Review: Untangling the influence of air-mass history in interpreting observed atmospheric composition

Fleming[1*], Z.L. and Monks[2], P.S., Manning[3], A.J.

1 National Centre for Atmospheric Science (NCAS), Department of Chemistry, University of Leicester, Leicester, UK

2 Department of Chemistry, University of Leicester, Leicester, UK

3 Met Office, Exeter, UK

* National Centre for Atmospheric Science (NCAS)
Department of Chemistry
University of Leicester
Leicester
UK
Tel: 0044 116 223 1289
Contact email: zf5@le.ac.uk

Table of Contents

Abstract

Is wind direction an adequate marker of air mass history? This review looks at the evolution of methods for assessing the effect of the origin and pathway of air masses on composition change and trends. The composition of air masses and how they evolve and the changing contribution of sources and receptors are key elements in atmospheric science. Source-receptor relationships of atmospheric composition can be investigated with back trajectory techniques, tracing forward from a defined geographical origin to arrive at measurement sites where the composition may have altered during transport.

The distinction between the use of wind sector analysis, trajectory models and dispersion models to interpret composition measurements are explained and the advantages and disadvantages of each is illustrated with examples. Historical uses of wind roses, back trajectories and dispersion models are explained as well as the methods for grouping and clustering air masses. The interface of these methods to the corresponding chemistry measured at the receptor sites is explored. The review does not detail the meteorological derivation of trajectories or the complexity of the models but focus on their application and the statistical analyses used to compare them with *in situ* composition measurements. A newly developed methodology for analysing atmospheric observatory composition data according to air mass pathways calculated with the NAME dispersion model is given as a detailed case study. The steps in this methodology are explained with relevance to the Weybourne Atmospheric Observatory in the UK.

1. Introduction

At a rudimentary level, clean, polluted, background and industrially influenced are all terms used to describe the composition of an air mass. These are clearly subjective in this context but in some way reflect an assessment of the air mass composition and its history. There has been an historical push in particular to classify air masses that are representative of direct anthropogenic influences nearby. In particular, this has been driven by a need to quantify the impact of the perturbation of continuing anthropogenic emissions on the overall state of the atmosphere.

Primary pollutants can originate from anthropogenic source regions (cities, industry, roads etc.) or stochastic events (biomass burning, volcanoes). Source-receptor relationships investigate composition over a receptor region produced by emission changes within a source region (Fiore et al., 2009; HTAP, 2007). Source-receptor relationships of atmospheric composition have been extensively investigated to look at the relationships between emission, transport and *in situ* measurements and reveal the influences pollutant emissions have not only on the local area but also on regions far from the source. Trends in meteorological variability can explain composition trends at some sites as well as the influence of transport pathways. In this review, methods for following source-receptor relationships on regional and global scales by using *in situ* measured composition data are explored.

The chemical and physical composition of an air mass is inherently related to its path through the atmosphere and in order to get the maximum information out of long term time series of composition measurements, data are often divided according to air mass history. Atmospheric composition measurements have been interpreted using wind speed and direction measurements as a marker of air mass history for many years but current science requires better attribution than that available from using wind direction. Trajectory models use a set of meteorological fields from within the domain of influence and dispersion models are one step on from trajectory models, in that the complexity of turbulence is included. A further complexity to transport-only or tracer models comes with the addition of emissions and chemistry of the atmospheric species making these Chemistry Transport Models (CTMs).

Using back trajectories it becomes possible to examine source-receptor relationships and the timescales of long-range and local transport and its effect on the observed composition. In

short-range transport, the airflow pathway is more influenced by emission source areas than in long-range transport, where various exchange and mixing processes (e.g. deposition and advection), physical losses and chemistry have more influence on the composition at the receptor location. The frequency and type of short-term pollution events can be tracked back to their source and seasonal and long term trends can be studied and compared to seasonal and long term air mass transport patterns. For example, Moody et al., 1989 carried out an analysis of atmospheric transport recognition techniques and found that as much as 30 % of chemical variability in the troposphere can be related to transport. Therefore, segregating chemical time series into periods receiving different air masses is important. Early work by Draxler and Taylor, 1982 explained long-range transport of pollutants by running trajectory models from a series of vertical layers as fractions of the boundary layer and investigated the effect of wind shear on pollutants represented by instantaneous puffs of particles and showed how wind shear must be incorporated into models to explain the dispersion.

Long term measurement stations at specific locations around the globe provide a vast amount of data on the chemical composition specific to their location. There are hundreds of permanent long term stations that form part of various national (e.g. AURN in UK, PAES in France, EPA's network in the US) or international (e.g. WMO's GAW (Global Atmospheric Watch), and the European AirBase and EMEP) networks as discussed in Laj et al., 2009. Measurement stations are generally denoted as urban, rural or marine as they are strategically positioned to sample representative segments of air masses, with the closest known pollution source (i.e. roads, industry, cities) or biogenic and natural emissions (e.g. oceans, forests and peat bogs). Background or "baseline" stations are located in geographical locations where they sample and detect any long term trends in the background atmosphere and determine the extent and effect of long range transport bringing pollution to the site. Much of the analysis of ground based composition data is accompanied by studies of what is known as "climatological pathways" with the aim of mapping the probability of hypothetical air masses reaching the station and identifying the emission sources and influences on the site.

Seasonal variations in composition at each station reflect changing meteorological patterns, temperatures, emission patterns, chemistry and physical loss processes within a year. Investigating long term trends (inter-annual or seasonal between years) is now possible, with many stations having continuous measurements for over 30 years. Intra-seasonal and short term changes in composition are very specific to the location of the station and can often be

explained by tracing back the air mass history during these events. Short-term measurement campaigns from aircrafts provide a vital picture of the three dimensional composition of the atmosphere and have been described as mobile *in situ* measurement platforms. Tropospheric-stratospheric exchange processes and advection, boundary layer height, convection, temperature and humidity and their relation to tropospheric composition have been investigated in these experiments. The campaigns are often planned with the idea of investigating specific pollution plumes and use prediction tools (often run using trajectory models) to enable flights to traverse particular air masses and follow composition changes within a plume's evolution (Blake et al., 1993). Aircraft campaigns provide a spatial scale and a representativeness to an area as opposed to a fixed point for ground based platforms but by their very nature they are temporally limited. To some extent this can be expanded by long term networks of aircraft measurements e.g. MOZAIC (Cammas et al., 2009), CARIBIC (Brenninkmeijer et al., 2007) and the HIPPO (HIAPER Pole-to-Pole Observations) Carbon Cycle and Greenhouse Gases Study (http://hippo.ucar.edu/).

Methods to attribute composition changes according to trajectory or dispersion models that do not include emissions are not subject to the errors and assumptions that would propagate from the emission inventories and can result in biased source attribution results. Chemistry Transport Models have many more assumptions in them, so even though both the inventories and the chemistry schemes within the models have been vastly improved, there is still an important role for using trajectories or dispersion models to interpret the measurements.

This review overviews previous studies in which atmospheric composition measurements have been analysed with respect to air mass history. It explains the various techniques used and looks at the classification of the various methods. A step by step explanation of a new technique that classifies the regional influences of a site is described (for the Weybourne Atmospheric Observatory in Norfolk, UK) as a detailed exemplar of the review topic.

2. Outline of methodologies used to assess air mass history and its influence on observed composition

2.1 *In situ* wind direction as a proxy for air mass history

Meteorological parameters have been used to derive statistically distinct meteorological regimes as a primary technique. *In situ* wind direction measurements at a given point have been extensively used to trace the direction of air arriving at a given site but this clearly does not take into account the synoptic scale of the flow field. Wind roses have been used to show the distribution of wind influences at a particular station and divide the composition data into the corresponding wind sectors. Trace gas data at Mace Head in Ireland from measurements in 1996 and 1997 were separated into five wind directions in order to isolate "clean" (Atlantic) and polluted (European) air masses as shown in Figure 1a (Salisbury et al., 2002). The wind rose method often tracks local wind influences (the last 2 or 3 hours before reaching the station), but in the longer term it can often be misleading. For example, local coastal sea-breeze effects can be different to the general circulation and the synoptic scale wind-field.

Radar wind profilers for surface winds and radiosonde data for vertical profiles were used in the MILAGRO campaign in Mexico city (de Foy et al., 2008) to give an extensive picture of the meteorological periods of the campaign and cluster analysis of the wind data was used to assign hourly air mass clusters for the whole campaign (Figure 1b) which were linked to composition measurement time series. During a study at La Réunion island Bhugwant et al., 2001 used a sectorised wind analysis (shown in Figure 1c) to confirm regional contamination of combustion by-products (and higher Black Carbon levels) during particular seasons of the year in the marine boundary layer on the island and confirmed this by comparing with the 5 day back trajectories for these periods. The influence of the meteorological component on the observed ozone and NO_2 trends was studied at an urban site in the Athens basin (Varotsos et al., 2003). Seasonal wind-roses, derived from both trajectory and meteorological data showed the air-transport effect on the air pollution of the Athens basin and cross correlations between surface ozone and the frequency of the air transport during different seasons were calculated.

Many wind rose studies measure wind direction in either 16 direction sectors (22.5º each) (Figure 2) or 36 sectors (10º each) (Figure 1c). Droppo and Napier, 2008 describe an

algorithm to interconvert and standardise analysis using various meteorological datasets with different directional sectors. Daily averaged wind roses of SO_2 and Particulate Matter (PM) at an industrial harbour computed using a Power-Ridge Pollutant (PRP) rose computational scheme (using ordinary least squares regression, outlier handling and weighted averages) (Figure 1d) were compared to half-hourly wind roses of these pollutants and showed that using daily averaged PRP values was as effective as using half-hourly wind roses (Cosemans et al., 2008).

Many statistical techniques have been used with wind speed and direction, humidity, temperature, pressure and cloud cover parameters to classify distinct meteorological regimes and periods. Crutcher et al., 1986 pioneered the use of cluster analysis to better elucidate the dependence of air quality on meteorology. Other studies include cluster analysis of ozone in St Louis, Missouri (Altshuller, 1986), Birmingham, Alabama (Eder et al., 1994), Houston, Texas (Davis et al., 1998) and ozone exceedances at Houston, Texas (Darby, 2005). The Darby, 2005 example is shown in Figure 1e, where wind barbs are used to represent wind speed and direction to show the wind direction in one air mass type separated by cluster analysis.

Neural networks have been used to remove the meteorological variability from datasets to discern temporal and spatial trends in response to changing precursor emissions. Turias et al, 2006 introduced a neural network approach to classify surface winds that could be used to improve air pollution forecasts. Long term ozone measurements and a suite of meteorological measurements have been analysed with neural networks (Gardner and Dorling, 2001; Gardner and Dorling, 1999, 2000) as well as NOx and PM (Kukkonen et al., 2003). The use of artificial neural networks in interpreting air quality data is discussed in a comprehensive review by Gardner and Dorling, 1998.

Beaver and Palazoglu, 2006 have used Principal Component Analysis (PCA) to cluster ozone measurements in San Francisco with wind measurements. PCA and Positive Matrix Factorization (PMF) (see section 4.1.5 for similar technique with back trajectories) were used to interpret the elemental composition and sources of aerosols arriving at Dunkirk, France (Alleman et al., 2010), associating each sample with distinctive emission sources. Hart et al. (Hart et al., 2006) have analysed ozone exceedance events over ten years in Sydney, Australia by using a suite of meteorological measurements combined with PCA and cluster

analysis to classify days into low and high ozone days. Oanh et al., 2005 have looked at SO_2 levels in the Mae Moh valley in the northern Thailand with PCA clustering of synoptic meteorological conditions.

The Conditional Probability Function (CPF) that represents the probability of an air mass arriving at a receptor site has been used on wind data in many studies (see Section 4.1.3 for its applications in trajectory residence time analysis). Wind direction and speed were used for sites measuring Black Carbon (BC) in New York state (Venkatachari et al., 2006) in order to identify likely locations of local point sources of BC, with CPF plots showing from which directions around the sites the highest 25 % BC levels occur. Extensive wind rose analyses deriving CPF of the concentration of the species of interest from various directions around the measurement station have been carried out: e.g to attribute PM levels to wind sectors at three sites in Ontario, Canada (Chan and Mozurkewich, 2007), to study sources of PM, O_3, NO, NO_2, CO and SO_2 arriving at Erfurt, Germany (Yue et al., 2008), to study VOC levels in Beijing (Song et al., 2008), to study $PM_{2.5}$, SO_2, CO, and O_3 in Rochester, NY, USA (Kasumba et al., 2009), to analyse PM_{10} levels arriving in Daejeon city, China (Lim et al., 2010) and to analyse VOC levels in Houston during the TexAQS-II campaign (Leuchner and Rappengluck, 2010) (as shown in Figure 2 with wind roses for two of the source types derived from PMF analysis).

The general weakness of using pollution roses is that one cannot assume that the wind direction measured at a point is consistent with the synoptic scale flow. The turbulent and synoptic nature of wind always leads to changes in the wind direction over a region and this is not shown from local or point wind direction measurements.

2.1.1 Meteorological data as synoptic weather patterns to classify composition

The classical large scale weather classification widely used in Europe is the Grosswetterlagen system, originally conceived by Baur et al., 1944, recently updated by Gerstengabe et al., 1999 and since maintained by the German Weather Service (DWD). The 29 Hess and Brezowsky Grosswetterlagen (HB-GWL) regimes can be viewed as readily identifiable large-scale circulation patterns involving the whole of Europe and the North-East Atlantic, with their primary focus on central Europe. Spichtinger et al., 1996 divided 3 years of ozone, NO

and meteorological measurements in Munich into 3 GWL. Linear multiple regression analysis was performed on each weather type to reveal how the ozone and NO concentrations are explained in terms of meteorological parameters, showing how there is a better correlation in the cyclonic compared to anticyclonic conditions, which shows how the increasing significance of photochemistry in cyclonic periods.

The classification of daily weather types (DWT) on a daily basis for the British isles between 1861 and 1997 has been carried out by Lamb, 1972, where the basic flow direction and level of anticyclonicity or cyclonicity over an area was determined. The grid over which the objective DWTs are calculated can be moved to different locations and a new catalogue, appropriate for the new location, can be constructed. Ohare and Wilby, 1995 used the Lamb weather types to show that ozone levels from a variety of measurement stations in the UK were strongly correlated with the prevailing weather type.

Meteorological measurements at Szeged, Hungary have been separated according to Factor Analysis (reduction of the dimensionality of the meteorological dataset) and objectively grouped into days with similar weather conditions and then compared to CO, NO, NO_2, SO_2, O_3 and Total Suspended Particles (TSP) measurements over 4 years (Makra et al., 2006). Factor analysis of meteorological data, followed by cluster analysis was applied to air masses arriving in Athens over 4 years and corresponding air pollutant measurements (Sindosi et al., 2003). Helimis et al., 2003 studied the connection of atmospheric circulation to trans-boundary air pollution by using a circulation-to-environment approach where 14 synoptic scale patterns were distinguished over Athens and SO_2, NOx and O_3 data were compared between the synoptic periods and additional modelling linked this to inflow and outflow of the Athens basin. Demuzere and Lipzig, 2010 used linear regression methods to explain O_3 and PM variations and found that classifying according to the automated “Lamb” weather type prior to the regression analysis was superior to just using the linear regression. Comparison of objective air mass types and the “Peczely” weather types to classify daily pollution levels over the Carpathian Basin (using 12 meteorological and eight pollutant parameters) has been conducted for a four year period (Makra et al., 2009).

2.2 Trajectory models

A modelled trajectory is an estimate of the transport pathway of an infinitesimally small air parcel and an estimate of the centreline of an advected air mass subject to vertical and horizontal dispersion. Back trajectories trace the path of a polluted air parcel over a period of time and have long been used to track the history and pathway of air parcels arriving at a specific location since they were first developed in the 1940s by Pettersen, 1940. The first trajectories were one dimensional lines calculated using recorded meteorological fields in a model dealing with the combination of wind field influences on the air. Computational advances in the 1960s allowed isentropic analysis and trajectory calculations to be performed graphically on computers (Danielsen, 1967). These back trajectories followed the path of release backwards on a Eulerian grid, where the flow of particles is depicted as a function of a fixed position and time. More information can be extracted from two-dimensional trajectories by colour-coding their pathways by height and by including markers corresponding to time of travel. Owing to errors and assumptions in the wind fields used to calculate the trajectories, the uncertainty of trajectories increases with time along the path and Seibert, 1993 investigated the accuracy of trajectories

Ashbaugh et al (Ashbaugh et al., 1985) were one of the first to use back trajectory analysis to identify source regions of pollutants (in this case sulphur) and then used air quality monitoring records to identify those regions from which high sulphur concentrations were most likely to arrive and the corresponding statistical associations between air mass history and above average concentrations.

A review of the types and uses for back trajectories and the associated errors and probabilities within them has been provided by Stohl, 1998, in which it is stated that there can be serious misinterpretations of a flow situation (represented as linear air mass movements) if the magnitude of the errors cannot be estimated. Kahl, 1993 investigated the errors within the calculation of trajectories by calculating trajectories at multiple levels and at regular grid intervals around a site to assess the extent of vertical and horizontal wind shears. The errors and issues associated with back trajectories (uncertainty arising from interpolation of sparse meteorological data, assumptions regarding vertical transport, observational errors, sub-grid-scale phenomenon, turbulence, convection, evaporation, and condensation) are explained in Polissar et al., 1999. Taking into account uncertainties, back trajectories are often better suited to large scale circulation studies, such as shown in the detailed global trajectory study

on air parcel circulation between the troposphere and stratosphere carried out by Jackson et al al., 2001.

Examples of some of the community's commonly used trajectory models are listed in Table 1. The FLEXTRA model has been used extensively and is described by Stohl et al., 1995. The UK Universities' Global Atmospheric Modelling Programme (UGAMP) trajectory model has been used extensively in many studies (e.g. (Cape et al., 2000; Methven et al., 2006)) as well as the HYbrid Single-Particle Lagrangian Integrated Trajectory (HYSPLIT) model. The HYSPLIT model, when run in particle mode advects a fixed number of particles about the model domain by the mean wind field and spread by a turbulent component (Draxler and Rolph, 2003). The Centre for Air Pollution Impact and Trends Analysis CAPITA Monte Carlo (CMC) model (Schichtel et al., 2006) uses meteorological wind fields to advect the particles in three dimensional space, while the intense vertical mixing that takes place within the atmospheric boundary layer is simulated using a Monte Carlo technique, which evenly distributes the particles between the surface and the mixing height. Other trajectory models include APTRA (Delcloo and De Backer, 2008) obtained from the ECMWF (European Centre for Medium-Range Weather Forecasts), the LAGRANTO model (Tarasova et al., 2009), the TRADOS model (long-range Trajectory, Dispersion and Dose Model of the Finnish Meteorological Institute) (Virkkula et al., 1999) and trajectory models run directly from meteorological datasets such as the ECMWF three dimensional isentropic model (Dogan et al., 2008), NMC (US National Meteorological Centre) (Merrill, 1994; Merrill and Moody, 1996), BADC (British Atmospheric Data Centre) (Walker et al., 2009), CMDC (Climate Monitoring and Diagnostics Laboratory) (Polissar et al., 1999; Polissar et al., 2001b) and NIES (Russian National Institute for Environmental Studies) (Pochanart et al., 2003) models.

A thorough comparison of various trajectory models with a number of meteorological parameters derived through different systems has been carried out by Gebhart et al., 2005. The 1980s era Atmospheric Transport and Dispersion (ATAD) model was compared to two of the more recent models; HYSPLIT and the CAPITA CMC model. Various data sets (ATAD, BRAVO, MM5 and GDAS) were used as meteorological fields. Tests on data at a field station in Texas showed that there was evidence for systematic differences between the average results of different back trajectory models. Depending on the meteorological input

data used in the calculations, the trajectories could be as different by as much as 180º during certain episodes.

Multiple trajectories are used to simulate the air mass history, since the air parcels arriving at a site could have followed different trajectories owing to turbulent atmospheric mixing and advective processes. Multiple trajectories also provide a measure of the uncertainty in the air mass transport pathway and the clustering of individual trajectories discussed later shows how grouping trajectories reduces uncertainty.

Example back trajectory analyses are shown in Figure 3 for Mace Head, Ireland (Cape et al., 2000), Weybourne, UK (Cardenas et al., 1998), the Cape Verde Atmospheric Observatory on the island of Sao Vicente (Müller et al., 2010) where many studies of correlations between chemistry and air masses have been carried out (see details in Table 1) and the TOR station at Porspoder, France.

2.3 Particle Dispersion Models

More complex methods to calculate air mass pathways and air mass footprints have come from Lagrangian Particle Dispersion Models (LPDM), that follow the chaotic pathways of air parcels as probability distributions. Lagrangian methods involve plotting the position of an individual parcel through time, giving the pathline of the parcel within a specific period of time. The Lagrangian dispersion concept is more accurate because individual particles move independently from each other and can thus carry additional information and the advection scheme is more accurate than those of trajectory models as it attempts to capture turbulence, which causes a more probabilistic and realistic growth in the volume of influence.

The 2-D maps created for backward runs illustrate which geographical regions have influenced the air arriving at a site. These can be seen as four-dimensional as opposed to quasi-one-dimensional back trajectories, making them an interesting extension to conventional trajectory models, allowing a more realistic representation of transport in the planetary boundary layer, where turbulence is important.

Stohl et al., 2002 explains the added accuracy and regional spread of dispersion models and how they can be a replacement for simple back trajectory models for comparing atmospheric

composition measurements at a given site. Flesch et al., 1995 describes the parameters within Lagrangian Stochastic models and their use in estimating atmospheric emissions and source-receptor relationships.

The HYSPLIT 4 model can be run with puffs instead of particles with an added dispersion scheme added to the initial trajectory computation (Cohen et al., 2004; Han et al., 2005). This was combined with emissions to extract source–receptor relationships in the area.

Typical models used in many studies include the FLEXPART and NAME models. The FLEXPART model (Stohl et al., 2005; Stohl et al., 1998) (an example is shown in Figure 7), has been used for a variety of research purposes and for emergency preparedness. The model is usually driven by ECMWF meteorological input data. FLEXPART evolved from the FLEXTRA back trajectory model but represents transport and dispersion by calculating the 3-D trajectories of a multitude of particles.

The UK Met. office's NAME (**N**umerical **A**tmospheric Dispersion **M**odelling **E**nvironment) (Jones et al., 2007; Ryall and Maryon, 1998) dispersion model described in detail in the case study in Section 5 can be used to give a footprint of a site on scales of hours to years. The model was developed in the late 1980s following the Chernobyl accident (and originally named (Nuclear Accident ModEl)) to give emergency response dispersion predictions for nuclear incidents (Maryon, 1991). NAME is usually run with the Met Office's operational global NWP model, the Unified Model (Cullen, 1993) meteorological data on a variety of regional or global scales.

Manning et al., 2003 and 2011, and Ryall et al., 2001 have used NAME to calculate the distribution of air masses arriving at Mace Head, Ireland, from different Atlantic Ocean and European regions and by combining with composition measurements at the station, they have derived emission inventories for Europe. In combination with satellite imagery and observational data from Mace Head the NAME model was used to investigate the origin of high particulate matter over the British isles during March 2000 and it showed that the most likely origin of the episode was long range transport of dust from the Sahara region of North Africa and not volcanic ash from an Icelandic volcano (Ryall et al., 2002). Ryall and Maryon, 1998 tested NAME using the ETEX database (European Tracer Experiment database of 168 ground-level sampling stations in Western and Eastern Europe) to assess its suitability to predict the overall spread and timing of a pollutant plume across Europe. The NAME model

has also been used in conjunction with satellite measurements (Hewitt, 2010) to develop a methodology to investigate regional scale carbon budgets.

2.4 Chemistry Transport Models

Chemistry Transport Models (CTM) combine meteorological fields, emissions and physical atmospheric processes with chemistry schemes and reaction kinetics to describe the chemistry and physical transformations within air masses. The NAME dispersion model has an option to add a chemistry scheme to the tracer transport calculations but is not as complete as a CTM. In this work we define CTMs as ones that by default include emission inventories and chemistry and transport schemes.

Emissions databases and chemistry reaction scheme databases are incorporated into models in order to quantitatively predict composition during transport. CTMs contain a large number of uncertainties as do the emissions databases that are used within them. They have been used in parallel with many of the studies detailed in this review to test how well they compare with the observed composition from various emission sources and influences and to add value to the trajectory analysis but CTMs are only mentioned in this review if they have been used to compare to simple trajectory or dispersion models.

