University of Reading

# *Statistical inference of OH concentrations and air mass dilution rates from successive observations of non-methane hydrocarbons in single air masses*

Article

Published Version

Arnold, S. R., Methven, J. ORCID: https://orcid.org/0000-0002-7636-6872, Evans, M. J., Chipperfield, M. P., Lewis, A. C., Hopkins, J. R., McQuaid, J., Watson, N. M., Purvis, R., Lee, J. D., Atlas, E., Blake, D. R. and Rappengluck, B. (2007) Statistical inference of OH concentrations and air mass dilution rates from successive observations of non-methane hydrocarbons in single air masses. Journal of Geophysical Research, 112. D10S40. ISSN 0148-0227 doi: 10.1029/2006JD007594 Available at https://centaur.reading.ac.uk/868/

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

Published version at: http://www.agu.org/journals/jd/

To link to this article DOI: http://dx.doi.org/10.1029/2006JD007594

Publisher: American Geophysical Union

University of Reading

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the End User Agreement.

www.reading.ac.uk/centaur

**CentAUR**

Central Archive at the University of Reading

Reading's research outputs online

# Statistical inference of OH concentrations and air mass dilution rates from successive observations of nonmethane hydrocarbons in single air masses

S. R. Arnold,[1] J. Methven,[2] M. J. Evans,[1] M. P. Chipperfield,[1] A. C. Lewis,[3] J. R. Hopkins,[3] J. B. McQuaid,[1] N. Watson,[3] R. M. Purvis,[3,4] J. D. Lee,[3] E. L. Atlas,[5] D. R. Blake,[6] and B. Rappenglück[7]

Received 31 May 2006; revised 20 December 2006; accepted 10 January 2007; published 3 May 2007.

[1] Bayesian inference has been used to determine rigorous estimates of hydroxyl radical concentrations ($\overline{[OH]}$) and air mass dilution rates ($K$) averaged following air masses between linked observations of nonmethane hydrocarbons (NMHCs) spanning the North Atlantic during the Intercontinental Transport and Chemical Transformation (ITCT)-Lagrangian-2K4 experiment. The Bayesian technique obtains a refined (posterior) distribution of a parameter given data related to the parameter through a model and prior beliefs about the parameter distribution. Here, the model describes hydrocarbon loss through OH reaction and mixing with a background concentration at rate $K$. The Lagrangian experiment provides direct observations of hydrocarbons at two time points, removing assumptions regarding composition or sources upstream of a single observation. The estimates are sharpened by using many hydrocarbons with different reactivities and accounting for their variability and measurement uncertainty. A novel technique is used to construct prior background distributions of many species, described by variation of a single parameter $\alpha$. This exploits the high correlation of species, related by the first principal component of many NMHC samples. The Bayesian method obtains posterior estimates of $\overline{[OH]}$, $K$ and $\alpha$ following each air mass. Median $\overline{[OH]}$ values are typically between 0.5 and 2.0 $\times$ $10^6$ molecules cm$^{-3}$, but are elevated to between 2.5 and 3.5 $\times$ $10^6$ molecules cm$^{-3}$, in low-level pollution. A comparison of $\overline{[OH]}$ estimates from absolute NMHC concentrations and NMHC ratios assuming zero background (the "photochemical clock" method) shows similar distributions but reveals systematic high bias in the estimates from ratios. Estimates of $K$ are $\sim$0.1 day$^{-1}$ but show more sensitivity to the prior distribution assumed.

**Citation:** Arnold, S. R., et al. (2007), Statistical inference of OH concentrations and air mass dilution rates from successive observations of nonmethane hydrocarbons in single air masses, *J. Geophys. Res.*, *112*, D10S40, doi:10.1029/2006JD007594.

## 1. Introduction

[2] The hydroxyl radical (OH) is a powerful oxidant, present ubiquitously in the daytime troposphere in small concentrations (typically $\sim$10$^6$ molecules cm$^{-3}$; <1pptv [e.g., *Bloss et al.*, 2005, and references therein]). It is chiefly formed through the photodissociation of ozone ($O_3$) in the presence of water vapor [*Levy*, 1971], however other sources include the photolysis of oxygenated hydrocarbon compounds, particularly in the upper troposphere [*Singh et al.*, 1995; *McKeen et al.*, 1997; *Jaeglé et al.*, 2000; *Arnold et al.*, 2004]. Despite its small abundance, the OH radical is responsible for the removal of the vast majority of trace gases emitted into the atmosphere, and is the driving force behind much of tropospheric photochemistry. The oxidation of carbon monoxide (CO), methane and nonmethane hydrocarbons (NMHCs) by OH produces peroxy radical species, which catalyze the production of ozone in the troposphere [*Crutzen*, 1973; *Chameides and Walker*, 1973]. The atmospheric lifetime of methane is controlled by its reaction with OH. The global distribution and abundance of OH therefore controls the climate impacts of methane on the Earth system. Inference of trends in OH abundances [*Prinn et al.*, 2001; *Spivakovsky et al.*, 2000], and

[1]Institute for Atmospheric Science, School of Earth and Environment, University of Leeds, Leeds, UK.
[2]Department of Meteorology, University of Reading, Reading, UK.
[3]Department of Chemistry, University of York, York, UK.
[4]Now at Facility for Airborne Atmospheric Measurements, Cranfield University, Bedford, UK.
[5]Division of Marine and Atmospheric Chemistry, Rosenstiel School of Marine and Atmospheric Science, University of Miami, Miami, Florida, USA.
[6]Department of Chemistry, University of California, Irvine, California, USA.
[7]Institute of Meteorology and Climate Research, Forschungszentrum Karlsruhe, Garmisch-Partenkirchen, Germany.

Copyright 2007 by the American Geophysical Union.
0148-0227/07/2006JD007594

quantification of its sources and sinks [e.g., *Jaeglé et al.*, 2000] are therefore central to our understanding of tropospheric chemistry, and its roles in air quality and climate.

[3] Despite small concentrations and a short lifetime (typically <1 s), sensitive techniques have been developed and improved over the past decade which allow direct in situ observations of OH [e.g., *Heard and Pilling*, 2003, and references therein]. Such techniques involve complex, cumbersome instrumentation, and this makes observations from aircraft platforms more challenging. In addition, such observations generally carry significant uncertainties, and the short OH lifetime means that they only probe the chemical system at the given instant in time and space. Alternative estimates of OH abundances can be inferred from the NMHC "photochemical clock" approach [e.g., *Calvert*, 1976; *Price et al.*, 2004]. This involves monitoring the change in concentrations of two NMHC species with different rates of OH reactivity. The evolution of the ratio of the two concentrations over time is then used to calculate the mean concentration of OH acting on the air mass over that time. This provides a time and space-averaged OH estimate, in contrast to direct in situ observation. However, there are significant uncertainties in this method, including assumptions regarding the mixing of the air mass with surrounding air masses, and assumptions made regarding NMHC ratios in the air mass upstream of the observation point (see section 2).

[4] During summer 2004, the International Consortium for Atmospheric Research on Transport and Transformation (ICARTT) [*Fehsenfeld et al.*, 2006] coordinated an effort to perform a pseudo-Lagrangian analysis of trans-Atlantic air mass transport called the ITCT-Lagrangian-2K4 (Intercontinental Transport and Chemical Transformation) experiment. This involved four separate aircraft: the NASA DC8 and NOAA WP-3D stationed on the U.S. East Coast, the FAAM BAe146 in the Azores, and the DLR Falcon in northern France. Successful sequences of multiple aircraft interceptions of the same air mass during the ICARTT experiment have been independently identified from meteorological models, nonmethane hydrocarbon signatures and long-lived tracer concentrations [*Methven et al.*, 2006]. Here, we exploit this novel Lagrangian aspect of the ICARTT experiment to statistically infer information concerning the photochemical and physical processing of NMHC signatures in the multiply sampled air masses. This leads to estimates of mean OH concentrations and rates of dilution with background air, with strong observational constraints up and downstream in the air masses, eliminating assumptions made in many previous studies.

## 2. NMHC Evolution and Observations

### 2.1. Simple Model for NMHC Evolution

[5] Following an air mass, the rate of change of concentration, $c_i$, of an NMHC species-$i$, for which oxidation by OH is the dominant chemical sink, can be represented by a simple dilution-chemistry model:

$$\frac{dc_i}{dt} = -K(c_i - C_i) - k_i[OH]c_i \tag{1}$$

where $C_i$ is the concentration in the background air which is diluting the air mass, [OH] is the instantaneous concentration of OH in the air mass, and $k_i$ and $K$ are respectively the reaction rate coefficient for species $i$ with OH, and a first-order rate coefficient representing all mixing processes which mix the air mass with background air.

[6] Integrating with respect to time, we can derive a downstream NMHC concentration $c_i^t$ in the air mass, from its upstream concentration, $c_i^0$ and the transport time $t$ between the upstream and downstream points:

$$c_i^t = \frac{KC_i}{K + k_i\theta} + \left(c_i^0 - \frac{KC_i}{K + k_i\theta}\right)e^{-(K+k_i\theta)t} \tag{2}$$

where $\theta$, $K$ and $C_i$ represent the OH concentration, $\overline{[OH]}$, mixing rate and background concentration averaged over time interval $t$.

[7] The goal of this analysis is the derivation of $\overline{[OH]}$ and $K$ from $c_i^t$ and $c_i^0$. However, this can only be achieved if $C_i$ is known, since this affects the rate at which the parcel is diluted with background air, and hence the balance between the pseudo-first-order rates, $K$ and $k_i\theta$. A simplification can be made if we assume that the background concentrations of the NMHCs are negligible. This removes the terms dependent on $C_i$ from equation (2) By considering the ratio ($r_i$) of $c_i$ to a longer-lived reference species, $y$, and assuming that background values $C_i$ are negligible, the simpler evolution equation:

$$r_i^t = r_i^0 e^{-k_i'\theta t} \tag{3}$$

is obtained where $k'_i = k_i - k_y$.

[8] In log-ratio, this becomes a simple translation:

$$\ln r_i^t = \ln r_i^0 - k_i'\theta t \tag{4}$$

[9] This allows the use of ratios of pairs of NMHC species with different OH reactivities to determine an estimate for $\overline{[OH]}$, without needing to account for $K$. This is often termed the hydrocarbon "photochemical clock" method.

[10] Estimates of $\overline{[OH]}$ using equation (3) were first reported by *Calvert* [1976], who used NMHC whole air samples to derive OH concentrations in the Los Angeles urban plume. Since then, many studies have reported estimates of $\overline{[OH]}$ using the "photochemical clock" approach [e.g., *Singh et al.*, 1981; *Roberts et al.*, 1984; *Rudolph and Johnen*, 1990; *Blake et al.*, 1993; *McKenna et al.*, 1995; *Volz-Thomas et al.*, 2000; *Dillon et al.*, 2002; *Price et al.*, 2004]. In using equation (3), many of these studies have identified and attempted to deal with the fact that background concentrations tend to be much higher relative to plume concentrations for the least reactive species, resulting in a change in ratio with time associated with mixing. This produces a bias in the estimates of $\overline{[OH]}$, since it is assumed to be the only agent changing the ratio [*Parrish et al.*,, 1992; *McKeen and Liu*, 1993; *Ehhalt et al.*, 1998]. *McKeen et al.* [1990] and *McKenna* [1997] investigated the impact of a continuous source for the inference of $\overline{[OH]}$ at points downstream in a plume. They described the impact of a coupling between mixing within the plume and chemical loss, and the bias in derived $\overline{[OH]}$ for species of different reactivities.

