Atmos. Chem. Phys. Discuss., 9, 6207–6245, 2009
www.atmos-chem-phys-discuss.net/9/6207/2009/


Atmospheric Chemistry and Physics Discussions

This discussion paper is/has been under review for the journal *Atmospheric Chemistry and Physics (ACP)*. Please refer to the corresponding final paper in *ACP* if available.

# Source-receptor relationships for airborne measurements of $CO_2$, CO and $O_3$ above Siberia: a cluster-based approach

**J.-D. Paris[1], A. Stohl[2], P. Ciais[1], P. Nédélec[3], B. D. Belan[4], M. Y. Arshinov[4], and M. Ramonet[1]**

[1]Laboratoire des Sciences du Climat et de l'Environnement/IPSL, joint unit CNRS-CEA-UVSQ, L'Orme des Merisiers, 91191 Gif sur Yvette, France
[2]Norwegian Institute for Air Research, Kjeller, Norway
[3]Laboratoire d'Aérologie, Observatoire Midi Pyrénées, CNRS-UPS, Toulouse, France
[4]Zuev Institute of Atmospherics Optics, SB RAS, Tomsk, Russia

Received: 22 January 2009 – Accepted: 19 February 2009 – Published: 9 March 2009

Correspondence to: J.-D. Paris (jean-daniel.paris@lsce.ipsl.fr)

Published by Copernicus Publications on behalf of the European Geosciences Union.





## Abstract

We analysed three intensive campaigns above Siberia resulting in a total of ~70 h of continuous $CO_2$, CO and $O_3$ measurements. The flight route consists of consecutive ascents and descents between Novosibirsk (55° N, 82° E) and Yakutsk (62° N, 129° E). Our data analysis uses clustering of footprints obtained with the Lagrangian particle dispersion model FLEXPART. The model-based technique was found to be able to separate efficiently tracers' concentrations. High CO and $O_3$ concentrations (median values 121 ppb and 54.5 pb respectively) were found in clusters associated with fires in Kazakhstan in September 2006. High correlation (as high as $R^2$=0.68) and robust linear relationships with regression slope between −0.10 and 0.24 ppb ppb$^{-1}$ were found in individual plumes. Summer (August 2007, September 2006) uptake of $CO_2$ was found to be largely (~50% of variance) explained by exposure to boreal and sub-arctic ecosystems, most likely by photosynthesis. This results in an average 5 to 10 ppm gradient in the August 2007 campaign. European emissions seem to contribute to high $O_3$ concentrations above Siberia in altitude were it is also near stratospheric inputs. Large-scale deposition processes reduce $O_3$ in the boreal and sub-arctic BL, resulting in a ~20 ppb gradient. This first attempt of Lagrangian footprint clustering is very promising and could also be advantageously applied to the interpretation of ground based measurements including calculation of tracers' sources and sinks.

## 1 Introduction

Three airborne campaigns were made across Siberia in April 2006, September 2006 and August 2007 respectively. Numerous vertical profiles of CO, $CO_2$, $O_3$ and aerosols were collected, allowing a "tomography" of the Siberian troposphere at different times of the year. After introducing the new data set, we present a clustering analysis of the tracers' concentrations based on partitioning of footprints obtained with a Lagrangian particle dispersion model (LPDM).



ACPD

9, 6207–6245, 2009



Atmospheric transport can occur both at low altitude in the boundary layer and in the free troposphere, where it can be much faster. Export of North-East Asian and North American emissions is associated with strong uplift in the warm conveyor belts of mid-latitude cyclones (e.g. Cooper et al., 2001; Hess and Vukicevic, 2003; Liang et al., 2004; Owen et al., 2006). On the opposite, European emissions have a tendency to remain in the lower troposphere and are most frequently channelled to the Arctic or to Siberia (Wild et al., 2004; Stohl et al., 2002, 2007a; Duncan and Bey, 2004; Law and Stohl, 2007; Liu et al., 2002). Faster zonal advection occurs in winter and early spring in association with the Siberian High (Newell and Evans, 2000; Wild et al., 2004; Liu et al., 2003). In April, European pollutants can contribute significantly to tracers' concentrations in the Asian outflow to the Pacific (Liu et al., 2003; Liang et al., 2004). With respect to ozone production, European export increase $O_3$ concentrations over Siberia by 2 to 4 ppb in spring and 2 to 6 ppb in summer (Wild et al., 2004; Duncan and Bey, 2004).

Pollutant export to or across Siberia, however, has been studied almost exclusively through modelling studies and lacks measurement-based assessment. Pochanart et al. (2003) has conducted ground based measurements supported by transport model analysis, showing that European emissions lead to a 1–4 ppb increase in $O_3$ near Baikal Lake. Eneroth et al. (2003), based on backtrajectories analysis of transport to the West-Siberian site of Zotino have shown that stagnant flow conditions were responsible of respiration build-up leading to high $CO_2$ concentrations (and ruled out European emissions). But both measurements of $CO_2$ and pollutant distribution are still required over Siberia, along with the possibility to relate them to remote emission sources i.e. source-receptor relationships (SRR).

Contrarily to single back-trajectories, the SRR obtained from a LPDM accounts for the atmospheric turbulence and convection (Stohl et al., 1998). Han et al. (2005) compared the result of regional source apportionment using single backtrajectories and backward dispersion. Their backward dispersion model added individual turbulence-related stochastic component to a large number of HYSPLIT backtrajectories. They

demonstrated the better ability of LPDM to identify regional point sources of reactive mercury. Use of LPDM has been proposed (Gerbig et al., 2003; Lin et al., 2003) and eventually demonstrated (Lauvaux et al., 2008) for meso- to regional-scale inversion of $CO_2$ surface fluxes.

Clustering of backtrajectories has been widely used for the analysis of atmospheric composition measurements at fixed observatories and have been found to be an efficient method for separating air masses with different properties (Moody and Galloway, 1988; Dorling et al., 1992; Sirois and Bottenheim, 1995; Eneroth et al., 2003). Moody and Galloway (1988) were the first to attempt clustering of back-trajectories in order to assess the wet deposition of acid compounds over the Bermuda. Clustering of back-trajectories has been used at the continental scale to derive SRRs for $O_3$ concentrations in central Siberia and supported the conclusion of elevated CO in air masses influenced by European emissions (Pochanart et al., 2003). Traub et al. (2003) applied a simple back-trajectories' partitioning technique to the extensive analysis of measurements obtained from an aircraft. No attempt has been made yet, to our knowledge, to use a clustering technique based on LPDM footprints.

Aircraft measurements only provide a snapshot of the atmosphere at a particular time, but they have the potential to explore transport processes and impact across transported plumes (Takegawa et al., 2004; Fehsenfeld et al., 2006; Stohl et al., 2007b; Methven et al., 2006; Lelieveld et al., 2002) or in the outflow of massively emitting regions (Jacob et al., 2003). Here we investigate data from three recent YAK-AEROSIB intensive airborne campaigns that sampled the Siberian troposphere at different times of the year (Paris et al., 2008). The present study includes new data from a campaign that took place in August 2007 that are analysed using a LPDM.

More specifically, we focus on the extent to which clusters of maps of potential emission sensitivity (PES) for receptor positions relate to the air masses chemical composition at these receptors. The PES is expressed here as a residence time. The clusters are used to generalize SRR throughout the aircraft trajectory through the Siberian air shed. We try to answer the following questions: (1) is it possible to identify the con-





tribution to $CO_2$ and CO concentrations of European or other remote anthropogenic sources predicted by models? (2) Is it possible to identify contributions from forest fire or from other types of biomass burning? (3) To what extent can we explain the seasonal variability in $CO_2$, CO and $O_3$ concentrations by variations in atmospheric transport patterns through intensive campaigns? (4) Is there a consistent signal about the regional carbon source/sink distribution emerging through model analysis of $CO_2$ concentrations?

Section 2 describes the experimental details and provides model and statistical tools description. Section 3 describes the variability of $CO_2$, CO and $O_3$ observed above Siberia during the three campaigns. In Sect. 4 we discuss, using cluster analysis, the tracer concentrations observed in selected flights. Section 5 investigates the "seasonal" (inter-campaign) variation of the connection between source regions and trace gas concentrations over Siberia.

