

We thank all three referees for their informative comments, which are addressed in turn below:

Anonymous Referee #1

General points

1) For the assessment of the OH reactivity or toluene equivalence for that matter it is critical to apply the correct rate coefficients. The reaction of toluene with OH has a weak negative T-dependence, i.e. the colder the temperature, the faster the reaction. For lower pressures, though, the amount of redissociation would go up a bit. The authors should specify exactly what rates have been used and state whether the reactivities calculated for the aircraft measurements have accounted for the pressure and temperature at the point of sampling i.e. low temp and pressure). Whether these compounds become more important to reactivity at higher tropospheric altitudes, would be an interesting discussion point to add. Perhaps a table can be added to give the coefficients used with references?

All aromatic + OH rate coefficients (at 298 K, 1 atm) are taken from the Calvert et al., aromatic oxidation review (Calvert, J. G.; Atkinson, R.; Becker, K. H.; Kamens, R. M.; Seinfeld, J. H.; Wallington, T. J.; Yarwood, G. *The Mechanisms of Atmospheric Oxidation of Aromatic Hydrocarbons*; Oxford University Press: New York, 2002.) unless otherwise stated in the text. The source of the monoaromatic rate data used has been added to the text.

The rate data are used at 298 K and 1 atm. We feel that this is appropriate as (i) monoaromatic OH rate coefficients generally show little dependence on decreasing temperature at temperatures < 325 K (Calvert et al., 2002); (ii) the majority of the flight data was taken at altitudes of less than 1 km (see Fig. 8).

2) Canister sampling of VOC is prone to artefacts, particularly for the larger semi-volatile species which are the focus of this study. Alkenes have been observed to increase over time in canisters while other species decrease. It would be good to include this point in the experimental section adding whatever longevity tests have been performed. This is a potential source of error for reactivity arguments that come later and should be mentioned. Have blanks been performed, i.e. canisters pumped up with scrubbed air and sampled ? Or perhaps the two systems have been used to measure ambient air directly while taking canisters in parallel for later analysis?

While the comment is indeed correct in that there is anecdotal evidence for an increase in alkene mixing ratios from canister samples. This is regarding very pristine samples, which is not the case in this study. The focus of this study is on aliphatic and in particular aromatic hydrocarbons. Stability tests for several hydrocarbon species have been performed covering the range of vapour pressures discussed and were found to be acceptable for the measurement timescales involved. (See supplementary data). A section of text has been added to highlight the reviewer's comments and draw the reader's attention to the supplementary data.

3) The authors claim that the normally unmeasured suite of monoaromatic species represent a significant fraction of the missing reactivity determined in the TORCH campaigns. In order to assess

this claim we need to take into account the total uncertainty of both the OH reactivity measurement itself and that of the summed individual measurement at these rather low levels. Error bars on Figure 9 would already help in this regard but better would be a bar chart breakdown of the contributions to reactivity (in s⁻¹) with error bars as this would show the relative importance of the “new” monoaromatic species to the other main species measurements from the campaign.

The errors associated with the MCM modelling (base case) of OH reactivity are discussed in detail in Lee et al., 2010 with respect to the TORCH 2 case study as well as the contributions of all measured and modelled species to the total OH reactivity. TORCH 1 flux calculations for the contribution of VOCs to the OH initiation chemistry (as well as radical propagation and termination) are also shown and discussed in detail in Emmerson et al., 2007. Relevant citations are made to these detailed modelling studies throughout the text.

Here, we are specifically looking at the increase in *modelled* OH reactivity afforded by the inclusion of a *Toluene Equivalent mixing ratio*, derived from the additional (OH) reactivity of the 10 monoaromatic species, relative to that of toluene, relative to a base model calculation of OH reactivity. Therefore, the uncertainties here mainly lie with the relative errors associated with the measured and SAR derived rate constants of the monoaromatics and their enhancement ratio, relative to toluene, as well as the applicability of the MCM toluene photooxidation chemistry as a proxy for all monoaromatics studied (with respect to the additional modelled oVOC chemistry contribution to OH reactivity).