Wind data have been used as input for many CTMs such as in the distance-weighted wind roses that were combined with measured ozone as input for the Edinburgh Lancaster Model for Ozone (ELMO) model. The model was used to predict ground level ozone concentrations in the UK (Strong et al., 2006) and it predicts the higher ozone episodes (98^{th} percentile) well. ELMO-2 used the HYSPLIT trajectory model as input and was seen to reproduce ozone episodes and diurnal cycles at several UK monitoring sites during summer 1995 (Strong et al., 2010). In the study by Gilliam et al (Gilliam et al., 2006), the Pennsylvania State University/National Center for Atmospheric Research Fifth Generation Mesoscale Meteorological Model (MM5) was used with seasonal subsets of pseudo-trajectories derived from radar wind profiler data and from simulated wind fields as input to provide an estimate of model errors in terms of wind transport. They found any inconsistencies in the meteorology were passed on to the air quality model. CO and SO_2 transport was well simulated with both MM5 and WRF (Weather Research and Forecast) meteorological parameters over the Mexico city basin during the MILAGRO study (de Foy et al., 2009). De

Foy et al., 2009 used FLEXPART back trajectories, in combination with measured air pollutant concentrations to identify potential source regions and the WRF and MM5 models were used to evaluate the simulated trajectories by comparing the potential source regions with emission inventories.

A Photochemical Trajectory Model (PTM) was used with emissions databases to simulate atmospheric composition, during the Pollution in the Urban Midlands Atmosphere (PUMA) campaign in Birmingham (Walker et al., 2009). The model described the photochemical ozone formation as well as inorganic and organic aerosol formation in north-western Europe and in this case followed advected air masses from source regions in Europe to the receptor location in the UK. The model was tested against 3 day back trajectories to test the model's ability to pull out the origin of all the pollutant episodes and it was not able to trace back the pollutant transport during conditions of slow moving anticyclonic conditions. This example shows how trajectory studies should be carried out independently to trajectory chemical modelling studies as in some cases the model cannot interpret all periods.

The Lagrangian chemistry transport model CiTTyCAT (Cambridge Tropospheric Trajectory model of Chemistry and Transport) (Evans et al., 2000) has been used in many studies investigating atmospheric composition. CiTTyCAT simulates chemical transformation following trajectories with a photochemistry scheme that includes the degradation of some hydrocarbons, a representation of the spread of surface emissions into the boundary layer (using emission inventories) and dry deposition. The chemical initial conditions at the trajectory origin are defined by interpolating concentrations from the TOMCAT CTM, which calculates the abundances of chemical species in the troposphere and stratosphere. CiTTyCAT was seen to accurately simulate 70 % of the variance in the relationship between chemical composition at Mace Head during field measurements in 1996 and the origin of the resolved flow when compared to time series of trajectory-origin-averaged measured ozone (Strong et al., 2006). Different processes influencing the evolution of pollutant levels in a trans-Atlantic plume have been analysed with the CiTTyCAT model and average trace gas concentrations and their correlations (O_3/CO and NOy/CO) calculated to study the factors governing ozone production (Real et al., 2008).

The GEOS-CHEM 3-D global CTM (using assimilated meteorological data compiled at the NASA Global Modeling and Assimilation Office (GMAO)) was used to calculate a long

range transport component to air masses arriving at Cheeka Peak Observatory (western US) from over Asia from trajectory analysis (Weiss-Penzias et al., 2004). The CO produced from Asian biomass burning, and Asian and European fossil fuel and biofuel sources was calculated and compared with measured CO. Reidmiller et al., 2009 used back trajectory analysis coupled with ground-based measurements from the Mount Bachelor Observatory (western US) to confirm GEOS-CHEM simulations, suggesting a significant change in long range transport between 2005 and 2006, owing to changing patterns of long range transport Asian air masses arriving at the site. Investigations of ozone and CO in biomass burning plumes were investigated using the GEOS-CHEM model during the AMMA campaign in West Africa campaign (Real et al., 2010) and the modelled mixing ratios at each successive day along the trajectory are shown in Figure 4, showing how modelled CO decreases downwind at all heights whereas O_3 decreases downwind only in the lowest trajectory.

Back trajectories calculated for analysis of cruise ship measurements in the Atlantic Ocean and were compared to the MATCH-MPIC (Model of Atmospheric Transport and Chemistry-Max Planck Institute for Chemistry version) model (Gros et al., 2004). MATCH is a global atmospheric offline model, driven by 3-D meteorological parameters with a CH_4-CO-HOx-NOx ‘‘background’’ chemistry, a simplified representation of isoprene and other VOC chemistry and emissions of CO and VOCs from energy and industrial activities taken from the EDGAR inventory (Olivier and Visschedijk, 1996) emissions database.

The MISTRA 1-D Lagrangian chemistry model has been used to simulate multiphase halogen cycling mechanisms and compared with the observed association between Cl_2 and pollutants at the Cape Verde islands. The model data was compared to the halogen chemistry during particular air transport pathways and found Cl to be involved in CH_4, DMS and O_3 cycling (Lawler et al., 2009).

During the Rocky Mountain Atmospheric Nitrogen and Sulfur Study (RoMANS), the Trajectory Mass Balance (TrMB) Model was used to probe source-receptor relationships. Hourly measured atmospheric NO_x, SO_2 and NH_3 concentrations were compared with model simulations using HYSPLIT trajectory residence times in various regions and modelled meteorology to determine correlations between measured and modelled chemistry and the transport within and into Colorado (Gebhart et al., 2011).

Another model to quantify source–receptor relationships was the Gaussian trajectory transfer-coefficient model (GTx), which has been used to model PM at Taichung City and Taipei City, Taiwan (Tsuang, 2003; Tsuang et al., 2003a; Tsuang et al., 2003b). It was able to simulate the daily variation of PM concentrations at these sites and was used to determine the source of particulates and dust from long range transport.

3. Applications of trajectory data to interpret composition observations

Many studies have been carried out to relate variability in chemical observations to variations in synoptic-scale circulation. A summary of previous research studies that combine trajectory or dispersion modelling with chemistry datasets from long term measurement stations and field campaigns are presented in Table 1. Examples of the techniques used, the composition measurements that were analysed, the locations, the type of cluster analysis or sector analysis techniques used (described in section 4) and if chemistry modelling was used to complement the study are listed for each study.

3.1 Characterisation of long-term *in situ* ground based measurements

Studies at the Northern hemispheric background measurement site of Mace Head, Ireland used trajectories to assign each 6 hourly air mass arriving at the site to one of eight 45º sectors (centred on north, northeast, and so on to northwest) (Simmonds et al., 1997). If all four trajectories within a day lay within the same sector, it was classified to that sector and if no such allocation was possible for a particular day, then that day was unclassified. Additional trajectory analysis sorted the data into the specific regions of USA and Canada, Greenland and Iceland, Europe, and southerly latitudes and with both of these sector analyses, so-called background or polluted conditions were separated for analysis of ozone and CO levels. The methodology has been shown to be robust for the determination of background trends. The Zeppelin station in the Arctic has also been classified with the same 8 sectors surrounding the site (see Figure 8d) and a transport sector allocated to each trajectory if at least 50% of the last 24 hours of the trajectory were closer than 850 km from the station (Solberg et al., 1996). Spring low ozone episodes were found to originate from a Westerly/Northerly Arctic oceanic direction.

The air arriving at Porspoder in Brittany, France was split into three kinds of oceanic air masses (North Atlantic northern and southern latitudes and North American continent) and seasonal variations of PAN, VOCs, O_3 and NOx according to air mass types over a period of 4 years were studied (Fenneteaux et al., 1999). Corresponding back trajectories for the 10^{th} and 90^{th} percentiles (see Figure 3d) and average levels of the species in winter and summer showed that there were strong seasonal and regional influences on these species. Air arriving at the remote mountain site of Mondy in Siberia were classified into 4 transport pathways (Europe, Siberia, High-latitude, and south-west air masses) and CO and O_3 levels at the site were then averaged for each trajectory type to reveal how European air masses had the highest O_3 and CO levels (Pochanart et al., 2003).

Various studies on many US National Parks and wilderness areas (through the IMPROVE sites) used residence time analysis of back trajectories to find source-receptor relationships and the source type and origin of high air-borne pollutants (Hopke and Allan, 2009). The online Combined Aerosol Trajectory (CATT) tool, uses the IMPROVE sites and 5 day back trajectories to colour weight the trajectories to individual or aggregated sites, in order to analyse long term patterns in pollution transport to these areas (Poirot and Allan, 2009).

In a review of recent aerosol studies in Europe it was found that 11 % of all studies used back trajectory methods to cluster aerosol levels according to their origin and transport pathways (Viana et al., 2008). Rozwadowska et al., 2010, Salvador et al., 2008 and 2010 and Sharma et al., 2006 show some of the latest aerosol source-receptor studies using back trajectory classification at a range of sites in Europe and the Arctic.

Tarasova et al., 2009 classified the vertical as well as geographical origin of air masses arriving at two mountain stations (Jungfraujoch and Kislovodsk) by using potential vorticity, altitude along the trajectory and boundary layer height to discriminate different vertical source areas, as well as the usual classification using the horizontal coordinates of back trajectories. This methodology allowed classification of the air masses according to their contact with the free troposphere and the stratosphere and showed how ozone levels varied significantly depending on their vertical pathway to the mountain tops. CO_2 measurements at the Jungfraujoch mountain station were compared using the FLEXPART dispersion model and the model was also used to calculate the residence times of the air masses in the boundary layer and relate this to CO_2 exchange processes (Tuzson et al., 2011). Delcloo and

De Backer, 2008 separated the trajectory analysis clustering of 32 years of trajectories arriving in Uccle, Belgium by elevation, into planetary boundary layer and free troposphere origins to understand how ozone levels have varied in both high and low elevation air masses.

3.2 Investigating long term trends and seasonality in composition measurements

Understanding the influence of transport patterns on long-term trends is essential to interpreting their changing chemical and physical signatures. Seasonal cluster analysis of trajectory types has been used to reveal the percentage of each trajectory type arriving in each season as well as the ozone trends in each season in Uccle, Belgium (Delcloo and De Backer, 2008). Monthly averaged CO and O_3 for different trajectory types were calculated for a station in Siberia (Pochanart et al., 2003). Junker et al., 2009 studied 12 years of O_3, CO, SO_2, PM and NO_x measurements at four measurement stations in Taiwan and China and separated the composition by their air mass history to show long term trends and also seasonal variations. Abdalmogith and Harrison, 2005 carried out a seasonal cluster analysis of PM levels at Harwell and Belfast in the UK and found small variations in the direction of the clusters and the levels and type of PM observed between the seasons. Solberg et al., 1997 found the background ozone to have a small seasonal variation and a spring maximum as opposed to the summer maximum and large seasonal variation for European polluted air masses at Birkenes in Norway (Figure 5a). Simmonds et al., 1997 looked at the relative ozone contribution from each of the source regions to the observed spring maximum at Mace Head. This was obtained by subtracting the mean ozone concentration for the Northern Hemisphere Marine Mid-Latitude Background (NHMLB) (34.8 ppbV at the time) from 5 years of ozone monthly means. These monthly differences were shown as ozone excesses or deficits relative to the NHMLB ozone concentration and displayed marked seasonal differences between Atlantic, European and American air masses in Figure 5b.

Long term records of composition variations separated by corresponding air mass history are rare due to the fact that running back trajectories on hourly or even daily timescales for over 10 years is very computer-intensive and the meteorological data needed for the model may have changed over the years. Trends in air mass climatology for a 40 year trajectory dataset have been investigated by Shadbolt et al., 2006 and discussed in section 4.1.2.

Jorba et al., 2004 have clustered the trajectory types arriving in Barcelona over a year and displayed this as monthly averaged contribution of each trajectory type that shows the seasonal influence on climatology (Figure 6a). A similar monthly distribution of regional influences was calculated for 15 years at both Alert and Barrow in the Arctic (Sharma et al., 2006) as shown in the monthly distribution of sector influences in Figure 6b. Furthermore Sharma et al.,2006 used a geometric time variation model to describe the temporal variation of 6-hourly Black Carbon (BC) concentrations, including a long-term trend, long-term cycles, seasonal variation, and an autoregressive component that described short-term temporal correlations. The long term trends in both winter and summer varied greatly between the different sectors, especially at Barrow.

Pochanart et al., 2001 found that there were strong seasonal differences in the relationship between residence times of air masses over Europe and ozone levels in Arosa, Switzerland and that ozone concentration depends significantly on European residence times in spring and summer. Eneroth et al.,2003 carried out a cluster analysis of ten years of trajectories arriving at Ny-Ålesand on a monthly basis, to study inter and intra-annual variations of CO_2 and found that there were seasonal differences in the prevalence of each trajectory type that lead to varying CO_2 levels but they found no conclusive linkage between CO_2 levels and transport pathways

3.3 Analysing the air mass history for aircraft based measurements

Combing trajectory and dispersion studies of air mass histories with aircraft measurements helps to build a 3-D picture of air mass movement and transport. This picture is often achieved by calculating back trajectories and dispersion pathways from multiple elevations along the aircraft flight track.

Back trajectories at 1, 2 and 3 km were run during the RAMMPP aircraft campaign over the mid-Atlantic US (Taubman et al., 2006) in order to pick out any variations in atmospheric circulation patterns in the lower atmosphere and identify the impacts on regional transport. Vertical composition measurements were combined with these back trajectories to look at the history of the air during summer ozone pollution episodes and the role of transport to the boundary layer.

Hains et al., 2008 calculated average ozone levels in vertical bins of the atmosphere and transposed this on the various trajectory types that were observed during flights over the north-east US in the INTEX-NA campaigns in 1993-2003. An 80 km radius circle was drawn around each hourly point on the trajectory and the NOx emissions from that area were compared to the ozone level at the corresponding altitude bin.

Clustering of air mass composition from the MOZAIC dataset (long-term composition measurements from an aircraft network) using the FLEXPART trajectory model was compared to a multivariate analysis technique that classifies ozone-rich layers observed in tropospheric profiles according to their origin by season (Colette et al., 2005). The ozone multivariate analysis technique was found to underestimate the long-range transport that the trajectory studies were able to identify.

Reverse Domain Filling (RDF) is a technique that defines regions of similar air mass origin with the use of back trajectories and is used to plan flight paths through regions of interest. Methven et al., 2003 used RDF trajectories arriving on a high-resolution three-dimensional grid (RDF3D) to simulate air mass structure accurately by colouring arrival grid points according to the specific humidity (or potential vorticity) at the origin of each trajectory. Back trajectories were calculated from every point on the 3-D grid from a reference time near the anticipated flight time and for each RDF forecast trajectories were integrated backwards in time for three days, using a combination of ECMWF forecasts and interpolating specific humidity from the forecasts to the origin of each trajectory. The flights were targeted at regions where there were neighbouring air masses with distinct origins. Chemical measurements during the flights were overlaid onto the RDF maps and variations in composition were compared with the pre-selected air mass types to see if composition did vary within the different air mass types. Hydrocarbon measurements during the IGAC Lagrangian 2K4 experiment in July 2004 (Methven et al., 2006) were analysed according to calculated back trajectories and RDF was used to identify Lagrangian matches between flight segments from different aircraft. Figure 7 shows examples of horizontal and vertical RDF plots used to plan flights that intercept biomass burning plumes during this study. This was the first experiment aiming to take measurements that were linked by trajectories over intercontinental distances through the free troposphere, where vertical motion is important and was described as a “pseudo-Lagrangian experiment”. Results from the FLEXPART model, run with CO tracers were also used to confirm matches.

Following on from this form of analysis, Real et al., 2008 isolated the chemistry of an anthropogenic pollutant plume transported across the North Atlantic at low altitudes. Similar use of flight data and back trajectories was used to track Alaskan wild fire plume transport during the ICARTT aircraft campaign (Real et al., 2007) and study the evolution of the composition and ozone formation a few days from emission. Schmale et al., 2011 used OFFLINE and LAGRANTO trajectory models as well as the FLEXPART dispersion model with EDGAR emissions to study the aerosol composition of a variety of air masses over Greenland during the POLARCAT (Polar Study using Aircraft, Remote Sensing, Surface Measurements and Models, of Climate, Chemistry, Aerosols, and Transport) summer campaign in Greenland.

During the AMMA campaign in West Africa RDF back trajectories were used to split the area over which flights had passed into boxes representing the 4 main regions of recent (10 day) origin (Law et al., 2010). Vertical differences in composition and the occurrence of convection were included in the separation and interpolation of the measured composition data.

4. Methods for deriving classifications of air mass pathways

In order to analyse the association between trajectories and concentrations of air arriving at a site, a multitude of methods to carry out trajectory classifications have been devised. These can generally be split into two different methodological groups. The first is to sort air masses by designated air mass sectors, representing a different influence on composition. Relationships between atmospheric composition and air mass origin are often analysed this way, by isolating the highest species concentrations or the exceedence levels and finding the main sectors of influence during those polluted periods. The second is to cluster the trajectories using a mathematical technique and then to analyse the concentrations at the receptor site for each trajectory classification to see whether each classification is chemically distinct. These fall broadly into sector classification and statistical classification (cluster analysis) and are described in detail in the following sections.

4.1 Geographical sector classification

The method of assigning trajectories according to regional influence and comparing the composition in those regions was developed by Ashbaugh , 1983. Siebert, 1994 added a way to deal with uncertainties followed by a subsequent refining of the method by Stohl, 1996 to allow for a non uniform concentration within each grid. The technique superimposes a grid over the domain of the trajectory computations and then a geometric mean concentration of all the trajectories passing through each specific gridded region is calculated. The assigning of the grid boxes is done either from prior knowledge of source regions, for example for western US influences in Figure 8a (Weiss-Penzias et al., 2004) where the region with high CO from the MODIS satellite was chosen to locate an Asian box or by geographical limits for Siberian and European masses (Paris et al., 2010) and (Salvador et al., 2008) respectively in Figures 8b and 8c or as an objective division of the radius around the site such as that done for Svalbard (Solberg et al., 1996) in Figure 8d.

Methven et al., 2001 studied the back trajectories for air arriving at Mace Head, Ireland using the technique of "Origin Averaging" by calculating a climatological density of origin (the region where the trajectories originate) and assigning a corresponding composition concentration for each area of origin. This technique identified chemical air masses associated with different ozone levels but the origin of CO and CH_4 levels proved harder to track.

4.1.2 Residence time analysis (trajectory regression analysis)

Residence time analysis is a qualitative source attribution technique (Ashbaugh, 1983) which generates a probability density function identifying the likelihood that an air mass will traverse a given region *en route* to the site of interest over a given time period. Air parcels that travel quickly through a pollutant source region have less time to accumulate pollutants than air parcels which remain in the source region for a long time.

The seasonal variation and climatological pathway of airflow to various stations around the Atlantic Ocean was studied by Merrill, 1994 in order to interpret the trace gas measurements. The number of hours spent by the trajectories in each area of a grid for a given season was shown as a cumulative probability field and the areas of high probability indicated that

trajectories which subsequently reached the site had spent more time in that geographical area. The same technique was used to help interpret composition measurements in the Pacific Exploratory Mission-West A (PEM-West A) flight experiment and ozone profiles above Bermuda (respectively Merril 2006a and b).

An example of residence time analysis for investigating long term air mass climatology which would affect long term changes in composition is shown in a 40 year study of airflow trajectories and residence time calculations for the lower peninsula of Michigan (Shadbolt et al., 2006) where monthly air mass climatology anomalies were plotted as standard deviations from the grid cell mean value. Here positive anomalies depicted airflow corridors that had more trajectories than average, and negative values depicted airflow corridors that had fewer trajectories than average.

Doddridge et al., 1994 carried out flow climatology studies to show the probability of the air over surrounding areas arriving at Mace Head and assigned a probability contour plot to represent the flow over a 3 month period that revealed predominantly Atlantic airflow in one year and a large amount of European anticyclonic Atlantic air in the other.

In order to compare aircraft composition measurements with corresponding back trajectories Traub et al., 2003 chose 4 areas to represent source regions over the Mediterranean. Each air parcel trajectory that passed over each of the defined regions was added to that particular source type. If a back trajectory resided over two or more defined regions, the residence time of the air parcel above the regions had to be above a critical residence time (2.75 days in this study), by which time it was thought that the air mass had adopted the chemical characteristics of that region.

Residence time analysis was carried out to attribute ammonium sulphate concentrations to source areas for the IMPROVE sites (US National Parks and wilderness areas) (Xu et al., 2006). The area around each station was divided into four quadrants and everything outside the site's US state was grouped into one of six larger regions, showing consistently considerable sulphate emissions from the Pacific Ocean arriving at these inland sites.

Air masses reaching the high altitude station of Arosa in Switzerland (Pochanart et al., 2001) were classified according to the influence of European regional pollution. Residence times of

air masses over Europe were derived on a monthly and seasonal basis to separate background ozone conditions (low European residence times) and the occurrence of accumulated high European ozone events (long residence times). Solberg et al., 2008 calculated residence times of air masses over a central European domain arriving at many European measurement stations from 7 day back trajectories for a 7 year period and linked this with the potential to form high ozone levels, such as the European heat-wave of 2003.

Studies at the Zeppelin station in Svalbard (Solberg et al., 1996) allocated a transport sector to a trajectory if at least 50% of the last 24 hours of the trajectory were closer than 850 km from the station as shown by the radius in Figure 8d and used this allocation to understand the origin of air masses causing ozone depletion events. Particulate Matter levels from Saharan and non Saharan air masses arriving at Castanya in Spain were analysed as well as a further gridding the Saharan region into 11 source areas (Escudero et al., 2011). Identification of a probability of detection for identifying dust events was done by linking the 11 source regions with 3 Spanish receptor regions.

Back trajectories run from Mount Bachelor and Cheeka Peak Observatories on the west coast of the US have been run during a variety of measurement periods to track how long each trajectory particle spends in an “East Asian box” (see Figure 8a) in order to estimate the magnitude of pollution transport from Asia (Weiss-Penzias et al., 2004; Weiss-Penzias et al., 2006; Wolfe et al., 2007). The number of trajectories passing through the box were weighted by the average amount of time each trajectory spends in the box and the trace gas measurements at their time of arrival showed the effect of long range transport of Asian emissions to the west coast of America.

In a study of nitrogen in precipitation and corresponding ambient PM measurements in North Carolina’s large agricultural corridor Occhipinti et al., 2008 used back trajectories to find air masses in which a minimum of 50% of the rainfall would have transited surrounding marine and agricultural source regions.

The spatial extent (footprint) for ozone from a given location can be examined by correlating the short-term component of the ozone time-series data at that location with that at all neighbouring stations. The short-term (weather-related) component was separated from the long-term (climate) and seasonal components embedded in ozone time-series data at

Whiteface Mountain in New York state with the Kolmogorov filter (Brankov et al., 1998). They carried out the correlations between each trajectory cluster type and the short-term component of ozone was found to be correlated with that measured at a number of other sites lying within a cluster envelope, lagged by up to 3 days. This form of analysis tested the hypothesis of whether when ozone was transported from a certain direction, the time-lagged correlation increases in that direction. The distance at which the time-lagged inter-site correlations reach a maximum is expected to be proportional to the distance the air mass can travel in that time and would suggest the transport of ozone pollution to the location of concern and reveal the spatial and temporal scales involved.

4.1.3 Trajectory statistical methods

Trajectory statistical methods can be divided roughly into those where residence time is not weighted by the concentration at the receptor point (e.g. potential source contribution function (PSCF)) and methods that do weight by the concentration (e.g. Conditional Probability Function (CPF)). Scheifinger and Kaiser, 2007 explains and compares CPF, PSCF and the Redistribution Concentration Field (RCF) methods. CPF and PSCF are less computationally intensive than residence time analysis as they only need to count the number of trajectory end points in each grid for all sampling days.

CPF is a more advanced form of residence time analysis where grid cells are superimposed over specific regions of interest and the CPF at each grid cell represents the probability of an air mass arriving at the receptor site, after having been observed to reside in this specific geographical region. Pioneering work by Ashbaugh et al., 1985 plotted the probability on a map, showing the influences on the air arriving at Grand Canyon National Park as shown in Figure 9a, indicating which regions would potentially contribute to high sulphur levels and Figure 9b shows the actual source contribution function, which are often different to each other.

When pollutant concentrations higher than a specified value are used to derive this probability, areas with high CPF can be considered as source regions of the pollutant under study and are frequently in the air mass pathway to the site. The CPF, otherwise known as Relative Residence Time is the ratio of the trajectory segment endpoint counts for the sorted

residence time to the everyday residence time and identifies the likelihood that if an air mass passes over a given area, it will arrive at the receptor with a high or low concentration.

Various tests to check the statistical significance of every grid cell CPF value have been used, such as one based on the binomial distribution for high CFC levels at Big Bend National park (Vasconcelos et al., 1996). The residence time of each trajectory arriving at 4 Alpine stations was weighted with the deviation of the actual concentration from the 3-monthly running mean and from that potential pollutant source regions were inferred (Kaiser et al., 2007).