## 2 Data and method

### 2.1 Campaigns overview and instruments

Intensive campaigns took place over Siberia between 11 and 14 April 2006 (YAK-1), between 7 and 10 September 2006 (YAK-2) and between 17 and 20 August 2007 (YAK-3). The airborne platform is an Antonov-30 dubbed "Optik-E" chartered by Tomsk Institute of Atmospheric Optics. It was equipped in collaboration between French and Russian laboratories for the measurement of $CO_2$, CO, $O_3$, aerosols and meteorological parameters. The flight route consists in a large, continental-scale loop from Novosibirsk in central Siberia to Yakutsk in Eastern Siberia (Fig. 1) that is done in four days. The flight route is repeated with minor modifications for each campaign. Flights were conducted in all weather conditions. Thick cloud decks were flown over instead of conducting normal ascent and descent through them. Profiles are collected as often as possible as the flight tracks consist mainly of ascents up to 7 km altitude and descents.



ACPD
9, 6207–6245, 2009



During each intensive campaign, the 4-days "loop" is divided into 4 flights (Fig. 1) of up to 8 h (limited by refuelling needs and airports availability). These flights are numbered 1–4 during the April 2006 campaign, 5–8 during the September 2006 and 9–12 during the August 2007 one.

The April campaign was dominated by stagnant flow conditions and low surface temperatures (Paris et al., 2008), with most of the ground covered by snow. In contrast, the September 2006 and August 2007 campaigns were characterized by westerly flow, warmer and highly variable temperatures, and significant frontal activity in the flight area.

Detailed instrument descriptions can be found in Paris et al. (2008); we give here only a short description. $CO_2$ is measured by a modified Non Dispersive Infrared Analyser based on a commercial Li-Cor 6262, with accuracy and precision of 0.15 ppm, obtained by periodical in-flight calibration against WMO-referenced reference gases bracketing atmospheric concentrations, and regulation of pressure, temperature and flow in the cells. CO is measured by IR absorption gas correlation with an accuracy of 5 ppb or 5%. The instrument is based on a commercial infrared absorption correlation gas analyser (Model 48C, TEI Thermo Environment Instruments, USA; Nedelec et al., 2003). $O_3$ is measured by a modified UV commercial fast response ozone analyser (Thermo Instruments Model 49) with a precision of 2 ppb, 2% for an integration time of 4 s.

### 2.2 Atmospheric backward transport model

Atmospheric transport was investigated using the FLEXPART v6.2 Lagrangian particle dispersion model. FLEXPART calculates the trajectories of tracer particles using the mean winds interpolated from the analysis fields plus random motions representing turbulence (Stohl and Thomson, 1999). Results presented here use ECMWF analysis fields, although both ECMWF and GFS (NOAA/NCEP) fields were used for a prior assessment of transport error. For moist convective transport, FLEXPART uses the scheme of (Emanuel, 1999), as described and tested by (Forster et al., 2007). A back-

ward simulation mode is available, described in more detail by Stohl et al. (2005) and Seibert and Frank (2004). Here the backward method is used to analyse transport pathways from potential flux regions to the receptor position. Each simulation consists of 40 000 particles released whenever the aircraft has moved 0.15° in latitude or longitude, or 10 hPa in altitude. This corresponds to about one minute of sampling or to a layer ~100 m thick during ascent or descent. Released particles were followed 10 days backward in time. Gridded PESs in three vertical levels (0–300 m, 300–3000 m, and 3000–50 000 m) are stored at 1°×1° resolution every 24 h. The fraction of stratospheric air is recorded as the percentage of all particles originating from the stratosphere. Results are available at http://zardoz.nilu.no/~andreas/YAK/.

### 2.3 Cluster analysis

Airborne campaigns can deliver detailed information on the atmospheric state at a particular time. To generalize this information and to organize the YAK-AEROSIB data set, a method is sought that groups the data according to common transport properties. Cluster analysis is such an exploratory tool, which sorts multivariate data into groups as dissimilar as possible but whose properties are not known a priori. We seek to investigate to which extent footprints can explain the air mass chemical composition in $CO_2$, CO and $O_3$. Footprints are implemented as 10-days summed spatial distribution of PES (proportional to potential emission sensitivity) obtained from FLEXPART, complemented by 10-days averaged relative contributions from the stratosphere, defined as the region with potential vorticity >2 PVU. To facilitate analysis we have reduced the number of variables from the gridded PES by further summing these PES over large regions of interest for our study (Fig. 2). A vector $\boldsymbol{x}_j = \left[\sum_{d=1}^{10} r_{1,j}, \sum_{d=1}^{10} r_{2,j}, \ldots, \sum_{d=1}^{10} r_{M,j}\right]^T$ of daily PES $r_k$ in box $m$ ($m$=1, 2, . . . , $M$) is associated with each consecutive receptor position $j$. This time series $\boldsymbol{x}_j$ will constitute our set of $j$ realizations of $M$ variables to cluster. Defining $j$ over a single campaign improves the separation capability of the



ACPD
9, 6207–6245, 2009



algorithm, whereas defining $j$ over the three campaigns allows inter-campaign comparison.

The $M$ boxes are identified a priori as regions with specific sources or sink for the species relevant to this study. Such identification is based either on case studies from the April and September 2006 campaigns in Paris et al. (2008), or on the questions to be addressed with this dataset, such as the transport of European emissions to Siberia. The regions identified are:

- Europe: European pollutant emissions and outflow;
- Western Russia: European Russian pollutant emissions and outflow;
- NE China, Japan and Koreas: extratropical cyclones can lift NE Asian pollutant over central or eastern Siberia (Paris et al., 2008);
- Central Asia: Central Asia has a different ecosystem and low anthropogenic sources but it is a potential source of biomass burning (e.g. van der Werf et al., 2006);
- Arctic: Arctic air from beyond the Arctic front can be observed over Siberia during cold air outbreaks;
- Local region: covers the area directly flown over by our aircraft;
- Stratosphere: an extra variable is added which averages the 10-day proportion of stratospheric particles.

The boxes' geographical extent is the result of a trade-off between maximizing differences between transport from the various boxes, and an insufficient number of particles if boxes are chosen to be too small. There are large regions not covered by any box. Theoretically, this could mean that a footprint is located entirely outside all of the boxes. However, the footprints normally cover quite large regions, so they always overlap with





one or more of the boxes. By separating the boxes from each other, differences between the various footprints can be maximized by the clustering algorithm. Choosing boxes that are contiguous make the separation more sensitive to horizontal transport error, and associated PES less contrasted.

More remote regions are less connected (few or no particles reaching the region) by transport to the measurements. As a result, remote regions' average footprint distributions are skewed toward 0. To account for this skewness and different region sizes, normalization is applied to the regions' average footprint time series $\boldsymbol{x}$ according to $\boldsymbol{x}_N = (\boldsymbol{x} - p_{0.05}) / \bar{\boldsymbol{x}}$ where $p_{0.05}$ is the 5th percentile of $\boldsymbol{x}$ and was found to optimally separate clusters. The robustness of data reduction into regions and data normalization was tested by running the clustering algorithm for varying region sizes, locations and number of regions (bootstrapping), and various normalization functions.

The K-means algorithm implemented in the MATLAB software's Statistics toolbox is used for clustering. K-means is a classic partitioning (non-hierarchical) algorithm (see e.g. Wilks, 2006) which attempts to separate observations into a fixed number of groups. (1) It defines K-centroids with initial random position vector of size $M$=7 (number of regions); then (2) associates each point $\boldsymbol{x}_j$ to the nearest (in Euclidian metrics) cluster centroid, (3) moves the centroid to the centre of the cluster and (4) repeats steps 2 and 3 until convergence is achieved. In the process, any empty cluster is discarded. In order to maintain the number of clusters, a point having maximum distance to its centroid is singled out and declared centroid of a new cluster.