**Table 1.** Characteristics of Observed NMHC Distributions From the FAAM BAe146 Aircraft During Flights Out of the Azores in the Mid-Atlantic During ICARTT[a]

| | Lifetime, days | Mean | $\sigma_i$ | $A_i$ |
|---|---|---|---|---|
| $C_2H_6$ | 33 | 6.99 | 0.37 | 0.93 |
| $C_2H_2$ | 9.2 | 4.94 | 0.73 | 0.84 |
| $C_3H_8$ | 6.5 | 4.98 | 0.60 | 0.87 |
| $C_6H_6$ | 5.0 | 3.65 | 0.93 | 0.84 |
| i-$C_4H_{10}$ | 2.9 | 2.29 | 0.96 | 0.71 |
| n-$C_4H_{10}$ | 2.8 | 2.91 | 0.82 | 0.91 |
| n-$C_5H_{12}$ | 1.6 | 1.68 | 0.77 | 0.81 |

[a]Mean and standard deviation are shown for $\ln c_i$ where $c_i$ is in pptv. $A_i$ is the amplitude of PC1 in the direction of the species (see text). Lifetimes are calculated for 273 K, 800 hPa, and [OH] = $2.0 \times 10^6$ molecules cm$^{-3}$.

[11] In addition to uncertainties regarding mixing, uncertainties are often introduced by estimating the upstream NMHC concentrations, or their ratio, and the transport time to the downstream observation point. In the absence of direct NMHC observations upstream, estimates of emission ratios for NMHC species are often used to constrain the NMHC ratio at the point of transport from the boundary layer [e.g., *Price et al.*, 2004]. This approach also makes the assumption that the sources of the NMHCs are coincident. Air mass back trajectories are often used to determine transport time from emission to the point of downstream observation [e.g., *Roberts et al.*, 1984], or local wind speed observations can be used to infer the transport time between two observations [e.g., *Volz-Thomas et al.*, 2000]. Both of these methods are subject to large uncertainties. *Volz-Thomas et al.* [2000] discuss the successful use of emissions of $SF_6$ as an inert tracer to minimize errors in tracking of the plumes and to aid in calculation of dilution rates.

[12] *Dillon et al.* [2002] used a nonlinear least-squares fit of observations of NMHCs in the Sacramento urban plume in a simple dilution-chemistry model to infer both $\overline{[OH]}$ and $K$ estimates. Fixed background concentrations were used in the analysis, taken from NMHC observations at a rural site. *Price et al.* [2004] estimated $\overline{[OH]}$ and $K$ values from long-range pollution transport episodes from Asia to the U.S. Pacific coast. They used methods based on both equations (2) and (3), and calculated several estimates of $K$ values based on dilution of long-lived chemical and aerosol tracers and from several meteorological models.

[13] Since both upstream and downstream concentrations ($c_i^0$ and $c_i^t$) were directly observed for the ICARTT Lagrangian match cases, we need not rely on estimates of upstream NMHC signatures from emissions or large-scale observation sets. We also have a good constraint on the transport time between observation points. The framework of the Lagrangian experiment therefore provides us with a unique opportunity to estimate $\overline{[OH]}$ and $K$ values consistent with observed changes in NMHC concentrations, constrained up and downstream on intercontinental scales. These estimates are dependent on the assumed background concentrations $C_i$, which still must be estimated.

### 2.2. NMHC Observations Over the Mid-Atlantic

[14] The environment with which air masses mix is not observed directly following the air mass. During such long time intervals an air mass comes into contact with neighbors of differing composition. In order to estimate possible environments that an air mass could experience over the Atlantic, the statistics of NMHC observations from the entire FAAM BAe146 data set were compiled. The BAe146 was based on Faial, Azores, from 12 July to 3 August 2004 during the ITOP (Intercontinental Transport of Ozone and Precursors) experiment, the European component of ICARTT [*Lewis et al.*, 2007].

[15] Whole air samples were analyzed by gas chromatograph for concentrations of NMHCs. A detailed description of this method is given by *Hopkins et al.* [2003]. The NMHC species used in the analysis are shown in Table 1, along with their chemical lifetimes at 273 K, 800 hPa and $\overline{[OH]} = 2 \times 10^6$ molecules cm$^{-3}$. In addition to air mass variability, each NMHC concentration has experimental uncertainty associated with the GC instrument and integration of its output. These errors represent a "top-hat" uncertainty around the quoted concentration, and are related to the concentration value by the fit:

$$\ln(Err_i) = 0.4753 \ln(c_i) + 0.2270 \quad (5)$$

where $Err_i$ and $c_i$ are in pptv. This relationship is derived for the entire FAAM BAe146 data from quoted error values and data values. These errors are mainly due to uncertainty in calibration and peak integration of the GC output, which becomes more uncertain as peak area diminishes with smaller concentrations.

[16] The sample probability distribution function (PDF) for each species is compiled in log-concentration, $X_i = lnC_i$, and convoluted with the instrument error top-hat function (half-width defined by equation (5)) to obtain an "observation PDF." The resulting distributions are approximately Gaussian, with some increased spread due to the instrument errors. Figure 1 shows the mean and standard deviation of

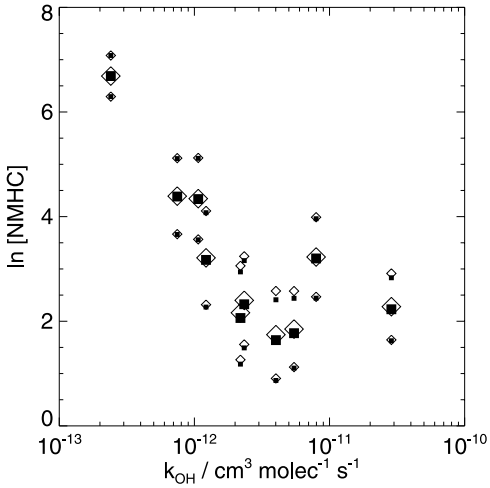


**Figure 1.** Mean and standard deviation of natural logarithms of NMHC species mixing ratios (pptv) from the FAAM BAe146 NMHC data set, shown as a function of species OH reactivity (solid squares). The mean and spread of observation PDFs obtained by convoluting the sample PDFs with top-hat errors characterized by equation (5) are shown by diamonds. The species shown are, in order of increasing reactivity, ethane, acetylene, propane, benzene, i-butane, n-butane, n-pentane, n-hexane, ethene, and propene.

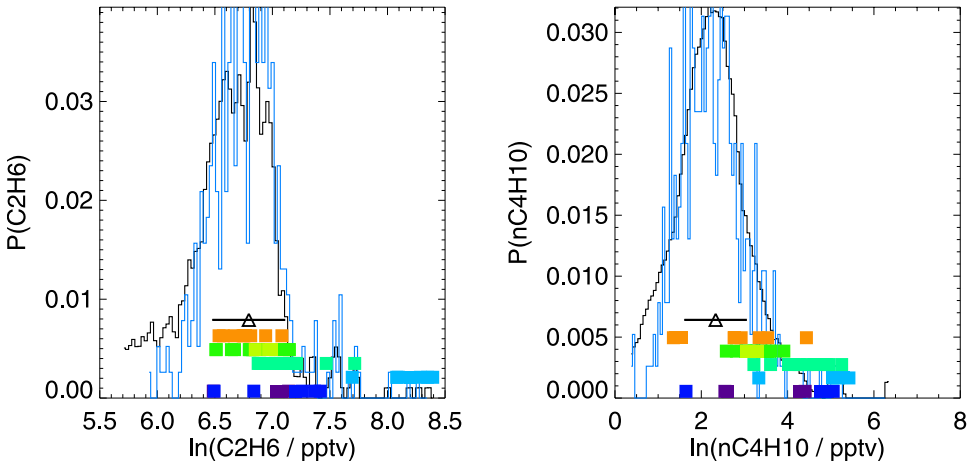


**Figure 2.** Observation PDFs of NMHC species (black) from the entire FAAM BAe146 data set, with PDFs reconstructed from the first principal component (see text) of the data set (blue). Triangles show mean concentrations, $\overline{X}_i$, with black horizontal bars indicating the concentrations reconstructed from adding and subtracting the first principal component eigenvector, $A_i\sigma_i$. Squares indicate the median concentrations from each of the match time windows for the Lagrangian cases, from case 1 (bottom row) to case 5 (top row).

the sample for each species (squares) and the mean and standard deviation of the observation PDFs (diamonds). Air mass variability dominates the instrument uncertainty which is only apparent at the lower concentrations typical of the more reactive species.

[17] The background distributions are constructed from these observation PDFs as follows. Each PDF is standardized by subtracting its mean, $\overline{X}_i$, and dividing by its standard deviation, $\sigma_i$. A principal component (PC) analysis is used to examine the covariation between species in the data set. The first principal component (PC1) explains 71% of the variance in the data, showing that the species cannot be considered as independent variables in the background. Example observation PDFs for ethane and n-butane are shown in Figure 2. These include convolution of the sample distributions with the top-hat measurement uncertainties. Distributions for the species constructed by projecting each sample onto PC1 are similar to the original PDFs showing how PC1 provides a representative contraction of the variability in the data set.

[18] The background PDF for each species is then defined as:

$$C_{i,n} = \alpha_n A_i \sigma_i + \overline{X}_i \tag{6}$$

where $A_i$ denotes the projection of PC1 in the direction of species-$i$ and the background parameter, $\alpha_n$, is assumed to be normally distributed with a mean of zero and standard deviation of unity. This ensures that the mean of the background distribution equals the mean of the observation PDF and the standard deviation of the background equals the amplitude of PC1. As the background parameter $\alpha_n$ is varied, all species concentrations vary together by an amount determined by PC1. This allows the variation of the background concentrations in many species using a single parameter by linking their variations through PC1 of the FAAM BAe146 data set for the Azores region. Inclusion of more PCs would introduce a corresponding number of parameters, which would introduce too many degrees of freedom into the subsequent analysis. The mean and standard deviations of the observation PDFs and values of the $A_i$ coefficients are shown in Table 1. This method allows a range of background scenarios to be included in the analysis, however it does not remove the need to make assumptions regarding the background concentrations.

### 2.3. NMHC Observations Within Lagrangian Windows

[19] More than 4000 whole air samples were collected on board the NASA DC8, NOAA WP-3D, FAAM BAe146 and DLR Falcon aircraft during flights over the northeast United States, mid-Atlantic and European west coast during the ICARTT intensive collaborative period 6 July to 3 August 2004. These were independently analyzed by each aircraft group by gas chromatography for NMHCs. This extensive NMHC data set coupled with the multiple interception events identified by *Methven et al.* [2006] allows the identification of linked pairs of NMHC concentrations, observed upstream and downstream in the same air mass (referred to as ''Lagrangian matches''). These linked NMHC observations allow inference of the rates of NMHC processing between observation platforms. *Methven et al.* [2006] identified five clear Lagrangian match cases, using a combination of Lagrangian models and NMHC ''fingerprint'' matches. We retain the nomenclature used in Table 1 of *Methven et al.* [2006] when referring to the five Lagrangian cases.

[20] Several whole air samples were usually taken within single target air masses. In practice, for most of the aircraft groups, sampling frequency was limited by the total number of canisters available during a flight as well as the canister fill period. However, enough samples we taken to derive an estimate of typical variabilities in the NMHC observations within an air mass. Equation (5) was applied to all NMHC data sets in the analysis, to produce estimates of instrument error for each sample. The resulting values are used to

define half-widths of top-hat error functions in linear concentration space, which are then convoluted with sample distributions defined as Gaussians in log-concentration space with means and standard deviations equal to that of the samples within the target air mass. The resulting "observation PDFs" are approximately Gaussian, with some increased spread at small concentrations due to the top-hat function. The observed match window variability dominates the instrument error for the species considered, as it did in the estimation of the background PDFs (see Figure 1).

## 3. Estimating OH and K From NMHC Data Using Bayesian Statistics

[21] The problem of inferring $\overline{[OH]}$ and $K$ values from NMHC data sets, is suited to a Bayesian inference method [e.g., *Lee*, 1997]. The aim of Bayesian inference is to estimate the PDF of a parameter, $\theta$, given data related to the parameter and prior beliefs about the PDF of the parameter. The parameter is assumed to be a random variable. Here, we describe the formulation of our statistical inference method and how observed distributions of NMHC concentrations can be used to infer the likely distribution of $\overline{[OH]}$ and $K$ values. The following notation will be used:

[22] $\theta_j$: parameter values discretized in bins, labeled $j$. Here, $\theta_j$ is used to denote regularly spaced $\overline{[OH]}$ values.