The incremental probability (IP) is the difference between the sorted and the everyday residence time and identifies regions that are more or less likely to be traversed during periods of high or low concentrations compared to an average day. Particulate sulphur measurements at Big Bend National park, Texas (Schichtel et al., 2006) used conditional probability analysis and Incremental Probability to segregate the 20^{th} and 80^{th} percentile sulphur levels according to regions with the results from 2 tracer release sites as shown in Figure 10.

The Potential Source Contribution Function (PSCF) can be interpreted as a conditional probability describing the spatial distribution of probable geographical source locations inferred by using trajectories arriving at the sampling site. Cells related to the high values of potential source contribution are the potential source areas and show those source areas whose emissions can be transported to the measurement site. It identifies locations more likely to be upwind if receptor concentrations are high and these upwind regions of highest probability are associated with emissions that contribute to impacts at the site, including areas where secondary formation is enhanced. The potential contribution function is the ratio of the high concentration probability divided by the everyday probability as opposed to the incremental probability which is the subtraction of the every day from the high.

Whilst interpreting the results of PSCF and Residence time ensemble trajectory techniques on PM levels detected at a site in Vermont, Poirot et al., 2001 stated that only qualitative indications of predominant transport patterns can be obtained and they can be highly sensitive to the subjective metrics used to define high pollution episodes or the scale of the gridded domain. However, the resulting maps indicating source regions can be a very powerful tool for understanding the air quality influences on a station.

Studies at Underhill, Vermont (Poirot et al., 2001; Polissar et al., 2001b) used PSCF analysis to combine the aerosol data with the air parcel backward trajectories. PSCF was applied to identify possible source areas and pathways that give rise to the observed high particulate mass concentrations from each of the emission sources (e.g. coal, smelters, salt). The PSCF plot for the black carbon factor shows high probabilities in the area surrounding the sampling site, indicating a strong local influence from residential wood combustion in northern New England and south western Quebec. Han et al., 2005 used PSCF to attribute high Mercury levels at 3 rural sites in New York state to nearby coal-fired power stations.

Polissar et al, 2001a and 1999 used PSCF analysis to study the origin of aerosol in the Arctic and found that long-range transport of anthropogenic aerosol to the Arctic is more effective in winter and spring than in the summer. Geographical Information System (GIS) software has shown to be useful for residence time analysis, such as the TrajStat software that has been developed by Wang et al., 2009 to compute Potential Source Contribution Function (PSCF) analysis with back trajectories.

Potential source regions give an idea of which directions/ areas pollutants are coming from but when assigning source regions by probability analysis the concentrations measured at the receptor locations are attributed equally to all segments of the related trajectory, while in reality emissions just take place in some segments. To account for this, an iterative scheme was developed by Stohl, 1996, redistributing the measured concentrations along the trajectories according to the estimated concentration field from the previous iteration. The Redistribution Concentration Field method (RCF) (Stohl, 1996) aims at extracting more information from the data, a step on from the Conditional Probability method where the concentration measured at the receptor sites is attributed with equal weight to all segments of the trajectory. Pollution sources are usually concentrated in ‘‘hot spots’’ so probes into concentrations at a smaller scale is needed. The concentration values along each trajectory are iteratively re-weighted according to the ratio of the concentration of that grid cell to the mean concentration of all grid cells along the path of that particular trajectory. The results are reported on maps where each grid square is assigned a weighted concentration of the component under study. Salvador et al., 2010 used this method to study transport pathways of particulate matter and aerosols to various locations in Europe. Emission maps for NO_y have been derived over Europe from Redistributed Concentration Field (RCF) derivations

(Wotawa and Kroger, 1999) and are shown in Figure 11b and (Wotawa et al., 2000) compared CPF and RCF for the 90th percentile value of O_3 for various stations in the Alps.

Mean concentration loadings from surface sources of Persistent Organic Pollutants for each grid cell within the pathway of Kosetice in the Czech Republic were calculated from the composition of the air masses that pass over each grid (Dvorska et al., 2009). Centres of gravity in defined sectors were determined to quantitatively compare mean loads in particular countries and the relative contribution of these countries to air pollution at the site. Potential source regions of dust and aerosols arriving at Tenerife in the Canary Islands were studied using Median Concentrations at Receptor (MCAR) plots to represent the median concentrations at the Izaňa station, recorded for various aerosol compounds when air masses passed above each grid box (Rodriguez et al., 2011). PSCF analysis of trajectories separated by the source apportionment technique of varimax-rotated factor analysis was used for aerosol measurements at Antalya, Turkey (Güllü et al., 2005), highlighting the areas that resulted in high aerosol levels at the site and the map for one factor is shown in Figure 10.

4.1.4 Footprint Emission Sensitivity

In situ greenhouse gas measurement data from three global networks (from nine measurement sites) were combined with back trajectories to extract emission information from global observed concentration increases over a baseline (that was objectively determined by the inversion algorithm) (Stohl et al., 2009). The plot of footprint emission sensitivity for this study is shown in Figure 11a for 20 day backward runs, showing the areas around the world that are most sampled from this station network. An NOy emission map over Europe has been derived from the concentration field calculated from trajectory statistics and the EMEP emission inventory (Wotawa and Kroger, 1999) and shown in Figure 11b.

The FLEXPART model was used to analyse transport pathways from potential flux regions towards Siberia as shown in Figure 8b. Ten day back trajectories released along the aircraft flight track were calculated and the data were grouped according to common transport properties with cluster analysis and this was used to investigate to which extent footprints can explain the air mass chemical composition (Paris et al., 2010). The footprints (relative residence times below 300 m) were calculated using Potential Emission Sensitivity (PES),

complemented by 10-day averaged relative contributions from the stratosphere to explain the source of the atmospheric composition along the flight track. Halse et al., 2011 used FLEXPART backward runs to plot the footprint emission sensitivity (the residence time of air masses per grid cell normalized by the volume) for persistent organic pollutants (POPs) from passive air samplers at 86 European background sites. This was also multiplied by emission inventories to calculate emission contributions to the European background.

Average footprint emission sensitivities were calculated around Arctic measurement stations using FLEXPART to study the origin of higher levels of aerosols, black carbon and ozone at Zeppelin, Alert, Barrow (Hirdman et al., 2010a) and the same stations and Summit (Hirdman et al., 2010b). The 10 % highest and lowest measured species concentrations were selected to calculate the average emission sensitivity for that data subset. The highest and lowest 10 % emission sensitivity (S_P) and the total emission sensitivity (S_T) peak near the observatory (emission sensitivities decrease with distance from the station) so this bias was removed by calculating a relative fraction, $R_p = L/M * S_P/S_T$ where M is the number of measured concentrations and L =M/10 highest or lowest concentrations. If the measured species were completely unrelated to air mass transport then the data subset and full dataset would look the same and the fraction would be 0.1 but if it was greater than 0.1, the cell would be a potential source (Hirdman et al., 2010b). The regions used in the footprint analysis, the average contribution of each region over a year and their annual average variation over 20 years are shown in Figure 12 (Hirdman et al., 2010a).

FLEXPART backward runs at 5 different elevations were used to assess the surface influence on CO_2 measurements from an aircraft over Spain (Font et al., 2011). The horizontal, vertical and temporal extent of the Regional Potential Surface Influence (RPSI) residence time on atmospheric CO_2 mixing ratios was calculated and a principal component analysis was carried out on the resulting residence times. Kuhn et al., 2010 looked at the transport of pollutant plumes from Russian and Alaskan forest fires by plotting the footprint emission sensitivities related to aerosol measurements at Ellesmere Island in the Canadian high Arctic and also used combined the footprint analysis with the EDGAR emissions database to isolate the sources of high aerosols.

4.1.5 Source identification with Positive Matrix Factorisation (PMF)

Positive Matrix Factorisation (PMF) is a multivariate mathematical model that has been used for source-receptor modelling that aims to determine the major sources of a sampled atmospheric species. PMF analyses are used in regulatory studies to assess pollution sources and divide the data into common sources such as various industrial activities or different types of fuel burning.

In an extensive study of the transport of Particulate Matter to a station in Vermont (Poirot et al., 2001) used both PSCF and residence time (Incremental Probability) analysis to help interpret and complement the results from multivariate mathematical models (Positive Matrix Factorization and UNMIX) which had identified seven common sources with different corresponding PM levels. PMF and UNMIX are used when source profiles are not known and are a form of factor analysis that is different from the traditional Principal Component Analysis (PCA, see section 4.2.3) Various maps showing the influence of nearly the whole of North America for different sources of PM (e.g. coal, woodsmoke, oil, soil, smelting etc.) were constructed and compared. Other examples of PMF studies that incorporate PSCF studies of actual sampled data are shown in Pekney et al., 2006 and Kocak et al., 2009 for Particulate Matter in Pittsburgh and Turkey respectively, Du and Rodenburg , 2007 for PCBs in New Jersey, Dogan et al., 2008 for aerosols in Turkey, Choi et al., 2010 for VOCs in Korea.

The source-apportionment Chemical Mass Balance (CMB), UNMIX, and Positive Matrix Factorization (PMF) models described in Song et al., 2008 are advanced multivariate receptor models that determine the number of sources and their chemical composition and contributions without knowing the source profiles.

4.2 Cluster analysis techniques and other statistical techniques to group air mass histories

Clusters are groups with similar distributions, in the case of back trajectories, similar directions and lengths or a combination of trajectory pathways and composition. Cluster analysis provides an objective means of clustering trajectories whilst giving information about the history of the air mass and the air pollution climatology of a site, helping to determine source-receptor relationships.

Cluster analysis is a multivariate statistical technique that groups individual trajectories of an ensemble into a smaller number of clusters, meaning that the errors in the individual trajectories tend to average out. Some examples of derived mean trajectories with which to carry out composition comparisons are shown in Figure 13 and a description of the techniques of the various techniques used in this area of research are described in the sections below.

A variety of cluster analysis techniques k-means, Dorling, and Partitioning around Medoids (PAM) and Two step algorithms have been tested on back trajectories arriving at Cape Verde to see how the average chemical composition varied in each cluster (Mace et al., 2011). The Dorling k-means method was shown to be the most efficient method for this study.

Kassomenos et al., 2010 reviews three of the commonly used cluster analysis techniques and their dependence on arrival height, with examples of PM_{10} from trajectory clusters around Athens. These were a hierarchical, non-heirarchical (k-means) and an artificial neural network known as Self Organising Maps (SOM). They recommended that a range of clustering techniques should be preferably used over one type even though their results from the various tests were similar.

There are two different types of clustering algorithms, namely hierarchical and non-hierarchical clustering.

4.2.1 Non-hierarchical clustering methods

Non-hierarchical clustering requires that the number of clusters is already known and that the objects are distributed between those. This algorithm is widely used in cases where *a priori* information on the nature of the measurements is available.

The k-means procedure is a non-hierarchical iterative algorithm that uses a specified number of clusters, k, to partition the data by comparing each object to the arithmetic mean of all the members of each of the k clusters (cluster centres). The selection of the optimal number of clusters that best describes the different air flow patterns is performed by computing the

percentage change in within-cluster variance, as a function of the number of clusters (Dorling et al., 1992). The assignment of members (trajectories) to a given group (cluster) was carried out by minimising the internal variability within the group of trajectories and maximising the external variability between different groups based on the trajectory co-ordinates. It uses the Root Mean Square Deviation (RMSD) of all individual clusters from their cluster mean trajectory against the number of clusters retained until a "break" was reached, indicating that two clusters have been merged which were unacceptably different. Alternatively, if a threshold percentage change in RMSD was exceeded at any particular point in the clustering process, this was also taken as an indication that an optimum number of clusters had been reached. The k-means clustering method is often quoted as the Dorling method in climatological clustering research and is well suited for large databases because of its relatively small computational requirements.

Dorling and Davies, 1995 analysed six European (3 UK and 3 Norwegian) measurement station's precipitation chemistry levels by cluster analysis of back trajectories using this method. Huang et al (Huang et al., 2010) have used the k-means cluster technique to separate 15 years of particulate matter concentrations at Alert in the Arctic into clusters (See Figure 12a). Average monthly ozone and mercury were calculated for cluster mean trajectories (calculated with k-means) around two Arctic stations to understand seasonal changes in transport to the Arctic and the effect on ozone levels (Eneroth et al., 2007) and the ensuing trajectory clusters for Zeppelin and their average length over 5 days is shown in Figure 14b. Sharma et al., 2006 used k-means clustering to derive 6 main clusters for Alert and Barrow stations for subsequent analysis of average composition in each sector. For the k-means clustering of Cape Verde air masses mentioned previously (Mace et al., 2011), the clustering was carried with and without chemistry (one of O_3, CO, NO_x, NO_y, ethane, propane, acetylene and CO:VOC ratios) to yield similar trajectory clusters but clustering with ethane appeared to be the best way to distinguish between the different chemistry regimes in each trajectory cluster. Figure 12c shows one of the 8 trajectory clusters for Cape Verde (African) from the clustering with ethane (Mace et al., 2011).

Makra et al., 2010 used the k-means algorithm for comparing trajectories with pollen levels at three European cities using a Mahalanobis metric, where the clustering of back trajectories was performed with a function which gathers the extreme trajectory positions belonging to a

cluster. This then encloses and creates the smallest convex hull (http://www.qhull.org) with minimum volume covering the backward trajectories of the clusters.

Fuzzy mean c-clustering is a technique that is very similar to k-means but each trajectory has a degree of belonging to several clusters, as in fuzzy logic, rather than belonging completely to one cluster. Fuzzy c-means uses an iterative algorithm to determine the grade of membership of each trajectory in each cluster, with 0 being no membership and 1 indicating full membership and in between is partial membership. Each trajectory was assigned to a single cluster for which it has the largest value. Xia et al., 2007 used this technique to derive seasonal clusters of trajectories for air arriving in Beijing, China, which managed to separate aerosol into physically distinct groups, explaining 47 % of the variance. However, this method could not separate fast from slowly moving trajectories.

Borge et al., 2007 applied a novel technique of two-stage clustering to capture the influence of relatively short, slow-moving trajectories on local air quality. An initial clustering (k-means) gave clusters that were influenced by trajectory length (wind speed) then the short trajectories with unclear directionality were re-analysed with the same methodology to create a further discrimination between them. Davis et al., 2010 also used this 2-stage technique on a series of stations in Virginia, USA using the distances between the horizontal and vertical trajectory endpoints and the station. Polluted air masses over Athens (Markou and Kassomenos, 2010) were studied by applying a second clustering method (using the Haversine formula, great-circle distance between two points) to separate the clusters already obtained by k-means clustering based on the length of their cluster-mean trajectories. This allowed them to distinguish between short slow moving and long fast moving trajectories.

Self Organising Maps (SOM) is considered an advanced approach of clustering (a type of Artificial Neural Network) that can produce reliable segregation even in difficult cases. They operate similarly to k-means, but instead of using a number of clusters they utilise a grid of nodes with predetermined shape and size. This grid iteratively adjusts to the data until it maps as close as possible their structure in space. The obtained nodes (or clusters) are also organized in a 2-D grid so that similar clusters are placed near each other. In that way, clustering is performed following a structured approach, in contrast with the unstructured k-means approach.

Kassomenos et al., 2010 used this technique on 4 years of back trajectories arriving in Athens, Greece and Karaca and Camci, 2010 for trajectories arriving at Istanbul, Turkey both with corresponding Particulate Matter measurements.

4.2.2 Hierarchical Clustering methods

The purpose of the hierarchical clustering is to join objects into successively larger clusters, using some measure of similarity or distance, by constructing clusters within clusters. All the classified objects are considered at each step of the hierarchical clustering and the process is determined by the construction of an agglomeration tree. This approach is usually used when the number of clusters is unknown. Kalkstein et al., 1987 has compared three hierarchical clustering procedures (Ward's, average-linkage and centroid) for climatological studies of back trajectories and showed that the average-linkage method was the most appropriate.

Hierarchical clustering partitions data following a series of steps either by grouping or by separating the objects one by one in each step. The two closest clusters are merged in each step, starting the procedure with singleton clusters and ending with a single cluster that contains all the objects. There are a number of different techniques to measure the distance (or the similarity) between the clusters, which may lead to different subsets.

Ozone measurements at Mace Head (Cape et al., 2000) for 3 years were divided into four 3-month periods and the derived trajectory clusters were used to classify the ozone (these are shown in Figure 3b as examples of linear trajectories). The squared distances (kilometres north and east and elevation from ground level (as pressure)) for each arrival time were calculated between pairs of trajectories, one trajectory from each cluster. Starting with each trajectory as a cluster, all possible pairs are evaluated and the two clusters with the smallest average distance between their members were joined. The procedure is designed to minimise within cluster variance and maximise between cluster variance. Figure 14 from (Cape et al., 2000) shows how the R^2, Route Mean Squared (RMS) distance and number of clusters changes as the number of clusters by iterations increases (Cape et al., 2000), a method commonly used to calculate the optimum number of clusters.

Aerosol Optical Depth measurements in Northern India were analysed using Cluster Spatial Variance (SPVAR) and their frequency of occurrence is shown in Figure 13d (Gogoi et al.,

2009). A combination of trajectory pairs were used to calculate the SPVAR and the 4 derived clusters. SPVAR is the sum of the squared distances between the endpoints of the cluster's component trajectories and the mean of the trajectories in that cluster. The Total Spatial Variance (TSV), the sum of all the SPVAR, is calculated and the pairs of clusters were combined (with the lowest increase in TSV (which is initially zero)). At each iteration, one more trajectory is joined to a cluster. The iterations are continued until the last two clusters are combined. The iterative step just before the large increase in the change of TSV gives the final number of clusters. Trajectories arriving in Hong Kong have been clustered using this method for analysis with CO and O_3 measurements (Wang et al., 2004) as well as trajectories arriving into Lamas d'Olo in Portugal (Carvalho et al., 2010).

Ward's method is a type of hierarchical cluster analysis that uses the sum of squares of the distance of each trajectory from the cluster's mean trajectory and has been used for classifying trajectory types at various Atlantic Ocean sites (Moody et al., 1989) and to study 10 years of tracer levels at Svalbard (Eneroth et al., 2003). The distances between each 5 day trajectory at every hourly time step along the trajectories were calculated and the spatial variance between two trajectories was quantified as the sum of all squared distances. The smaller the distances, the more similar were the trajectories and they were grouped together until the spatial variance increases rapidly.

4.2.3 Principal Component Analysis

For the purposes of data dimension reduction in large datasets, Principal Component Analysis (PCA) has been used to group trajectories. Riccio et al., 2007 examined the role exerted by meteorology on air quality through the classification of atmospheric circulation patterns as a function of air mass origin for 10 years of back trajectories arriving into Naples. They used 116,896 trajectories, embedded in a 144 dimensional space. It was found that the first eight components, i.e. the reduction of the (116896×144) data matrix to a (116896×8) matrix, explained almost the total (>98%) portion of initial variance, without sacrificing accuracy and without significantly affecting the classification procedure, but with a large speed-up in computations.

4.2.4 Significance tests between air mass types and composition

The Kruskal-Wallis test (Miller, 1981) is a nonparametric technique that has been used in many studies e.g. (Mace et al., 2011; Salvador et al., 2008; Sharma et al., 2006; Sharma et al., 2004) for non normally distributed data to identify whether the median species concentrations were different between all the air mass sectors. If the Kruskal-Wallis test leads to the rejection of the null hypothesis and, thus, to the conclusion that not all samples are identical, it is appropriate to use a multiple comparison procedure to find out which clusters were different from the others. For example, the Dunn test was used as a multiple comparison procedure after the Kruskal-Wallis test in a study of ozone levels in a few stations in Northeast USA (Brankov et al., 1998). The Spearman rank-order correlation coefficients were used in (Han et al., 2005) and the Kendall τ and Pearson's correlation coefficient were used in (Paris et al., 2010) to test for significant differences in the chemistry between the clusters or between trajectory and dispersion model clusters as in (Han et al., 2005). Mace et al., 2011 also used the Behrens-Fisher test to test differences in composition between different clusters.

Analysis of Variance (ANOVA) is a technique to test whether or not the means of several groups are all equal. Borge et al., 2007 used an analysis of variance analysis on 24 hour mean NOx and PM_{10} to test if the cluster averaged concentrations for each pollutant were statistically significant. Occhipintu et al., 2008 used ANOVA statistics to study the influence of agricultural areas on PM and nitrogen deposition and Makra et al., 2006 and Xia et al., 2007 used ANOVA analysis to test for differences in composition between various air mass clusters in the Carpathian basin and Beijing China respectively.

4.2.5 Cluster analysis on dispersion models

Cluster analysis has been carried out on dispersion models on a number of occasions. One such study used a k-means cluster technique based on FLEXPART dispersion model footprints where regions were chosen prior to clustering analysis (regions of specific sources or sinks relevant to the site) (Paris et al., 2010) as shown in Figure 8c. Stohl et al., 2002 used Mace Head FLEXPART runs to compare the accuracy of classical trajectory techniques against dispersion models by comparing representative single trajectories and trajectory clusters (retroplumes). Particles were released in each dispersion run and their position data was used for deriving a condensed model output. Cluster analysis was used as a semi-

objective method applied to best characterise the position and shape of the entire retroplume, by calculating the retroplume centroid, followed by on-line cluster analyses of the particle positions at selectable time intervals that minimizes the root-mean-square distance between the particles of each of the clusters and their respective retroplume cluster centroids, and maximizes the distance between the cluster centroids. As the clustering is performed independently each time, subsequent retroplume cluster centroids do not lie on a trajectory and thus cannot be connected by a line in a trajectory plot.

Potential Source Contribution Function (PSCF) multi-receptor (MURA) residence time probability analysis has been used to simulate sulphur concentrations in the South West US (Lee and Ashbaugh, 2007a) on HYSPLIT 4 backwards runs with puffs of particles rather than individual particles. Back trajectory Conditional Probability analysis uses a single receptor at a time (see section 4.1.3) whereas the MURA method uses several receptors at once in order to detect sources with greater accuracy. The MURA method designates potential source regions by counting trajectories for each grid cell and then examines them to see how often each region affects each receptor. Lee and Ashbaugh, 2007b developed the single receptor forward conditional probability (SIRA) method, which is a conditional probability method with the second step of the MURA method added to it. A high CP indicates a higher probability that that location contains a source or is on the pathway to the source. To calculate the SIRA back trajectories are divided into 1 hour segments and a grid is superimposed on the area of interest and the number of trajectory segment points located in each grid cell was counted for both sample days and high incident days. Lee and Ashbaugh, 2007c looked at the impact of running trajectories at various elevations on the MURA method and found that it was best to run ensembles of trajectories in the MURA method so as to average out most of the biases found from different trajectory starting heights.

5. Case study: Using the NAME model for classification of air mass types and corresponding composition variations at a site

In order to illustrate the use of dispersion models to untangle the regional influences of an atmospheric observatory, a step by step description of a new methodology that has been developed using the NAME model output is detailed for an observatory on the UK North Sea coast as the case study.

The Weybourne Atmospheric Observatory (52°57'N 1°07'E) on the North Sea coast (Figure 15) is at a strategic location for receiving a variety of Atlantic, Arctic, European, UK and North Sea air masses. In previous work, air masses have been divided according to wind direction or manually classified trajectories (Cardenas et al., 1998; Penkett et al., 2007) but because of the rapidly changing wind directions and the variety of influences close by, a more detailed technique is required.

In this case study the NAME model was run in backwards mode, 10 days backwards in time at 3 hourly intervals (related to the timescale of the Unified Model meteorological fields) for 4 years of Weybourne station data (2006-2009). The particles were released from the height of the station's tower where the instruments sample (10 m). All instances when the particles were near to the ground (0-100 m) were recorded to indicate when surface emissions from different geographical regions (marine or land) will have been picked up by the air mass and transported to the observation station. The horizontal spatial resolution was 0.25º x 0.25º for the 10 day regional domain used.

Figure 16 shows how single run outputs can be combined to produce integrated plots (monthly and annual) that illustrate the seasonality of the site footprints. Monthly averaged footprint plots for 2008 for Weybourne show seasonal changes in the air masses histories. The 12 monthly Weybourne footprints show how there is a subtle seasonal pattern of more Arctic air in the spring and summer months and a wider range of the footprint in winter.

For the station of interest, the domain of influence of the NAME run is split subjectively into the main geographical areas that have differing source characteristics, especially differentiating between land and marine sectors. This geographical sector map for Weybourne is shown in Figure 17.