Figure 3 shows the "silhouette" index ($S \in [-1, 1]$), a measure of the separation capability of the clustering, as a function of the chosen number of clusters. The S index is defined as:

$$S = \frac{1}{N} \sum_{i=1}^{N} \frac{d_{\text{nearest}} - d_{\text{centroid}}}{\max(d_{\text{nearest}}, d_{\text{centroid}})} \quad (1)$$

with N number of points, and where point $i$ has a distance of to the nearest cluster $d_{\text{nearest}}$ and distance to its associated centroid $d_{\text{centroid}}$. The S index informs about how





close a point is to its cluster's centroid, relative to the other nearest cluster centroid (Matlab R2006b documentation, Statistics toolbox, http://www.mathworks.com/access/helpdesk/help/toolbox/stats/silhouette.html). The minimum number $k$ of clusters was found to be optimally set to 6 in the global clustering option, and 4 in the campaign by campaign option (Fig. 3).

As the clustering algorithm is sensitive to a priori (random) position of the centroids, the process is repeated 20 times and the result optimizing the separation between centroids is retained. This also ensures a very high confidence in reproducibility of cluster centroids determination. The normalization and the metrics were also found to have a strong impact on the result. In this respect, cluster analysis is not an objective classification technique as the criteria set for the clustering have to be determined subjectively.

## 3 Campaign observations

### 3.1 $CO_2$ measurements over Siberia

Figure 4a, d and g shows the observed $CO_2$ concentrations averaged for each of the 4 flights of each campaign. Among the seasons chosen for our surveys, the minimum BL $CO_2$ concentrations have been observed in August, namely 366 ppm on average in the lowest 1 km during flight 12 (Fig. 4g). Flight 12 occurred mostly over a forested region, with sparse agricultural landscapes and industrial centres (Kemerovo, Novosibirsk) nearby. This $CO_2$ concentration minimum would be even lower compared to the other campaigns if we would correct the data for the trend in atmospheric $CO_2$ due to global anthropogenic emissions of $\sim$1.4 ppm $yr^{-1}$, since this campaign was carried out in August 2007, and the others in April and September 2006. This minimum $CO_2$ concentration is coincident with a very high BL top (up to 4 km a.s.l., about 3.5 km a.g.l.), as deduced from humidity and $CO_2$ gradients. In the free troposphere, $CO_2$ was 377–379 ppm in August 2007, comparable (up to the interannual trend) to the 377–378 ppm

in September 2006 but much lower than April 2006 (387–388 ppm), reflecting the hemispheric $CO_2$ seasonal cycle. $CO_2$ in April 2006 was as high as 392 ppm in polluted filaments encountered at 5–6 km altitude (Paris et al., 2008).

### 3.2 CO and $O_3$ measurements over Siberia

Figure 4b shows the average profiles of CO concentrations for each flight during campaigns 1. Throughout the YAK-1 campaign the vertical average CO concentration was 175 ppb. Elevated CO concentrations were repeatedly observed and a warm conveyor belt uplifted a polluted air mass with CO up to 220 ppb from NE China above Siberia (Paris et al., 2008). During campaign 1, the $O_3$ profile shows a rather homogeneous gradient from ~60 ppb at 6 km to ~44 ppb at 1 km (Fig. 4c). Occasionally, significantly lower CO and higher $O_3$ values up to 85 ppb, have been observed near flight ceiling.

During YAK-2 the average CO concentration was about 100 ppb, with a large variability (Fig. 4e). During Flight 5 the aircraft crossed a warm front exhibiting strong tracer gradients across its surface with CO concentration up to 150 ppb immediately above the front, between 2 and 3 km altitude.

High ozone values were found at the highest aircraft altitude during YAK-3 campaign, up to 90 ppb (Fig. 4i). The $O_3$ vertical gradient between upper and lower altitudes was therefore the highest of the three campaigns. Except for local emissions in the lower troposphere, the YAK-3 campaign has shown low CO values (background 100±5 ppb) with little variation, although satellite fire detection has shown numerous fires in temperate and boreal Eurasia. $O_3$ exhibits a nearly linear gradient up to 7 km altitude were the concentrations reach 90 ppb.

## 4 Cluster-based SRR relationships: case studies

In this section we apply and discuss the clustering results to examine the measurements from one flight from each of the three campaigns (Flight 1 on 11 April 2006,





Flight 5 on 7 September 2006 and Flight 9 on 17 August 2007). All three flights are going from Novosibirsk to Myrni and correspond to the green route in Fig. 1). The clustering is done on a single campaign-basis. The clustering being independent from the observed concentrations, statistical separation of the concentrations across different clusters is interpreted as a validation of the SRR.

### 4.1 Flight 1: pollution from North-Eastern China

Figure 5 shows the average footprint for each cluster in logarithmic color scale. Footprints classified as Cluster C (n=255; Fig. 5c) are on average dominantly located over South East and East Siberia but have less average sensitivity to potential surface emissions than the other clusters. A marginal number of footprints are associated to Cluster A (n=5; Fig. 5a) with sensitivity to European emissions. Cluster B (n=68; Fig. 5b) gathers a significant amount of data having resided over Northern Siberia and the Arctic. Footprints in Cluster D (n=19; Fig. 5d) are exposed to surface exchange in Southern Siberia and the Kazakhstan. The four well-differentiated footprint maps illustrate the validity of cluster-based partitioning.

Figure 6 shows the distribution (quartiles and median) of $CO_2$, CO, $O_3$ and water vapour measurements associated to each cluster. Associated altitude is given in Fig. 6e, and the four (A–D) clusters' centroid (in std dev-normalized coordinates) position vector is given in Fig. 6f. Each cluster's centroid position vector can be related to the cluster average footprint map in Fig. 5.

The data associated to Cluster D (sensitive to NE China emissions) are encountered between 5 and 6.5 km altitude (Fig. 6e) and have a median CO concentration of 180 ppb (Fig. 6b), well differentiated from other clusters. On the opposite, $CO_2$ values (median 388.6 ppm; Fig. 6a) are not distinguishably higher. This cluster identifies the transport of North Eastern China emissions to Eastern Siberia on 12 April 2006 (Paris et al., 2008). The ability of the LPDM to resolve thin layers associated to synoptic features is a useful advantage over Eulerian models.

Cluster C is ubiquitous throughout Flight 1 and has a higher stratospheric influence



ACPD
9, 6207–6245, 2009



reflected in its highest $O_3$ concentrations (up to 85 ppb). Cluster B has high CO concentrations (median 174 ppb) corresponding to local emission sensitivity of the lower troposphere.

### 4.2 Flight 5: fires and $CO_2$ uptake

Figure 7 shows the average footprint for each cluster for the 7 September 2006 flight. As in the previous section, clustering was performed on the whole campaign (here, the September 2006 campaign). Data within Cluster B are the most ubiquitous (n=148, Fig. 7b) and were mostly associated to FT zonal flow, resulting in a generally weak footprint mainly over European Russia and the Black Sea. This cluster has the highest stratospheric signature of the dataset (Fig. 8f). Cluster D is also ubiquitous in this flight (n=132; Fig. 7d) but has a high footprint density region located over Kazakhstan, between Caspian and Aral Seas. A large number of fires detected by ATSR fire count in this region, probably of agricultural origin, were injected in the FT as the region was swept by a front (Paris et al., 2008). Cluster A (n=47, Fig. 7a) is clearly associated to an Arctic air mass channelled southward in the BL. Cluster C (n=14, Fig. 7c) reflects advection of air masses in contact with potential European emissions.

The European (Cluster C) air mass has a median CO concentration of 109.5 ppb (Fig. 8b) at a median altitude of 5 km, to be compared to 105.0 ppb for the more ubiquitous Cluster B. Although this is only a single case, it shows that European pollution does not always follow the low level advection pathway (Wild et al., 2003; Stohl et al., 2002; Duncan and Bey, 2004). Other characteristic of this air mass include relatively high $CO_2$ (median 376.3 ppm) and an $O_3$ median concentration of 53 ppb.