[23] $f(\theta_j)$: "probability mass function" of $\theta_j$. For constant bin size $\delta\theta$, the PDF, $P(\theta_j) = f(\theta_j)/\delta\theta$ such that $\sum_j f(\theta_j) = \sum_j P(\theta_j)\,\delta\theta = 1$. Since $f \propto P$, we will refer to $f$ as a PDF. $f(\theta_j)$ also denotes the prior distribution of $\theta_j$, an estimate of the distribution of $\overline{[OH]}$ before examining the NMHC data.

[24] $f(\theta|r)$: PDF of $\theta$ given (conditional on) measured data $r$. This is the posterior distribution of $\theta_j$.

[25] In the Bayesian method, the posterior distribution is inferred from the prior distribution using data which is related to the parameter through a likelihood model:

[26] $L(r|\theta_j)$: the likelihood of the observed data given a specified value of the parameter $\theta$. This function is constructed using a data model, in this case either equation (2) or (3).

### 3.1. Hydrocarbon Concentrations Method

[27] Three parameters can be estimated using the simple evolution model (2) and observed distributions of NMHC concentrations upstream and downstream: $\theta$, $K$ and $\alpha$, the OH-concentration, mixing rate and background parameter.

#### 3.1.1. Prior Beliefs

[28] Bayesian inference relies on specifying a prior distribution for each parameter based on the state of the knowledge and physical constraints. For example, we know that [OH] $\geq 0$ and that typical values are $\sim 10^6$ molecules cm$^{-3}$. Here the prior for OH is specified as a gamma distribution:

$$f(\theta) = Aq^p\theta^{p-1}e^{-q\theta} \tag{7}$$

which has mean $p/q$, variance $p/q^2$ and peaks at $(p-1)/q$. This function cannot be less than zero, and is continuous in $\theta$. We have chosen a peak value of [OH] $\approx 2 \times 10^6$ molecules cm$^{-3}$, which is the value used to identify matches between NMHC signatures in the Lagrangian analysis of *Methven et al.* [2006]. The aim here is to refine this overall estimate for individual cases. The prior distribution of $K$ is also a gamma function with a peak at 0.1 days$^{-1}$, in accordance with dilution rates inferred from CO loss by *Methven et al.* [2006]. The prior for $\alpha$ is Gaussian with a mean of zero and standard deviation of one (see section 2.2).

#### 3.1.2. Forward Model

[29] The Lagrangian match events allow us to relate upstream and downstream concentrations, using a forward model and values of $\theta$, $K$ and $\alpha$. The forward model for concentration at time $t$ relative to measurement at $t = 0$ is given by equation (2). The following labeling is used for species and the discretized prior distributions:

[30] $x_i = \ln c_i$: the log-concentration of hydrocarbon species $i$.

[31] $f(\theta_j)$: prior distribution for OH.

[32] $f(K_m)$: prior distribution for K.

[33] $f(\alpha_n)$: prior distribution for background parameter.

[34] An equation for $x_i^t$ is obtained by taking the natural log of equation (2). The values of $k_i$ are known accurately and will be treated as constants. $t$ is also assumed to be known accurately and can be positive or negative depending on whether we are adjusting upstream or downstream measurements respectively.

[35] Values of $k_i$ are calculated for each species using observed temperature/pressure and standard kinetic parameters [*Atkinson et al.*, 2002]. For a good Lagrangian match, the difference between upstream and downstream potential temperature will be small [*Methven et al.*, 2006]. The absolute temperature used to calculate the rate constants, $k_i$, in each case is derived from mean pressure and the mean of the upstream and downstream potential temperature.

#### 3.1.3. Likelihood Model

[36] The likelihood model is an estimate of the probability that the hydrocarbon fingerprints of the upstream and downstream samples are the same when adjusted to a common reference time using the forward model. The likelihood varies with the parameter values and is greatest for the choice with gives the best fit between the two time points. The PDF of the upstream observations, $f(x_i^u)$, is defined on the basis of observed variabilities and instrument error (see section 2.3). It is normalized such that its sum over log-concentrations bins equals unity. It is adjusted using the forward model to a common reference time, to produce an "adjusted upstream distribution" for species $i$, $f(x_i^{u'}|\theta_j, K_m, \alpha_n)$, which also sums to unity. The same procedure is used to obtain an adjusted downstream distribution.

[37] The likelihood that upstream and downstream air masses are the same allowing for OH loss, mixing and background concentrations is given by the product of the PDFs of upstream and downstream observations, adjusted by the forward model to a common time $t$:

$$L(x_i^t|\theta_j, K_m, \alpha_n) = f\left(x_i^{u'}|\theta_j, K_m, \alpha_n\right) f\left(x_i^{d'}|\theta_j, K_m, \alpha_n\right) \tag{8}$$

$t$ is defined to be the midpoint of the Lagrangian match trajectory interval. This ensures that the posterior $\overline{[OH]}$ values are not biased toward the upstream or downstream observations. Note that the adjusted PDFs and the resulting likelihood function share the same discretization over log-

concentration values. The likelihood is obtained by summing over log-concentration bins. The resulting PDF is three dimensional for each species with array indices $j$, $m$, $n$.

#### 3.1.4. Bayes Formula to Obtain the Posterior Distributions

[38] The posterior estimate of the (discrete) PDF in 3-D parameter space, given measurements of species $i$ is obtained by Bayes formula:

$$f(\theta_j, K_m, \alpha_n \,|\, x_i^t) = \frac{f(\theta_j, K_m, \alpha_n) L(x_i^t|\theta_j, K_m, \alpha_n)}{\sum_j \sum_m \sum_n f(\theta_j, K_m, \alpha_n) L(x_i^t|\theta_j, K_m, \alpha_n)} \tag{9}$$

where the likelihood $L(x_i^t|\theta_j,\ K_m,\ \alpha_n)$ has already been integrated over concentration. Note that $f(\theta_j,\ K_m,\ \alpha_n) = f(\theta_j)f(K_m)f(\alpha_n)$ since the prior distributions of the three parameters are independent.

#### 3.1.5. Combining the Posterior Distributions

[39] The posterior distribution of parameter values dependent on NMHC observations, $x$, is obtained by multiplying the results from individual species together:

$$f(\theta_j, K_m, \alpha_n|x) = \Pi_i f(\theta_j, K_m, \alpha_n|x_i^t) \tag{10}$$

[40] Note that if the posterior estimates from each species are similar then the combined estimate will become more peaked and increasingly so for a larger number of species, and will be less dependent on the prior distributions of the variables.

[41] The resulting distribution is a 3-D joint PDF using the data from all NMHC species. 1-D marginal distributions for each variable are defined by summing over the other two directions in parameter space. For OH:

$$f(\theta_j \,|\, x) = \sum_m \sum_n f(\theta_j, K_m, \alpha_n \,|\, x) \tag{11}$$

for mixing rate:

$$f(K_m \,|\, x) = \sum_j \sum_n f(\theta_j, K_m, \alpha_n \,|\, x) \tag{12}$$

and for background parameter:

$$f(\alpha_n \,|\, x) = \sum_j \sum_m f(\theta_j, K_m, \alpha_n \,|\, x). \tag{13}$$

[42] These distributions are refined estimates of each parameter given the NMHC observations and their uncertainties.

### 3.2. Hydrocarbon Ratios Method

[43] Using ratios of observed NMHC distributions, $\overline{[OH]}$ can be estimated by assuming that the plume is diluted with air containing negligible concentrations of the NMHCs such that equation (3) is valid. The prior distribution for $\overline{[OH]}$ is constructed as in the concentrations method.

[44] Distributions of observed NMHC ratios up and downstream are constructed by first applying the convolution of the top-hat error function with Gaussians constructed using observed variabilities as in section 2.3. Then, means and standard deviations from the resulting PDFs for two species $\ln c_i$ and $\ln y$ are combined in quadrature to produce Gaussian distributions of NMHC ratios in log space, $\ln r_i = \ln c_i - \ln y$. For this method, the relevant data variable is the log of the ratio, $x_i = \ln r_i$, rather than the log of concentration.

[45] Application of the forward model for log ratios (equation (4)), produces modified distributions of upstream and downstream NMHC ratios, adjusted to a common time at the center of the Lagrangian interval, by oxidation with a given $\overline{[OH]}$ concentration, $\theta_j$. Multiplication of these adjusted distributions then produces the likelihood that the adjusted upstream and downstream NMHC ratio distributions are the same allowing for OH oxidation, $L(x_i(^t|\theta_j)$.

[46] The posterior estimate of the (discrete) PDF for variable OH concentration $\theta$, given measurements of ratio $x_i$ upstream and downstream, is then obtained from Bayes formula (as equation (9)):

$$f(\theta_j|x_i^t) = \frac{f(\theta_j) L(x_i^t|\theta_j)}{\sum_j f(\theta_j) L(x_i^t|\theta_j)} \tag{14}$$

where likelihood $L(x_i|\theta_j)$ has already been summed over ratio bins. The resulting posterior distributions for each species are then multiplied to obtain the combined posterior distribution for each case as in equation (10).

[47] It is instructive to consider the limit $t \to 0$ (no OH adjustment). $L(x_i^t|\theta_j)$ becomes independent of $\theta_j$ and therefore:

$$f(\theta_j|x_i^0) = \frac{f(\theta_j) L(x_i^0)}{\sum_j f(\theta_j) L(x_i^0)} = \frac{f(\theta_j)}{\sum_j f(\theta_j)} = f(\theta_j) \tag{15}$$

[48] This means that the hydrocarbon observations could not alter the $\overline{[OH]}$ distribution from the prior. Therefore, in order for both the upstream and downstream observations to influence the OH inference, they must each be adjusted toward the midpoint of the time interval.

[49] Moreover, it is clear that if there is insufficient time for OH loss and mixing to take effect, that the combined posterior distribution will be given by the prior for OH to the power of the number of species used in the numerator of the ratios, $I$. This "self-multiplied prior" must have a peak located at the same OH value as the prior itself, but as more species are combined the distribution becomes more strongly peaked. Therefore the posterior results should always be compared with the self-multiplied prior in order to demonstrate that the NMHC observations have added useful information to the PDF of OH.

## 4. Sensitivity of the Bayesian Method

[50] In this section the sensitivity of the inferred OH distributions to details of the methodology is explored. The first investigation uses upstream observations of NMHCs and other trace gases important for photochemistry (including $O_3$, CO, $NO_x$) as initial conditions for the Lagrangian photochemical model CiTTyCAT (Cambridge Tropospheric Trajectory model of Chemistry and Transport) [*Evans et al.*,

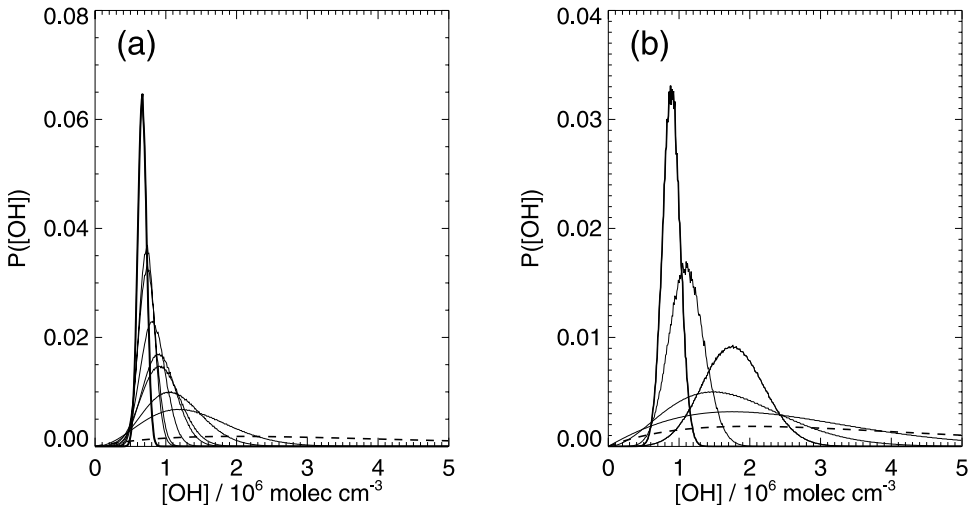


**Figure 3.** Posterior distributions of $\overline{[OH]}$ values for an increasing number of NMHC ratios for (a) a model initialized with upstream observations simulating downstream ratios and $\overline{[OH]} = 0.62 \times 10^6$ molecules cm$^{-3}$ and (b) the same case using ICARTT observations of upstream and downstream NMHC ratios (but exact OH unknown). See text for details. The prior $\overline{[OH]}$ distribution is shown by the dashed curve. The function P([OH]) sums to unity and is discretized with a bin size of $0.01 \times 10^6$ molecules cm$^{-3}$.