Errors in k_{OH} 's (@ 298K) from the measured monoaromatics in general are very similar (ca. ± 20-30 %, Toluene ± 20 % (Calvert et al., 2002)). Only two of the k_{OH} 's are SAR derived as discussed in the text. The main sources of error associated with calculating the additional modelled monoaromatic OH reactivity here are the applicability of the toluene photochemistry scheme to calculating the contribution of the additional measured monoaromatic photochemistry to the oVOC reactivity (which have the greatest effect on the modelled reactivity – see Lee et al., 2010) as well as applying the derived RONOCO EQ_{Tol} value to the TORCH datasets. A caveat with respect to this is added to the text. *“The simulations presented here serve to show that the addition of only ten, low abundance aromatics to the model can potentially provide a significant extra sink for OH. However, the major assumptions are (i) that the MCM toluene photooxidation scheme can be used as a surrogate mechanism to estimate the contribution of the secondary photochemistry of the 10 unmeasured monoaromatics to the additional OH reactivity, and (ii) that the EQ_{Tol} values derived are applicable to those found over the UK/European region”*

On page 32434, the authors suggest that the ratio of toluene to the heavier monoaromatic VOC loading reported here could be used by others (who have perhaps measured only toluene) to extrapolate an estimate of the monoaromatic loading which they claim it is a significant fraction of the free tropospheric reactivity. Although this is discussed later I think it would be prudent to insert the caveat immediately “assuming the ratios measured off the UK coast are globally applicable” or “assuming similar emission ratios to those in the UK/European region”

Done, the following statement has been added “This would potentially allow techniques that cannot analyse these species (GC-FID, PTR-MS etc) to make a prediction of the additional aromatic content **assuming similar emission ratios to those in the UK.**”

It seems to me unfortunate that the authors have chosen J to denote the equivalent mixing ratio of toluene (Jtol) as it could in future studies become confused with photolysis rates. Perhaps Toleq or another alternative could be used?

Done, changed to EQtol, it should be noted that this nomenclature was used to be consistent with the original paper (Chameides1992) that used propylene equivalents as Jprop

Page 32436, line 22. Surely it depends on the time since emission rather than the distance which works assuming a wind to be present. How about “with airmass age”?

Specific comments

All minor typographical and referencing corrections have been done. More specific addition/comments are given below.

Experimental - The canisters used have been coated with a layer of silico steel. Were the lines leading to the canisters also silico steel or only stainless steel as line 18, page 32428 implies. Please specify, text presently unclear.

The lines leading to the WAS bottles were 316 stainless as stated in the text “All fittings, connectors and gauges used on the WAS case and aircraft assembly are 316 stainless steel”

Page 32430 for both instrument descriptions it would be informative to give how long each sample takes to analyze on each system. Likewise although detailed LOD information is given later in Table 1, it would be good to include overall accuracy and precision information for both systems in the experimental section so that the reader can already appreciate the general comparative differences.

Done, a sentence summarising each systems performance has been added to the end of the GCxGC-TOFMS and GC-FID subsections.

Page 32438, line 11 The error on the OH reactivity measurement of Ingham et al. should be given here i.e. $1.9 \pm ?$ so the reader may assess how much reactivity really remains open.

There is no quoted error for the reactivity measurement quoted in the Lee 2009 paper

Figure 9 needs error bars.

Point addressed above

Anonymous Referee #2

All minor typographical and referencing corrections have been done. More specific addition/comments are given below.

Specific comments:

A general comment is that the introduction seems much more about the measurement technique than the application to monoterpenes in the North Sea. Make sure the introduction is aligned with the title of the paper.

Page 32427 Line 11: Mention of analysis on the rural edge of London is surprising here, given the title's emphasis on the North Sea.

The title has been changed to include coastal UK boundary layer

Page 32427 Line 20: Give the reader a general sense of where the East Midlands Airport is. Identify it in Figure 8. Also change Figure 8 to Figure 1 since it's the first Figure to be discussed. Axis labeling on Figure 8 is too small to read.