The NAME output represents the 10,000 inert tracer particles released during each 3 hourly period and where they are likely to have travelled on their way to the site and the output is an integration of the number of particles per grid cell over the 10 day period and represents a probability that the air passed over that area near to the ground, similar to the emission sensitivities in FLEXPART (Hirdman et al., 2010b). The particle distribution output from NAME is used to extract the information about how many particles have passed over each sector during the 10 days of travel for each 3 hour period by counting the total number of particles that pass over each grid box in each geographical area. The distribution of air

particles passing over each sector can be shown as a percentage of the total number of particles in the domain from each 3 hour release period (Figure 18). The first plot in Figure 18 shows 4 years of distributions (2006-2009) and the second shows a smaller time period where the frequency and size of the variations of influence are clearer and can be used to visually isolate events and changes in air mass sector influences. Nearly all trajectories pass over multiple geographical areas. The monthly averaged regional distribution of air masses arriving at Weybourne illustrates the seasonality of the Arctic and North Sea air masses that are more frequent in spring as shown in Figure 19a. Figure 19b shows a similar plot for the Cape Verde observatory, showing how this method can pick up the seasonality of the regions that influence a site (the synoptic climatology), with much greater Saharan influence in winter (Carpenter et al., 2011). These are similar to the seasonal distribution of trajectory types in the studies shown in Figure 6 and Figure 12.

To account for the fact that each trajectory in its ten day passage to the station will have passed over more than one geographical area, various permutations of combinations of these regions have been fitted into convenient classifications. Seven subjective air-mass classifications were defined for Weybourne and are shown in Figure 20. They are denoted as Arctic only, Arctic and Europe, Atlantic, European, Local (UK and North Sea), Scandinavian and Greenland, America (no European).

The time integrated particle concentration or dosage (gs/m^3) in each sector shows a wide variation, with sporadic peaks. Various tests were carried out to test for the most statistically robust method for assigning a threshold for the amount of particles in a sector that would make that sector contribute significant influence on the air mass arriving at the station. The final chosen method was to select a threshold of 10 % of the maximum dosage for each sector. Obviously, being closer to the station, there were more particles in the UK sector than the American so that made the concentration threshold for assigning the American sector as significant more sensitive.

Thresholds for each sector were derived for assigning each trajectory the sectors it passed through. These thresholds were chosen by plotting dosage (in gs/m^3) against % time in that region and finding the % thresholds for a given sector a dosage of 0.001 gs/m^3 (for the UK and North Sea the threshold dosage was chosen to be higher at 0.002 and for the American

sector a lower threshold dosage of 0.0003 was chosen). These thresholds came to: Arctic (10%), Scandinavia (10%), Europe (10 %), Atlantic (13%), Greenland and Iceland (5%), America (3 %), UK (24 %) and North Sea (17%). This form of classification flags up the more distant sectors like America as being labelled as important when the same % in other sectors would not have been significant. The local UK sector was only flagged as particularly important when the trajectory spent over 24 % of the time over the UK.

Permutations of the seven sectors were combined to derive a trajectory classification (as shown in Figure 20) for the entire time series, in this case as three hourly time series. The final result is a time series with flags assigned to trajectory types numbered 1 to 7.

The station composition data time series were averaged into 3 hourly intervals so as to be comparable with the 3 hourly air mass trajectory type time series from the NAME-based classification. Figure 21 shows how the chemistry can vary when the regions that the air masses pass over before reaching the station change, showing how ozone and NO_2 increase during the period of a high European and UK influence (e.g. 25-28th July and 30th July-1st August). The average values for each of the air mass trajectory type was calculated as well as the standard deviation and mean for winter (DJF), summer (JJA), spring (MAM) and autumn (SON) data for January 2006-September 2009. Figure 22 shows the three year seasonally averaged composition in each air mass type. SO_2 was highest in the local sector and lowest in Arctic air masses and O_3 was highest in European, local and Scandinavian air masses in summer but highest in American and Arctic air masses in winter.

The methodology of using NAME to classify composition time series can be analysed on various timescales other than as shown in Figure 22 for the 4 year average of each air mass type. It would be of interest to calculate yearly averages so as to follow annual trends in composition for each air mass type. Alternatively, this technique is useful for focussing in on particular periods of interest to understand sudden changes in composition or to isolate the origin of pollution spikes. This described methodology illustrates the use of dispersion models for building up a database of station footprints which can be used in a number of ways to extract the regional influences on average levels of a variety of atmospheric species at any timescale from hours to years. Trends in meteorological and synoptic scale influences on a site of interest can also be tracked with this method, to track seasonal variations in air mass origin and assess long term variations. Comparing the use of the NAME dispersion model to other techniques described in this review, it is most similar to the regional

assignment of trajectories as passing through particular geographical sectors discussed in section 4.1 and Figure 8 (especially the FLEXPART dispersion model in Figure 8b). However, owing to the spread of influence in the NAME dispersion model from turbulence, the distribution over multiple regions is easier to track than with a single line trajectory. Using the NAME model in this way is very similar to the FLEXPART regional assignment and relative residence times calculated by (Halse et al., 2011) and particularly (Hirdman et al., 2010a) (Figure 12) and (Hirdman et al., 2010b).

6. Conclusions

Interpreting air mass history and the role of transport remains an important tool for interpreting observed atmospheric composition which, owing to meteorology, is influenced by a variety of local and long range transport processes. Local wind direction and speed have been mostly replaced by a number of computational techniques (trajectory and dispersion models) that follow the far field transport influence to be assessed and quantified. The increase in the accuracy of trajectory and dispersion models is related to improvements in the resolution of available meteorological fields, which leads to a better resolution of the atmospheric physics and the ability to interpret the movement mixing and transport of atmospheric constituents.

This review paper has detailed the evolution of methodologies to interpret atmospheric composition measurements according to air mass history. *In situ* wind measurements or the use of agglomerated meteorological fields in trajectory or dispersion models can be used to assess the influence of changing air masses history on composition.

The historical use of local wind direction or daily meteorological patterns is still of use for obtaining a short term picture of the variation of air masses but is not as accurate to detect long range transport as seen in trajectory models and dispersion models. The pros and cons of trajectory and dispersion models have been explained that is they are. less computationally intense and lower output volume for trajectories, added accuracy of simulating the atmosphere by the inclusion of turbulence and the various uses of these for the interpretation of ground based station or aircraft data have been illustrated.

The technique of linking trajectories and dispersion models with sector analysis describes the pathway and links the residence time over zones of interest with composition levels with

time. Cluster analysis is an objective method for clustering trajectories according to similar characteristics and then composition measurements can be related to the various cluster types.

A step by step example of the procedures used in this form of analysis is detailed by demonstrating a new methodology (using the NAME atmospheric dispersion model) for assessing the effect of air mass origin on the atmospheric composition at a long term measurement station. The case study for four years for the Weybourne Atmospheric Observatory illustrates how any dataset can be exploited and how this is an especially effective technique for stations with a variety of composition influences (e.g. a combination of marine and continental influences or clean and polluted sectors). The NAME model has been used to calculate a multi-year time series of trajectory types and sectors of influence. This methodology can provide a long term view of the type of air masses and conditions that affect each site and the inter-annual and intra-annual variability of the air mass types reaching the site as well as the composition of those air masses.

Understanding and being unable to unpick the history of air-masses in the atmosphere remains an important tool for assessing influence of emissions and change in the atmosphere.

Acknowledgements

We would like to thank the authors of the work referenced herein for allowing reproduction of their various figures. We would like to thank the UK Met Office for supplying the Unified Model Meteorological data and the use of the NAME model. We would also like to thank University of Leicester's High Performance Computing services for supplying the necessary computing power for running, plotting and storing the model output. We would like to thank Roland Leigh at Leicester University for developing the code to plot and interpret the NAME model output on the global grid scale. We would like to thank Brian Bandy, William Sturges and Stuart Penkett and many past and present scientists at UEA for providing the Weybourne dataset and the BADC for hosting it, as well as NCAS and UEA for funding the observatory. Finally we would like to thank the National Centre for Atmospheric Science (NCAS) and NERC for funding.

References

T. Aalto *et al.*, Tropospheric carbon dioxide concentrations at a northern boreal site in Finland: basic variations and source areas, *Tellus Series B-Chemical and Physical Meteorology* **54**(2002), pp. 110-126.

S.S. Abdalmogith and R.M. Harrison, The use of trajectory cluster analysis to examine the long-range transport of secondary inorganic aerosol in the UK, *Atmospheric Environment* **39**(2005), pp. 6686-6695.

L.Y. Alleman, L. Lamaison, E. Perdrix, A. Robache and J.C. Galloo, PM10 metal concentrations and source identification using positive matrix factorization and wind sectoring in a French industrial zone, *Atmospheric Research* **96**(2010), pp. 612-625.

A.P. Altshuller, Relationships between direction of wind flow and ozone inglow concentrations at rural locations outside of St Louis, MO *Atmospheric Environment* **20**(1986), pp. 2175-2184.

F. Apadula *et al.*, Localization of source and sink regions of carbon dioxide through the method of the synoptic air trajectory statistics, *Atmospheric Environment* **37**(2003), pp. 3757-3770.

L.L. Ashbaugh, A statistical trajectory technique for determining air-pollution source regions, *Journal of the Air Pollution Control Association* **33**(1983), pp. 1096-1098.

L.L. Ashbaugh, W.C. Malm and W.Z. Sadeh, A residence time probability analysis of sulphur concentrations at Grand-Canyon National Park, *Atmospheric Environment* **19**(1985), pp. 1263-1270.

J. Baker, A cluster analysis of long range air transport pathways and associated pollutant concentrations within the UK, *Atmospheric Environment* **44**(2010).

F. Baur, P. Hess and H. Nagel, Kalendar der Grosswetterlagen Europas 1881–1939. Bad Homburg (DWD), (1944).

S. Beaver and A. Palazoglu, Cluster analysis of hourly wind measurements to reveal synoptic regimes affecting air quality, *Journal of Applied Meteorology and Climatology* **45**(2006), pp. 1710-1726.

B.A. Begum, E. Kim, C.H. Jeong, D.W. Lee and P.K. Hopke, Evaluation of the potential source contribution function using the 2002 Quebec forest fire episode, *Atmospheric Environment* **39**(2005), pp. 3719-3724.

C. Bhugwant, M. Bessafi, E. Riviere and J. Leveau, Diurnal and seasonal variation of carbonaceous aerosols at a remote MBL site of La Reunion island, *Atmospheric Research* **57**(2001), pp. 105-121.

S.R. Biegalski and P.K. Hopke, Total potential source contribution function analysis of trace elements determined in aerosol samples collected near Lake Huron, *Environ. Sci. Technol.* **38**(2004), pp. 4276-4284.

N.J. Blake *et al.*, Estimates of Atmospheric Hydroxyl Radical Concentrations From the Observed Decay of Many Reactive Hydrocarbons in Well-Defined Urban Plumes, *J. Geophys. Res.* **98**(1993), pp. 2851-2864.

R. Borge, J. Lumbreras, S. Vardoulakis, P. Kassomenos and E. Rodriguez, Analysis of long-range transport influences on urban PM10 using two-stage atmospheric trajectory clusters, *Atmospheric Environment* **41**(2007), pp. 4434-4450.

E. Brankov, S.T. Rao and P.S. Porter, A trajectory-clustering-correlation methodology for examining the long-range transport of air pollutants, *Atmospheric Environment* **32**(1998), pp. 1525-1534.

C.A.M. Brenninkmeijer *et al.*, Civil Aircraft for the regular investigation of the atmosphere based on an instrumented container: The new CARIBIC system, *Atmos. Chem. Phys.* **7**(2007), pp. 4953-4976.

J.D. Burley and J.D. Ray, Surface ozone in Yosemite National Park, *Atmospheric Environment* **41**(2007), pp. 6048-6062.

J.P. Cammas *et al.*, Injection in the lower stratosphere of biomass fire emissions followed by long-range transport: a MOZAIC case study, *Atmospheric Chemistry and Physics* **9**(2009), pp. 5829-5846.

J.N. Cape, J. Methven and L.E. Hudson, The use of trajectory cluster analysis to interpret trace gas measurements at Mace Head, Ireland, *Atmospheric Environment* **34**(2000), pp. 3651-3663.

L.M. Cardenas *et al.*, Correlations between CO, NOy, O-3, and non-methane hydrocarbons and their relationships with meteorology during winter 1993 on the North Norfolk Coast, UK, *Atmospheric Environment* **32**(1998), pp. 3339-3351.

L.J. Carpenter *et al.*, Seasonal characteristics of tropical marine boundary layer air measured at the Cape Verde Atmospheric Observatory, *Journal of Atmospheric Chemistry (submitted)*(2011).

A. Carvalho *et al.*, High ozone levels in the northeast of Portugal: Analysis and characterization, *Atmospheric Environment* **44**(2010), pp. 1020-1031.

T.W. Chan and M. Mozurkewich, Application of absolute principal component analysis to size distribution data: identification of particle origins, *Atmospheric Chemistry and Physics* **7**(2007), pp. 887-897.

M.-D. Cheng and C.-J. Lin, Receptor modeling for smoke of 1998 biomass burning in Central America, *J. Geophys. Res.* **106**(2001), pp. 22871-22886.

E. Choi, J.B. Heo and S.M. Yi, Apportioning and Locating Nonmethane Hydrocarbon Sources to a Background Site in Korea, *Environ. Sci. Technol.* **44**(2010), pp. 5849-5854.

D.D. Cohen, J. Crawford, E. Stelcer and V.T. Bac, Long range transport of fine particle windblown soils and coal fired power station emissions into Hanoi between 2001 to 2008, *Atmospheric Environment* **44**(2010), pp. 3761-3769.

M. Cohen *et al.*, Modeling the atmospheric transport and deposition of mercury to the Great Lakes, *Environmental Research* **95**(2004), pp. 247-265.

A. Colette, G. Ancellet and F. Borchi, Impact of vertical transport processes on the tropospheric ozone layering above Europe. Part I: Study of air mass origin using multivariate analysis, clustering and trajectories, *Atmos. Environ.* **39**(2005), pp. 5409-5422.

G. Cosemans, J. Kretzschmar and C. Mensink, Pollutant roses for daily averaged ambient air pollutant concentrations, *Atmospheric Environment* **42**(2008), pp. 6982-6991.

H.L. Crutcher, R.C. Rhodes, M.E. Graves, B. Fairbairn and A.C. Nelson, Application of cluster-analysis to aerometric data, *Journal of the Air Pollution Control Association* **36**(1986), pp. 1116-1122.

M.J.P. Cullen, The Unified Forecast/Climate Model, Meteorol. Mag., 1449, 81–94., (1993).

E.F. Danielsen, and R. Bleck, Research in four-dimensional diagnosis of cyclonic storm cloud system, *AFCRL Rep. 67-0617, Bedford, Mass., 1967. (Available as NTIS AD 670 847, from Natl. Tech. Inf. Serv., Springfield, Va.)*(1967).

L.S. Darby, Cluster analysis of surface winds in Houston, Texas, and the impact of wind patterns on ozone, *Journal of Applied Meteorology* **44**(2005), pp. 1788-1806.

J.M. Davis, B.K. Eder, D. Nychka and Q. Yang, Modeling the effects of meteorology on ozone in Houston using cluster analysis and generalized additive models, *Atmospheric Environment* **32**(1998), pp. 2505-2520.

R.E. Davis *et al.*, A comparison of trajectory and air mass approaches to examine ozone variability, *Atmospheric Environment* **44**(2010), pp. 64-74.

B. de Foy *et al.*, Basin-scale wind transport during the MILAGRO field campaign and comparison to climatology using cluster analysis, *Atmos. Chem. Phys.* **8**(2008), pp. 1209-1224.

B. de Foy, M. Zavala, N. Bei and L.T. Molina, Evaluation of WRF mesoscale simulations and particle trajectory analysis for the MILAGRO field campaign, *Atmospheric Chemistry and Physics* **9**(2009), pp. 4419-4438.

A.W. Delcloo and H. De Backer, Five day 3D back trajectory clusters and trends analysis of the Uccle ozone sounding time series in the lower troposphere (1969-2001), *Atmospheric Environment* **42**(2008), pp. 4419-4432.

M. Demuzere and N.P.M. van Lipzig, A new method to estimate air-quality levels using a synoptic-regression approach. Part I: Present-day O-3 and PM10 analysis, *Atmospheric Environment* **44**(2010), pp. 1341-1355.

R.G. Derwent, C.S. Witham, S.R. Utembe, M.E. Jenkin and N.R. Passant, Ozone in Central England: the impact of 20 years of precursor emission controls in Europe, *Environmental Science & Policy* **13**(2010), pp. 195-204.

B.G. Doddridge, P.A. Dirmeyer, J.T. Merrill, S.J. Oltmans and R.R. Dickerson, Interannual variability over the Eastern North-Atlantic Ocean – Chemical and meteorological evidence for tropical influence on regional-scale transport in the extratropics, *Journal of Geophysical Research-Atmospheres* **99**(1994), pp. 22923-22935.

G. Dogan, G. Gullu and G. Tuncel, Sources and source regions effecting the aerosol composition of the Eastern Mediterranean, *Microchemical Journal* **88**(2008), pp. 142-149.

S.R. Dorling and T.D. Davies, Extending cluster-analysis- Synoptic meteorology links to characterize chemical climates at 6 northwest European monitoring stations, *Atmospheric Environment* **29**(1995), pp. 145-167.

S.R. Dorling, T.D. Davies and C.E. Pierce, Cluster-analysis - A technique for estimating the synoptic meteorological controls on air and precipitation chemistry - Method and applications, *Atmospheric Environment Part a-General Topics* **26**(1992), pp. 2575-2581.

R.R. Draxler and G.D. Rolph, HYSPLIT (HYbrid Single-Particle Lagrangian Integrated Trajectory) Model *Model access via NOAA ARL READY Website (http://ready.arl.noaa.gov/HYSPLIT.php). NOAA Air Resources Laboratory, Silver Spring, MD*(2003).

R.R. Draxler and A.D. Taylor, Horizontal parameters for long-range transport modelling, *Journal of Applied Meteorology* **21**(1982), pp. 367-372.

J.G. Droppo and B.A. Napier, Wind direction bias in generating wind roses and conducting sector-based air dispersion modeling, *Journal of the Air & Waste Management Association* **58**(2008), pp. 913-918.

S.Y. Du and L.A. Rodenburg, Source identification of atmospheric PCBs in Philadelphia/Camden using positive matrix factorization followed by the potential source contribution function, *Atmospheric Environment* **41**(2007), pp. 8596-8608.

C. Dueñas *et al.*, Air mass origin and its influence on radionuclide activities (^{7}Be and ^{210}Pb) in aerosol particles at a coastal site in the western Mediterranean, *Atmospheric Research* **In Press, Accepted Manuscript**(2011).

A. Dvorska, G. Lammel and I. Holoubek, Recent trends of persistent organic pollutants in air in central Europe - Air monitoring in combination with air mass trajectory statistics as a tool to study the effectivity of regional chemical policy, *Atmospheric Environment* **43**(2009), pp. 1280-1287.

B.K. Eder, J.M. Davis and P. Bloomfield, An automated classification scheme designed to better elucidate the dependence of ozone on meteorology, *Journal of Applied Meteorology* **33**(1994), pp. 1182-1199.

K. Eneroth, K. Holmen, T. Berg, N. Schmidbauer and S. Solberg, Springtime depletion of tropospheric ozone, gaseous elemental mercury and non-methane hydrocarbons in the European Arctic, and its relation to atmospheric transport, *Atmospheric Environment* **41**(2007), pp. 8511-8526.

K. Eneroth, E. Kjellstrom and K. Holmen, A trajectory climatology for Svalbard; investigating how atmospheric flow patterns influence observed tracer concentrations, *Physics and Chemistry of the Earth* **28**(2003), pp. 1191-1203.

M. Escudero *et al.*, Source apportionment for African dust outbreaks over the Western Mediterranean using the HYSPLIT model, *Atmospheric Research* **99**(2011), pp. 518-527.

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

I. Fenneteaux *et al.*, Influence of continental sources on oceanic air composition at the eastern edge of the North Atlantic Ocean, TOR 1992-1995, *Journal of Atmospheric Chemistry* **32**(1999), pp. 233-280.

A.M. Fiore *et al.*, Multimodel estimates of intercontinental source-receptor relationships for ozone pollution, *Journal of Geophysical Research-Atmospheres* **114**(2009).

T.K. Flesch, J.D. Wilson and E. Yee, Backward-time Lagrangian stochastic dispersion models and their application to estimate gaseous emissions, *Journal of Applied Meteorology* **34**(1995), pp. 1320-1332.

A. Font, J.A. MorguÃ and X. RodÃ³, Assessing the regional surface influence through Backward Lagrangian Dispersion Models for aircraft CO2 vertical profiles observations in NE Spain, *Atmos. Chem. Phys.* **11**(2011), pp. 1659-1670.

C. Forster *et al.*, Transport of boreal forest fire emissions from Canada to Europe, *Journal of Geophysical Research-Atmospheres* **106**(2001), pp. 22887-22906.

M. Gardner and S. Dorling, Artificial neural network-derived trends in daily maximum surface ozone concentrations, *Journal of the Air & Waste Management Association* **51**(2001), pp. 1202-1210.

M.W. Gardner and S.R. Dorling, Artificial neural networks (the multilayer perceptron) - A review of applications in the atmospheric sciences, *Atmospheric Environment* **32**(1998), pp. 2627-2636.

M.W. Gardner and S.R. Dorling, Neural network modelling and prediction of hourly NOx and NO2 concentrations in urban air in London, *Atmospheric Environment* **33**(1999), pp. 709-719.

M.W. Gardner and S.R. Dorling, Meteorologically adjusted trends in UK daily maximum surface ozone concentrations, *Atmospheric Environment* **34**(2000), pp. 171-176.

K.A. Gebhart, B.A. Schichtel and M.G. Barna, Directional biases in back trajectories caused by model and input data, *Journal of the Air & Waste Management Association* **55**(2005), pp. 1649-1662.

K.A. Gebhart *et al.*, Back-trajectory-based source apportionment of airborne sulfur and nitrogen concentrations at Rocky Mountain National Park, Colorado, USA, *Atmospheric Environment* **45**(2011), pp. 621-633.

F.-W. Gerstengabe, P. Werner and U. Rüge, Katalog der Grosswetterlagen Europas 1881–1998 nach P. Hess und H. Brezowsky. 5. Aufl. – Potsdam-Inst. F. Klimafolgenforschung, Potsdam, Germany, 138 pp, (1999).

R.C. Gilliam, C. Hogrefe and S.T. Rao, New methods for evaluating meteorological models used in air quality applications, *Atmospheric Environment* **40**(2006).

M.M. Gogoi, K.K. Moorthy, S.S. Babu and P.K. Bhuyan, Climatology of columnar aerosol properties and the influence of synoptic conditions: First-time results from the northeastern region of India, *Journal of Geophysical Research-Atmospheres* **114**(2009).

G.L. Gregory *et al.*, Chemical signatures of aged Pacific marine air: Mixed layer and free troposphere as measured during PEM-West A, *Journal of Geophysical Research-Atmospheres* **101**(1996), pp. 1727-1742.

V. Gros *et al.*, Tracing the origin and ages of interlaced atmospheric pollution events over the tropical Atlantic Ocean with in situ measurements, satellites, trajectories, emission inventories, and global models, *Journal of Geophysical Research-Atmospheres* **109**(2004).

G. Güllü, G. Dogan and G. Tuncel, Atmospheric trace element and major ion concentrations over the eastern Mediterranean Sea: Identification of anthropogenic source regions, *Atmospheric Environment* **39**(2005), pp. 6376-6387.

J.C. Hains *et al.*, Origins of chemical pollution derived from Mid-Atlantic aircraft profiles using a clustering technique, *Atmospheric Environment* **42**(2008), pp. 1727-1741.

A.K. Halse *et al.*, Spatial variability of POPs in European background air, *Atmos. Chem. Phys.* **11**(2011), pp. 1549-1564.

Y.J. Han, T.M. Holsen, P.K. Hopke and S.M. Yi, Comparison between back-trajectory based modeling and Lagrangian backward dispersion Modeling for locating sources of reactive gaseous mercury, *Environ. Sci. Technol.* **39**(2005), pp. 1715-1723.

R.M. Harrison *et al.*, Influence of airmass back trajectory upon nitrogen compound composition, *Atmospheric Environment* **34**(2000), pp. 1519-1527.

R.M. Harrison *et al.*, Measurement and modelling of air pollution and atmospheric chemistry in the UK West Midlands conurbation: Overview of the PUMA Consortium project, *Science of the Total Environment* **360**(2006), pp. 5-25.

M. Hart, R. De Dear and R. Hyde, A synoptic climatology of tropospheric ozone episodes in Sydney, Australia, *International Journal of Climatology* **26**(2006), pp. 1635-1649.

C.G. Helmis *et al.*, Estimation of transboundary air pollution on the basis of synoptic-scale weather types, *International Journal of Climatology* **23**(2003), pp. 405-416.

A.J. Hewitt, Investigating land-air carbon fluxes using a Lagrangian model and satellite retrieved carbon dioxide, Thesis, University of Leicester, (2010).