Figure 8 shows that Cluster D has the most elevated CO concentration, up to 158 ppb (median 121.0 ppb, third quartile 132.2 ppb). It is also associated to high $O_3$ (median 54.5 ppb, inter-quartile range (IQR) 51.2–57.0 ppb) but not to high $CO_2$ (median 375.0 ppm). High CO and $O_3$ are likely to reflect the sensitivity of Cluster D (Fig. 7d) to the biomass burning detected in the Caspian region (especially in Northern Kazakhstan at ~50° N, between 50 and 70° E). The ratio of $O_3$ to CO enhancements reflects



ACPD
9, 6207–6245, 2009



a balance between ozone production and loss in the plume (Pfister et al., 2006; Val Martin et al., 2006). A plume with the highest CO concentration (144.3±14.0, $O_3$ concentration 58.9±3.8 ppb) exhibits a high CO-$O_3$ correlation ($R^2$=0.68, n=90 10-s samples). The robust linear regression slope is 0.24 mol mol$^{-1}$, which is usually associated to ozone production within the plume from combustion-generated precursors like CO. The regression slope in another plume with high CO (134.8±9.1 ppb) was found to be negative (−0.10 mol mol$^{-1}$). Both of these plumes are encountered at 2–2.5 km altitude. The $O_3$-CO correlation coefficients are comparable to those found in the literature for boreal forest fire plumes (range 0.05–0.30 mol mol$^{-1}$; see Val Martin et al., 2006; Real et al., 2007 and references therein).

Both of these plumes have a high CO-$CO_2$ correlation (respectively $R^2$=0.80 and 0.78) with negative CO-$CO_2$ regression slopes of −0.06 and −0.09 mol mol$^{-1}$ respectively. These negative slopes reflect competing loss processes such as CO oxidation within the plume and $CO_2$ assimilation by intact vegetation.

Cluster A gathers observations affected by Arctic air (Fig. 8f) with low $CO_2$, CO and $O_3$ concentrations values ($CO_2$ median 373.06 ppm with IQR 370.81–373.94 ppm, CO median 103.0 and $O_3$ median 34.5 ppb). It is associated to a low altitude (1 km) within the BL and with a high water vapour mixing ratio (Fig. 8d). Advection of Arctic air in the campaign domain sampled during September 2006 occurred in the boundary layer. FLEXPART results indicate that total PES over the Arctic was 5 days or more in the last 10 days for the majority of data in Cluster A. The low CO and low $CO_2$ concentrations are suggestive of a strong isolation from sources in the recent air mass history. During the five days of meridian transport from the Arctic, exposure to uptake by high latitude vegetation with permanent daylight period and to surface deposition must also have contributed to low $CO_2$ and low $O_3$ values respectively.

### 4.3 Flight 9: $CO_2$ uptake, forest fire and stratospheric input

Figure 9 shows the average footprint for each of the 4 clusters for the 14 August 2007 flight. Cluster B indicates the dominant footprint pattern for this flight (n=194; Fig. 9b). It has a strong stratospheric component (Fig. 10f) and the weak footprint density is spread over Western Siberia. Cluster C (n=50; Fig. 9c) has a high sensitivity north of Lake Baikal. Cluster A (n=26; Fig. 9a) is centred over North European Russia and Scandinavia while Cluster D (n=26; Fig. 9d) reflects air of Arctic origin, comparable to Cluster A in the September 2006 case.

Figure 10b–c shows that Cluster C has the highest CO median (122.0 ppb) with a wide IQR within 110.0–190.9 ppb. Such high CO values point to regional fire influence, as seen on ATSR fire atlas for the period of the campaign. The altitude range is very large (1–6 km). The highest CO concentrations (300 ppb) were observed in two layers in the lower FT (Fig. 11).

Comparing $CO_2$ concentrations in the low-altitude Arctic clusters of September 2006 (Cluster A) and August 2007 (Cluster D), the latter (median 369.31 ppm) is lower by ~4 ppm despite anthropogenic inter-annual increase of ~1.5 ppm $yr^{-1}$. The same is found for $O_3$ with 25.0 ppb in August 2007 vs. 34.5 ppb in September 2006. In Siberia FT $CO_2$ concentrations are minimum in August (Ramonet et al., 2002), and September concentrations increase by about 2 ppm in the FT and 5 ppm in the BL relative to August.

Cluster B shows the highest stratospheric component for this flight. The highest $O_3$ values (up to 91 ppb) associated to it are consistent with the "stratospheric" classification. High $O_3$ concentrations (up to 90 ppb; Fig. 11) are associated to >50% of freshly (2 days) exported stratospheric particles.

## 5 Statistical relation between source regions and species across campaigns

In order to compare between campaigns and generalize the above case studies we resort to pair-wise correlations statistics associating individual trace gases and each region as defined in Fig. 2.

### 5.1 Variations of $CO_2$ associated to source regions

Figure 12 shows the influence of simulated surface influence on $CO_2$ measurements with a focus on the Arctic and Siberian boxes. Linear regression lines are fitted to the data and results are reported in Table 1, as well as correlation coefficients. Pearson correlation coefficients reported in Table 1 were calculated after excluding data associated to residence time<100 s grid$^{-1}$. Since $CO_2$ fluxes are highly variable at the sub-day timescale, whereas simulated exposure to fluxes is averaged over 10 days, the relations between $CO_2$ concentrations and PES over source or sink regions are not expected to be strongly linear. The Kendall $\tau$ parameter equally reported in Table 1 is a more robust assessment of the strength of such a non-linear relationship (Wilks, 2006).

Figure 12 shows that $CO_2$ are significantly ($p$<0.01) correlated to and decrease with PES over local ecosystems both in September 2006 ($R$=−0.20, $\tau$=−0.37, see Table 1) and August 2007 ($R$=−0.41, $\tau$=−0.43). Robust linear regression slopes are $-1.0\,10^{-4}$ ppm s$^{-1}$ and $-3.9\,10^{-4}$ ppm s$^{-1}$ respectively. Similarly, for the Arctic box in summer, significant correlation is observed between $CO_2$ concentrations and the PES (in September 2006, $R$=−0.35 and $\tau$=−0.29; in August 2007, $R$=−0.37 and $\tau$=−0.40). As discussed in Sect. 4, high Arctic signal is typically associated to advection in the BL, and therefore integrates between 1 and 10 days of exposure to uptake by Siberian sub-arctic ecosystems. Together, 10-day cumulated exposure to local and Arctic fluxes in August 2007 explain ∼50% of the total $CO_2$ variance, where boreal and sub-arctic ecosystem uptake is at its annual maximum.

In August 2007 the North Eastern China region seem to has a positive contribution



ACPD
9, 6207–6245, 2009



to $CO_2$ concentrations with a significant correlation ($R$=0.54, slope=5.0 $10^{-2}$ ppm $s^{-1}$, $\tau$=0.14). In the clustering analysis has produced no cluster with dominant NE China signal (Fig. 10). These two diverging results suggest that NE Chinese $CO_2$ emissions enhanced August $CO_2$ concentration over Siberia but without a well-defined filamentary structure. In April 2006 both a highly polluted cluster of dominant NE China footprint, and a consistent positive relationship ($\tau$=0.31) between $CO_2$ and NE China PES were found.

### 5.2 Variations of CO and $O_3$ associated to source regions

Tables 2 and 3 show the pairwise statistics associating possible source regions and CO and $O_3$ concentrations respectively. CO budget over Siberia seems to include anthropogenic sources (NE China in April 2006: $\tau$=0.32, as well as local sources in all campaigns, see Table 2) and Western Kazakhstan sources in September 2006 ($\tau$=0.29) and August 2007 ($\tau$=0.26), where it is most likely attributed to biomass burning. A regression slope of 3.6 $10^{-3}$ ppb $s^{-1}$ is found for both periods. Our calculations fail to identify $O_3$ production in these fire plumes (Table 3) in relation to the Western Kazakhstan source region.

$O_3$ statistics exhibit good correlations with most regions excepted in September 2006. As can be expected the stratosphere is a significant contributor to the $O_3$ concentration ($\tau$ ranges between 0.41 and 0.53) in all campaigns with linear regression slopes of 0.1 in April and 0.58 in September. Paralleling $CO_2$ patterns, exposure to ground influence in the Siberian BL consistently reduce $O_3$ concentrations in all campaigns. Linear correlation is stronger ($R$ between −0.76 and −0.45, see Table 3) and the slopes are steeper by almost one order of magnitude in summer months. Influence of Europe (Western Europe and European Russia) seems to result in overall higher $O_3$ concentrations, confirming the case studies of Sect. 4. This influence is clear ($p$<0.01) for Western Europe in April ($R$=0.35) and August ($R$=0.34), and for Western Russia in April only ($R$=0.24).