This has been done

Page 32428 Line 8: What is the sampling duration and frequency? Though the canisters are coated inside, what are the canisters themselves made of? Are "canister" and "bottle" being used interchangeably? Suggest just using one term.

Bottles has been changed to canisters and the following sentence was added "The filling duration and frequency of the sampling is user controlled, filling times are increased at higher altitudes to ensure adequate fill pressure but range from 30 - 120 s for altitudes 50 - 7000 m."

Page 32428 Line 18: Describe what sensitivity tests you have done to ensure minimal losses.

The main purpose of the stainless steel assembly is to eliminate contamination from the sampling line. Previous studies from the AMMA (D. Stone et al 2010, Atmos. Chem. Phys., 10, 9415-9429) and BORTAS campaigns have directly compared certain species (acetone, isoprene and MVK/MACR and Benzene) using WAS and measured in situ by PTR-MS which have shown good agreement.

The statement has been modified to read "All fittings, connectors and gauges used on the WAS case and aircraft assembly are 316 stainless steel to eliminate contamination and has been shown to provide minimal losses for NMHCs also measured in situ (Stone et al 2010)."

Page 32430 Line 18: What is meant by "as frequently as possible"? How many times per flight? Is this statement applying more to the ozone and terpene standards than the AR54 hydrocarbon standard (which seems to be run 4-8 times a flight)?

As stated in the text the AR54 and AR74 Standard were used for quantification during the RONOCO campaign, the AR54 standard was ran every 8 – 16 samples. The AR74 standard was ran every 24 samples or every 24 hours depending on which occurred first.

The terpene standard was run during method development (precision, linearity, identification and limits of detection) however, not used for calibration during the RONOCO campaign as no terpene data was reported and is not the focus of this paper.

The NPL30 standard was used during development due to it being readily available. Stability and cold finger recovery tests were performed with this (discussed elsewhere) as it covered the range of volatilities found in both the AR54 and AR74 standards but was more readily available.

A few modifications have been made to clarify the section which now reads;-

“Calibration of the TD-GCxGC-TOFMS system during the RONOCO flying campaign was performed using a combination of the AR54 and the AR74 gas standards. The standards were run as frequently as possible to ensure accurate quantification and to minimise uncertainty due to MS detector drift. The AR54 hydrocarbon standard was analysed every time new WAS cases were connected to the instrument (every 8 to 16 samples). The AR74 hydrocarbon standard was run less frequently, once every 24 hours or every 24 samples depending on which occurred first, due to limited standard availability.”

Page 32432 Line 20: The number of samples needs to be clarified. Earlier it seemed like there were 64 WAS bottles in 5 flights, for a total of about 320. The number here (191) is much smaller.

191 samples is the correct number of samples analysed during the 5 flights as some flights did not fill a full set of 64 bottles which is the maximum number of samples that can be taken during one flight

Page 32432 Line 22: Name the seven compounds (benzene, toluene, heptane, octane. . . what else?)

The sentence now reads “In total seven compounds (pentane, hexane, heptane, octane, benzene, toluene and ethylbenzene) were common between both analyses. Good correlations were observed ($R^2 = 0.93 - 0.99$) for all species”

Page 32434 Line 17: The use of ‘J’ is initially confusing because in atmospheric chemistry J is associated with photodissociation rate constants.

We agree and have changed the acronym to EQtol. Originally we used the J to be consistent with the original work on propylene equivalent by Chameides et al, 1992 who used Jprop but it is clear this is confusing.

Page 32434 Line 13: “For the correlation plots presented in Fig. 5, the intercept in each was set to zero.” This doesn’t always seem to be the case. There are also some plots showing negative concentrations.

The plots in Fig 5 have two y-axis to correspond to two aromatic species. Toluene is plotted on the x-axis the scales of each y-axis are offset to ensure the scatter plots overlap as little as possible. There are no negative concentrations and the intercept has been forced through zero.

This confusion may arise from the size of the figure, which has been reduced in the discussion copy and was much larger in the originally submitted version.