D. Hirdman *et al.*, Transport of mercury in the Arctic atmosphere: Evidence for a springtime net sink and summer-time source, *Geophysical Research Letters* **36**(2009).

D. Hirdman *et al.*, Long-term trends of black carbon and sulphate aerosol in the Arctic: changes in atmospheric transport and source region emissions, *Atmospheric Chemistry and Physics* **10**(2010a), pp. 9351-9368.

D. Hirdman *et al.*, Source identification of short-lived air pollutants in the Arctic using statistical analysis of measurement data and particle dispersion model output, *Atmospheric Chemistry and Physics* **10**(2010b), pp. 669-693.

P.K. Hopke and H.L. Allan, Chapter 1 Theory and Application of Atmospheric Source Apportionment. *Developments in Environmental Sciences*, Elsevier (2009), pp. 1-33.

HTAP, Task Force on Hemispheric Transport of Air Pollution, Hemispheric transport of air pollution 2007 interim report, edited by T. J. Keating and A. Zuber, Air Pollut. Stud. 16, U.N. Econ. Comm. for Europe, New York, (2007).

L. Huang, S.L. Gong, S. Sharma, D. Lavoue and C.Q. Jia, A trajectory analysis of atmospheric transport of black carbon aerosols to Canadian high Arctic in winter and spring (1990-2005), *Atmospheric Chemistry and Physics* **10**(2010), pp. 5065-5073.

D.R. Jackson, J. Methven and V.D. Pope, Transport in the low-latitude tropopause zone diagnosed using particle trajectories, *Journal of the Atmospheric Sciences* **58**(2001), pp. 173-192.

A. Jones, D. Thomson, M. Hort and B. Devenish, The UK Met Office's next-generation atmospheric dispersion model, NAME III, *Air Pollution Modeling and Its Applications XVII* **17**(2007), pp. 580-589.

O. Jorba, C. Perez, F. Rocadenbosch and J.M. Baldasano, Cluster analysis of 4-day back trajectories arriving in the Barcelona area, Spain, from 1997 to 2002, *Journal of Applied Meteorology* **43**(2004), pp. 887-901.

C. Junker *et al.*, Measurement and analysis of aerosol and black carbon in the southwestern United States and Panama and their dependence on air mass origin, *Journal of Geophysical Research-Atmospheres* **109**(2004).

C. Junker, J.L. Wang and C.T. Lee, Evaluation of the effect of long-range transport of air pollutants on coastal atmospheric monitoring sites in and around Taiwan, *Atmospheric Environment* **43**(2009), pp. 3374-3384.

J.D. Kahl, A cautinary note on the use of air trajectories in interpreting atmospheric chemistryu measurements, *Atmospheric Environment Part a-General Topics* **27**(1993), pp. 3037-3038.

A. Kaiser *et al.*, Transport of nitrogen oxides, carbon monoxide and ozone to the Alpine Global Atmosphere Watch stations Jungfraujoch (Switzerland), Zugspitze and Hohenpeissenberg (Germany), Sonnblick (Austria) and Mt. Krvavec (Slovenia), *Atmospheric Environment* **41**(2007), pp. 9273-9287.

L.S. Kalkstein, G.R. Tan and J.A. Skindlov, An evaluation of 3 clustering procedures for use in synoptic climatological classification, *Journal of Climate and Applied Meteorology* **26**(1987), pp. 717-730.

C.M. Kang, B.W. Kang and H.S. Lee, Source identification and trends in concentrations of gaseous and fine particulate principal species in Seoul, South Korea, *Journal of the Air & Waste Management Association* **56**(2006), pp. 911-921.

F. Karaca, I. Anil and O. Alagha, Long-range potential source contributions of episodic aerosol events to PM10 profile of a megacity, *Atmospheric Environment* **43**(2009), pp. 5713-5722.

F. Karaca and F. Camci, Distant source contributions to PM10 profile evaluated by SOM based cluster analysis of air mass trajectory sets, *Atmospheric Environment* **44**(2010), pp. 892-899.

P. Kassomenos *et al.*, Comparison of statistical clustering techniques for the classification of modelled atmospheric trajectories, *Theoretical and Applied Climatology* **102**(2010), pp. 1-12.

J. Kasumba, P.K. Hopke, D.C. Chalupa and M.J. Utell, Comparison of sources of submicron particle number concentrations measured at two sites in Rochester, NY, *Science of the Total Environment* **407**(2009), pp. 5071-5084.

M. Kocak, N. Mihalopoulos and N. Kubilay, Origin and source regions of PM10 in the Eastern Mediterranean atmosphere, *Atmospheric Research* **92**(2009), pp. 464-474.

T. Kuhn, R. Damoah, A. Bacak and J.J. Sloan, Characterising aerosol transport into the Canadian High Arctic using aerosol mass spectrometry and Lagrangian modelling, *Atmos. Chem. Phys.* **10**(2010), pp. 10489-10502.

J. Kukkonen *et al.*, Extensive evaluation of neural network models for the prediction of NO2 and PM10 concentrations, compared with a deterministic modelling system and measurements in central Helsinki, *Atmospheric Environment* **37**(2003), pp. 4539-4550.

P. Laj *et al.*, Measuring Atmospheric Composition Change, *Atmospheric Environment*(2009).

H.H. Lamb, British Isles Weather types and a register of daily sequence of circulation patterns, 1861-1971, *Geophysical Memoir 116, HMSO, London, 85pp*(1972).

K.S. Law *et al.*, Air mass origins influencing TTL chemical composition over West Africa during 2006 summer monsoon, *Atmos. Chem. Phys.* **10**(2010), pp. 10753-10770.

M.J. Lawler *et al.*, Pollution-enhanced reactive chlorine chemistry in the eastern tropical Atlantic boundary layer, *Geophysical Research Letters* **36**(2009).

J.D. Lee *et al.*, Year-round measurements of nitrogen oxides and ozone in the tropical North Atlantic marine boundary layer, *Journal of Geophysical Research-Atmospheres* **114**(2009).

S. Lee and L. Ashbaugh, Comparison of multi-receptor and single-receptor trajectory source apportionment (TSA) methods using artificial sources, *Atmospheric Environment* **41**(2007a), pp. 1119-1127.

S. Lee and L. Ashbaugh, Comparison of the MURA and an improved single-receptor (SIRA) trajectory source apportionment (TSA) method using artificial sources, *Atmospheric Environment* **41**(2007b), pp. 4466-4481.

S. Lee and L. Ashbaugh, The impact of trajectory starting heights on the MURA trajectory source apportionment (TSA) method, *Atmospheric Environment* **41**(2007c), pp. 7022-7036.

M. Leuchner and B. Rappengluck, VOC source-receptor relationships in Houston during TexAQS-II, *Atmospheric Environment* **44**(2010), pp. 4056-4067.

A.C. Lewis *et al.*, Chemical composition observed over the mid-Atlantic and the detection of pollution signatures far from source regions, *Journal of Geophysical Research-Atmospheres* **112**(2007).

J.M. Lim, J.H. Lee, J.H. Moon, Y.S. Chung and K.H. Kim, Source apportionment of PM10 at a small industrial area using Positive Matrix Factorization, *Atmospheric Research* **95**(2010), pp. 88-100.

A. Lupu and W. Maenhaut, Application and comparison of two statistical trajectory techniques for identification of source regions of atmospheric aerosol species, *Atmospheric Environment* **36**(2002), pp. 5607-5618.

A.R. Mace *et al.*, Cluster Analysis of the Origin and Transport of Atmospheric Chemical Species, *Atmospheric Chemistry and Physics Discussions (submitted)*(2011).
L. Makra *et al.*, An objective classification system of air mass types for Szeged, Hungary, with special interest in air pollution levels, *Meteorology and Atmospheric Physics* **92**(2006), pp. 115-137.
L. Makra, J. Mika, A. Bartzokas, R. Beczi and Z. Sumeghy, Comparison of objective air-mass types and the Peczely weather types and their ability to classify levels of air pollutants in Szeged, Hungary, *International Journal of Environment and Pollution* **36**(2009), pp. 81-98.
L. Makra *et al.*, Airborne pollen in three European cities: Detection of atmospheric circulation pathways by applying three-dimensional clustering of backward trajectories, *J. Geophys. Res.* **115**(2010), p. D24220.
A.L. Malcolm, R.G. Derwent and R.H. Maryon, Modelling the long-range transport of secondary PM10 to the UK, *Atmospheric Environment* **34**(2000), pp. 881-894.
A.L. Malcolm and A.J. Manning, Testing the skill of a Lagrangian dispersion model at estimating primary and secondary particulates, *Atmospheric Environment* **35**(2001), pp. 1677-1685.
A.J. Manning, S. O'Doherty, A.R. Jones, P.G. Simmonds and R.G. Derwent, Estimating UK methane and nitrous oxide emissions from 1990 to 2007 using an inversion modeling approach, *J. Geophys. Res.* **116**(2011), p. D02305.
A.J. Manning, D.B. Ryall, R.G. Derwent, P.G. Simmonds and S. O'Doherty, Estimating European emissions of ozone-depleting and greenhouse gases using observations and a modeling back-attribution technique, *Journal of Geophysical Research-Atmospheres* **108**(2003).
M.T. Markou and P. Kassomenos, Cluster analysis of five years of back trajectories arriving in Athens, Greece, *Atmospheric Research* **98**(2010), pp. 438-457.
R.H. Maryon, Smith, F. B., Conway, B. J. and Goddard, D. M. , The U.K. Nuclear accident model, *Progress in Nuclear Energy 26(2), 85Ð104.*(1991).
C.L. McConnell *et al.*, Seasonal variations of the physical and optical characteristics of Saharan dust: Results from the Dust Outflow and Deposition to the Ocean (DODO) experiment, *Journal of Geophysical Research-Atmospheres* **113**(2008).
J.T. Merrill, Isentropic air-flow probability analysis, *Journal of Geophysical Research-Atmospheres* **99**(1994), pp. 25881-25889.
J.T. Merrill, Trajectory results and interpretation for PEM-West A, *Journal of Geophysical Research-Atmospheres* **101**(1996), pp. 1679-1690.
J.T. Merrill and J.L. Moody, Synoptic meterology and transport during the North Atlantic Regional Experiment (NARE) intesnive: Overview, *J. Geophys. Res.* **101**(1996), pp. 28903-28921.
J. Methven *et al.*, Estimating photochemically produced ozone throughout a domain using flight data and a Lagrangian model, *Journal of Geophysical Research-Atmospheres* **108**(2003).
J. Methven *et al.*, Establishing Lagrangian connections between observations within air masses crossing the Atlantic during the International Consortium for Atmospheric Research on Transport and Transformation experiment, *J. Geophys. Res.* **111**(2006), p. D23S62.
J. Methven, M. Evans, P. Simmonds and G. Spain, Estimating relationships between air mass origin and chemical composition, *Journal of Geophysical Research-Atmospheres* **106**(2001), pp. 5005-5019.
R.G. Miller, Jr. , Simultaneous Statistical Inference, *2nd ed., 299 pp., Springer, New York.*(1981).
J.L. Moody, J.A. Galusky and J.N. Galloway, The use of atmospheric transport pattern recognition techniques in understanding variation in precipitation chemistry, *Atmospheric Deposition (Proceedings of the Baltimore Symposium, May 1989). IAHS Publ. No. 179*(1989).
L.A. Moy, R.R. Dickerson and W.F. Ryan, Relationship between back trajectories and tropospheric trace gas concentrations in rural Virginia, *Atmospheric Environment* **28**(1994), pp. 2789-2800.
K. Müller *et al.*, Particle characterization at the Cape Verde atmospheric observatory during the 2007 RHaMBLe intensive, *Atmospheric Chemistry and Physics* **10**(2010), pp. 2709-2721.

N.T.K. Oanh, P. Chutimon, W. Ekbordin and W. Supat, Meteorological pattern classification and application for forecasting air pollution episode potential in a mountain-valley area, *Atmospheric Environment* **39**(2005), pp. 1211-1225.

C. Occhipinti, V.P. Aneja, W. Showers and D. Niyogi, Back-trajectory analysis and source-receptor relationships: Particulate matter and nitrogen isotopic composition in rainwater, *Journal of the Air & Waste Management Association* **58**(2008), pp. 1215-1222.

G.P. Ohare and R. Wilby, A review of ozone pollution in the United-Kingdom and Ireland with an analysis using Lamb weather types, *Geographical Journal* **161**(1995), pp. 1-20.

J. Olivier, Bouwman, A., Maas, C. v. d., Berdowski, J., Veldt, C., Bloos, J., and A. Visschedijk, Zandveld, P., Haverlag, J., Description of EDGAR Version 2.0: A set of global emission inventories of greenhouse gases and ozone-depleting substances for all anthropogenic and most natural sources on a per country basis and on 1 degree x 1 degree grid. , *Natl. Inst. of Public Health and the Environ., Bilthoven, Netherlands.*(1996).

J.D. Paris *et al.*, Source-receptor relationships for airborne measurements of CO2, CO and O-3 above Siberia: a cluster-based approach, *Atmospheric Chemistry and Physics* **10**(2010), pp. 1671-1687.

N.J. Pekney, C.I. Davidson, L.M. Zhou and P.K. Hopke, Application of PSCF and CPF to PMF-modeled sources of PM2.5 in Pittsburgh, *Aerosol Science and Technology* **40**(2006), pp. 952-961.

S.A. Penkett *et al.*, Evidence for large average concentrations of the nitrate radical (NO_3) in Western Europe from the HANSA hydrocarbon database, *Atmospheric Environment* **41**(2007), pp. 3465-3478.

S. Petterssen, Weather Analysis and Forecasting, (1940), pp. pp. 221-223.

P. Pochanart, H. Akimoto, Y. Kajii, V.M. Potemkin and T.V. Khodzher, Regional background ozone and carbon monoxide variations in remote Siberia/East Asia, *Journal of Geophysical Research* **108**(2003), p. 18.

P. Pochanart, H. Akimoto, S. Maksyutov and J. Staehelin, Surface ozone at the Swiss Alpine site Arosa: the hemispheric background and the influence of large-scale anthropogenic emissions, *Atmospheric Environment* **35**(2001), pp. 5553-5566.

R.L. Poirot and H.L. Allan, Chapter 2 Use of Trace Metals as Source Fingerprints. *Developments in Environmental Sciences*, Elsevier (2009), pp. 35-60.

R.L. Poirot, P.R. Wishinski, P.K. Hopke and A.V. Polissar, Comparative application of multiple receptor methods to identify aerosol sources in northern Vermont, *Environ. Sci. Technol.* **35**(2001), pp. 4622-4636.

L. Poissant, Potential sources of atmospheric total gaseous mercury in the St. Lawrence River valley, *Atmospheric Environment* **33**(1999), pp. 2537-2547.

A.V. Polissar, P.K. Hopke and J.M. Harris, Source regions for atmospheric aerosol measured at Barrow, Alaska, *Environ. Sci. Technol.* **35**(2001a), pp. 4214-4226.

A.V. Polissar *et al.*, The aerosol at Barrow, Alaska: long-term trends and source locations, *Atmospheric Environment* **33**(1999), pp. 2441-2458.

A.V. Polissar, P.K. Hopke and R.L. Poirot, Atmospheric aerosol over Vermont: Chemical composition and sources, *Environ. Sci. Technol.* **35**(2001b), pp. 4604-4621.

E. Real *et al.*, Lagrangian analysis of low altitude anthropogenic plume processing across the North Atlantic, *Atmospheric Chemistry and Physics* **8**(2008), pp. 7737-7754.

E. Real *et al.*, Processes influencing ozone levels in Alaskan forest fire plumes during long-range transport over the North Atlantic, *J. Geophys. Res.* **112**(2007), p. D10S41.

E. Real *et al.*, Cross-hemispheric transport of central African biomass burning pollutants: implications for downwind ozone production, *Atmospheric Chemistry and Physics* **10**(2010), pp. 3027-3046.

D.R. Reidmiller *et al.*, Interannual variability of long-range transport as seen at the Mt. Bachelor observatory, *Atmospheric Chemistry and Physics* **9**(2009), pp. 557-572.

A. Riccio, G. Giunta and E. Chianese, The application of a trajectory classification procedure to interpret air pollution measurements in the urban area of Naples (Southern Italy), *Science of the Total Environment* **376**(2007), pp. 198-214.

S. Rodriguez *et al.*, Transport of desert dust mixed with North African industrial pollutants in the subtropical Saharan Air Layer, *Atmos. Chem. Phys. Discuss.* **11**(2011), pp. 8841-8892.

A. Rozwadowska, T. Zielinski, T. Petelski and P. Sobolewski, Cluster analysis of the impact of air back-trajectories on aerosol optical properties at Hornsund, Spitsbergen, *Atmospheric Chemistry and Physics* **10**(2010), pp. 877-893.

D.B. Ryall *et al.*, The origin of high particulate concentrations over the United Kingdom, March 2000, *Atmospheric Environment* **36**(2002), pp. 1363-1378.

D.B. Ryall, R.G. Derwent, A.J. Manning, P.G. Simmonds and S. O'Doherty, Estimating source regions of European emissions of trace gases from observations at Mace Head, *Atmospheric Environment* **35**(2001), pp. 2507-2523.

D.B. Ryall and R.H. Maryon, Validation of the UK Met. Office's name model against the ETEX dataset, *Atmospheric Environment* **32**(1998), pp. 4265-4276.

G. Salisbury, P.S. Monks, S. Bauguitte, B.J. Bandy and S.A. Penkett, A seasonal comparison of the ozone photochemistry in clean and polluted air masses at Mace Head, Ireland, *Journal of Atmospheric Chemistry* **41**(2002), pp. 163-187.

P. Salvador *et al.*, Evaluation of aerosol sources at European high altitude background sites with trajectory statistical methods, *Atmospheric Environment* **44**(2010), pp. 2316-2329.

P. Salvador, B. Artinano, X. Querol and A. Alastuey, A combined analysis of backward trajectories and aerosol chemistry to characterise long-range transport episodes of particulate matter: The Madrid air basin, a case study, *Science of the Total Environment* **390**(2008), pp. 495-506.

H. Scheifinger and A. Kaiser, Validation of trajectory statistical methods, *Atmospheric Environment* **41**(2007), pp. 8846-8856.

B.A. Schichtel, K.A. Gebhart, M.G. Barna and W.C. Malm, Association of airmass transport patterns and particulate sulfur concentrations at Big Bend National Park, Texas, *Atmospheric Environment* **40**(2006), pp. 992-1006.

J. Schmale *et al.*, Source identification and airborne chemical characterisation of aerosol pollution from long-range transport over Greenland during POLARCAT summer campaign 2008, *Atmos. Chem. Phys. Discuss.* **11**(2011), pp. 7593-7658.

P. Seibert, Convergence and accuracy of numerical methods for trajectory calculations *Journal of Applied Meteorology* **32**(1993), pp. 558-566.

P. Seibert and A. Frank, Source-receptor matrix calculation with a Lagrangian particle dispersion model in backward mode, *Atmos. Chem. Phys.* **4**(2004), pp. 51-63.

P. Seibert, Kromp-Kolb H., Baltensperger U., Jost D. T., Schwikowski M., Kasper A. and Puxbaum H., Trajectory analysis of aerosol measurements at high alpine sites, *In Transport and Transformation of Pollutants in the Troposphere (edited by Borrell P.M., Borrell P., Cvitai T. and Seiler W.), pp. 689-693. Academic Publishing, Den Haag*(1994).

R.P. Shadbolt, E.A. Waller, J.P. Messina and J.A. Winkler, Source regions of lower-tropospheric airflow trajectories for the lower peninsula of Michigan: A 40-year air mass climatology, *Journal of Geophysical Research-Atmospheres* **111**(2006).

S. Sharma, E. Andrews, L.A. Barrie, J.A. Ogren and D. Lavoue, Variations and sources of the equivalent black carbon in the high Arctic revealed by long-term observations at Alert and Barrow: 1989-2003, *Journal of Geophysical Research-Atmospheres* **111**(2006).

S. Sharma, D. Lavoue, H. Cachier, L.A. Barrie and S.L. Gong, Long-term trends of the black carbon concentrations in the Canadian Arctic, *Journal of Geophysical Research-Atmospheres* **109**(2004).

P.G. Simmonds, R.G. Derwent, A.L. Manning and G. Spain, Significant growth in surface ozone at Mace Head, Ireland, 1987-2003, *Atmos. Environ.* **38**(2004), pp. 4769-4778.

P.G. Simmonds, S. Seuring, G. Nickless and R.G. Derwent, Segregation and interpretation of ozone and carbon monoxide measurements by air mass origin at the TOR Station Mace Head, Ireland from 1987 to 1995, *Journal of Atmospheric Chemistry* **28**(1997), pp. 45-59.

O.A. Sindosi, B.D. Katsoulis and A. Bartzokas, An objective definition of air mass types affecting Athens, Greece; The corresponding atmospheric pressure patterns and air pollution levels, *Environmental Technology* **24**(2003), pp. 947-962.

S. Solberg *et al.*, European surface ozone in the extreme summer 2003, *Journal of Geophysical Research-Atmospheres* **113**(2008).

S. Solberg, N. Schmidbauer, A. Semb, F. Stordal and O. Hov, Boundary-layer ozone depletion as seen in the Norwegian Arctic in Spring, *Journal of Atmospheric Chemistry* **23**(1996), pp. 301-332.

S. Solberg, F. Stordal and O. Hov, Tropospheric ozone at high latitudes in clean and polluted air masses, a climatological study, *Journal of Atmospheric Chemistry* **28**(1997), pp. 111-123.

Y. Song *et al.*, Comparison of receptor models for source apportionment of volatile organic compounds in Beijing, China, *Environmental Pollution* **156**(2008), pp. 174-183.

N. Spichtinger, M. Winterhalter and P. Fabian, Ozone and grosswetterlagen analysis for the Munich metropolitan area, *Environmental Science and Pollution Research* **3**(1996), pp. 145-152.

A. Stohl, Trajectory statistics - A new method to establish source-receptor relationships of air pollutants and its application to the transport of particulate sulfate in Europe, *Atmospheric Environment* **30**(1996), pp. 579-587.

A. Stohl, Computation, accuracy and applications of trajectories - A review and bibliography, *Atmospheric Environment* **32**(1998), pp. 947-966.

A. Stohl *et al.*, Pan-Arctic enhancements of light absorbing aerosol concentrations due to North American boreal forest fires during summer 2004, *J. Geophys. Res.* **111**(2006), p. D22214.

A. Stohl *et al.*, Arctic smoke - record high air pollution levels in the European Arctic due to agricultural fires in Eastern Europe in spring 2006 *Atmos. Chem. Phys.* **7**(2007), pp. 511-534.

A. Stohl *et al.*, Forecasting for a Lagrangian aircraft campaign, *Atmos. Chem. Phys.* **4**(2004), pp. 1113-1124.

A. Stohl *et al.*, A replacement for simple back trajectory calculations in the interpretation of atmospheric trace substance measurements, *Atmospheric Environment* **36**(2002), pp. 4635-4648.

A. Stohl, C. Forster, A. Frank, P. Seibert and G. Wotawa, Technical note: The Lagrangian particle dispersion model FLEXPART version 6.2, *Atmos. Chem. Phys.* **5**(2005), pp. 2461-2474.

A. Stohl, M. Hittenberger and G. Wotawa, Validation of the Lagrangian particle dispersion model FLEXPART against large-scale tracer experiment data, *Atmospheric Environment* **32**(1998), pp. 4245-4264.

A. Stohl *et al.*, An analytical inversion method for determining regional and global emissions of greenhouse gases: Sensitivity studies and application to halocarbons, *Atmospheric Chemistry and Physics* **9**(2009), pp. 1597-1620.

A. Stohl *et al.*, The influence of stratospheric intrusions on alpine ozone concentrations, *Atmos. Environ.* **34** (2000), pp. 1323-1354.

A. Stohl, G. Wotawa, P. Seibert and H. Krompkolb, Interpolation errors in wind fields as a function of spatial and temporal resoloution and their impact on different types of kinematic trajectories, *Journal of Applied Meteorology* **34**(1995), pp. 2149-2165.

J. Strong, J.D. Whyatt and C.N. Hewitt, Application of multiple wind-roses to improve the modelling of ground-level ozone in the UK, *Atmospheric Environment* **40**(2006), pp. 7480-7493.

J. Strong, J.D. Whyatt, C.N. Hewitt and R.G. Derwent, Development and application of a Lagrangian model to determine the origins of ozone episodes in the UK, *Atmospheric Environment* **44**(2010), pp. 631-641.

O.A. Tarasova *et al.*, Surface ozone at the Caucasian site Kislovodsk High Mountain Station and the Swiss Alpine site Jungfraujoch: data analysis and trends (1990â€“2006), *Atmos. Chem. Phys.* **9**(2009), pp. 4157-4175.