## 6 Discussion and conclusion

We analysed three intensive campaigns above Siberia resulting in a total of ~70 h of $CO_2$, CO and $O_3$ measurements. The sampling strategy was to perform consecutive ascents and descents to obtain a tomography of the atmosphere above Siberia. Our data analysis was based on a novel application of clustering to Lagrangian particle dispersion model footprint. The technique was found to be able to separate tracers' concentrations although it was based solely on atmospheric transport modelling. CO concentrations of ~109 ppb were observed in clusters associated to a dominant European emissions sensitivity in September 2006 and ~105 ppb in August 2007. Dominant NE China PES was associated to concentrations of 180 ppb in April 2006. High CO and $O_3$ concentrations (median values 121 ppb and 54.5 pb respectively) were found in clusters associated with fires in Kazakhstan in September 2006. High correlation ($R^2$=0.68) and robust linear relationships with regression slopes ranging between −0.10–0.24 ppb ppb$^{-1}$ where found in individual plumes.

Summer (August 2007, September 2006) uptake of $CO_2$ was found to be largely (~30%) explained by exposure to boreal and sub-arctic ecosystems, most likely by photosynthesis. It is likely that permanent sunlight exposure of vegetation above 66° N contributes to a more constant uptake in this region, therefore strongly depleting $CO_2$ in transiting air masses.

Discernable influence of air exposed to European emissions was observed toward higher $O_3$ concentrations, besides large removal in the Siberian BL and constant stratospheric contribution.

Airborne measurements across Siberia during three YAK campaigns have been capable to detect concentrations originating from sources located in remote regions, thus validating the sampling strategy. The cluster-based method proposed here was successful at separating air masses with different chemical compositions, although the classification scheme is based entirely on simulated transport properties and eventually independently matched against air chemical composition measurements. Single-flight





cluster analysis exhibited a variety of air masses origin, related to a large variability in $CO_2$, CO and $O_3$. The technique used here, although it has sufficient resolving power, would benefit from more refinements. The cluster-based method should be able to ensure comparability between different campaigns at different locations and/or periods (given no strong concentrations trend between campaigns). Another possible application of this technique to surface observatories or regular aircraft profiles is to contribute to the definition of background conditions and to provide a systematic classification of air masses origin.

*Acknowledgements.* The measurement campaigns were funded under the project GDRE YAK-AEROSIB by the CNRS (France), the French Ministry of Foreign Affairs, CEA (France), RAS (Russia) and RFBR (Russia). A. S. contribution was supported by POLARCAT Norway. J. L. Teffo, G. Golytsin and I. G. Granberg are acknowledged for their support to the project.

The publication of this article is financed by CNRS-INSU.

## References

Cooper, O. R., Moody, J. L., Parrish, D. D., Trainer, M., Ryerson, T. B., Holloway, J. S., Hubler, G., Fehsenfeld, F. C., Oltmans, S. J., and Evans, M. J.: Trace gas signatures of the airstreams within North Atlantic cyclones: Case studies from the North Atlantic Regional Experiment (NARE '97) aircraft intensive, J. Geophys. Res., 106(D6), 5437–5456, 2001.

Dorling, S. R., Davies, T. D., and Pierce, C. E.: Cluster-Analysis – a Technique for Estimating the Synoptic Meteorological Controls on Air and Precipitation Chemistry – Method and Applications, Atmos. Environ., 26(14), 2575–2581, 1992.





Duncan, B. N. and Bey, I.: A modeling study of the export pathways of pollution from Europe: Seasonal and interannual variations (1987–1997), J. Geophys. Res., 109, D08301, doi:10.1029/2003JD004079, 2004.

Emanuel, K. A. and Zivkovic-Rothman, M.: Development and evaluation of a convection scheme for use in climate models, J. Atmos. Sci., 56(11), 1766–1782, 1999.

Eneroth, K., Kjellstrom, E., and Holmen, K.: Interannual and seasonal variations in transport to a measuring site in western Siberia and their impact on the observed atmospheric $CO_2$ mixing ratio, J. Geophys. Res., 108(D21), 4660, doi:10.1029/2002JD002730, 2003.

Fehsenfeld, F. C., Ancellet, G., Bates, T. S., Goldstein, A. H., Hardesty, R. M., Honrath, R., Law, K. S., Lewis, A. C., Leaitch, R., McKeen, S., Meagher, J., Parrish, D. D., Pszenny, A. A. P., Russell, P. B., Schlager, H., Seinfeld, J., Talbot, R., and Zbinden., R.: International Consortium for Atmospheric Research on Transport and Transformation (ICARTT): North America to Europe – Overview of the 2004 summer field study. J. Geophys. Res., 111, D23S01, doi:10.1029/2006JD007829, 2006.

Forster, C., Stohl, A., and Seibert, P.: Parameterization of convective transport in a Lagrangian particle dispersion model and its evaluation. J. Appl. Meteorol. Clim., 46(4), 403-422, 2007.

Gerbig, C., Lin, J. C., Wofsy, S. C., Daube, B. C., Andrews, A. E., Stephens, B. B., Bakwin, P. S., and Grainger, C. A.: Toward constraining regional-scale fluxes of $CO_2$ with atmospheric observations over a continent: 2. Analysis of COBRA data using a receptor-oriented framework, J. Geophys. Res., 108(D24), 4757, doi:10.1029/2003JD003770, 2003.

Han, Y. J., Holsen, T. A., Hopke, P. K., and Yi, S. M.: Comparison between back-trajectory based modeling and Lagrangian backward dispersion modeling for locating sources of reactive gaseous mercury, Environ. Sci. Technol., 39(10), 3887–3887, 2005.

Hess, P. G. and Vukicevic, T.: Intercontinental transport, chemical transformations, and baroclinic systems, J. Geophys. Res., 108(D12), 4354, doi:10.1029/2002JD002798, 2003.

Jacob, D. J., Crawford, J. H., Kleb, M. M., Connors, V. S., Bendura, R. J., Raper, J. L., Sachse, G. W., Gille, J. C., Emmons, L., and Heald, C. L.: Transport and Chemical Evolution over the Pacific (TRACE-P) aircraft mission: Design, execution, and first results, J. Geophys. Res., 108(D20), 1–19, 2003.

Lauvaux, T., Uliasz, M., Sarrat, C., Chevallier, F., Bousquet, P., Lac, C., Davis, K. J., Ciais, P., Denning, A. S., and Rayner, P. J.: Mesoscale inversion: first results from the CERES campaign with synthetic data, Atmos. Chem. Phys., 8, 3459–3471, 2008, http://www.atmos-chem-phys.net/8/3459/2008/.





Law, K. S. and Stohl, A.: Arctic air pollution: Origins and impacts, Science, 315(5818), 1537–1540, 2007.

Lelieveld, J., Berresheim, H., Borrmann, S., Crutzen, P. J., Dentener, F. J., Fischer, H., Feichter, J., Flatau, P. J., Heland, J., Holzinger, R., Korrmann, R., Lawrence, M. G., Levin, Z., Markowicz, K. M., Mihalopoulos, N., Minikin, A., Ramanathan, V., de Reus, M., Roelofs, G. J., Scheeren, H. A., Sciare, J., Schlager, H., Schultz, M., Siegmund, P., Steil, B., Stephanou, E. G., Stier, P., Traub, M., Warneke, C., Williams, J., and Ziereis, H.: Global air pollution crossroads over the Mediterranean, Science, 298(5594), 794–799, 2002.

Liang, Q., Jaegle, L., Jaffe, D. A., Weiss-Penzias, P., Heckman, A., and Snow, J. A.: Long-range transport of Asian pollution to the northeast Pacific: Seasonal variations and transport pathways of carbon monoxide, J. Geophys. Res., 109, D23S07, doi:10.1029/2003JD004402, 2004.