Page 32435 Line 19: Include the standard errors of the slope here as well as in the Figure.

Done, Added 2.03 +/- 0.04 and 7.30 +/- 0.15

Page 32436 Line 7: What is meant by “differential plume sampling”?

Differential plume sampling refers to the collection of samples within and outside of pollution plumes, resulting in sharp spikes in hydrocarbon mixing ratios.

The sentence has been changed to clarify the statement

“Variation at lower altitudes is the result of sampling various pollution plumes which result in a sharp rise in hydrocarbon mixing ratios over the background levels.”

Page 32437 Line 10: Quantify ‘significant portion’.

The statement refers to previous studies and so the subjective term “significant” has been removed.

The sentence now reads “consistently find that measured OH reactivity is unaccounted for when compared to the directly measured OH sinks, which is often attributed to unmeasured VOCs.”

Page 32439 Lines 16, 18: ‘Significant’ is used too frequently.

We agree with the reviewers here. The section now reads “although only a small percentage difference, is still a significant value at a maximum of 1 s⁻¹. This shows that the addition of only ten, low abundance aromatics to the model can provide an appreciable extra sink for OH”

Page 32440 Lines 12-14: “Quantification of many of the species detected within this study was not possible owing to a lack of appropriate standards and unknown stability within the WAS canisters.” This sentence needs to be clarified. What species are you referring to? It seems to contradict Page 32428 Line 18 on the minimal losses in the WAS canisters. Page 32440 Lines 12, 19: The statements about lack of standards are confusing because other groups measure many compounds discussed in this paper and so standards are indeed available. Specify what species or classes of compounds you mean by ‘many of the species’.

This statement refers to the species highlighted in Figure 10, which were observed in most air samples however there is limited or no rate constant data, standards for quantification or to test sampling efficiency and stability of these species. This section is to highlight the resolving power of GCxGC-TOFMS for Atmospheric analysis but further development work is still needed and the technique has the same limitations as other 1D systems.

The final section has been modified to clarify and now reads “GCxGC-TOFMS has the ability to detect and resolve many other atmospherically relevant species such as higher MW multifunctional volatile oxygenates, halocarbons and alkyl nitrates shown in Figure 10. Quantification of many of the species detected within this study was not possible owing to a lack of appropriate standards and unknown stability within the WAS canisters. Their impact is also uncertain since in many cases kinetic data is not available. In the future it would be possible to develop an atmospheric sampling method that has the potential to target all these species, within a single analysis, if standards for these species were available.”

Figure 2: The figure is important but the font size is almost impossible to read. Also red lettering against a blue background is difficult to read. In box 'C' 1-8 can't be distinguished from one another; does this improve if this corner of the chromatogram is enlarged?

During the original submission the figure was much larger and was reduced for the discussion paper output. In the amended copy it has been increased to its original size

Figure 5a, e: What does a negative concentration mean? What limitation is this showing?

There are no negative concentrations, different coloured data points correspond to different y-axis. The size of the figure has been reduced during submission which may have led to confusion, it has now been increased to original size.

Figure 7: There are too many significant figures in the cited percentages. To aid in clarity make the legend order the same as the pie chart order. Also explain the left hand versus right hand pie chart in each panel; the red average isn't included in the legend though the blue total toluene is.

Significant figures have been changed, however I believe the figure is self-explanatory in that the chart on the right is a breakdown of the red section of the chart on the left.

Figures 1-6, 8-10: The font size in each figure is too small.

These should be increased in the final version and were much larger on original submission

Figure 5 and other Figures: Label each panel.

The species plotted in each panel is given in both the axes and the legend box. We think this is sufficient to explain the content of each panel.

Anonymous Referee #3

All minor typographical and referencing corrections have been completed.

Major comments:

All ambient air samples from the WAS must pass through the cold finger at -30 °C. This is at a sufficiently cold temperature to remove a significant percentage of the higher aromatic species that are targeted in this analysis due to their relatively low vapor pressures. Were calibrations conducted with the gas standards going through the cold finger?