B.F. Taubman *et al.*, Aircraft vertical profiles of trace gas and aerosol pollution over the mid-Atlantic United States: Statistics and meteorological cluster analysis, *Journal of Geophysical Research-Atmospheres* **111**(2006).

C. Toledano *et al.*, Airmass Classification and Analysis of Aerosol Types at El Arenosillo (Spain), *Journal of Applied Meteorology and Climatology* **48**(2009), pp. 962-981.

M. Traub *et al.*, Chemical characteristics assigned to trajectory clusters during the MINOS campaign, *Atmospheric Chemistry and Physics* **3**(2003), pp. 459-468.

W. Tscherwenka, P. Seibert, A. Kasper and H. Puxbaum, On-line measurements of sulfur dioxide at the 3 km level over central Europe (Sonnblick Observatory, Austria) and statistical trajectory source analysis, *Atmospheric Environment* **32**(1998), pp. 3941-3952.

B.J. Tsuang, Quantification on the source/receptor relationship of primary pollutants and secondary aerosols by a Gaussian plume trajectory model: Part I theory, *Atmospheric Environment* **37**(2003), pp. 3981-3991.

B.J. Tsuang *et al.*, Quantification on the source/receptor relationship of primary pollutants and secondary aerosols by a Gaussian plume trajectory model: Part II. Case study, *Atmospheric Environment* **37**(2003a), pp. 3993-4006.

B.J. Tsuang *et al.*, Quantification on the source/receptor relationship of primary pollutants and secondary aerosols by a Gaussian plume trajectory model: Part III-Asian dust-storm periods, *Atmospheric Environment* **37**(2003b), pp. 4007-4017.

I.J. Turias, F.J. Gonzalez, M.L. Martin and P.L. Galindo, A competitive neural network approach for meteorological situation clustering, *Atmospheric Environment* **40**(2006), pp. 532-541.

B. Tuzson *et al.*, Continuous isotopic composition measurements of tropospheric CO2 at Jungfraujoch (3580 m a.s.l.), Switzerland: real-time observation of regional pollution events, *Atmos. Chem. Phys.* **11**(2011), pp. 1685-1696.

C.A. Varotsos, M.N. Efstathiou and K.Y. Kondratyev, Long-term variation in surface ozone and its precursors in Athens, Greece - A forecasting tool, *Environmental Science and Pollution Research* **10**(2003), pp. 19-23.

L.A.D. Vasconcelos, J.D.W. Kahl, D.S. Liu, E.S. Macias and W.H. White, A tracer calibration of back trajectory analysis at the Grand Canyon, *Journal of Geophysical Research-Atmospheres* **101**(1996), pp. 19329-19335.

P. Venkatachari *et al.*, Spatial and temporal variability of black carbon in New York City, *Journal of Geophysical Research-Atmospheres* **111**(2006).

M. Viana *et al.*, Source apportionment of particulate matter in Europe: A review of methods and results, *Journal of Aerosol Science* **39**(2008), pp. 827-849.

A. Virkkula *et al.*, Chemical composition of atmospheric aerosol in the European subarctic: Contribution of the Kola Peninsula smelter areas, central Europe, and the Arctic Ocean, *Journal of Geophysical Research-Atmospheres* **104**(1999), pp. 23681-23696.

H.L. Walker, R.G. Derwent, R. Donovan and J. Baker, Photochemical trajectory modelling of ozone during the summer PUMA campaign in the UK West Midlands, *Science of the Total Environment* **407**(2009), pp. 2012-2023.

T.J. Wang, K.S. Lam, C.W. Tsang and S.C. Kot, On the variability and correlation of surface ozone and carbon monoxide observed in Hong Kong using trajectory and regression analyses, *Advances in Atmospheric Sciences* **21**(2004), pp. 141-152.

Y.Q. Wang, X.Y. Zhang and R.R. Draxler, TrajStat: GIS-based software that uses various trajectory statistical analysis methods to identify potential sources from long-term air pollution measurement data, *Environmental Modelling & Software* **24**(2009), pp. 938-939.

Y.X. Wang *et al.*, Year round measurements of O-3 and CO at a rural site near Beijing: variations in their correlations, *Tellus Series B-Chemical and Physical Meteorology* **62**(2010), pp. 228-241.

P. Weiss-Penzias, D.A. Jaffe, L. Jaegle and Q. Liang, Influence of long-range-transported pollution on the annual and diurnal cycles of carbon monoxide and ozone at Cheeka Peak Observatory, *Journal of Geophysical Research-Atmospheres* **109**(2004).

P. Weiss-Penzias *et al.*, Observations of Asian air pollution in the free troposphere at Mount Bachelor Observatory during the spring of 2004, *Journal of Geophysical Research-Atmospheres* **111**(2006).

G.M. Wolfe *et al.*, Influence of trans-Pacific pollution transport on acyl peroxy nitrate abundances and speciation at Mount Bachelor Observatory during INTEX-B, *Atmos. Chem. Phys.* **20**(2007), pp. 5309-5325.

G. Wotawa and H. Kroger, Testing the ability of trajectory statistics to reproduce emission inventories of air pollutants in cases of negligible measurement and transport errors, *Atmospheric Environment* **33**(1999), pp. 3037-3043.

G. Wotawa, H. Kroger and A. Stohl, Transport of ozone towards the Alps - results from trajectory analyses and photochemical model studies, *Atmospheric Environment* **34**(2000), pp. 1367-1377.

Z.J. Wu, M. Hu, K.S. Shao and J. Slanina, Acidic gases, NH3 and secondary inorganic ions in PM10 during summertime in Beijing, China and their relation to air mass history, *Chemosphere* **76**(2009), pp. 1028-1035.

X. Xia, H. Chen and W. Zhang, Analysis of the dependence of column-integrated aerosol properties on long-range transport of air masses in Beijing, *Atmospheric Environment* **41**(2007), pp. 7739-7750.

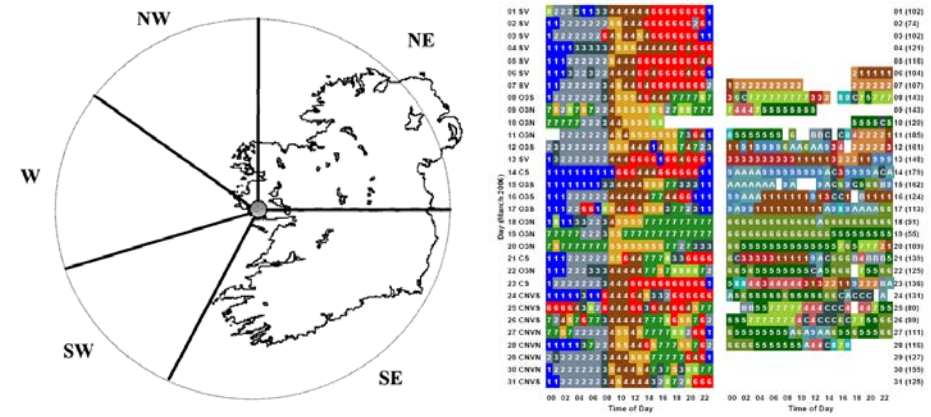
H. Xiao *et al.*, Transport of semivolatile organic compounds to the Tibetan Plateau: Monthly resolved air concentrations at Nam Co, *Journal of Geophysical Research-Atmospheres* **115**(2010).

J. Xu, D. DuBois, M. Pitchford, M. Green and V. Etyemezian, Attribution of sulfate aerosols in Federal Class I areas of the western United States based on trajectory regression analysis, *Atmospheric Environment* **40**(2006), pp. 3433-3447.

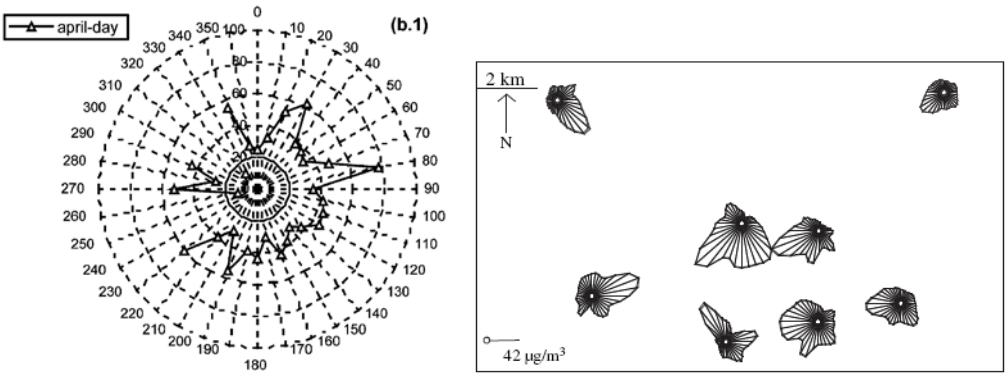
P. Yan *et al.*, The measurement of aerosol optical properties at a rural site in Northern China, *Atmos. Chem. Phys.* **8**(2008), pp. 2229-2242.

W. Yue *et al.*, Source apportionment of ambient fine particle size distribution using positive matrix factorization in Erfurt, Germany, *Science of the Total Environment* **398**(2008), pp. 133-144.

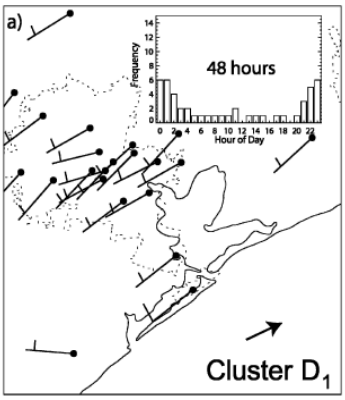
L. Zhu, X. Huang, H. Shi, X. Cai and Y. Song, Transport pathways and potential sources of PM10 in Beijing, *Atmospheric Environment* **45**(2011), pp. 594-604.



a) b)



c) d)



e)

Figure 1- Visualisation of wind data and separation into wind sectors:

a) Mace Head wind sector divisions used for a range of trace gas measurements in 1996 and 1997. Taken from (Salisbury et al., 2002).

b) Meteorological episodes derived from surface wind measurements arranged as a timeseries meteorological flags during the MILAGRO campaign in Mexico city. Taken from (de Foy et al., 2008).

c) Wind roses of Black Carbon levels from various directions for a month at La Reunion island. Taken from (Bhugwant et al., 2001).

d) High-resolution pollutant roses (known as Power-Ridge Pollutant (PRP) roses) for SO_2 arriving at the Antwerp harbour area. Taken from (Cosemans et al., 2008).

e) The use of wind barbs representing the wind direction and speed used to show one of the 16 meteorological regime clusters for studying ozone exceedances in Houston, Texas. . Taken from (Darby, 2005).

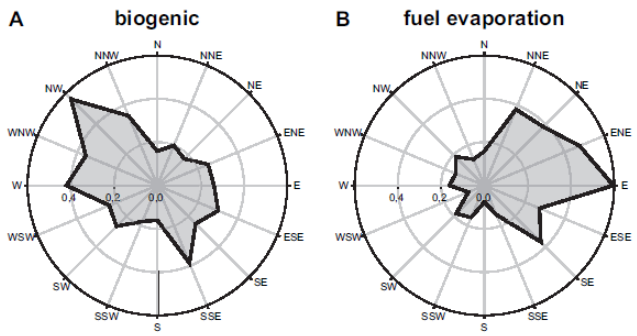
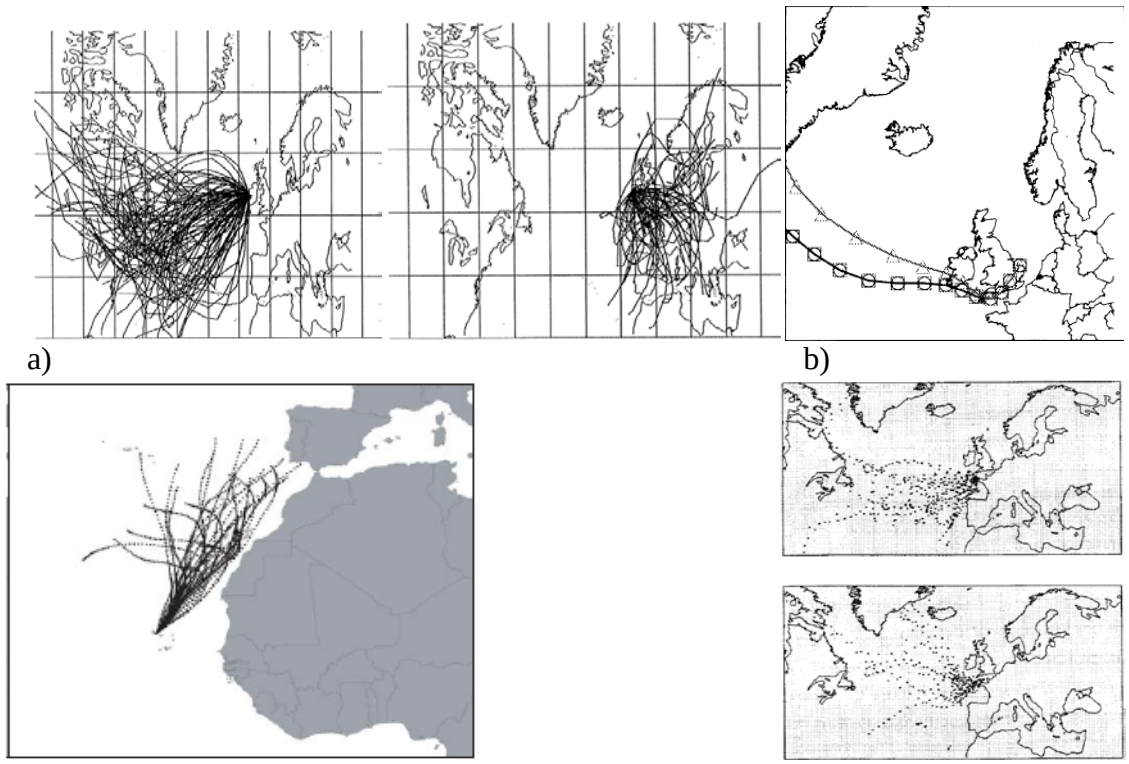


Figure 2- Wind roses of the 75th percentile PM_{10} levels from biogenic and fuel evaporation sources (out of a total of 8 different sources separated by positive Matrix Factorisation (PMF) analysis) at Houston, Texas. Taken from (Leuchner and Rappengluck, 2010).



a) b)

c) d)

Figure 3- Trajectory examples:
a) Westerly and Easterly Mace Head, Ireland 6 hourly 5 day back trajectory clusters. Taken from (Cape et al., 2000).
b) Weybourne, UK 6 hourly trajectories arriving during 1993. Taken from (Cardenas et al., 1998).
c) 4 day backward Cape Verde Observatory daily trajectories from mid May to mid June 2007. Taken from (Müller et al., 2010).
d) Back trajectories for the TOR station at Porspoder, France for summers 1992-1995 associated with 10th and 90th percentile ozone levels respectively. Taken from (Fenneteaux et al., 1999).

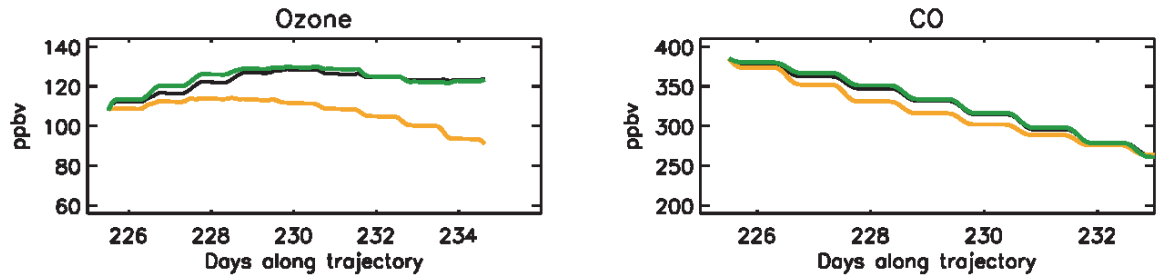


Figure 4- Chemistry modelling along a trajectory: GEOS-CHEM model simulations of ozone and CO evolution in a biomass burning plume arriving into Western Africa from South America during the AMMA aircraft campaign. Black represents the reference trajectory from the flight altitude, green a higher level trajectory and orange a lower level trajectory. Taken from (Real et al., 2010).

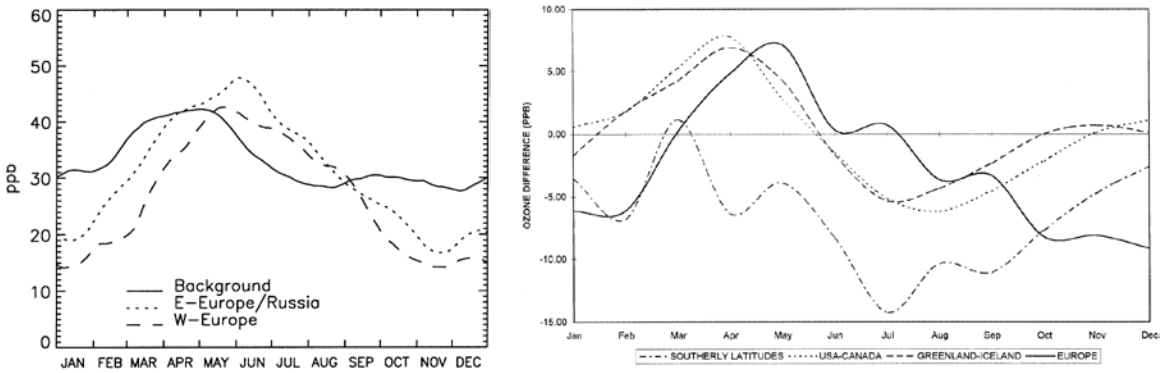


Figure 5- Seasonal composition cycles in different trajectory types:
a) Seasonal O_3 cycles at Birkenes, Norway in 3 air mass sector types derived from trajectories. Taken from (Solberg et al., 1997).
b) The difference between the Northern Hemisphere Mid Latitude background (NHMLB) average ozone at Mace Head for the various sectors around the site. Taken from (Simmonds et al., 1997).

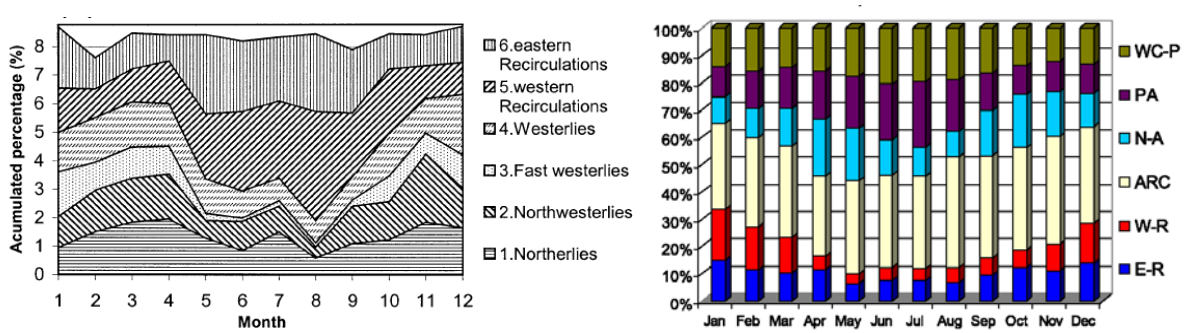


Figure 6- Seasonality of trajectory types:
a) Monthly % distribution of sectors for all 3565 back trajectories arriving in Barcelona over 4 years. Taken from (Jorba et al., 2004).
b) Monthly distribution of trajectory types representing various sectors (Russia, Europe, America, Arctic in various combinations) influencing the Barrow Arctic station over 15 years and subsequently used for interpreting Black Carbon measurements. Taken from (Sharma et al., 2006).

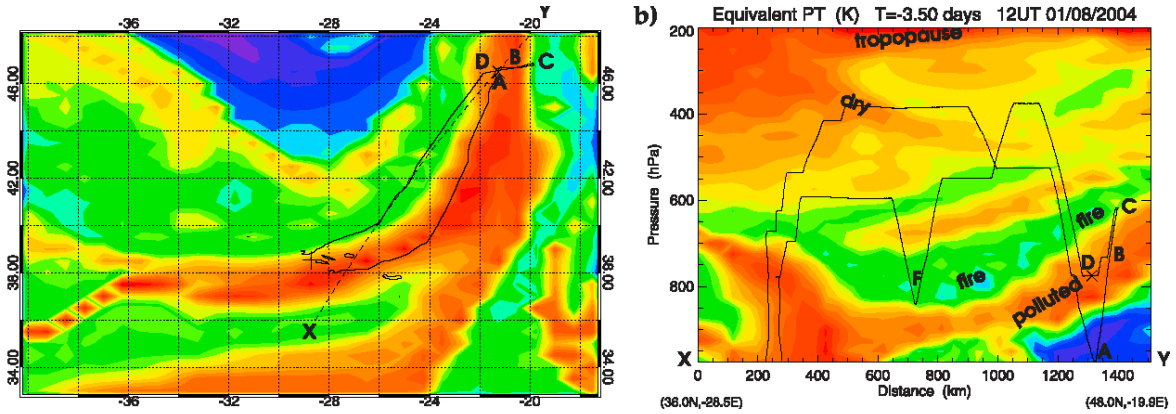
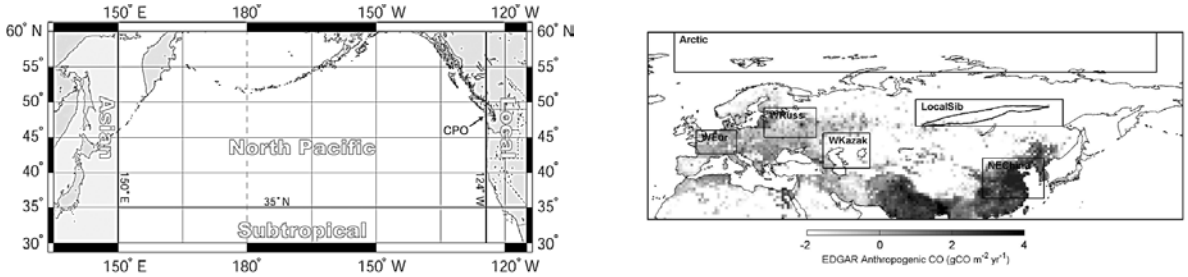
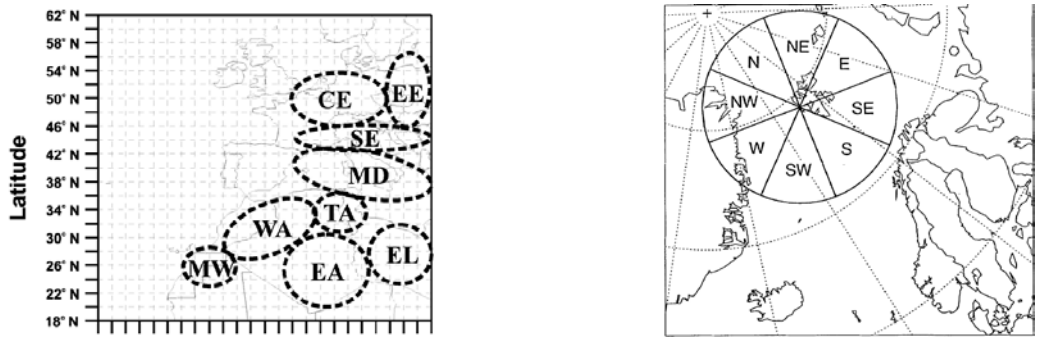


Figure 7- Reverse Domain Filling (RDF) using FLEXPART plots used in planning the ICARTT flights over the Atlantic and the flight tracks marked in black (different colours represent different source humidities, indicating varying source types). Taken from (Methven et al., 2006).



a) b)



c) d)

Figure 8- Regional emission zones and defining sector influences:
a) Cheeka Peak Observatory (CPO), WA regions of influence used for classifying trajectory types. Taken from (Weiss-Penzias et al., 2004).
b) Source regions assigned for allocating influences to trajectories arriving in Siberia for analysis of aircraft measurements. Taken from (Paris et al., 2010).
c) European and north African potential emission zones used for classifying European and north African PM transport to Madrid, Spain. Taken from (Salvador et al., 2008).
d) Zeppelin, Svalbard transport sectors with which to allocate trajectories. Allocation to each sector is done if at least 50% of the last 24 hours of the trajectory has travelled less than 850 km. Taken from (Solberg et al., 1996).