2000; *Arnold et al.*, 2004]. The model integrates the chemical equations following the pressure, temperature and humidity of the air mass trajectory to the time of the matching downstream observations, assuming zero mixing. The model simulates an OH concentration that is then averaged along the trajectory to obtain $\overline{[OH]}$. The NMHC ratios simulated at the downstream end of the trajectory are combined with the observed upstream ratios as input data and the Bayesian method for ratios is used to infer $\overline{[OH]}$. Since mixing is not simulated, the ratios method should be able to retrieve the exact $\overline{[OH]}$ that is known from the simulation, given sufficient data and time interval for OH loss to affect NMHC ratios.

[51] The particular case studied is a highly polluted biomass burning plume from forest fires over Alaska, since it was much more polluted that its environment and had a very distinctive hydrocarbon fingerprint. It is referred to as Lagrangian case 2 by *Methven et al.* [2006]. The upstream measurements were collected by the NASA DC8 on 18 July in the upper troposphere near Newfoundland. These are used as initial conditions for the CiTTyCAT model. The downstream time is given by the interception of the plume over France by the DLR Falcon on 23 July. The plume was also intercepted on 20 July by the FAAM BAe146 but the longest time interval is used to maximize the effects of OH loss on NMHC ratios.

[52] Figure 3 shows the sensitivity of the posterior $\overline{[OH]}$ distribution to the number of ratios used in the inference. Ethane is used as the species in the denominator, *y*, since it is the longest lived. Ratios of species with increasing reactivity are added in the following order: acetylene, propane, benzene, i-butane, n-butane, n-pentane, n-hexane and ethene. The posterior distribution becomes more peaked, and shifts away from the prior peak with an increasing number of species. As more species are included, the sample size increases but in addition, as the species become more reactive, they allow greater differentiation between $\overline{[OH]}$ values that provide the best fit between the upstream and downstream ratios, allowing for OH loss. The populated tails of the prior distribution become flattened, as these $\overline{[OH]}$ values are discounted by the likelihood model. The posterior distribution increasingly converges toward the known model $\overline{[OH]}$ value of $0.62 \times 10^6$ molecules cm$^{-3}$.

[53] Note that the shift in the peak location can only occur through information drawn from the NMHC data. As discussed in section 3.2, if the time interval is short compared with the OH loss rate, the adjustment of upstream and downstream distributions by OH reaction is small with the effect that the posterior OH distribution inferred from each species is almost equal to the OH prior. The posterior obtained by combining all species is similar to the prior to the power of the number of (numerator) species. This self-multiplied prior must peak at the same place as the prior, although the peak is sharper.

[54] A second investigation was conducted using the actual data collected by the NASA DC8 upstream and the DLR Falcon downstream. In this case OH is not known a priori and can only be inferred from the posterior results. Figure 3b shows the results of the ratios method. Acetylene was used in the denominator of the ratios because the case 2 plume is much more concentrated than the background acetylene and the ratios method assumes zero background. In contrast, the background for ethane is much higher relative to plume concentrations because of its longer lifetime and is best omitted for the ratios method. The ratios for numerator species are added progressively in the analysis in order of reactivity as for the model case. Note that n-hexane is not added since its values are often close to the detection limit in the ICARTT data set and many samples report missing data for n-hexane. A similar sharpening of the posterior is obtained for an increasing number of species. The posterior distribution shifts away

**Table 2.** The 25th, 50th, and 75th Percentile Values of Posterior $\overline{[OH]}$, $K$, and $\alpha$ Distributions Obtained by Bayesian Inference Applied to Observed Distributions of NMHC Concentrations for Each of the ICARTT Lagrangian Match Events[a]

| | Upstream | | Downstream | | $\overline{[OH]}/10^6$ molecules cm$^{-3}$ | | | $K$/day$^{-1}$ | | | $\alpha$ | | | Consistency | | Number of Species |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Case | Aircraft | Day | Aircraft | Day | 25% | 50% | 75% | 25% | 50% | 75% | 25% | 50% | 75% | Prior/$10^{-5}$ | Post/Prior | |
| 1A | dc8 | 197.84441 | faam | 201.54474 | 1.6 | 1.8 | 2.0 | 0.07 | 0.10 | 0.12 | −2.3e-01 | −6.1e-03 | 2.2e-01 | 2.23 | 1.30 | 7 |
| 1A | dc8 | 197.84441 | fa | 204.53053 | 0.93 | 1.2 | 1.4 | 0.05 | 0.08 | 0.11 | −8.9e-01 | −5.8e-01 | −2.6e-01 | 1.77 | 0.82 | 4 |
| 1B | dc8 | 197.84639 | fa | 204.54443 | 1.1 | 1.4 | 1.7 | 0.08 | 0.11 | 0.15 | −5.6e-01 | −2.3e-01 | 9.6e-02 | 2.21 | 0.81 | 4 |
| 1B | dc8 | 197.84639 | faam | 207.77194 | 1.7 | 2.0 | 2.2 | 0.09 | 0.11 | 0.13 | −1.1e+00 | −8.3e-01 | −5.8e-01 | 1.93 | 1.63 | 6 |
| 2 | dc8 | 200.78537 | faam | 202.52014 | 0.61 | 0.78 | 0.95 | 0.05 | 0.07 | 0.08 | −3.0e-01 | −3.5e-02 | 2.2e-01 | 0.98 | 2.63 | 7 |
| 2 | dc8 | 200.78537 | fa | 205.54562 | 1.0 | 1.2 | 1.3 | 0.08 | 0.10 | 0.12 | −4.0e-01 | −1.3e-01 | 1.4e-01 | 1.86 | 1.81 | 6 |
| 3 | p3 | 202.88708 | p3 | 203.74834 | 1.1 | 1.4 | 1.7 | 0.07 | 0.09 | 0.12 | −3.1e-01 | −5.3e-02 | 2.1e-01 | 1.05 | 1.19 | 7 |
| 3 | p3 | 202.88708 | p3 | 204.72479 | 0.47 | 0.59 | 0.7 | 0.05 | 0.07 | 0.09 | −3.1e-01 | −5.4e-02 | 2.0e-01 | 1.50 | 1.48 | 7 |
| 3 | p3 | 202.88708 | fa | 207.67188 | 0.22 | 0.29 | 0.37 | 0.04 | 0.06 | 0.07 | −3.9e-01 | −1.5e-01 | 7.8e-02 | 2.41 | 0.97 | 7 |
| 3 | p3 | 202.88708 | fa | 208.71902 | 0.26 | 0.35 | 0.43 | 0.05 | 0.07 | 0.08 | −3.5e-01 | −8.8e-02 | 1.7e-01 | 2.50 | 0.89 | 6 |
| 4A | dc8 | 207.77319 | faam | 211.52438 | 0.52 | 0.69 | 0.88 | 0.08 | 0.10 | 0.13 | 1.7e-01 | 3.6e-01 | 5.5e-01 | 0.97 | 4.19 | 6 |
| 4A | p3 | 209.67917 | faam | 211.52438 | 0.88 | 1.1 | 1.4 | 0.08 | 0.10 | 0.13 | −7.5e-02 | 1.6e-01 | 3.9e-01 | 1.31 | 1.72 | 7 |
| 4A | dc8 | 210.79105 | faam | 211.52438 | 1.5 | 1.9 | 2.4 | 0.08 | 0.10 | 0.13 | −3.0e-01 | −4.4e-02 | 2.1e-01 | 1.14 | 1.02 | 7 |
| 4A | faam | 211.52438 | fa | 213.52423 | 0.76 | 0.98 | 1.2 | 0.04 | 0.06 | 0.08 | −9.3e-01 | −6.7e-01 | −3.8e-01 | 1.05 | 1.08 | 6 |
| 4B | dc8 | 207.75230 | faam | 211.53818 | 0.56 | 0.74 | 0.95 | 0.09 | 0.11 | 0.13 | 2.4e-01 | 4.3e-01 | 6.2e-01 | 1.18 | 4.21 | 6 |
| 5 | p3 | 209.78229 | p3 | 210.83926 | 3.1 | 3.6 | 4.1 | 0.08 | 0.10 | 0.13 | −4.9e-01 | −2.4e-01 | −5.2e-04 | 0.89 | 1.53 | 7 |
| 5 | p3 | 209.78229 | faam | 213.48645 | 2.5 | 2.7 | 2.9 | 0.07 | 0.09 | 0.12 | −5.7e-01 | −3.6e-01 | −1.3e-01 | 1.53 | 1.60 | 7 |
| 5 | p3 | 209.78229 | faam | 214.48874 | 2.2 | 2.5 | 2.8 | 0.07 | 0.09 | 0.12 | −5.2e-01 | −2.4e-01 | 2.6e-02 | 1.31 | 1.49 | 5 |

[a]The "consistency measure" ($\Gamma$) is shown for the prior self-multiplied 3-D parameter distribution, and for the posterior 3-D distribution as a ratio to the prior. See text for details. The number of NMHC species distributions used in each case is shown in the final column.

from the self-multiplied prior distribution, demonstrating the modification of the prior estimate by the likelihood function based on NMHC ratio observations.

[55] The posterior estimate of $\overline{[OH]}$ inferred directly from the NMHC data for case 2 is very similar to that simulated by the CiTTyCAT model. This suggests that the $\overline{[OH]}$ inference is a physically reasonable value. In addition, it implies that our understanding of photochemistry as represented by the CiTTyCAT model is capable of simulating [OH]. Our confidence in the Lagrangian match events is also reinforced, since the observed up and downstream NMHC signatures are consistent with NMHC processing diagnosed from a photochemical model free-running along the trajectory linking them.

[56] On the basis of these sensitivity scenarios, we present results for which at least 6 species were measured in the Lagrangian match window, including at least two of n-butane, i-butane and n-pentane. For the cases presented in the following sections, hexane observations are unavailable for most match windows, and alkene data is problematic, since signatures frequently appear "younger" downstream compared with the alkanes (with the exception of case 2). Hence the range of species from ethane to n-pentane shown in Table 1 are used for the ICARTT match cases.

## 5. Results From ICARTT Lagrangian Matches

### 5.1. Ratios Method

#### 5.1.1. Application of the Bayesian Inference

[57] The PDFs of possible OH concentrations that could explain the upstream and downstream NMHC measurements, given their variability and uncertainty, are inferred using the ratios method described in section 3.2. Acetylene is used as the least reactive reference species in the denominator, $y$, and the more reactive species shown in Table 1, are used as the numerator species. Ethane is not used as the reference species, since it has appreciable background concentrations, and so the assumption of negligible background concentrations implicit in the ratios method would not be reasonable. The number of NMHC species present for each match case is shown in Table 2, with a maximum number of 7. This leaves a maximum of 5 NMHC ratios (since acetylene and ethane are not numerator species).

[58] The resulting posterior distributions of $\overline{[OH]}$ are summarized in Table 3 for each of the ICARTT cases described by *Methven et al.* [2006], and selected posterior distributions are plotted in Figure 4 along with the prior $\overline{[OH]}$ distribution.