Cold fingers have been used extensively in previous atmospheric studies (Hopkins et al 2006, 2011, Apel et al 2008, Lewis et al 2013) and have been found to be suitable method of water removal for a range of compound functionalities and vapour pressures

Tests were performed to investigate cold finger losses using the NPL30 standard which has species that correspond to the functionalities (aliphatics and aromatics) and range of vapour pressures investigated in this study. The investigation showed no observable losses outside of injector reproducibility (see supplementary information).

The section is the experimental now reads;-

“To prevent excessive trapping of water, a 200 ml cold finger submerged in an ethylene glycol water chiller set at -30 degC was used during all of the atmospheric gas analysis. Cold fingers have been used extensively in previous atmospheric studies and have shown to be a suitable water removal method for the analysis of a range of compounds of different functionalities and vapour pressures. (Hopkins2006, Hopkins2011, Apel2008, Lewis2013) Tests were also performed as part of this study and found there to be no issue in the use of a cold finger for the compounds discussed in this investigation (See supplementary data)”

Were tests run to measure the transmission efficiency of the WAS (i.e., fill the canister with a known calibration standard to ensure that none of the heavier aromatics are lost to the walls or other surfaces)? The fact that two different measurement techniques were able to accurately measure what comes out of the WAS does not directly address the question of what was actually in the WAS to begin with. Comparison to other in-situ measurements would be useful if available. Otherwise, laboratory tests or an appropriate reference to prior lab tests should be included.

See response to Reviewers 1's 2's comments regarding minimal line losses (comparison between WAS measurements and in situ PTR-MS data) and stability studies performed involving aromatic species of lower vapour pressures (e.g. 1,2,3-trimethyl benzene)

What procedure was used to quantify individual peaks? For example, are the individual components quantified via 1-D maximum peak height, 2-D peak area, or 3-D volume?

Sentence added “Peak quantification was performed using the summed peak area from each modulation pulse (3-D volume).”

P32434 L13 and Figure 5: Why were the intercepts set to 0? Since toluene is the longer lived species,

One would expect the y-intercept of the VOC-to-toluene correlation to be small and negative but not negligible.

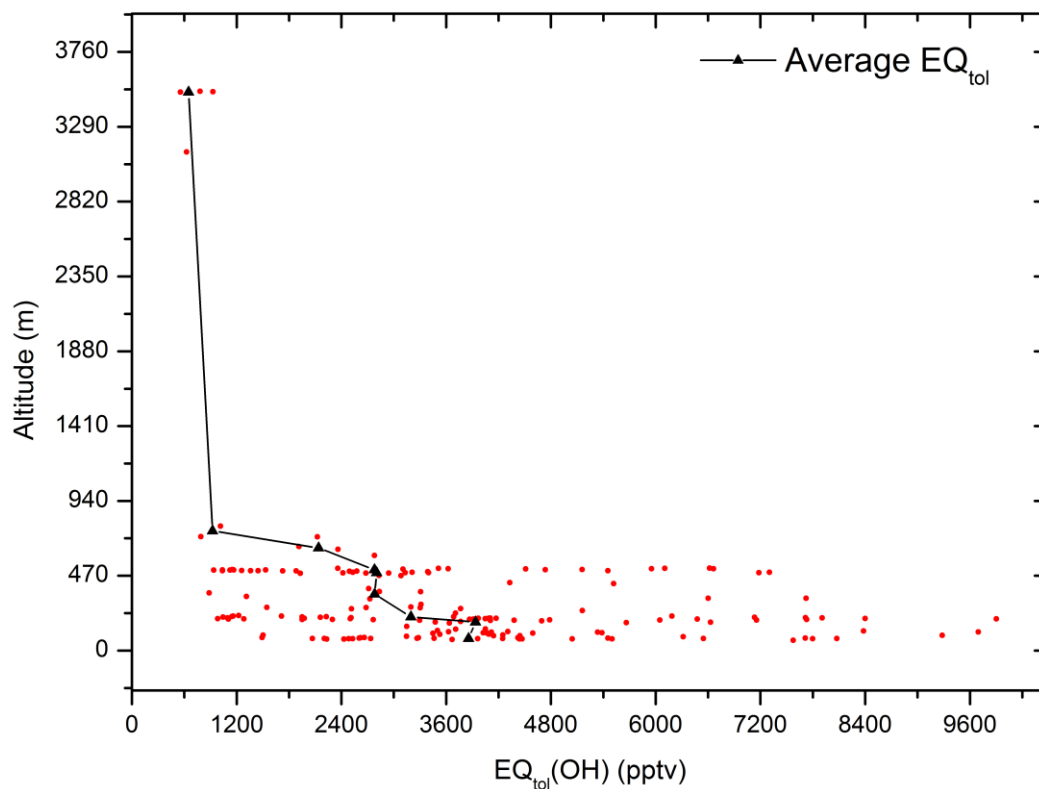
In order to simplify the application of an EQtol value, it was decided to use a simple scaling factor rather than include the intercept. This allowed us to use a single scaling factor in the modelling studies *i.e.* 0.5 here. In addition, this also reduces the possibility of obtaining a negative value during the calculation.