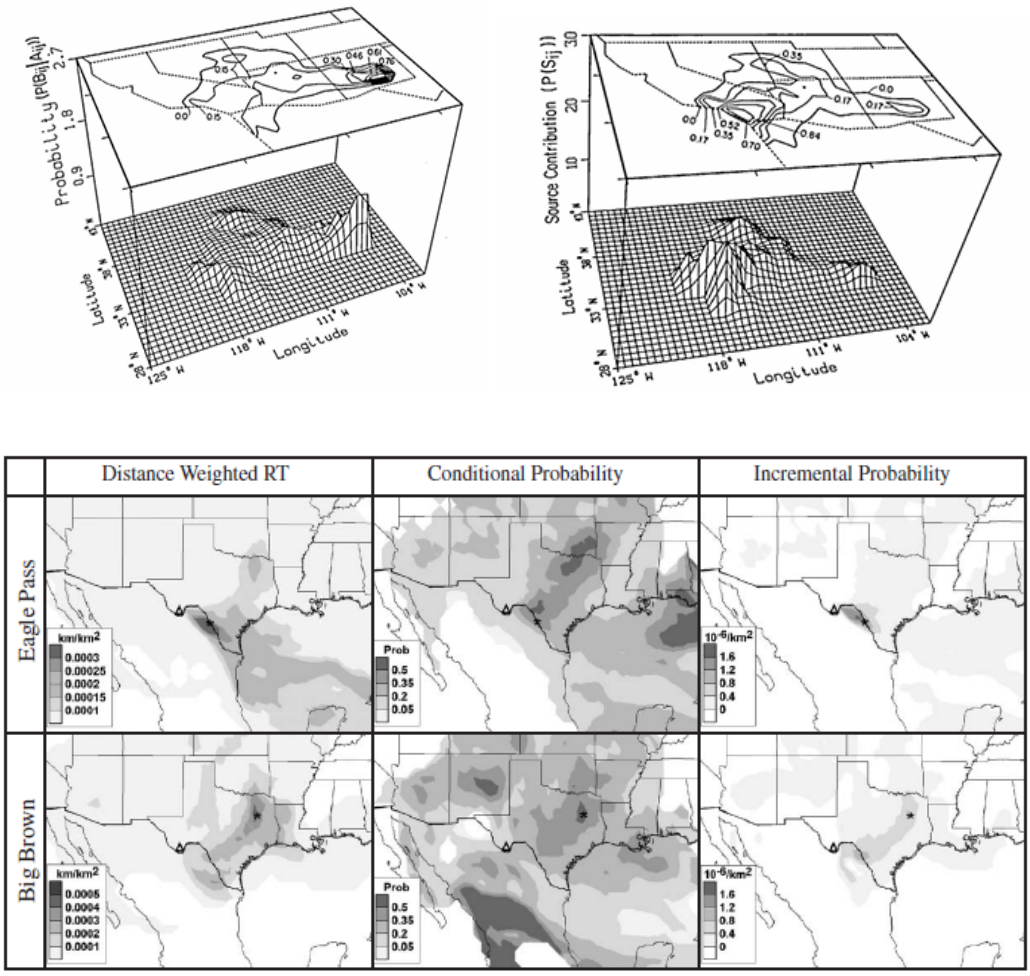


Figure 9- Conditional probability maps:
a) Conditional Probability Function (CPF) for highest *potential* for contributing to high sulphur levels arriving at Grand Canyon National Park (highest from New Mexico and south California). Taken from (Ashbaugh et al., 1985).
b) The source Contribution Function (SCF) shows that the region that *actually* contributes most to high sulphur levels is southern California. Taken from (Ashbaugh et al., 1985).
c) Conditional probability techniques (residence time, conditional probability and incremental probability) used for Big Bend National Park, Texas (small triangle). Tracer release from 2 sites (Eagle Pass and Big Brown) were combined with 5 day back trajectories used to study the upper 20th percentile particulate sulphur sources to the National Park. Taken from (Schichtel et al., 2006).

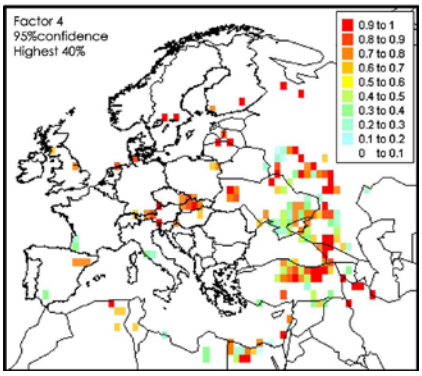


Figure 10- Potential Source Contribution Factor (PSCF) analysis for aerosol levels at Antalya, Turkey. The highest 40% of each of 4 Factor scores from factor analysis for source apportionment of the aerosol were selected as polluted trajectories. Taken from (Güllü et al., 2005).

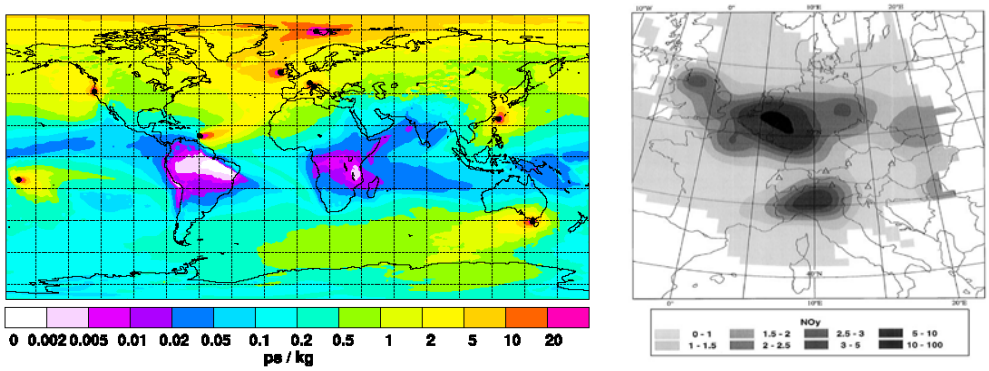
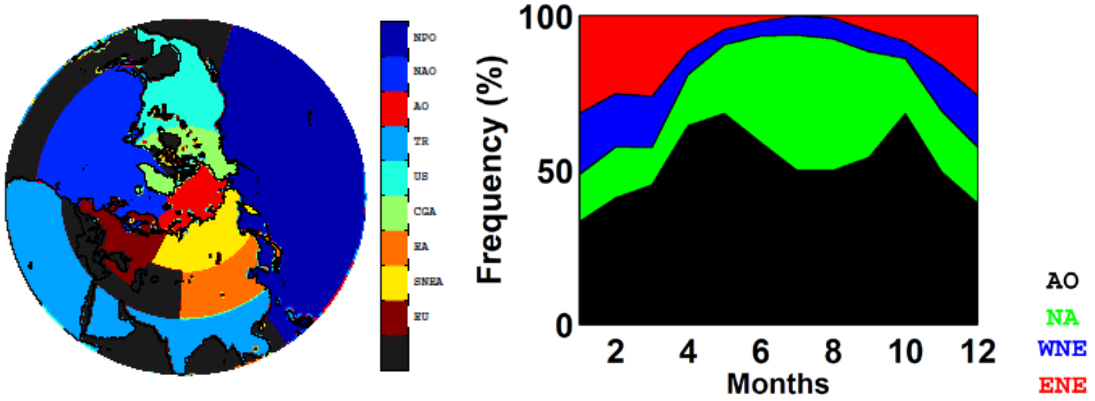


Figure 11- Footprint emission sensitivity maps:
a) Footprint emission sensitivity map of hydrofluorocarbons (in picoseconds per kilogram) obtained from FLEXPART 20 day backward calculations (Januaryy 2005- March 2007) . Taken from (Stohl et al., 2009).
b) NOy emission (from 0-100 x 10^{-10} kgm^{-2} s^{-1}) map over Europe derived from the concentration field calculated from trajectory statistics and the EMEP emission inventory. Taken from (Wotawa and Kroger, 1999).



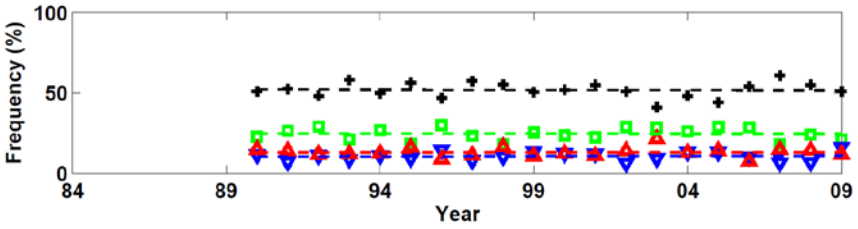


Figure 12- Source region classifications for Zeppelin, Spitzbergen. Taken from (Hirdman et al., 2010):
a) Regions for clustering of the footprint emission sensitivities, b) Monthly mean distribution of the trajectory types (AO: Arctic Ocean, NA: North America, WNE: Western Northern Eurasian cluster, ENE: Eastern Northern Eurasian), c) Annual mean trajectory type distribution between 1990 and 2009

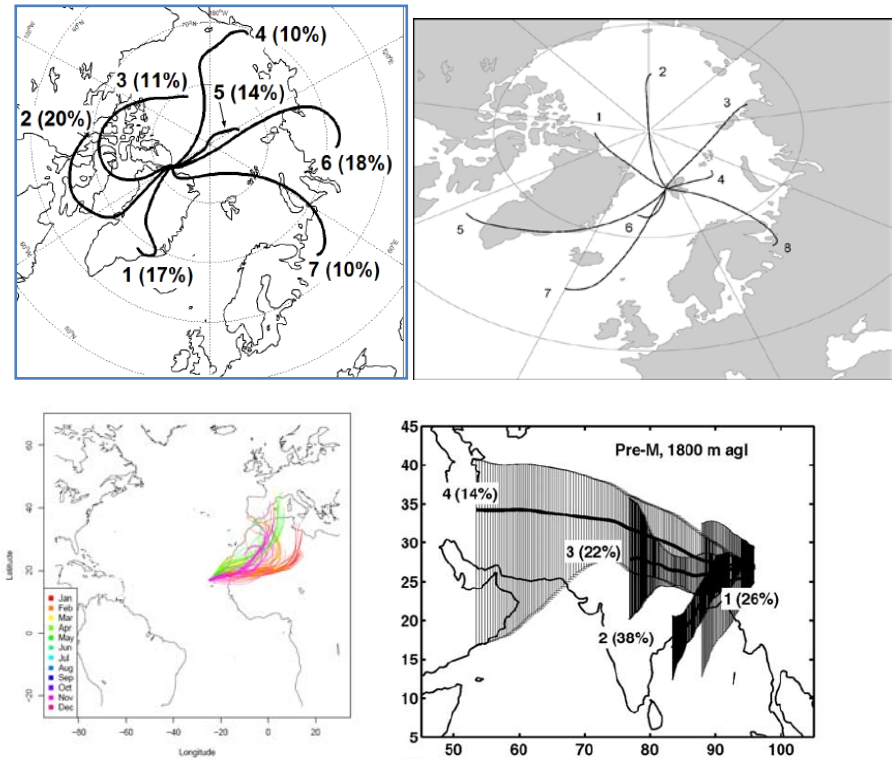


Figure 13- Cluster analysis of trajectories:
a) Mean trajectory pathways for Alert in January (1990-2005), 10 days backwards with frequency of occurrence in % for each cluster shown. Taken from (Huang et al., 2010).
b) Mean trajectory pathways for Zeppelin (5 day backwards) Trajectory lengths in the different clusters were 1, 3180 km; 2, 3920 km; 3, 3880 km; 4, 3200 km; 5, 4850 km; 6, 4330 km; 7, 4330km and 8, 3880 km. Taken from (Eneroth et al., 2007).
c) One of 8 trajectory clusters (African) of 5 day back trajectories arriving at Cape Verde in 2007 (coloured by month) with ethane included in cluster analysis. Taken from (Jorba et al., 2004).
d) Cluster mean trajectories arriving at the mountain site of Dibrugarhin India in the pre-monsoon season. Taken from (Gogoi et al., 2009).

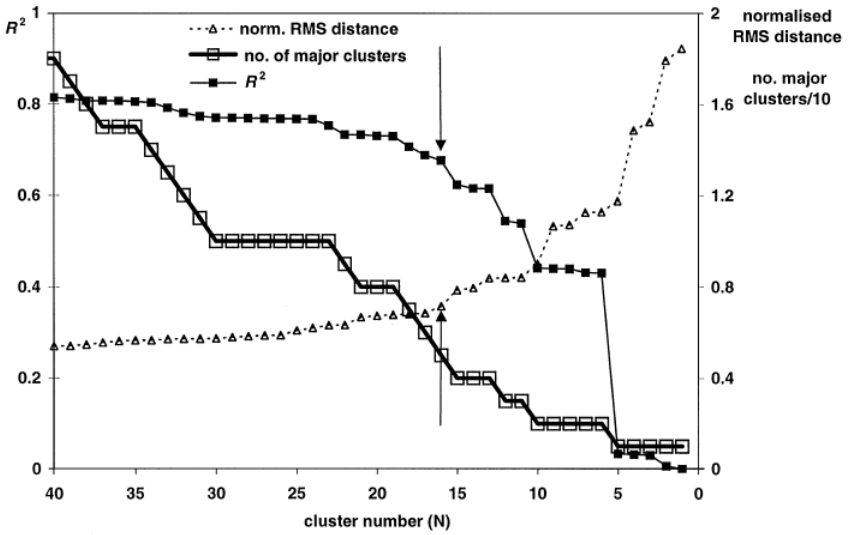


Figure 14- Changes in the normalised RMS distance between clusters, and the total variance (R^2) as a function of the number of clusters for the daily back trajectories at Mace Head 1995 to 1997. Step changes in these statistics represent logical points for defining an optimum number of clusters to retain in the analysis, in this case, 5 major trajectory clusters. Taken from (Cape et al., 2000).

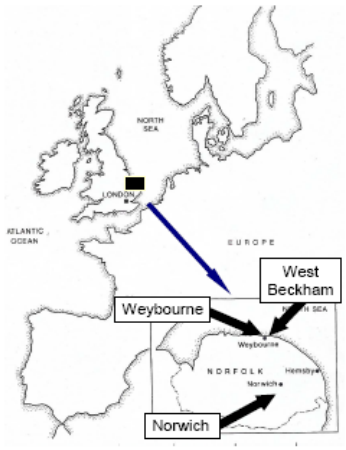


Figure 15- The Weybourne Atmospheric Observatory

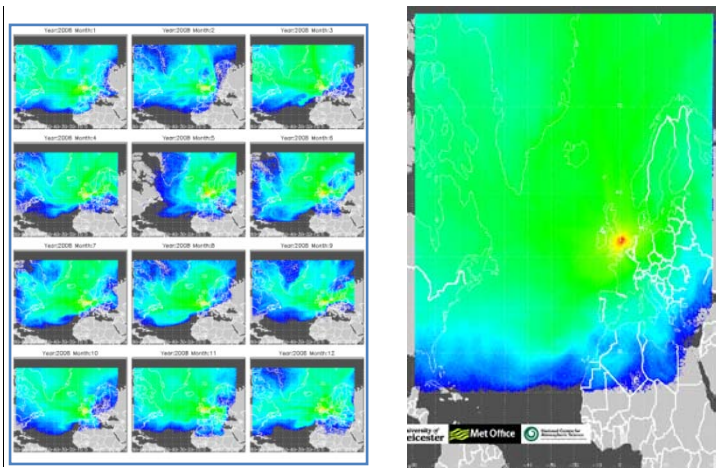


Figure 16- Weybourne 10 day 2008 integrated footprints a) Monthly b) Annual

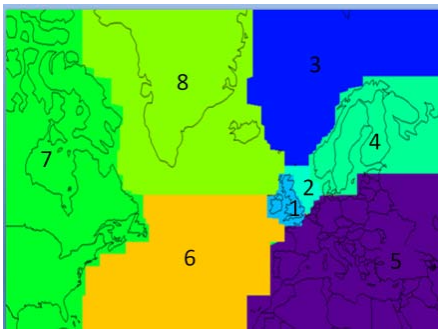


Figure 17- Regional grid divisions for Weybourne for 10 day domain

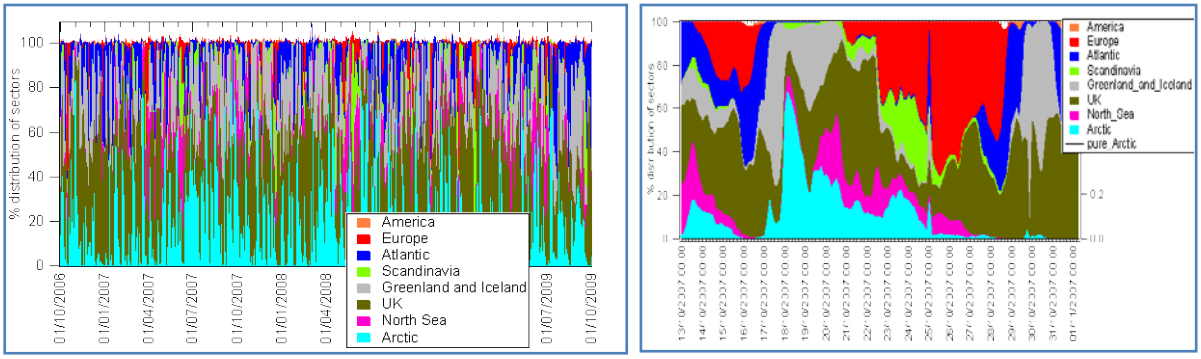


Figure 18- Division of Weybourne sector influences for each 3 hourly period (January 2006-September 2009) and zoomed in section to show the small scale changes in regional influences

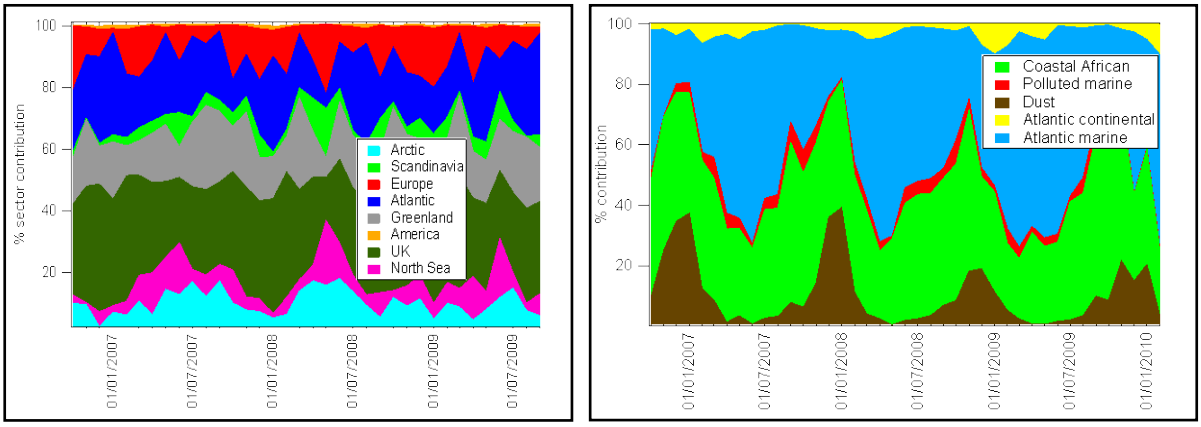


Figure 19- Monthly averaged regional distribution at a) Weybourne and b) Cape Verde Observatory (see (Carpenter, 2011) for detailed analysis for Cape Verde)

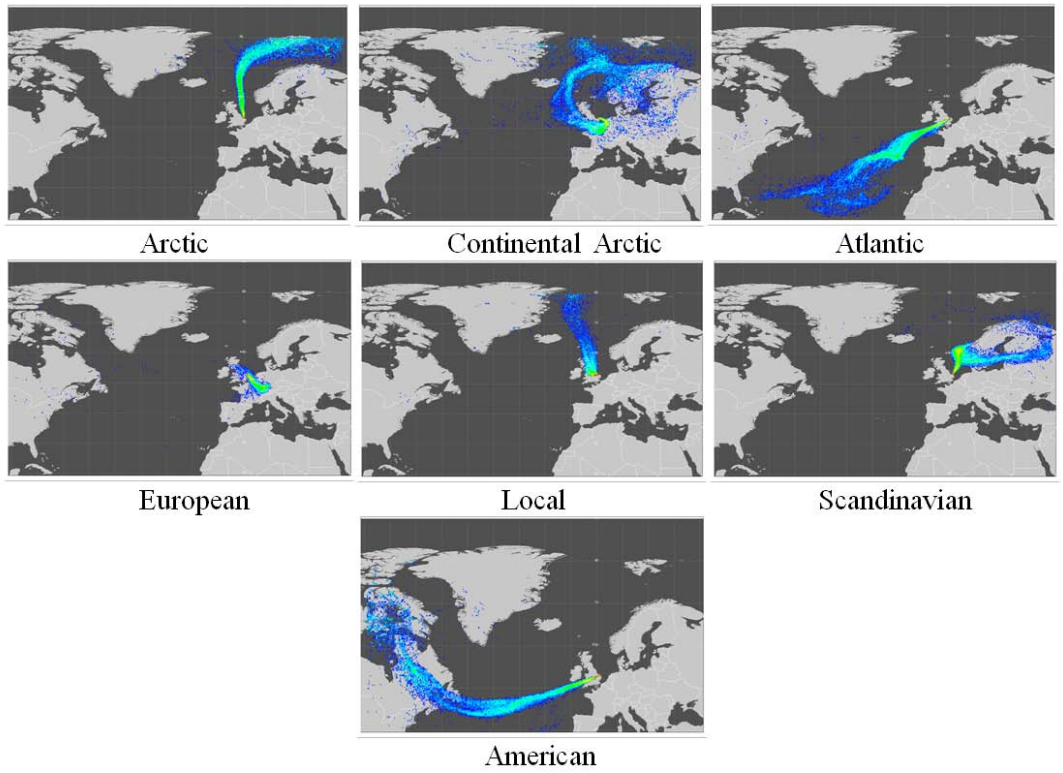


Figure 20- Examples of the 7 Weybourne air mass classifications (3 hourly trajectories)

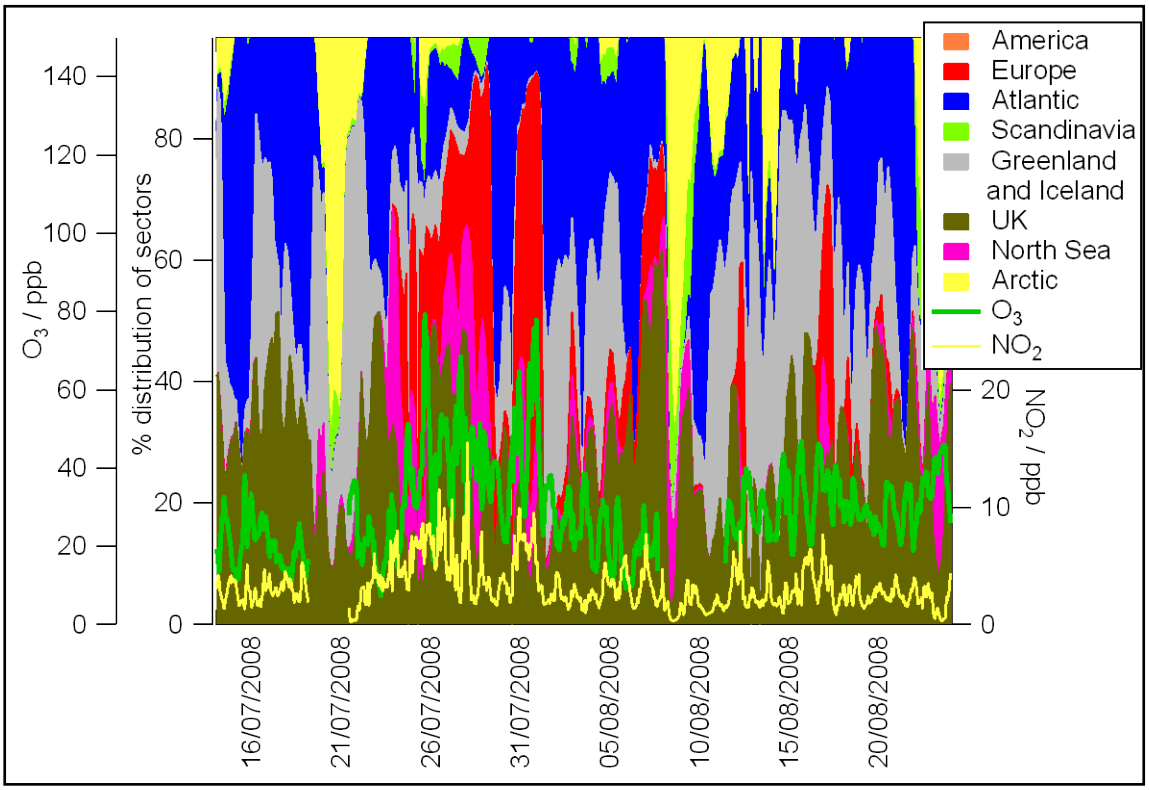


Figure 21- Weybourne composition variations with regional distribution of air mass history. As European and UK regional influence increases, ozone increases as well as NO_2.

