uk/archive/reports/cat05/0612111534-421_AirPollutionUK2005text.pdf, 2006.

Carslaw, D. C. and Carslaw, N.: Detecting and characterising small changes in urban nitrogen dioxide concentrations, Atmos. Environ., 41, 4723–4733, 2007.

Carslaw, D. C., Beevers, S. D., and Tate, J. E.: Modelling and assessing trends in traffic-related emissions using a generalised additive modelling approach, Atmos. Environ., 41, 5289–5299, 2007.

Chatfield, C.: The Analysis of Time Series: An Introduction, Chapman and Hall/CRC, London, UK, 4th edn., 62 pp., 1992.

Creasey, D. J., Evans, G. E., Heard, D. E., and Lee, J. D.: Measurements of OH and HO$_2$ concentrations in the Southern Ocean marine boundary layer, J. Geophys. Res., 108(D15), 4475, doi:10.1029/2002JD003206, 2003.

Emmerson, K. M., Carslaw, N., Carpenter, L. J., Heard, D. E., Lee, J. D., and Pilling, M. J.: Urban Atmospheric Chemistry During the PUMA Campaign 1: Comparison of Modelled OH and HO$_2$ Concentrations with Measurements, J. Atmos. Chem., 52, 143–164, 2005.

Emmerson, K. M., Carslaw, N., Carslaw, D. C., Lee, J. D., McFiggans, G., Bloss, W. J., Gravestock, T., Heard, D. E., Hopkins, J., Ingham, T., Pilling, M. J., Smith, S. C., Jacob, M., and Monks, P. S.: Free radical modelling studies during the UK TORCH Campaign in Summer 2003, Atmos. Chem. Phys., 7, 167–181, 2007, http://www.atmos-chem-phys.net/7/167/2007/.

European Parliament and Council: Decision No 1600/2002/EC of the European Parliament and of the Council of 22 July 2002 lay-

ing down the Sixth Community Environment Action Programme, http://ec.europa.eu/environment/newprg/intro.htm, 2002.

Freund, R. J. and Wilson, W. J.: Regression Analysis: Statistical Modeling of a Response Variable, Academic Press, San Diego and London, 192 pp., 1998.

Goldstein, A. H. and Schade, G. W.: Quantifying biogenic and anthropogenic contributions to acetone mixing ratios in a rural environment, Atmos. Environ., 34, 4997–5006, 2000.

Hastie, T. J. and Tibshirani, R.: Generalized additive models, Chapman and Hall, London, 352 pp., 1990.

Heard, D. E. and Pilling, M. J.: Measurement of OH and HO_2 in the troposphere, Chem. Rev, 103, 5163–5198, 2003.

Jenkin, M. E., Saunders, S. M., Wagner, V., and Pilling, M. J.: Protocol for the development of the Master Chemical Mechanism, MCM v3 (Part B): tropospheric degradation of aromatic volatile organic compounds, Atmos. Chem. Phys., 3, 181–193, 2003, http://www.atmos-chem-phys.net/3/181/2003/.

Konrad, S., Schmitz, T., Buers, H. J., Houben, N., Mannschreck, K., Mihelcic, D., Musgen, P., Patz, H. W., Holland, F., Hofzumahaus, A., Schafer, H. J., Schroder, S., Volz-Thomas, A., Bachmann, K., Schlomski, S., Moortgat, G., and Großmann, D.: Hydrocarbon measurements at Pabstthum during the BERLIOZ campaign and modeling of free radicals, J. Geophys. Res, 108, 8251, doi:10.1029/2001JD000866, 2003.

Lee, J. D., Lewis, A. C., Monks, P. S., Jacob, M., Hamilton, J. F., Hopkins, J. R., Watson, N. M., Saxton, J. E., Ennis, C., Carpenter, L. J., Carslaw, N., Fleming, Z., Bandy, B. J., Oram, E. D., Penkett, S. A., Slemr, J., Norton, E., Rickard, A. R., Whalley, L. K., Heard, E. D., Bloss, W. J., Gravestock, T., Smith, S. C., Stanton, J., Pilling, M. J., and Jenkin, M. E.: Ozone photochemistry and elevated isoprene during the UK heatwave of August 2003, Atmos. Environ., 40, 7598–7613, 2006.

Lelieveld, J., Butler, T., Crowley, J., Dillon, T., Fischer, H., Ganzeveld, L., Harder, H., Lawrence, M., Martinez, M., Taraborrelli, D., and Williams, J.: Atmospheric oxidation capacity sustained by a tropical forest, Nature, 452, 737–740, 2008.

McCullagh, P. and Nelder, J. A.: Generalized Linear Models, Chapman&Hall/CRC, Boca Raton, Florida, USA, 532 pp., 1989.

Montgomery, D. C. and Peck, E. A.: Introduction to Linear Regression Analysis, 4th edn., WileyBlackwell, New York, USA, 117–120, 2006.

R Development Core Team: R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, 3-900051-07-0, http://www.R-project.org, access: 5the June, 2007, 2007.

Reiss, R.: Temporal trends and weekend-weekday differences for benzene and 1,3-butadiene in Houston, Texas, Atmos. Environ., 40, 4711–4724, 2006.

Ren, X., Harder, H., Martinez, M., Lesher, R. L., Oliger, A., Simpas, J. B., Brune, W. H., Schwab, J. J., Demerjian, K. L., He, Y., Zhoud, X., and Gao, H.: OH and HO_2 Chemistry in the urban atmosphere of New York City, Atmos. Environ., 37, 3639–3651, 2003.

Rohrer, F. and Berresheim, H.: Strong correlation between levels of tropospheric hydroxyl radicals and solar ultraviolet radiation, Nature, 442, 184–187, 2006.

Saunders, S. M., Jenkin, M. E., Derwent, R. G., and Pilling, M. J.: Protocol for the development of the Master Chemical Mechanism, MCM v3 (Part A): tropospheric degradation of nonaromatic volatile organic compounds, Atmos. Chem. Phys., 3, 161–180, 2003, http://www.atmos-chem-phys.net/3/161/2003/.

Smith, S. C., Lee, J. D., Bloss, W. J., Johnson, G. P., Ingham, T., and Heard, D. E.: Concentrations of OH and HO_2 radicals during NAMBLEX: measurements and steady state analysis, Atmos. Chem. Phys., 6, 1435–1453, 2006, http://www.atmos-chem-phys.net/6/1435/2006/.

Warneck, P.: Photodissociation of acetone in the troposphere: an algorithm for the quantum yield, Atmos. Environ., 35, 5773–5777, 2001.

Westmoreland, E., Carslaw, N., Carslaw, D., Gillah, A., and Bates, E.: Analysis of air quality within a street canyon using statistical and dispersion modelling techniques, Atmos. Environ., 41, 9195–9205, 2007.

Wood, S. N.: Generalized Additive Models: An Introduction with R, Chapman and Hall/CRC,Boca Raton, Florida, USA, 154–160, 2006.