P32435 L20-22: The authors state that the additional monoaromatic species may have a considerable effect on the OH reactivity because the combined reactivity is twice that of toluene, however, what is the potential effect on the total OH reactivity (OH + CO, CH₄, NO₂, all other VOCs)? Is toluene and therefore the monoaromatic VOCs large contributors overall?

In the case studies modelled here, the measured aromatic species in general are not the dominant VOC contributors to the overall primary OH reactivity. This is dominated by oxygenated VOCs, NO₂, CO, CH₄, as the Referee suggests, and is shown in detail in Lee et al., 2010 for TORCH 2. TORCH 1 model flux calculations for the contribution of VOCs to the OH initiation chemistry (as well as radical propagation and termination) are shown and discussed in detail in Emmerson et al., 2007. This shows that monoaromatics contribute about 10% to the primary daytime OH initiation chemistry during TORCH 1. However, both these modelling studies show that the oVOCs make a large contribution to the modelled OH reactivity and this is reflected here in the additional modelled contribution of the oVOCs formed in the photooxidation of the monoaromatics. The point here is to show that only 10 extra compounds (a small fraction of missing VOC) *does* improve the model to measured comparison.

Figure 8, Right panel: A vertical profile would be much more useful. This would include the altitude on the y-axis and the Jtol on the x-axis. The flight track (non-colored) could be added to the panel on the left.

A figure like this was considered for the paper however due to the sampling of different pollution plumes a vertical profile is less meaningful and is shown below. The current plot allows the reader to see the areas in which the flights took place (and areas of low level flying). This can then be visually matched up with the corresponding EQtol values on the left plot.



Section 2.2: One suggestion is to structure the gas sampling section so that the description is in the order that the gas sampling actually occurs. For example, start with pre-sampling steps such as the line purging. Second would be the control of the WAS as detailed on P 32429 L8-19. Next would be the description of the cold finger and finally the Unity 2 system.

We feel that the current structure is suitable and have left it as it is.

All legends and fits should have larger font in all figures whenever possible.

Sizing was altered after submission, figure sizes have been restored to original submitted sizes

Figure 1: I am confused on the designations of “no flow” and “pressurized, but no flow.” Clearly, there must be flow from the WAS to the cold trap. Perhaps this is the schematic for the final phase of sampling rather than the entire process? If so, this should be clearly stated or the entire flow path for the entire sampling process should be shown. What is the unlabeled yellow box at the top of the figure?

The figure shows the flow schematic for the final desorption step of sampling this has been added to the caption, the unlabelled yellow box is an on/off valve to allow a helium blank to be performed through the cold finger if required, the figure has been adjusted.

Figure 6: Black markers should be labelled as GCxGC-TOFMS. The colors blue and black are particularly hard to distinguish in this figure.

Done, Blue markers changed to red, legend changed to read “TD-GCxGC-TOFMS” and “Calculated from FGAM DC-GC-FID Toluene Data”