Lin, J. C., Gerbig, C., Wofsy, S. C., Andrews, A. E., Daube, B. C., Davis, K. J., and Grainger, C. A.: A near-field tool for simulating the upstream influence of atmospheric observations: The Stochastic Time-Inverted Lagrangian Transport (STILT) model, J. Geophys. Res., 108(D16), 4493, doi:10.1029/2002JD003161, 2003.

Liu, H., Jacob, D. J., Chan, L. Y., Oltmans, S. J., Bey, I., Yantosca, R. M., Harris, J. M., Duncan, B. N., and Martin, R. V.: Sources of tropospheric ozone along the Asian Pacific Rim: An analysis of ozonesonde observations, J. Geophys. Res., 107(D21), 4573, doi:10.1029/2001JD002005, 2002.

Liu, H. Y., Jacob, D. J., Bey, I., Yantosca, R. M., Duncan, B. N., and Sachse, G. W.: Transport pathways for Asian pollution outflow over the Pacific: Interannual and seasonal variations, J. Geophys. Res., 108(D20), 8786, doi:10.1029/2002JD003102, 2003.

Methven, J., Arnold, S. R., Stohl, A., Evans, M. J., Avery, M., Law, K., Lewis, A. C., Monks, P. S., Parrish, D. D., Reeves, C. E., Schlager, H., Atlas, E., Blake, D. R., Coe, H., Crosier, J., Flocke, F. M., Holloway, J. S., Hopkins, J. R., McQuaid, J., Purvis, R., Rappengluck, B., Singh, H. B., Watson, N. M., Whalley, L. K., and Williams, P. I.: 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.-Atmos., 111(D23), D23S62, doi:10.1029/2006JD007540, 2006.

Moody, J. L. and Galloway, J. N.: Quantifying the relationship between atmospheric transport and the chemical composition of precipitation on Bermuda, Tellus B, 40(5), 463–479, 1988.





Nedelec, P., Cammas, J.-P., Thouret, V., Athier, G., Cousin, J.-M., Legrand, C., Abonnel, C., Lecoeur, F., Cayez, G., and Marizy, C.: An improved infrared carbon monoxide analyser for routine measurements aboard commercial Airbus aircraft: technical validation and first scientific results of the MOZAIC III programme, Atmos. Chem. Phys., 3, 1551–1564, 2003, http://www.atmos-chem-phys.net/3/1551/2003/.

Newell, R. E. and Evans, M. J.: Seasonal changes in pollutant transport to the North Pacific: the relative importance of Asian and European sources, Geophys. Res. Lett., 27(16), 2509–2512, 2000.

Owen, R. C., Cooper, O. R., Stohl, A., and Honrath, R. E.: An analysis of the mechanisms of North American pollutant transport to the central North Atlantic lower free troposphere, J. Geophys. Res., 111, D23S58, doi:10.1029/2006JD007062, 2006.

Paris, J.-D., Ciais, P., Nédélec, P., Ramonet, M., Belan, B. D., M. Y., Arshinov, Golitsyn, G. S., Granberg, I., Stohl, A., Cayez, G., Athier, G., Boumard, F., and Cousin, J.-M.: The YAK-AEROSIB transcontinental aircraft campaigns: new insights on the transport of $CO_2$, CO and $O_3$ across Siberia, Tellus B, 60(4), 551–568, 2008.

Pfister, G. G., Emmons, L. K., Hess, P. G., Honrath, R., Lamarque, J. F., Martin, M. V., Owen, R. C., Avery, M. A., Browell, E. V., Holloway, J. S., Nedelec, P., Purvis, R., Ryerson, T. B., Sachse, G. W., and Schlager, H.: Ozone production from the 2004 North American boreal fires, J. Geophys. Res., 111, D24S07, doi:10.1029/2006JD007695, 2006.

Pochanart, P., Akimoto, H., Kajii, Y., Potemkin, V. M., and Khodzher, T. V.: Regional background ozone and carbon monoxide variations in remote Siberia/East Asia, J. Geophys. Res., 108(D1), 4028, doi:10.1029/2001JD001412, 2003.

Ramonet, M., Ciais, P., Nepomniachii, I., Sidorov, K., Neubert, R. E. M., Langendorfer, U., Picard, D., Kazan, V., Biraud, S., Gusti, M., Kolle, O., Schulze, E. D., and Lloyd, J.: Three years of aircraft-based trace gas measurements over the Fyodorovskoye southern taiga forest, 300 km north-west of Moscow, Tellus B, 54(5), 713–734, 2002.

Real, E., Law, K. S., Weinzierl, B., Fiebig, M., Petzold, A., Wild, O., Methven, J., Arnold, S., Stohl, A., Huntrieser, H., Roiger, A., Schlager, H., Stewart, D., Avery, M., Sachse, G., Browell, E., Ferrare, R., and Blake, D.: Processes influencing ozone levels in Alaskan forest fire plumes during long-range transport over the North Atlantic, J. Geophys. Res., 112, D10S41, doi:10.1029/2006JD007576, 2007.





Seibert, P. and Frank, A.: Source-receptor matrix calculation with a Lagrangian particle dispersion model in backward mode, Atmos. Chem. Phys., 4, 51–63, 2004, http://www.atmos-chem-phys.net/4/51/2004/.

Sirois, A. and Bottenheim, J. W.: Use of Backward Trajectories to Interpret the 5-Year Record of Pan and O-3 Ambient Air Concentrations at Kejimkujik National-Park, Nova-Scotia. J. Geophys. Res., 100(D2), 2867–2881, 1995.

Stohl, A., Forster, C., Frank, A., Seibert, P., and Wotawa, G.: Technical note: The Lagrangian particle dispersion model FLEXPART version 6.2, Atmos. Chem. Phys., 5, 2461–2474, 2005, http://www.atmos-chem-phys.net/5/2461/2005/.

Stohl, A., and Thomson, D. J.: A density correction for Lagrangian particle dispersion models, Bound.-Lay. Meteorol., 90(1), 155–167, 1999.

Stohl, A., Eckhardt, S., Forster, C., James, P., and Spichtinger, N.: On the pathways and timescales of intercontinental air pollution transport, J. Geophys. Res., 107(D23), 4684, doi:10.1029/2001JD001396, 2002.

Stohl, A., Berg, T., Burkhart, J. F., Fjæraa, A. M., Forster, C., Herber, A., Hov, Ø., Lunder, C., McMillan, W. W., Oltmans, S., Shiobara, M., Simpson, D., Solberg, S., Stebel, K., Ström, J., Tørseth, K., Treffeisen, R., Virkkunen, K., and Yttri, K. E.: 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, 511–534, 2007a, http://www.atmos-chem-phys.net/7/511/2007/.

Stohl, A., Forster, C., Huntrieser, H., Mannstein, H., McMillan, W. W., Petzold, A., Schlager, H., and Weinzierl, B.: Aircraft measurements over Europe of an air pollution plume from Southeast Asia - aerosol and chemical characterization, Atmos. Chem. Phys., 7, 913–937, 2007b, http://www.atmos-chem-phys.net/7/913/2007/.

Stohl, A., Hittenberger, M., and Wotawa, G.: Validation of the Lagrangian particle dispersion model FLEXPART against large-scale tracer experiment data, Atmos. Environ., 32(24), 4245–4264, 1998.

Takegawa, N., Kondo, Y., Koike, M., Chen, G., Machida, T., Watai, T., Blake, D. R., Streets, D. G., Woo, J. H., Carmichael, G. R., Kita, K., Miyazaki, Y., Shirai, T., Liley, J. B., and Ogawa, T.: Removal of $NO_x$ and $NO_y$ in Asian outflow plumes: Aircraft measurements over the western Pacific in January 2002, J. Geophys. Res., 109, D23S04, doi:10.1029/2004JD004866, 2004.





Traub, M., Fischer, H., de Reus, M., Kormann, R., Heland, H., Ziereis, H., Schlager, H., Holzinger, R., Williams, J., Warneke, C., de Gouw, J., and Lelieveld, J.: Chemical characteristics assigned to trajectory clusters during the MINOS campaign, Atmos. Chem. Phys., 3, 459–468, 2003, http://www.atmos-chem-phys.net/3/459/2003/.