[59] Lagrangian case 1 involved transport of anthropogenic pollution at low levels from the New England coast across the Atlantic to the coast of Portugal (see *Methven et al.* [2006] for detailed information on individual cases). The median posterior $\overline{[OH]}$ value for the shorter match window between the NASA DC8 on day 197 and the FAAM BAe146 on day 201 is $\sim 1.9 \times 10^6$ molecules cm$^{-3}$, which is larger than values inferred for other Lagrangian links in case 1. This is consistent with larger [OH] in fresher pollution nearer the U.S. East Coast. Posterior distributions of $\overline{[OH]}$ have shifted downward from the prior distribution, and small and large $\overline{[OH]}$ values in the tails of the prior have been eliminated completely (see Figure 4). The longest time

**Table 3.** The 25th, 50th, and 75th Percentile Values of Posterior $\overline{[OH]}$ Distributions Obtained by Bayesian Inference Applied to Observed Distributions of NMHC Ratios for Each of the ICARTT Lagrangian Match Events[a]

| | Upstream | | Downstream | | $\overline{[OH]}/10^6$ molecules cm$^{-3}$ | | | Consistency | |
|---|---|---|---|---|---|---|---|---|---|
| Case | Aircraft | Day | Aircraft | Day | 25% | 50% | 75% | Prior/$10^{-2}$ | Post/Prior |
| 1A | dc8 | 197.84441 | faam | 201.54474 | 1.6 | 1.9 | 2.1 | 1.40 | 1.69 |
| 1A | dc8 | 197.84441 | fa | 204.53053 | 0.80 | 1.2 | 1.6 | 1.08 | 0.24 |
| 1B | dc8 | 197.84639 | fa | 204.54443 | 0.91 | 1.3 | 1.8 | 0.89 | 0.15 |
| 1B | dc8 | 197.84639 | faam | 207.77194 | 1.1 | 1.3 | 1.4 | 1.39 | 2.35 |
| 2 | dc8 | 200.78537 | faam | 202.52014 | 0.71 | 0.94 | 1.2 | 1.22 | 1.85 |
| 2 | dc8 | 200.78537 | fa | 205.54562 | 0.91 | 1.1 | 1.2 | 1.49 | 2.07 |
| 3 | p3 | 202.88708 | p3 | 203.74834 | 1.4 | 1.9 | 2.4 | 1.17 | 1.18 |
| 3 | p3 | 202.88708 | p3 | 204.72479 | 0.43 | 0.61 | 0.79 | 1.24 | 1.68 |
| 3 | p3 | 202.88708 | fa | 207.67188 | 0.14 | 0.20 | 0.29 | 1.30 | 1.24 |
| 3 | p3 | 202.88708 | fa | 208.71902 | 0.15 | 0.23 | 0.30 | 1.44 | 1.36 |
| 4A | dc8 | 207.77319 | faam | 211.52438 | 0.77 | 1.0 | 1.3 | 1.36 | 2.39 |
| 4A | p3 | 209.67917 | faam | 211.52438 | 1.0 | 1.4 | 1.7 | 1.28 | 1.70 |
| 4A | dc8 | 210.79105 | faam | 211.52438 | 1.6 | 2.2 | 2.8 | 1.12 | 1.10 |
| 4A | faam | 211.52438 | fa | 213.52423 | 0.25 | 0.34 | 0.48 | 1.29 | 2.74 |
| 4B | dc8 | 207.75230 | faam | 211.53818 | 0.96 | 1.2 | 1.5 | 1.23 | 1.61 |
| 5 | p3 | 209.78229 | p3 | 210.83926 | 2.5 | 3.1 | 3.8 | 1.03 | 1.10 |
| 5 | p3 | 209.78229 | faam | 213.48645 | 2.3 | 2.7 | 3.0 | 1.17 | 1.03 |
| 5 | p3 | 209.78229 | faam | 214.48874 | 2.0 | 2.5 | 2.9 | 0.96 | 1.24 |

[a]The "consistency measure" ($\Gamma$) is shown for the prior $\overline{[OH]}$ distribution and for the posterior $\overline{[OH]}$ distribution as a ratio to the prior. See text for details.

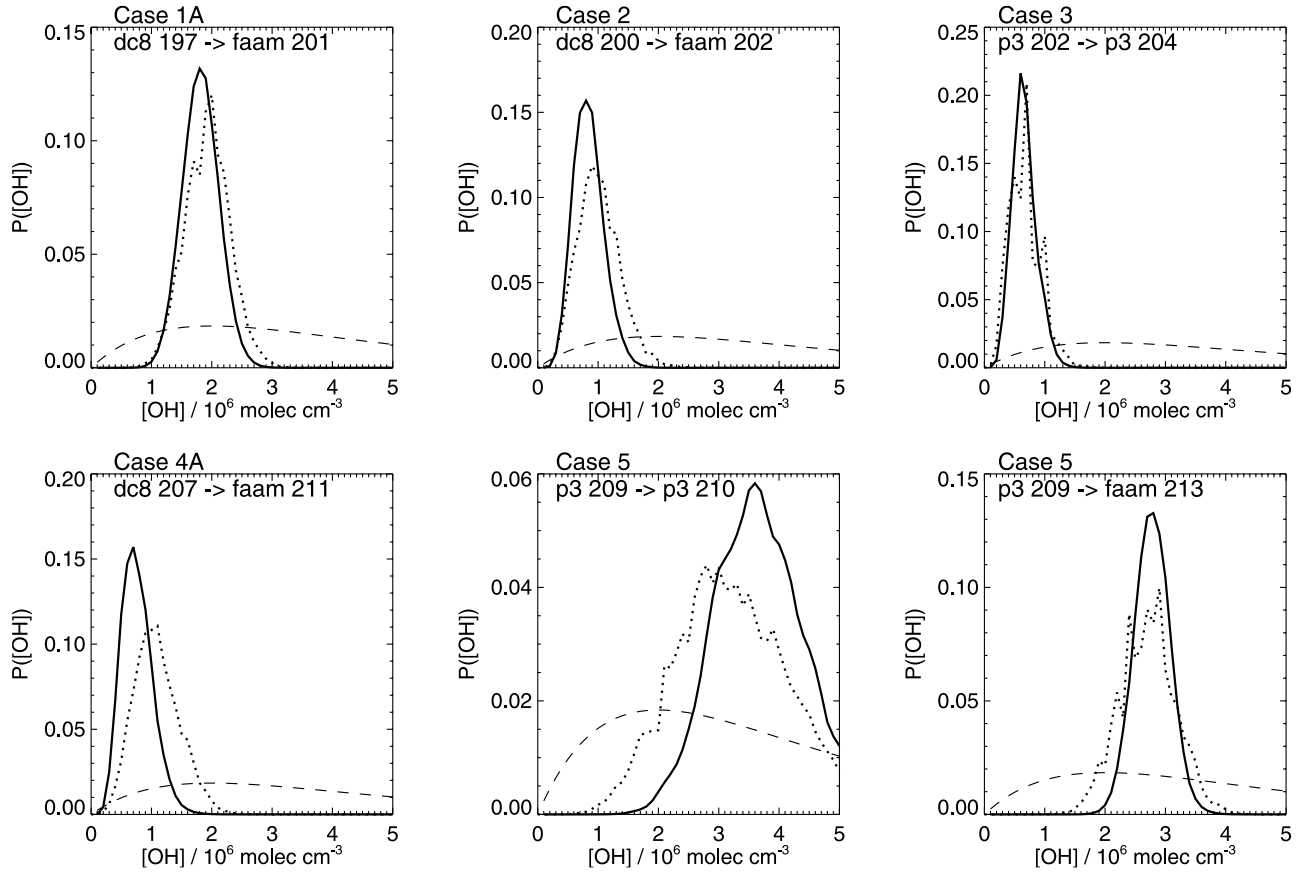


**Figure 4.** Posterior $\overline{[OH]}$ distributions obtained by applying the Bayesian inference model to NMHC ratios (dotted) and absolute NMHC concentrations (solid) from selected ICARTT Lagrangian matches. The prior $\overline{[OH]}$ distribution is shown by the dashed line. Case number and upstream/downstream aircraft and day number are shown on each panel. The function P([OH]) sums to unity and is discretized with a bin size of $0.1 \times 10^6$ molecules cm$^{-3}$.

window between interceptions is ~10 days, between the NASA DC8 on day 197 and the FAAM BAe146 on day 207. The posterior $\overline{[OH]}$ distribution is sharply peaked for this case, with a median value of $\sim 1.3 \times 10^6$ molecules $cm^{-3}$, demonstrating that even over such timescales with very dilute NMHC signatures downstream, the method can still retrieve a preference for a narrow range of $\overline{[OH]}$ values.

[60] In case 2, a strongly concentrated biomass burning plume originating from forest fires in Alaska was intercepted in the upper troposphere near Newfoundland, the Azores and France. There is evidence for slow processing in this plume from studies of both photochemistry [*Real et al.*, 2007] and aerosol [*Petzold et al.*, 2007]. Median $\overline{[OH]}$ values inferred for this case are $\sim 1.0 \times 10^6$ molecules $cm^{-3}$, with the posterior distributions showing a significant shift from the prior distribution. The posterior distributions are also sharply peaked, with a small spread, indicating that elevated concentrations of more reactive NMHCs has allowed the Bayesian method to differentiate strongly between the likelihood of different prior $\overline{[OH]}$ values. Because of the strongly elevated concentrations in the plume relative to the background air, the assumptions implicit in the ratios method are likely to be most applicable to this case.

[61] During case 3, the NOAA WP-3D sampled a plume transported downstream of the New York urban area. Its transport ahead of a cold front was tracked across the Atlantic at low levels, and the same air was eventually sampled over Europe by the DLR Falcon on days 207 and 208. An identified match between the two NOAA WP-3D flights on days 202 and 203 produces a broad posterior $\overline{[OH]}$ distribution, with a similar shape to the self-multiplied prior function (see Figure 4). The short time interval between observations has not allowed the likelihood model to weight parts of the prior distribution, since the timescale is not long enough for application of different $\overline{[OH]}$ values in the forward model to shift the observed NMHC ratio distributions by significantly different amounts. This is an important limitation on using changes in NMHC concentrations to infer the rate of photochemical processing. The time between samples must be sufficiently long for the change in an NMHC distribution due to one $\overline{[OH]}$ value to be statistically separable from that produced by a different $\overline{[OH]}$ value. The match between WP-3D flights on days 202 and 204 provides sufficient time for the prior to be adjusted, and a relatively tight $\overline{[OH]}$ distribution is produced with a median value of $0.61 \times 10^6$ molecules $cm^{-3}$. Posterior $\overline{[OH]}$ distributions for matches with the DLR Falcon are very strongly peaked at small $\overline{[OH]}$ values, with very narrow spread. These estimates are likely hampered by downstream NMHC signatures being "younger" than expected for a true Lagrangian event, because of the influence of mixing from more polluted parts of plume during transit across the Atlantic.

[62] Case 4A sampled upper level export from a warm conveyor belt, associated with a frontal system along the U.S. East Coast. The short case 4A match interval between the NASA DC8 on day 210 and the FAAM BAe146 on day 211, results in a broad $\overline{[OH]}$ distribution, dominated by the prior, since the window is again too short for the likelihood model to differentiate the $\overline{[OH]}$ prior values. The case 4A and 4B events linking the FAAM BAe146 on day 211 downstream with the NASA DC8 on day 207 and the NOAA WP-3D on day 209, produce remarkably similar posterior $\overline{[OH]}$ distributions. Posterior $\overline{[OH]}$ estimates for the match event to the DLR Falcon downstream on day 213 are strongly peaked at small values, again indicating influence from mixing with more polluted signatures after upstream sampling.

[63] In case 5 air was exported more slowly at low levels ahead of the same cold front as in case 4, but originating below and west of the warm conveyor belt. Posterior distributions of $\overline{[OH]}$ are shifted toward larger values, consistent with the low-level, moist environment, which is conducive to efficient production of OH. Largest values are inferred for a linked event between two interceptions by the NOAA WP-3D above the U.S. East Coast, in the fresh polluted outflow. Median $\overline{[OH]}$ values remain elevated well above $2 \times 10^6$ molecules $cm^{-3}$ for the subsequent matches with the FAAM BAe146, further downstream. For each of the events in case 5, the inference method eliminates the possibility that $\overline{[OH]}$ values were less than $\sim 1.5 \times 10^6$ molecules $cm^{-3}$.