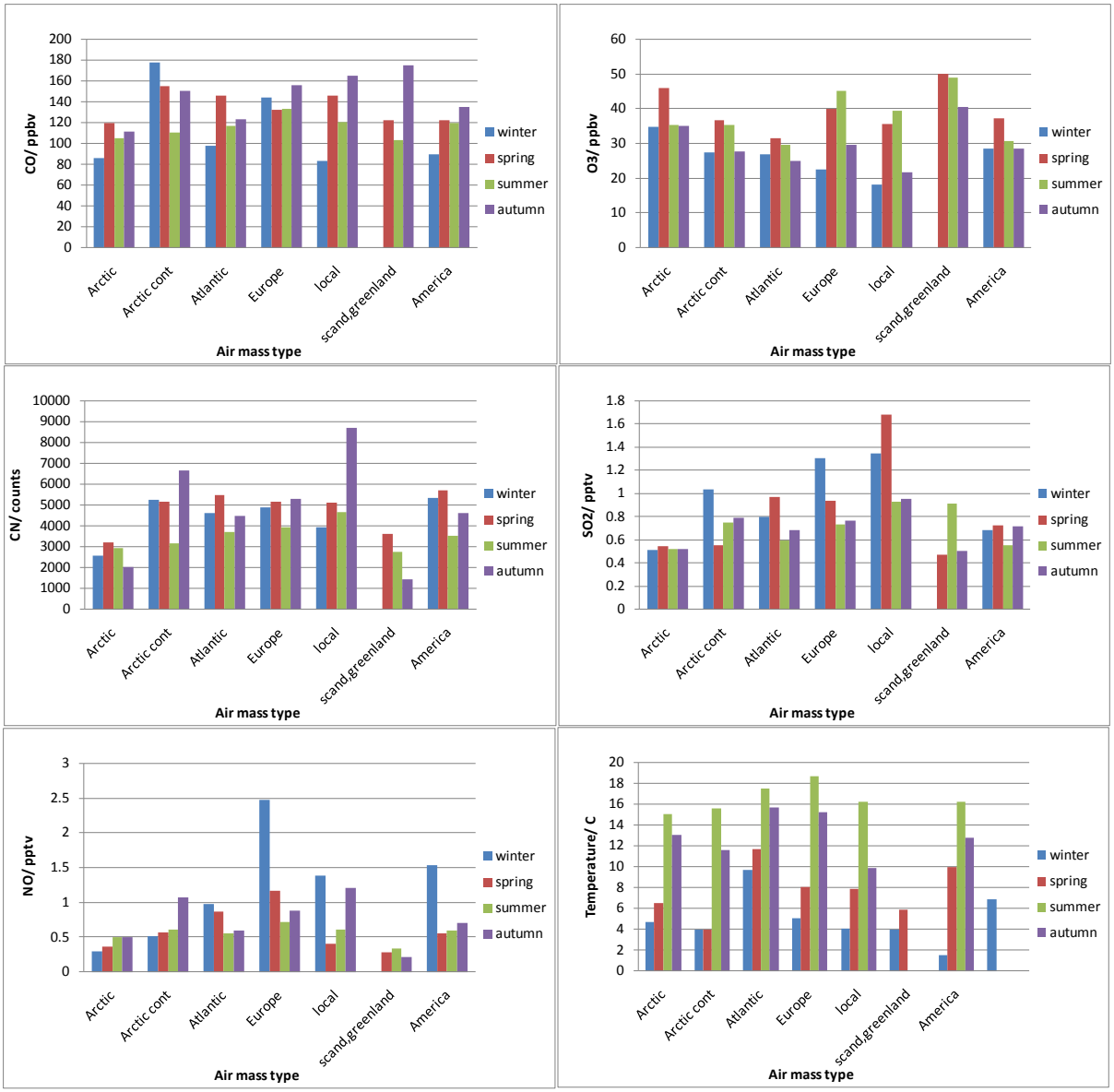


Figure 22- Composition (CO, O_3, CN, SO_2, NO) and temperature distribution in different trajectory types 4 year average 2006-2009 winter (DJF), spring (MAM), summer (JJA) and autumn (SON)

Table 1- Overview of studies dealing with using back trajectories or dispersion models to segregate composition measurements. If the trajectory is not specified, it is just left as "Trajectory". Abbreviations in the cluster analysis or modelling section are explained in Sections 4.2 and 2.4 but a summary is included here; CF (concentration Field), CPF (Conditional Probability Function), ANOVA (analysis of variance), PSCF (Potential Source Contribution Function), RCF (Redistribution Concentration Field), SOM (Self Organising Maps), RDF (Reverse Domain Filling), MCAR (Median Concentrations at Receptor), PMF (Positive Matrix Factorisation), PCA (Principal Component Analysis), MVA (Multivariate analysis).

Reference	Location	Species measured	Trajectory/ dispersion model type	Cluster analysis methods	Residence time Analysis	Chemistry Modelling?
(Aalto, Hatakka et al. 2002)	Pallas, Finland	CO_2, O_3, aerosols	TRADOS Trajectory		Residence time, CF	
(Abdalmogith and Harrison 2005)	Belfast, Harwell, UK	PM_{10}	HYSPLIT Trajectory	k-means		
(Apadula, Gotti et al. 2003)	Plateau Rosa, Monte Cimone and Zugspitze	CO_2	TRAIET Trajectory		CF	Source-receptor model
(Ashbaugh, Malm et al. 1985)	Grand Canyon NP, USA	PM	Trajectory		CPF, PSCF	
(Baker 2010)	Birmingham, Harwell, UK	O_3, NO_x, PM, CO, SO_2, benzene	HYSPLIT Trajectory	k-means		
(Begum, Kim et al. 2005)	Philadelphia, US	PM	HYSPLIT Trajectory		PSCF	
(Biegalski and Hopke 2004)	Burnt Island, Canada	As, In, Sb, Se, Sn, Zn	AES trajectory		PSCF	
(Borge, Lumbreras et al. 2007)	Athens, Madrid and Birmingham	PM	HYSPLIT Trajectory	k-means (2-stage)		
(Brankov, Rao et al. 1998)	3 stations in NE USA	O_3, PM	HYSPLIT Trajectory	k-means		
(Burley and Ray 2007)	Yosemite NP, USA	O_3	HYSPLIT Trajectory		PSCF	
(Cape, Methven et al. 2000)	Mace Head, Ireland	O_3	UGAMP Trajectory	Hierarchical		
(Cardenas, Austin et al. 1998)	Weybourne, UK	CO, O_3, VOC	Trajectory			
(Carvalho, Monteiro et al. 2010)	Lamas d'Olo, Portugal	O_3	HYSPLIT Trajectory	k-means	Hierarchical	
(Cheng and Lin 2001)	Lamont, Oklahoma, US	Aerosols	HYSPLIT Trajectory		PSCF	
(Choi, Heo et al. 2010)	Sukmo Island, Korea	VOC	HYSPLIT Trajectory		PSCF	PMF
(Cohen, Artz et al. 2004)	Lake Michigan, USA	Mercury	HYSPLIT Trajectory			Emissions data
(Cohen, Crawford et al. 2010)	Hanoi, Vietnam	PM	HYSPLIT Trajectory		CPF	PMF
(Colette, Ancellet et al. 2005)	European stations and MOZAIC	O_3	FLEXPART Dispersion			MVA

	(aircraft)					
(Davis, Normile et al. 2010)	Virginia, USA	O_3	HYSPLIT Trajectory	k-means (2-stage)		
(de Foy, Zavala et al. 2009)	Mexico city	O_3	FLEXPART Dispersion	k-means (2-stage)		PCA
(Delcloo and De Backer 2008)	Uccle, Belgium	O_3	APTRA Trajectory	k-means		
(Derwent, Witham et al. 2010)	UK	O_3	NAME Dispersion			PTM
(Doddridge, Dirmeyer et al. 1994)	Mace Head, Ireland	O_3, CO	NMC Trajectory		CPF	
(Dogan, Gullu et al. 2008)	Antalya, Turkey	Aerosols	ECMWF Trajectory		PSCF	PMF
(Dorling and Davies 1995)	3 Norwegian and 3 UK stations	sulphate and precipitation chemistry	Trajectory	k-means		
(Du and Rodenburg 2007)	Camden, NJ, USA	PCB	HYSPLIT Trajectory		PSCF	PMF
(Dueñas, Orza et al. 2011)	Málaga, Spain	Aerosols, ^{7}Be and ^{210}Pb	HYSPLIT Trajectory	k-means		
(Dvorska, Lammel et al. 2009)	Kosetice, Czech Republic	Persistant Organic Pollutants	HYSPLIT Trajectory		Concentration loadings per country	
(Eneroth, Kjellstrom et al. 2003)	Ny Alesand, Svaalbard	CO_2	HYSPLIT Trajectory	Hierarchical		
(Eneroth, Holmen et al. 2007)	Zeppelin, Ny Alesand, Svaalbard	O_3, VOC, Mercury	HYSPLIT Trajectory	Hierarchical		
(Escudero, Stein et al. 2011)	La Castanya, Spain	PM, TSP	HYSPLIT Trajectory		Regional	
(Evans, Shallcross et al. 2000)	Mace Head, Ireland	O_3, NOx, CO	UGAMP Trajectory			CiTTyCAT
(Fenneteaux, Colin et al. 1999)	Porspoder, France	O3, VOC, PAN, and NO*x*	NMC Trajectory			
(Font, MorguÃ et al. 2011)	Northern Spain (aircraft)	CO_2	FLEXPART Dispersion		Footprint Emission Sensitivity	PCA
(Forster, Wandinger et al. 2001)	Mace Head and Europe	O_3, aerosols	FLEXTRA Trajectory			FLEXPART with CO tracer
(Gebhart, Schichtel et al. 2011)	Rocky Mountain NP, USA	SO_2, NH_3, NO, NO_2	HYSPLIT Trajectory			TrMB
(Gogoi, Moorthy et al. 2009)	Dibrugarh, India	Aerosol Optical Depth	HYSPLIT Trajectory	Hierarchical		
(Gregory, Bachmeier et al. 1996) and (Merrill 1996)	Pacific (aircraft)	O_3, CO, NO_{xy}, VOCs, minerals, CFCs	NMC Trajectory			
(Gros, Williams et al. 2004)	Atlantic ocean (ship)	O_3, CO, propane	FLEXTRA Trajectory			MATCH-MPIC
(Güllü, Dogan et al. 2005)	Antalya, Turkey	Aerosols	ECMWF Trajectory		PSCF	Factor analysis
(Hains, Taubman et al. 2008)	Mid-Atlantic USA (aircraft)	O_3, CO, SO_2	HYSPLIT Trajectory			
(Halse, Schlabach et al. 2011)	86 European background stations	Persistant Organic Pollutants	FLEXPART Dispersion		Footprint Emission Sensitivity	
(Han, Holsen et al. 2005)	New York state	Mercury	HYSPLIT Trajectory + Dispersion		PSCF	
(Harrison, Grenfell et al. 2000)	Weybourne, UK	NO_x and NO_y, NH_3, NH_4^+	Trajectory			
(Harrison, Yin et al. 2006)	Birmingham, UK	NO_{xy}, OH	NAME Dispersion (forward with chemistry)			

(Hirdman, Burkhart et al. 2010),(Hirdman, Sodemann et al. 2010)	Zeppelin, Alert, Barrow (+ Summit for b)	O_3, BC, aerosols	FLEXPART Dispersion		Footprint emission sensitivity	
(Hirdman, Aspmo et al. 2009)	Zeppelin	Mercury	FLEXPART Dispersion		Footprint emission sensitivity	
(Huang, Gong et al. 2010)	Alert, Greenland	BC	HYSPLIT Trajectory	k-means		
(Junker, Sheahan et al. 2004)	Orogrande, New Mexico, US	BC	HYSPLIT Trajectory		Regional	
(Junker, Wang et al. 2009)	4 stations in Taiwan and China	O_3, CO, SO_2, NO_x, PM	HYSPLIT Trajectory		Regional	
(Kaiser, Schelfinger et al. 2007)	5 Alpine stations	O_3, CO, NO_x	FLEXTRA Trajectory		PSCF, CPF and RCF	
(Kang, Kang et al. 2006)	Seoul, Korea	PM, HNO_3, HONO, SO_2	HYSPLIT Trajectory		PSCF	
(Karaca and Camci 2010)	Istanbul, Turkey	PM	HYSPLIT Trajectory	SOM		
(Karaca, Anil et al. 2009)	Istanbul, Turkey	PM	HYSPLIT Trajectory		PSCF	
(Kassomenos, Vardoulakis et al. 2010)	Athens, Greece	PM_{10}	HYSPLIT Trajectory	k-means, Hierarchical, SOM		
(Kocak, Mihalopoulos et al. 2009)	Erdemli, Turkey	PM_{10}	HYSPLIT Trajectory		PSCF	PMF
(Kuhn, Damoah et al. 2010)	Ellesmere island, Canada	Aerosols	FLEXTRA Trajectory		footprint emission sensitivity	EDGAR emissions
(Law, Fierli et al. 2010)	West Africa (aircraft)	O_3, CO, CO_2, NO_{xy}	UGAMP Trajectory		RDF	
(Lawler, Finley et al. 2009)	Cape Verde and Atlantic Ocean	O_3, NO_x, Chlorine species	Trajectory			MISTRA
(Lee and Ashbaugh 2007; Lee and Ashbaugh 2007; Lee and Ashbaugh 2007)	Grand Canyon NP, USA	SO_2	HYSPLIT Trajectory		CPF, PSCF, MURA	
(Lee, Moller et al. 2009)	Cape Verde	NO_x, NO_y	BADC Trajectory			
(Lewis, Evans et al. 2007)	Atlantic Ocean (aircraft)	CO, PAN, alkanes	UGAMP Trajectory			
(Lupu and Maenhaut 2002)	3 sites in Scandinavia and 1 in Israel	PM, BC	HYSPLIT Trajectory		PSCF, CF	
(Mace, Wang et al. 2011)	Cape Verde	O_3, CO, NO_{xy}, VOCs	HYSPLIT Trajectory	k-means (and others)		
(Makra, Sánta et al. 2010)	Thessaloniki, Szeged, and Hamburg	Pollen counts	HYSPLIT Trajectory	k-means (Mahalanobis, Convhull)		
(Malcolm, Derwent et al. 2000; Malcolm and Manning 2001)	UK and European sites	PM	NAME Dispersion			
(Manning, Ryall et al. 2003)	Mace Head, Ireland	CFCs, CH_4 and NO	NAME Dispersion			
(McConnell, Highwood et al. 2008)	West Africa (aircraft)	O_3, Ca and Al (dust), aerosols	NAME Dispersion			
(Merrill 1994; Merrill and Moody 1996)	Barbados, Bermuda, Mace Head, Tenerife	-	NMC Trajectory		CPF	

(Methven, Evans et al. 2001)	Mace Head, Ireland	O_3	UGAMP Trajectory		CPF	CiTTyCAT
(Methven, Arnold et al. 2003)	Atlantic and Arctic Oceans (aircraft)	O_3	UGAMP Trajectory		RDF	CiTTyCAT
(Methven, Arnold et al. 2006)	Atlantic Ocean (aircraft)	O_3, CO, NO_y	UGAMP Trajectory		RDF	CiTTyCAT
(Methven, Arnold et al. 2006)	Atlantic Ocean (aircraft)	O_3, CO, NO_{xy}, VOCs,	UGAMP Trajectory, FLEXPART Dispersion		RDF	
(Moody, Galusky et al. 1989)	Virginia, Bermuda, Cape point, Amsterdam island	Precipitation chemistry	ATAD Trajectory	Hierarchical		GAMBIT
(Moy, Dickerson et al. 1994)	Shenandoah NP, Virginia, US	O_3, CO, NO_y	HYSPLIT Trajectory	Hierarchical (Average-linkage, centroid, Ward's method)		
(Müller, Lehmann et al. 2010)	Cape Verde	PM, Ca, K, Fe	HYSPLIT Trajectory			
(Occhipinti, Aneja et al. 2008)	North Carolina, USA	PM, nitrogen precipitation	HYSPLIT Trajectory		Regional	
(Paris, Stohl et al. 2010)	Siberia (aircraft)	O_3, CO_2, CO	FLEXPART Dispersion	k-means	footprint emission sensitivity	
(Pekney, Davidson et al. 2006)	Pittsburgh, USA	PM	HYSPLIT Trajectory		PSCF	PMF
(Pochanart, Akimoto et al. 2001)	Arosa, Switzerland	O_3	Trajectory		Regional	
(Pochanart, Akimoto et al. 2003)	Mondy, Siberia	O_3, CO	NIES Trajectory			
(Poirot, Wishinski et al. 2001)	Underhill (Vermont), USA	PM	CAPITA Trajectory		PSCF, CPF	PMF, UNMIX
(Poissant 1999)	St. Lawrence River valley, Canada	Mercury	AES Trajectory		PSCF	
(Polissar, Hopke et al. 1999; Polissar, Hopke et al. 2001)	Barrow, Alaska	CN, BC, aerosols	CMDC Trajectory		PSCF	
(Polissar, Hopke et al. 2001)	Underhill (Vermont), USA	Na, Br, Ca, Mg, BC, Sulphate	CAPITA Trajectory		PSCF	
(Real, Law et al. 2007)	Canada, Arctic, Atlantic (aircraft)	O_3, NOx, CO, PAN	UGAMP Trajectory			CiTTyCAT
(Real, Law et al. 2008)	Atlantic Ocean (aircraft)	O_3, CO, NO_{xy}, VOCs, aerosols	FLEXTRA Trajectory			CiTTyCAT
(Real, Orlandi et al. 2010)	Africa and Atlantic (AMMA)	O_3, CO, NO_x	FLEXTRA Trajectory			CiTTyCAT
(Reidmiller, Jaffe et al. 2009)	Mount Bachelor, WA, USA	O_3, CO, Mercury	HYSPLIT Trajectory		Residence Time	GEOS-CHEM
(Riccio, Giunta et al. 2007)	Naples, Italy	O_3, PM	HYSPLIT Trajectory	k-means		PCA
(Rodriguez, Alastuey et al. 2011)	Izaña, Canary islands	Aerosols	HYSPLIT Trajectory		MCAR	
(Rozwadowska, Zielinski et al. 2010)	Hornsund, Svaalbard	Aerosol Optical Thickness	HYSPLIT Trajectory	k-means		
(Ryall, Derwent et al. 2001)	Mace Head, Ireland	CFC	NAME Dispersion			
(Ryall, Derwent et al. 2002)	UK	PM	NAME Dispersion			
(Salvador, Artinano et al. 2008)	Campisabalos, Spain	PM, aerosols	FLEXTRA Trajectory	k-means		

(Salvador, Artinano et al. 2010)	Schauinsland (Germany), Puy de Dôme (France) Sonnblick (Austria)	PM, aerosols	FLEXTRA Trajectory	k-means	RCF	
(Scheifinger and Kaiser 2007)	Austrian stations	O_3, CO, NO_x	FLEXTRA Trajectory		CPF, PSCF, RCF	
(Schichtel, Gebhart et al. 2006)	Big Bend NP, Texas	PM	CAPITA Trajectory		IP, CPF	
(Schmale, Schneider et al. 2011)	Greenland (aircraft)	Aerosols	OFFLINE and LAGRANTO Trajectory FLEXPART Dispersion			EDGAR emissions
(Seibert and Frank 2004)	Stockholm, Sweden	^{137}Cs	FLEXPART Dispersion		Residence Time	
(Sharma, Lavoue et al. 2004)	Alert, Greenland	PM, aerosols	CMC Trajectory	k-means	CPF	
(Sharma, Andrews et al. 2006)	Alert, Barrow, Arctic	PM, aerosols	CMC Trajectory	k-means	CPF	
(Simmonds, Derwent et al. 2004)	Mace Head, Ireland	O_3	NAME Dispersion			
(Simmonds, Seuring et al. 1997)	Mace Head, Ireland	O_3, CO	Trajectory		Regional	
(Solberg, Schmidbauer et al. 1996)	Zeppelin, Svaalbard	O_3	EMEP Trajectory		Regional	
(Solberg, Stordal et al. 1997)	Zeppelin and Norwegian stations	O_3	EMEP Trajectory		Regional	
(Solberg, Hov et al. 2008)	Europe	-	FLEXTRA Trajectory		Residence Time	OSLO CTM EMEP model
(Stohl 1996)	European stations	sulphate	FLEXTRA Trajectory		RCF	
(Stohl, Spichtinger-Rakowsky et al. 2000)	Jungfraujoch, Sonnblick, Zugspitze, Mt. Cimone (Alps)	O_3	FLEXPART Dispersion			FLEXPART tracers
(Stohl, Eckhardt et al. 2002)	Mace Head, Ireland	-	FLEXPART Dispersion	Retroplume clusters		
(Stohl, Cooper et al. 2004)	Zeppelin, Svaalbard	Halocarbons CO_2, O_3, CO, Mercury	FLEXPART Dispersion			FLEXPART tracers
(Stohl, Andrews et al. 2006)	(Aircraft)	Aerosol Optical Depth	FLEXPART Dispersion		Footprint emission sensitivity	
(Stohl, Berg et al. 2007)	Zeppelin, Svaalbard	O_3, CO, aerosol	FLEXPART Dispersion		Footprint emission sensitivity	FLEXPART tracers
(Stohl, Seibert et al. 2009)	Various stations	HFC	FLEXPART Dispersion			Inversion model
(Strong, Whyatt et al. 2010)	14 UK rural sites	O_3	HYSPLIT Trajectory			ELMO
(Tarasova, Senik et al. 2009)	Jungfraujoch (Switzerland), Kislovodsk (Russia)	O_3	LAGRANTO Trajectory			
(Taubman, Hains et al. 2006)	US/ Atlantic Ocean (aircraft)	O_3, CO, SO_2	HYSPLIT Trajectory	Hierarchical	Residence time, CPF	
(Toledano, Cachorro et al. 2009)	El Arenosillo, Spain	Aerosol Optical Depth	Trajectory	k-means	Regional	
(Traub, Fischer et al. 2003)	Mediterranean (aircraft)	O_3, NO_{xy}, CO, CO_2, HCHO, CH_4, VOCs	FLEXTRA Trajectory		Regional	

(Tscherwenka, Seibert et al. 1998)	Snnblick, Austria	SO_2	FLEXTRA Trajectory		Residence time	
(Tuzson, Henne et al. 2011)	Jungfraujoch, Switzerland	CO_2	FLEXPART Dispersion		Residence time	
(Vasconcelos, Kahl et al. 1996)	Grand Canyon NP, USA	CH_3CCl_3	ATAD Trajectory		CPF	
(Virkkula, Aurela et al. 1999)	Sevettijärvi, Finland	Aerosols	TRADOS Trajectory		Regional	
(Walker, Derwent et al. 2009)	Birmingham, UK	CO, O_3, NO_x, VOC	HYSPLIT and BADC Trajectory			PTM
(Wang, Lam et al. 2004)	Hong Kong, China	O_3, CO	HYSPLIT Trajectory	Hierarchical		
(Wang, Hao et al. 2010)	Beijing, China	O_3, CO	HYSPLIT Trajectory			
(Weiss-Penzias, Jaffe et al. 2004)	Cheeka Peak, WA, USA	O_3, CO	HYSPLIT Trajectory		Residence Time	GEOS-CHEM
(Weiss-Penzias, Jaffe et al. 2006)	Mount Bachelor, WA, USA	O_3, CO, Mercury	HYSPLIT Trajectory		Residence Time	
(Wolfe, Thornton et al. 2007)	Mount Bachelor, WA USA	O_3, CO, PAN	HYSPLIT Trajectory		Residence Time	
(Wotawa and Kroger 1999)	central Europe	NO_{xy}	FLEXTRA Trajectory		CPF, RCF	IMPO model
(Wotawa, Kroger et al. 2000)	11 stations in the Alps	O_3	FLEXTRA Trajectory		CPF, RCF	
(Wu, Hu et al. 2009)	Beijing, China	PM, NH_3, acidic gases	HYSPLIT Trajectory		PSCF	
(Xia, Chen et al. 2007)	Beijing, China	Aerosol Optical Depth	HYSPLIT Trajectory	k-means (fuzzy c-means)		
(Xiao, Kang et al. 2010)	Tibetan plateau	PCBs, PAH	HYSPLIT Trajectory		CPF	
(Xu, DuBois et al. 2006)	US National Parks	PM	HYSPLIT Trajectory		Residence Time	
(Yan, Tang et al. 2008)	Shangdianzi, China	Aerosol Optical Depth	Trajectory	k-means		
(Zhu, Huang et al. 2011)	Beijing, China	PM	HYSPLIT Trajectory	k-means	PSCF	