Val Martin, M., Honrath, R. E., Owen, R. C., Pfister, G., Fialho, P., and Barata, F.: Significant enhancements of nitrogen oxides, black carbon, and ozone in the North Atlantic lower free troposphere resulting from North American boreal wildfires, J. Geophys. Res., 111, D23S60, doi:10.1029/2006JD007530, 2006.

van der Werf, G. R., Randerson, J. T., Giglio, L., Collatz, G. J., Kasibhatla, P. S., and Arellano Jr., A. F.: Interannual variability in global biomass burning emissions from 1997 to 2004, Atmos. Chem. Phys., 6, 3423–3441, 2006, http://www.atmos-chem-phys.net/6/3423/2006/.

Wild, O., Pochanart, P., and Akimoto, H.: Trans-Eurasian transport of ozone and its precursors, J. Geophys. Res., 109, D11302, doi:10.1029/2003JD004501, 2004.

Wilks, D. S.: Statistical methods in the atmospheric sciences, Academic Press, Amsterdam, Boston, 627 pp., 2006.





**Table 1.** Region-concentration pairwise statistics for $CO_2$.

| | April 2006 | | | September 2006 | | | August 2007 | | |
|---|---|---|---|---|---|---|---|---|---|
| | $\tau^a$ | $R^b$ | $s^c$ | $\tau^a$ | $R^b$ | $s^c$ | $\tau^a$ | $R^b$ | $s^c$ |
| W Europe | *−0.07* | *0.03* | – | 0.31 | *0.03* | – | 0.42 | *0.19* | – |
| W Russia | *0.05* | −0.29 | −3.7e-4 | 0.15 | 0.11 | 4.9e-4 | 0.11 | −0.12 | −1.9e-3 |
| W Kazak | 0.13 | *−0.04* | – | *0.01* | −0.15 | −2.1e-4 | 0.25 | −0.13 | −3.0e-3 |
| Arctic | *0.06* | −0.25 | −1.3e-4 | −0.29 | −0.35 | −5.0e-4 | −0.40 | −0.37 | −4.7e-4 |
| NE China | 0.31 | *−0.01* | – | *0.00* | *0.68* | - | 0.14 | 0.54 | 5.0e-2 |
| Local | 0.40 | 0.14 | 5.0e-5 | −0.37 | −0.20 | −1.0e-4 | −0.43 | −0.41 | −3.9e-4 |
| Strato | *−0.01* | −0.35 | −4.0e-2 | 0.37 | *0.00* | - | 0.37 | *−0.21* | −1.4e-2 |

[a] Kendall's tau. Values are given in italics if not statistically robust ($p$>0.01).

[b] Pearson correlation coefficient. Values are given in italics if not statistically robust ($p$>0.01).

[c] robust linear regression slope (ppm $s^{-1}$). No values is given if both Pearson and Kendall correlations are not robust, or if only Pearson correlation is not robust with $|R|$ <0.2.



ACPD

9, 6207–6245, 2009



**Table 2.** Region-concentration pair-wise statistics for CO.

| | April 2006 | | | September 2006 | | | August 2007 | | |
|---|---|---|---|---|---|---|---|---|---|
| | $\tau^a$ | $R^b$ | $s^c$ | $\tau^a$ | $R^b$ | $s^c$ | $\tau^a$ | $R^b$ | $s^c$ |
| W Europe | −0.08 | *−0.04* | – | *0.02* | *−0.01* | – | 0.18 | −0.28 | −7.9e-02 |
| W Russia | *0.03* | −0.25 | −5.3e-03 | 0.09 | *0.02* | – | *0.03* | *−0.07* | – |
| W Kazak | 0.17 | *−0.11* | – | 0.29 | 0.21 | 3.6e-03 | 0.26 | 0.17 | 3.6e-03 |
| Arctic | *0.03* | *0.04* | – | 0.15 | −0.18 | −4.1e-03 | −0.06 | *−0.13* | – |
| NE China | 0.32 | *0.00* | – | *−0.03* | *−0.86* | – | 0.17 | −0.22 | −1.7e-01 |
| Local | 0.29 | *0.01* | – | 0.27 | *0.06* | – | 0.14 | 0.36 | −3.8e-05 |
| Strato | *−0.04* | −0.36 | −6.3e-01 | −0.15 | *−0.42* | −1.1e+00 | *0.03* | 0.26 | – |

[a] Kendall's tau. Values are given in italics if not statistically robust ($p$>0.01).
[b] Pearson correlation coefficient. Values are given in italics if not statistically robust ($p$>0.01).
[c] robust linear regression slope (ppm $s^{-1}$). No values is given if both Pearson and Kendall correlations are not robust, or if only Pearson correlation is not robust with $|R|$ <0.2.

**Table 3.** Region-concentration pair-wise statistics for $O_3$.

| | April 2006 | | | September 2006 | | | August 2007 | | |
|---|---|---|---|---|---|---|---|---|---|
| | $\tau^a$ | $R^b$ | $s^c$ | $\tau^a$ | $R^b$ | $s^c$ | $\tau^a$ | $R^b$ | $s^c$ |
| W Europe | 0.28 | 0.35 | 8.5e-03 | 0.31 | *0.01* | – | 0.39 | 0.34 | 1.2e-01 |
| W Russia | 0.18 | 0.24 | 1.1e-03 | 0.07 | *−0.06* | – | 0.13 | −0.10 | −8.2e-03 |
| W Kazak | 0.12 | −0.20 | −3.7e-04 | 0.07 | *0.06* | – | 0.17 | *−0.05* | – |
| Arctic | −0.44 | −0.36 | −6.3e-04 | −0.33 | −0.47 | −3.8e-03 | −0.45 | −0.45 | −1.8e-03 |
| NE China | 0.08 | −0.36 | −1.7e-03 | −0.10 | *0.15* | – | 0.16 | 0.44 | 2.1e-01 |
| Local | −0.47 | −0.53 | −9.8e-04 | −0.44 | −0.76 | −2.5e-03 | −0.50 | −0.69 | −2.3e-03 |
| Strato | 0.45 | 0.43 | 1.0e-01 | 0.41 | *0.30* | 5.8e-01 | 0.52 | *−0.02* | – |

[a] Kendall's tau. Values are given in italics if not statistically robust ($p>0.01$).

[b] Pearson correlation coefficient. Values are given in italics if not statistically robust ($p>0.01$).

[c] robust linear regression slope (ppm $s^{-1}$). No values is given if both Pearson and Kendall correlations are not robust, or if only Pearson correlation is not robust with $|R| < 0.2$.