#### 5.1.2. Evaluation of the Posterior Estimates

[64] The aim of the Bayesian analysis is to refine the prior estimate of the $\overline{[OH]}$ distribution for each of the cases. It is important to demonstrate that these refined posterior distributions are improved estimates of the $\overline{[OH]}$ concentrations acting in each case compared with the prior distributions. This is achieved by comparing the output of the forward model between upstream and downstream aircraft using the prior and posterior $\overline{[OH]}$ distributions, with the observed NMHC ratio change. We use the likelihood function (equation (8)), as a measure of overlap of the adjusted upstream and downstream hydrocarbon ratio distributions for each $\overline{[OH]}$ value. The overlap ($\gamma(\theta)$) is defined as the mean of the likelihoods obtained from individual species ratios. This indicates how well each $\overline{[OH]}$ value reproduces the observed changes in all concentration ratios. The overall ability of the $\overline{[OH]}$ distributions to reproduce the observed ratio changes is given by a sum of the overlap function over $\theta$, weighted by the normalized self-multiplied prior and posterior $\theta$ distributions:

$$\Gamma_{prior} = \sum_j \{\Pi_i f(\theta_j)\}\gamma(\theta_j) \tag{16}$$

$$\Gamma_{posterior} = \sum_j \{\Pi_i f(\theta_j|x_i^t)\}\gamma(\theta_j) \tag{17}$$

[65] The $\Gamma$ values are a "consistency measure," allowing evaluation of how consistent the prior or posterior $\overline{[OH]}$ distributions are with the observed NMHC evolution and the forward model. The prior values and posterior/prior ratios of $\Gamma$ are shown for each case in Table 3. The increase in the value of the consistency measure provides an indication of the extent to which the hydrocarbon ratio observations have provided a useful constraint on the $\overline{[OH]}$ distribution for each case. A posterior/prior $\Gamma$ ratio of less than or close unity indicates that the Lagrangian match is not good enough to infer useful information about the $\overline{[OH]}$ distribution. Case 1 matches with the DLR Falcon down-

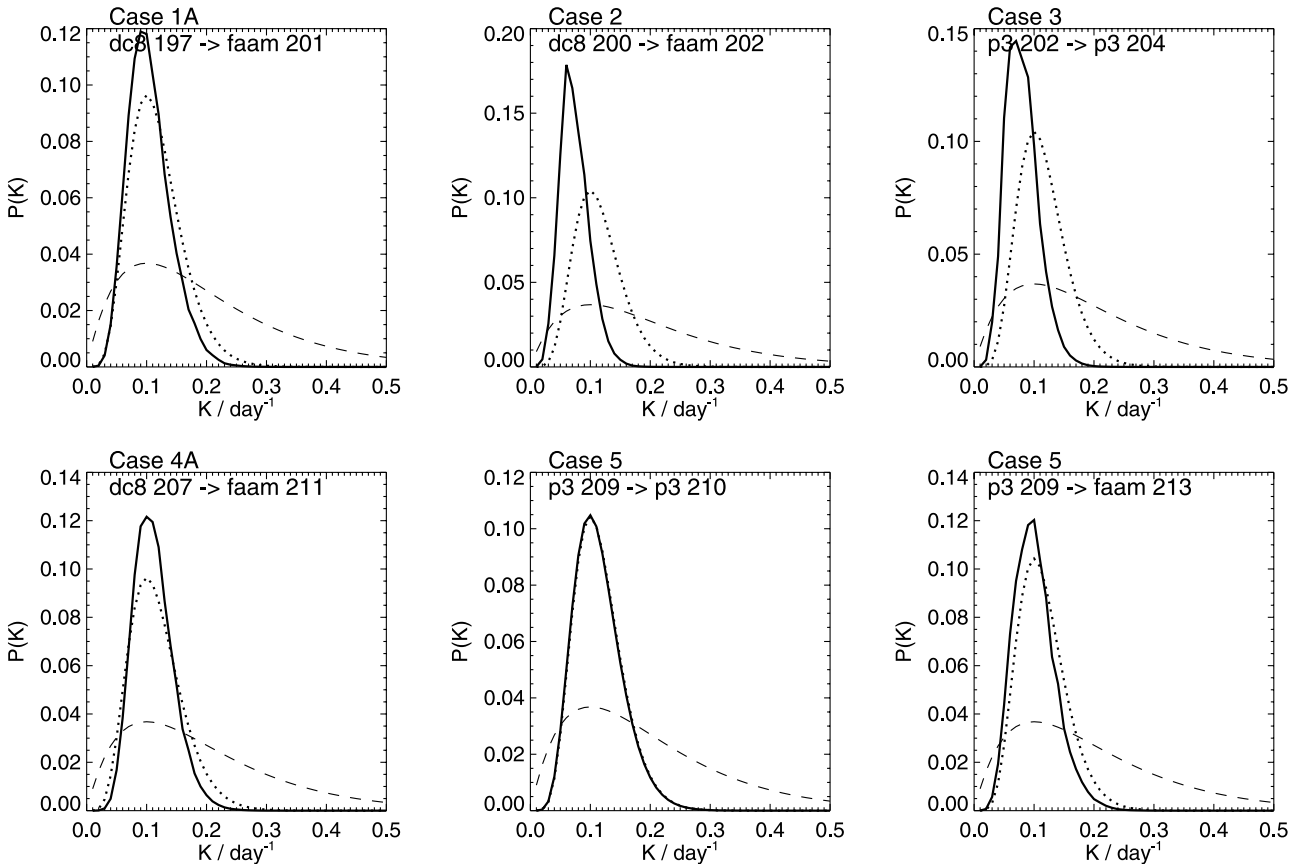


**Figure 5.** Posterior $K$ distributions obtained by applying the Bayesian inference model to absolute NMHC concentrations (solid) from selected ICARTT Lagrangian matches. The prior $K$ distribution is shown by the dashed line. For comparison, the prior $K$ distribution self-multiplied over the number of species used is also shown (dotted). Case number and upstream/downstream aircraft and day number are shown on each panel. The function P($K$) sums to unity and is discretized with a bin size of 0.01 day$^{-1}$.

stream fall into this category, where the reliability of matching samples between the aircraft is dubious. Also short time window matches, such as the case 4A 1-day match window between the NASA DC8 and BAe146, do not allow a significant improvement to the prior distribution, since there is not enough time for the forward model to adequately differentiate the different $\overline{[OH]}$ values. For the other more reliable, longer time window matches, the posterior $\overline{[OH]}$ distributions provide an improvement in simulating the observed evolution of the NMHC ratios, indicating that the NMHC observations have provided a useful constraint on the distribution of $\overline{[OH]}$ values for these cases. The posterior parameter distributions are more consistent with the observed distribution of NMHC composition upwind and downwind and their evolution encapsulated by the forward model.

### 5.2. Absolute Concentrations Method

[66] Application of the Bayesian method to observed upstream and downstream distributions of absolute NMHC concentrations (section 3.1) yields posterior distributions for three parameters; $\overline{[OH]}$, the mixing rate, $K$, and background parameter, $\alpha$. This method has been applied to absolute concentration observations from each of the ICARTT Lagrangian events. The prior distributions for $\overline{[OH]}$, $K$ and $\alpha$ are shown in Figures 4, 5, and 6 respectively, and are constructed as described in section 3.1. Posterior distributions of the three parameters are summarized in Table 2, and are shown for selected examples in Figures 4, 5, and 6.

#### 5.2.1. Posterior Distributions of $\overline{[OH]}$

[67] Comparison of the inferred $\overline{[OH]}$ distributions with those inferred from the ratios method (Figure 4) demonstrates similar estimates from the two methods for each case. For many cases, there is evidence of a systematic shift toward larger $\overline{[OH]}$ values for the ratios method compared with the concentrations method. This bias arises because of the implicit assumption of zero background in the ratios method, as noted by several previous studies [e.g., *McKeen et al.*, 1990; *McKeen and Liu*, 1993]. Typically, the less reactive NMHCs have less elevated plume concentrations relative to their background so that mixing with the background results in slower rate of loss of these species compared with the shorter-lived species. Dilution therefore acts to decrease ratios of short-lived to longer-lived species in a similar fashion to OH reaction. Since the ratios method assumes that mixing cannot change the ratios, $\overline{[OH]}$ is overestimated to explain the observed decrease in ratios with time. Where the plume concentrations are large relative to those in the background (e.g., case 2), the ratios and concentration methods are expected to give similar results, since the assumption of a negligible background is more realistic. However, since more species were used as data for

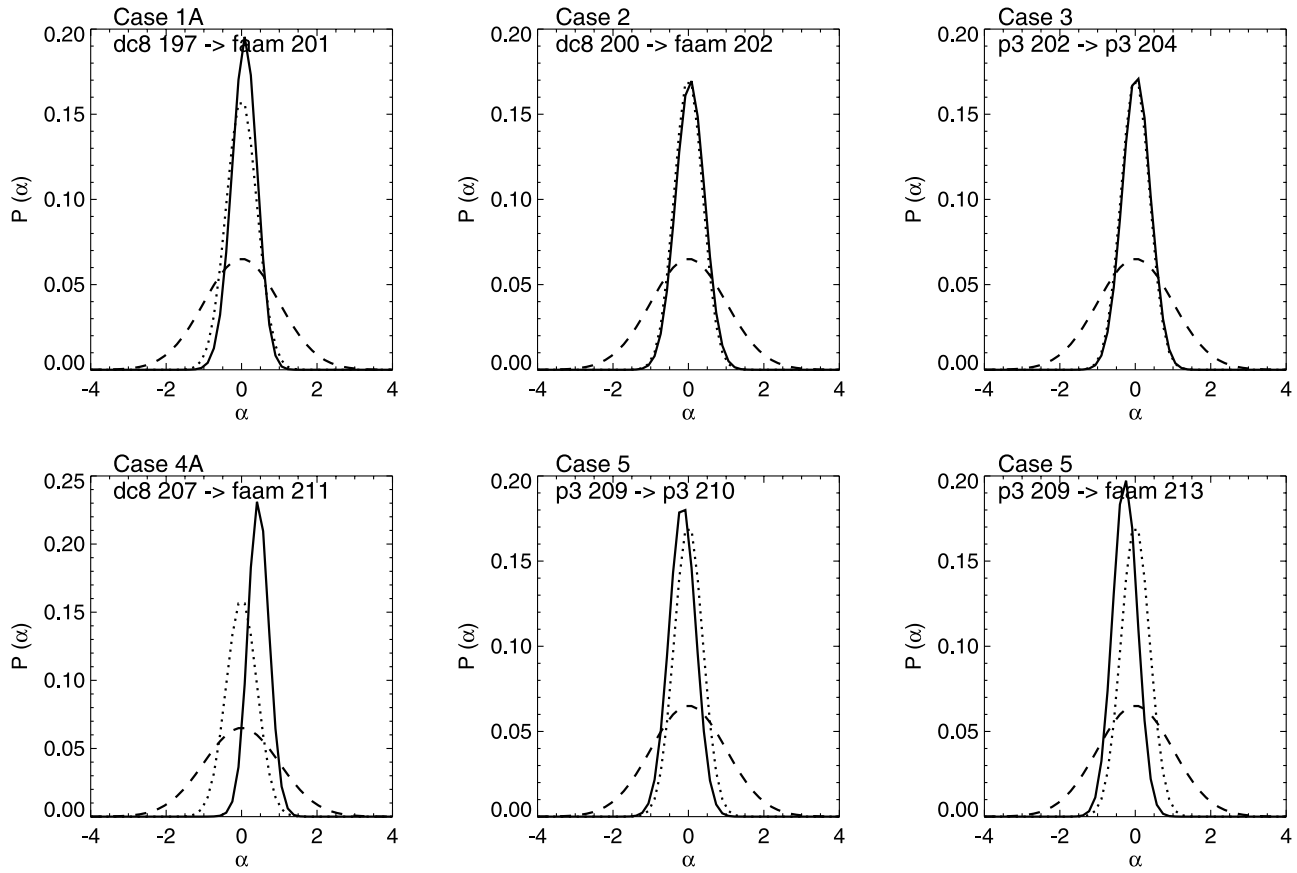


**Figure 6.** Posterior $\alpha$ distributions obtained by applying the Bayesian inference model to absolute NMHC concentrations (solid) from selected ICARTT Lagrangian matches. The prior $\alpha$ distribution is shown by the dashed line. For comparison, the prior $\alpha$ distribution self-multiplied over the number of species used is also shown (dotted). Case number and upstream/downstream aircraft and day number are shown on each panel. The function P($\alpha$) sums to unity and is discretized with a bin size of 0.16.

the concentrations method, this enables the posterior distributions to shift further from their priors.

[68] In cases where the target air mass has concentrations downstream that are lower than the mean of the prior background PDFs, mixing in the forward model can increase concentrations in the air mass, especially at late times because of the assumption of time-invariant background. Hence, in these cases larger $\overline{[OH]}$ values are inferred using the concentrations method to account for oxidation of extra pollution being mixed into the air masses. Examples include case 1B and case 5 (see Figure 2), particularly for the longest time intervals between upstream and downstream samples (Table 2).

### 5.2.2. Evaluation of the Posterior Estimates

[69] Values of the consistency measure, $\Gamma$, are calculated according to equations (16) and (17), weighted by the three-dimensional self-multiplied prior and posterior PDFs (equation (10)). These overlap values are now dependent on all three parameters, $\overline{[OH]}$ , $K$ and $\alpha$, and are described in Table 2. This gives a measure of the consistency of the prior and posterior distributions of $\overline{[OH]}$ , $K$ and $\alpha$, with the observed NMHC concentration changes in each case. If the three parameters had independent effects on the NMHC concentrations, the $\Gamma$ values would be approximately equal to the single-parameter ratio method $\Gamma$ values to the power 3. This accounts for the smaller consistency measure values in Table 2 compared with those in Table 3.

[70] For cases 4 and 5, there is a tendency for the posterior/prior to be increased relative to the ratios method results (Table 3). While this may not be general behavior, it seems to be the case for those cases that are expected to work well because of a sufficiently long match interval and a reliable NMHC match fingerprint. For these successful cases, mixing is sufficiently active to be able to refine the estimate of the background parameter (see section 5.2.3). The absolute concentrations method is not capable of refining the parameter estimates for the shorter time intervals ($<\sim$2 days), in common with the ratios method. Unreliable matches with the DLR Falcon in case 1 produce posterior distributions that show no improvement on the prior distributions, as in the ratios method. In addition, case 3 matches with the DLR Falcon show a $\Gamma$ ratio of just less than unity, indicating that no useful constraint on the prior parameter distributions is provided by these matches.

### 5.2.3. Posterior Distributions of $K$ and $\alpha$

[71] Inferred $K$ and $\alpha$ distributions are shown in Figures 5 and 6, respectively. These two parameters are intimately linked through the mixing term. Posterior distributions for $\alpha$ provide an indication of the level of background concentrations mixed into the air mass averaged across the Lagrangian time interval. This is a unique aspect of the

Bayesian method allowing the background parameter to be treated as a variable, compared to previous methods for estimating $\overline{[OH]}$ and $K$ from NMHCs, which imposed a fixed set of best estimate background concentrations. Shifting of the posterior $\alpha$ distribution from the prior produces a modified distribution of background concentrations which are more consistent with the observed NMHC concentration changes and the dilution-chemistry model.

[72] The posterior distributions of $K$ typically have median values close to the prior distribution peak of 0.1 day$^{-1}$. However, they are more peaked than the self-multiplied prior indicating that the NMHC data for the Lagrangian match has improved the estimate of mixing rate. Using a prior peaking at $K = 0.2$ day$^{-1}$, the posterior distributions shift more significantly to $K$ values between 0.1 and 0.2 day$^{-1}$. Since the action of varying a prior is to drag the posterior estimate toward it, it is clear that $K = 0.2$ day$^{-1}$ would be a significant overestimate of mixing rate and that 0.1 day$^{-1}$ is a better estimate for the prior. Nevertheless, the posterior distributions of $K$ are more sensitive to the choice of prior than the OH estimates.

[73] The median posterior $K$ values compare well with calculated exponential loss rates of CO, based on the change in plume CO concentration over the time between observations [*Methven et al.*, 2006]. These loss rates only correspond to the mixing rate, $K$, in the limit that background CO and photochemical loss are negligible. For most of the cases, the $K$ inferred from NMHCs is larger than the CO loss rate, consistent with the weak elevation of plume CO concentrations compared with its background and therefore much slower CO decrease than if the background were zero. For case 2, the rates should be most comparable, because of the strong elevation of NMHC and CO concentrations compared to background concentrations for the Alaskan fire plume. Median $K$ estimates of 0.07 and 0.10 day$^{-1}$ compare well with CO dilution rates of 0.04 and 0.13 day$^{-1}$. The estimates combining observations of many NMHC species are expected to be tighter and more robust than those obtained from a single species (CO).

[74] In most cases, the posterior distribution of $\alpha$ is barely shifted relative to the prior of zero, but is more peaked than the self-multiplied prior indicating that the NMHC data has added information to refine the background estimate. However, a shift in background parameter does occur in three cases. In case 4 between the BAe146 downstream on day 211 and the DC8 on day 207 (and WP-3D on day 209), the posterior distributions of $\alpha$ are shifted by almost one half a standard deviation to larger values, implying that the NMHC data points to mixing with a more polluted background than the prior estimates based on the mid-Atlantic ITOP data set. This is consistent with CO observations which showed little decrease between these links. It is known that deep convection was embedded within the warm conveyor belt over the Eastern USA in the region flown through by the DC8 and WP-3D and downstream of them. It is likely that some further emissions were mixed into the air mass by convection from below over the USA. Also, mixing within the WCB itself would imply a more polluted background than the prior distribution.

[75] In many cases the $\alpha$-distribution is shifted to lower values, indicating that the data is consistent with mixing to a more dilute background than the prior estimate. This is particularly marked for the furthest downstream matches in case 1B and case 4A, implying that the typical conditions in the Azores region during ICARTT were more polluted than the environment experienced in these cases over the Eastern Atlantic near Europe.

[76] In general, the posterior estimates for OH are tighter and more robust to changes in the prior distributions than the mixing rate and background parameter. This can be understood from consideration of the forward model equation (2). The action of the OH reaction term is simply to translate the upstream and downstream observation PDFs toward each other in log-concentration space by a shift $k_i$ $[OH]$ $(t - t_u)$ and $k_i$ $[OH]$ $(t - t_d)$ respectively, without affecting their spread. The most likely value of OH achieves the greatest overlap in these adjusted PDFs, as quantified by the likelihood function. However, the action of the mixing term is to increase the spread of the downstream PDF as well as modify the shift as it is adjusted to the reference time. The increase in spread is determined by the term $KC_i/(K + k_i\theta)$ which is zero if either the mixing rate or background concentration is zero. Therefore the spread of the upstream PDF is much less affected because the upstream concentrations are typically much greater than those of the background. The effect of the spreading is to reduce the likelihood resulting from summing the overlap over log-concentration bins, but obtain similar likelihoods for a wide range of $K$ and $\alpha$ values. As a result, the posterior 3-D PDF can be much broader along the $K$ and $\alpha$ directions than in the OH direction. Furthermore, the slow variation in the likelihood with $K$ and $\alpha$ implies that the posterior distributions of these parameters must be similar to the prior, as argued below equation (15).

[77] *Parrish et al.* [2007] showed that the relationships between observed NMHC ratios near the U.S. East Coast during ICARTT can only be explained by considering atmospheric mixing and chemical aging together. The NMHC signature of a given air mass is determined by its history of mixing with air masses of different ages during its transport from source. In this sense, it is clear that the use of a distribution of fixed background concentrations in a simple model such as equation (2) is inadequate for treatment of the NMHC evolution. By using a variable distribution of background concentrations, which can be modified by $\alpha$, we are inferring the most consistent distribution of background concentrations for each case. The resulting background concentration distributions represent the spectrum of air mass NMHC signatures which best explain the evolution of each air mass assuming that the forward model is correct. For example, in case 4A a shift of the $\alpha$ posterior to larger concentrations suggests that the target air mass mixed with more polluted surroundings than implied by the prior background distribution.

## 6. Summary and Conclusions

[78] We have applied Bayesian inference to the problem of retrieving $\overline{[OH]}$ values and mean dilution rate values ($K$) from successive observations of nonmethane hydrocarbons (NMHCs) linked by Lagrangian trajectories. To our knowledge, these are the first estimates of these parameters, averaged following air masses for several days, accounting for uncertainties in NMHCs observed upstream and down-

stream and in background concentrations. Although we are unable to remove weaknesses in the "photochemical clock" approach because of simplifications inherent in the derivation of the forward models (equations (2) and (3)), we have presented new methods for deriving key parameters in their formulation. We have described a method for formulating distribution functions for observed NMHC concentrations within air masses, based on observed variability and instrument error. In addition, we have presented a novel method for constructing a distribution of background concentrations based on principal components of a multispecies NMHC data set. This allows the variation of the background concentrations in many species to be controlled using a single parameter ($\alpha$) by linking their variations through the first principal component.

[79] The inference method has been applied to both absolute concentrations of NMHCs, based on a dilution-chemistry model, and to ratios of NMHCs, based on a chemistry-only model derived from the first by assuming dilution with a zero background (the "photochemical clock" method). It is important that the change in NMHC concentrations associated with OH loss and mixing over the trajectory is sufficiently large that the downstream concentrations are significantly more dilute than the upstream concentrations. Specifically, the shift in the peak of the PDFs of each species with time should exceed the spread of the PDF associated with variability within air masses and measurement uncertainty. To achieve this, it is necessary to include more reactive NMHCs (butanes, pentanes, alkenes) and to ensure the time interval between observations is of sufficient length (typically > 1 day) to produce differential oxidation of NMHCs by $\overline{[OH]}$.

[80] The consistency of the prior and posterior distributions of $\overline{[OH]}$, $K$ and $\alpha$ with the observed NMHC ratio and concentration changes have been compared quantitatively using a weighted overlap of observed NMHCs upstream and downstream, adjusted by the forward models. These comparisons indicate that in most cases the Bayesian inference produces improved posterior estimates of the parameter distributions. For a small number of cases, unreliability of the Lagrangian match does not allow a good constraint to be placed on the prior estimates. In addition, the lack of time for forward model adjustment of NMHCs during short matches means posterior distributions are not a significant improvement on the prior estimates.

[81] The posterior $\overline{[OH]}$ distribution derived from NMHC ratios in a strongly concentrated biomass burning plume (case 2) is very similar to the $\overline{[OH]}$ simulated by a photochemical model initialized upstream in the plume. This suggests that the inference method produces physically reasonable estimates of $\overline{[OH]}$, and implies that the representation of photochemistry in the photochemical model can successfully simulate [OH]. This gives us confidence that the photochemical model is a sufficient representation of reality for further studies to be carried out using the ICARTT Lagrangian match events. Future work will investigate the sensitivities of ozone photochemistry in the plumes to model and observation uncertainties, using the same observation-model framework.

[82] Posterior $\overline{[OH]}$ distributions from the NMHC ratios and absolute concentrations are similar, with values typically between 0.5 and 2.0 $\times$ $10^6$ molecules cm$^{-3}$. Median values for case 5 are elevated to between 2.5 and 3.5 $\times$ $10^6$ molecules cm$^{-3}$, consistent with advection of the plume at low levels in the warm, moist air just ahead of a cold front. In cases where the downstream samples are still more concentrated than the prior background with which it is assumed to mix, the concentrations method obtains lower estimates for OH than the ratios method. This is because typically long-lived species are less elevated relative to the background than shorter-lived ones so that mixing results in a decrease of NMHC ratios with time (the longest lived species is in the denominator), similar to the effect of OH reaction. The ratios method assumes that the decrease in ratio is only associated with OH reaction and therefore overestimates OH in order to account for the observed decrease in ratios.

[83] Median posterior $K$ distributions are typically $\sim$0.1 day$^{-1}$, and are similar in magnitude to estimates of CO dilution rates by *Methven et al.* [2006]. In general, the posterior estimates for $K$ are less tight than for OH and are less independent of the prior distribution assumed. This occurs because the OH reaction term in the rate equation for the NMHCs simply results in a translation of the PDF of log-concentration without affecting its spread, while the mixing term increases the spread of the downstream PDF as it is adjusted backward in time toward the midpoint of the Lagrangian interval. The result is that the likelihood that the upstream and downstream data match when shifted to the midtime varies much less with $K$ than with OH. Consequently, the posterior estimate for $K$ is broader and more strongly bound to the prior.

[84] An important aspect of our method is the treatment of the background parameter, $\alpha$, as a variable in the Bayesian inference. For example, the background PDF shifts to higher values if the upstream and downstream samples can be better linked by assuming dilution with a more polluted background from the prior. In a case of pollution exported from the U.S. East Coast in a warm conveyor belt, the posterior $\alpha$ distribution suggests an optimal mixing background significantly more polluted than the prior estimate constructed from the mid-Atlantic NMHC data set. Conversely, in cases where the downstream samples are close to the European western seaboard the $\alpha$ distribution is shifted to lower values indicating that mixing is more likely to have occurred with a background that is more dilute than the prior (averaged over the trajectory). Generally, the posterior background is less readily changed from its prior than for OH, for the same reason as the mixing rate distribution discussed above. However, in this example the prior is based on a large data set collected in the mid-Atlantic during the same period and represents a reasonable estimate of the environment experienced by air masses averaged along their trajectories between the USA and Europe.

[85] Despite the lack of information on the history of air masses between Lagrangian-match samples and the uncertainty inherent in the veracity of those matches, it has been demonstrated that it is possible to infer the distribution of mean OH concentrations and dilution rates that could explain the observed NMHC evolution. It was necessary to consider carefully the uncertainties arising from the measurement technique, variability within the plume and the composition of neighboring air masses with which the

plume mixes. Despite this careful treatment, large uncertainties remain in the posterior estimates, and the impact of uncertainties in background concentrations are especially problematic, particularly for retrieval of $K$ values. These results have important implications for previous $\overline{[OH]}$ and $K$ estimates derived from less well-constrained NMHC concentration changes, and ad hoc choices of constant background concentrations.

[86] **Acknowledgments.** Thanks to the whole of the ICARTT team: the scientists, pilots and support staff associated with all the measurement platforms and their campaign bases. The Natural Environment Research Council (NERC) funded the ITOP-UK project and flying time as part of the Upper Troposphere–Lower Stratosphere Ozone (UTLS) thematic programme. The ITOP-UK team would also like to give thanks to the dedication of the staff from the Facility for Airborne Atmospheric Measurements and, in particular, John Reid. Steve Arnold gratefully acknowledges funding from NERC, and John Methven is grateful for support through an Advanced Fellowship jointly sponsored by NERC and the Environment Agency. The ITCT Lagrangian 2K4 campaign was conducted under the framework of the IGAC (International Global Atmospheric Chemistry) Project (http://www.igac.noaa.gov/). We thank the two anonymous reviewers for suggestions that greatly improved the manuscript.

## References

Arnold, S. R., M. P. Chipperfield, M. A. Blitz, D. E. Heard, and M. J. Pilling (2004), Photodissociation of acetone: Atmospheric implications of temperature-dependent quantum yields, *Geophys. Res. Lett.*, *31*, L07110, doi:10.1029/2003GL019099.

Atkinson, R., et al. (2002), Summary of evaluated kinetic and photochemical data for atmospheric chemistry, IUPAC Subcomm. on Gas Kinet. Data Eval., Research Triangle Park, N. C.

Blake, N. J., S. A. Penkett, K. C. Clemintshaw, P. Anwyl, P. Lightman, A. R. W. Marsh, and G. Butcher (1993), Estimates of atmospheric hydroxyl radical concentrations from the observed decay of many reactive hydrocarbons in well-defined urban plumes, *J. Geophys. Res.*, *98*, 2851–2864.

Bloss, W. J., M. J. Evans, J. D. Lee, R. Sommariva, D. E. Heard, and M. J. Pilling (2005), The oxidative capacity of the troposphere: Coupling of field measurements of OH and a global chemistry transport model, *Faraday Disc.*, *130*, 425–436.

Calvert, J. (1976), Hydrocarbon involvement in photochemical smog formation in the Los Angeles atmosphere, *Environ. Sci. Technol.*, *10*, 256–262.

Chameides, W. L., and J. C. G. Walker (1973), A photochemical theory of tropospheric ozone, *J. Geophys. Res.*, *78*, 8751–8760.

Crutzen, P. J. (1973), A discussion of the chemistry of some minor constituents in the stratosphere and troposphere, *Pure Appl. Geophys.*, *106–107*, 1385–1399.

Dillon, M. B., M. S. Lamanna, G. W. Schade, A. H. Goldstein, and R. C. Cohen (2002), Chemical evolution of the Sacramento urban plume: Transport and oxidation, *J. Geophys. Res.*, *107*(D5), 4045, doi:10.1029/2001JD000969.

Ehhalt, D. H., F. Rohrer, A. Wahner, M. J. Prather, and D. R. Blake (1998), On the use of hydrocarbons for the determination of tropospheric OH concentrations, *J. Geophys. Res.*, *103*, 18,981–18,998.

Evans, M. J., et al. (2000), Evaluation of a Lagrangian box model using field measurements from EASE (Eastern Atlantic Summer Experiment) 1996, *Atmos. Environ.*, *34*, 3843–3863.

Fehsenfeld, F. C., et al. (2006), International Consortium for Atmospheric Research on Transport and Transformation (ICARTT): North America to Europe-Overview of the 2004 summer field study, *J. Geophys. Res.*, *111*, D23S01, doi:10.1029/2006JD007829.

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

Hopkins, J. R., A. C. Lewis, and K. A. Read (2003), A two-column method for long-term monitoring of non-methane hydrocarbons (NMHCs) and oxygenated volatile organic compounds (o-VOCs), *J. Environ. Monit.*, *5*, 8–13.

Jaeglé, L., et al. (2000), Photochemistry of $HO_x$ in the upper troposphere at northern midlatitudes, *J. Geophys. Res.*, *105*, 3877–3892.

Lee, P. M. (1997), *Bayesian Statistics: An Introduction*, 2nd ed., 344 pp., Edward Arnold, London.

Levy, H., II (1971), Normal atmosphere: Large radical and formaldehyde concentrations predicted, *Science*, *173*, 141–143.

Lewis, A. C., et al. (2007), Chemical composition observed over the mid-Atlantic and the detection of pollution signatures far from source regions, *J. Geophys. Res.*, *112*, D10S39, doi:10.1029/2006JD007584.

McKeen, S. A., and S. C. Liu (1993), Hydrocarbon ratios and photochemical history of air masses, *Geophys. Res. Lett.*, *20*, 2363–2366.

McKeen, S. A., M. Trainer, E. Y. Hsie, R. K. Tallamraju, and S. C. Liu (1990), On the indirect determination of atmospheric OH radical concentrations from reactive hydrocarbon measurements, *J. Geophys. Res.*, *95*, 7493–7500.

McKeen, S. A., et al. (1997), The photochemistry of acetone in the upper troposphere: A source of odd-hydrogen radicals, *Geophys. Res. Lett.*, *24*, 3177–3180.

McKenna, D. S. (1997), Analytic solutions of reaction diffusion equations and implications for the concept of an air parcel, *J. Geophys. Res.*, *102*, 13,719–13,726.

McKenna, D. S., C. J. Hord, and J. M. Kent (1995), Hydroxyl radical concentrations and Kuwait oil fire emission rates for March 1991, *J. Geophys. Res.*, *100*, 26,005–26,026.

Methven, J., et al. (2006), Establishing Lagrangian connections between observations within air masses crossing the Atlantic during the ICARTT experiment, *J. Geophys. Res.*, *111*, D23S62, doi:10.1029/2006JD007540.

Parrish, D. D., C. J. Hahn, E. J. Williams, R. B. Norton, and F. C. Fehsenfeld (1992), Indications of photochemical histories of air masses from measurements of atmospheric trace species at Point Arena, California, *J. Geophys. Res.*, *97*, 15,883–15,901.

Parrish, D. D., A. Stohl, C. Forster, E. L. Atlas, D. R. Blake, P. D. Goldan, W. C. Kuster, and J. A. de Gouw (2007), Effects of mixing on evolution of hydrocarbon ratios in the troposphere, *J. Geophys. Res.*, *112*, D10S34, doi:10.1029/2006JD007583.

Petzold, A., et al. (2007), Perturbation of the European free troposphere aerosol by North American forest fire plumes during the ICARTT-ITOP Experiment in summer 2004, *Atmos. Chem. Phys. Disc.*, *7*, 4925–4979.

Price, H. U., D. A. Jaffe, O. R. Cooper, and P. V. Doskey (2004), Photochemistry, ozone production and dilution during long-range transport episodes from Eurasia to the northwest United States, *J. Geophys. Res.*, *109*, D23S13, doi:10.1029/2003JD004400.

Prinn, R. G., et al. (2001), Evidence for substantial variations of atmospheric hydroxyl radicals in the past two decades, *Science*, *292*, 1882–1888.

Real, E., et al. (2007), Processes influencing $O_3$ levels in Alaskan forest fire plumes during long-range transport over the North Atlantic, *J. Geophys. Res.*, doi:10.1029/2006JD007576, in press.

Roberts, J. M., F. C. Fehsenfeld, S. C. Liu, M. J. Bollinger, C. Hahn, D. L. Albritton, and R. E. Sievers (1984), Measurements of aromatic hydrocarbon ratios and NOx concentrations in the rural troposphere: Observations of air mass photochemical aging and NOx removal, *Atmos. Environ.*, *18*, 2421–2432.

Rudolph, J., and F. J. Johnen (1990), Measurements of light atmospheric hydrocarbons over the Atlantic in regions of low biological activity, *J. Geophys. Res.*, *95*, 20,583–20,591.

Singh, H. B., J. R. Martinez, D. G. Hendry, R. J. Jaffe, and W. B. Johnson (1981), Assessment of the oxidant-forming potential of light saturated hydrocarbons in the atmosphere, *Environ. Sci. Technol.*, *15*, 113–119.

Singh, H. B., M. Kanakidou, P. J. Crutzen, and D. J. Jacob (1995), High concentration and photochemical fate of oxygenated hydrocarbons in the global troposphere, *Nature*, *378*, 50–54.

Spivakovsky, C. M., et al. (2000), Three-dimensional climatological distribution of tropospheric OH: Update and evaluation, *J. Geophys. Res.*, *105*, 8931–8980.

Volz-Thomas, A., H. Geiß, and N. Kalthoff (2000), Schauinsland Ozone Precursor Experiment (SLOPE96): Scientific background and main results, *J. Geophys. Res.*, *105*, 1553–1562.

S. R. Arnold, M. P. Chipperfield, M. J. Evans, and J. B. McQuaid, Institute for Atmospheric Science, School of Earth and Environment, University of Leeds, Woodhouse Lane, Leeds LS2 9JT, UK. (sra@env. leeds.ac.uk)

E. L. Atlas, Division of Marine and Atmospheric Chemistry, Rosenstiel School of Marine and Atmospheric Science, University of Miami, Miami, FL 33149, USA.

D. R. Blake, Department of Chemistry, University of California, Irvine, CA 92697, USA.

J. R. Hopkins, J. D. Lee, A. C. Lewis, and N. Watson, Department of Chemistry, University of York, York YO10 5DD, UK.

J. Methven, Department of Meteorology, University of Reading, Reading RG6 6BB, UK.

R. M. Purvis, Facility for Airborne Atmospheric Measurements, Cranfield University, Bedford MK43 0AL, UK.

B. Rappenglück, Institute of Meteorology and Climate Research, Forschungszentrum Karlsruhe, D-82467 Garmisch-Partenkirchen, Germany.