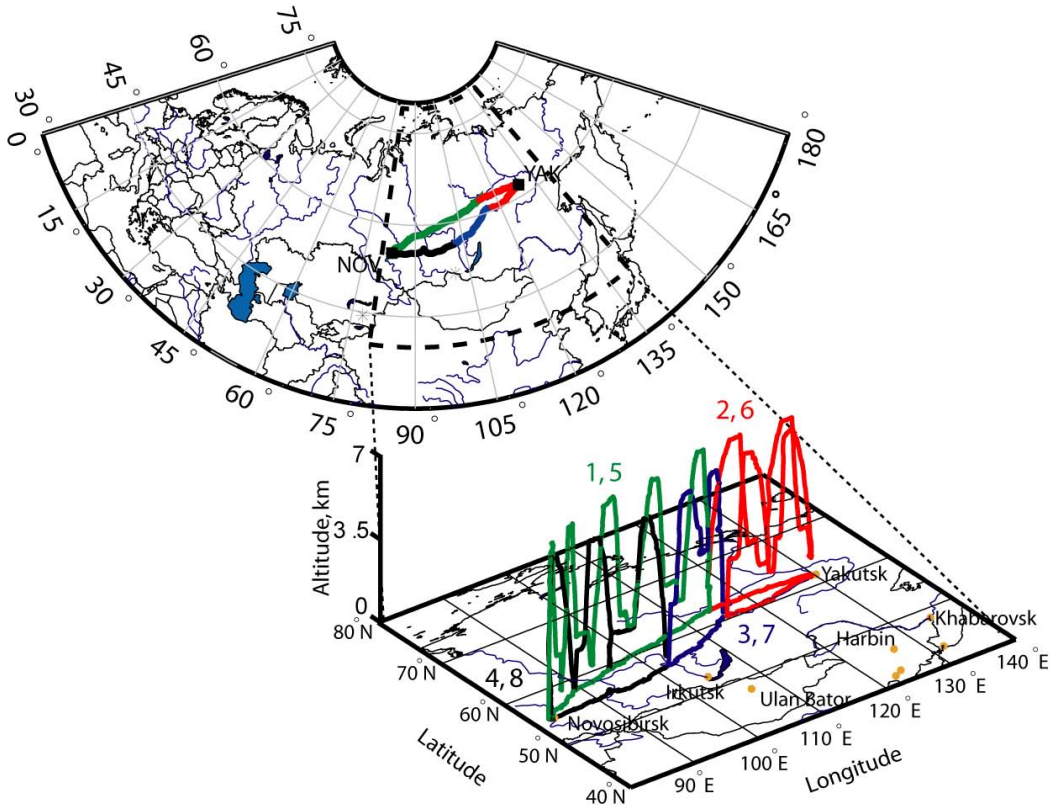


**Fig. 1.** Aircraft itinerary (top panel) and vertical flight pattern (bottom panel). Each campaign followed a similar itinerary. The different flights of each campaign (four each, typically one per day) are shown by different colors.



ACPD
9, 6207–6245, 2009



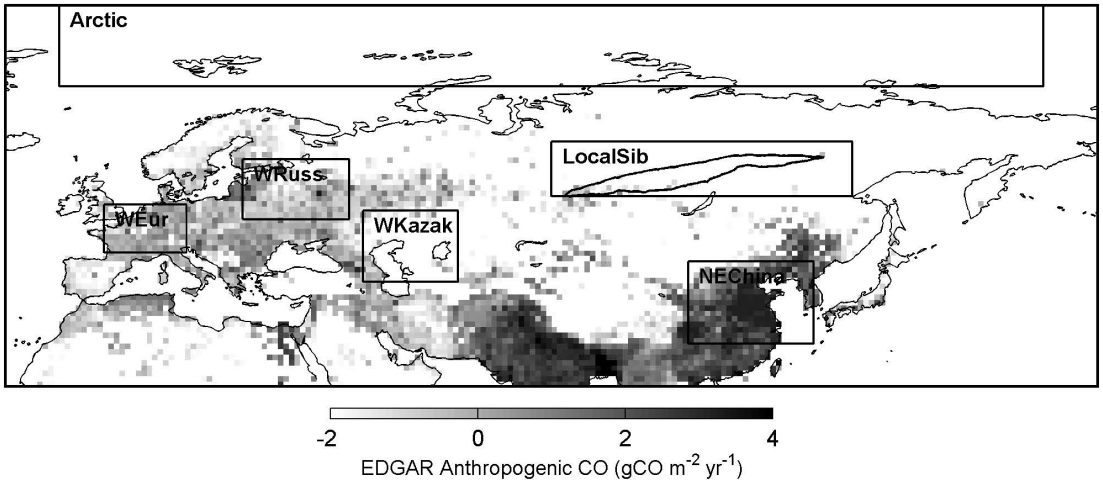


**Fig. 2.** Regions chosen for data reduction prior to the clustering analysis. The black line shows the itinerary of the aircraft during each campaign. Anthropogenic CO emissions from the EDGAR database are also shown in logarithmic grey scale, in g(CO) $m^{-2}$ $yr^{-1}$. to illustrate difference between regions.



ACPD
9, 6207–6245, 2009

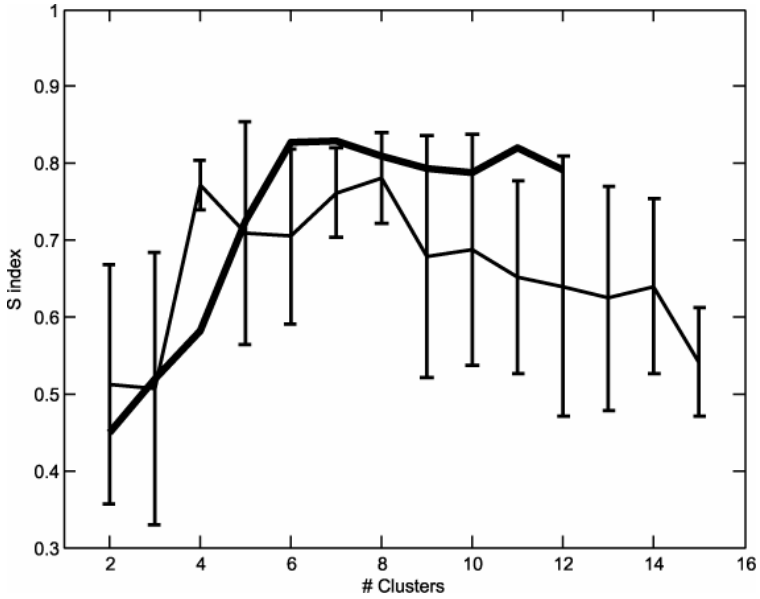


**Fig. 3.** S index of the clustering algorithm separation capability as a function of the number of clusters for all data (thick line) and separated campaign by campaign (thin line with error bars showing±1 std dev).



ACPD
9, 6207–6245, 2009

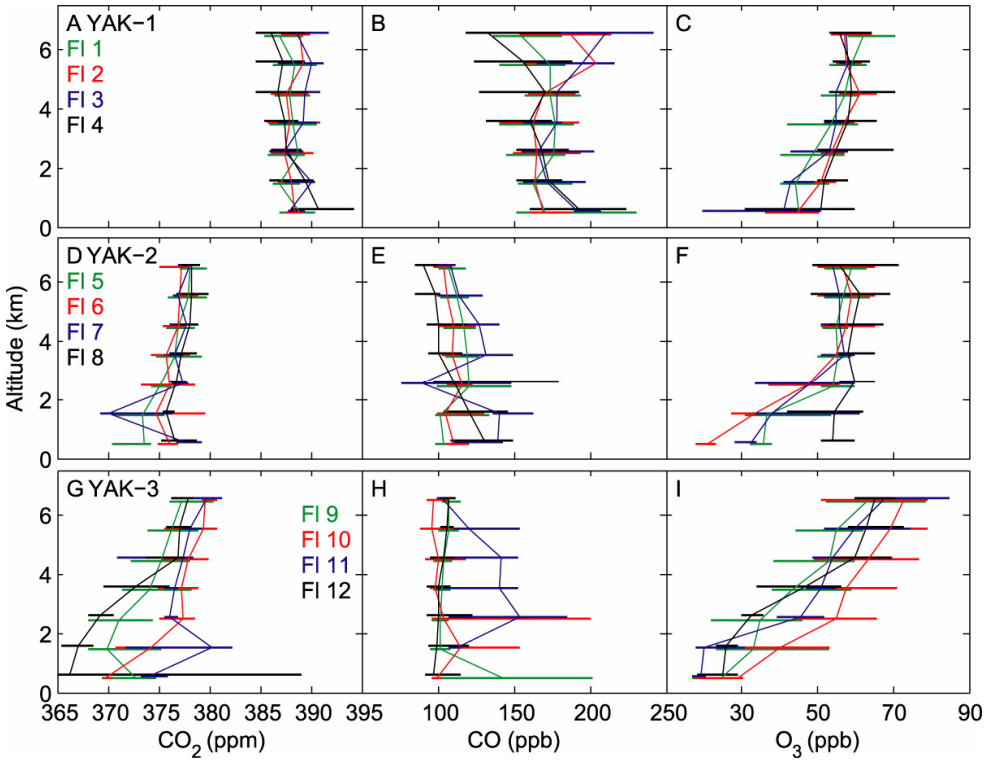


**Fig. 4.** Profiles of $CO_2$, CO and $O_3$ over the 3 campaigns. Each flight is averaged in 1000 m bins. The vertical line joins the median of each bin, while the horizontal lines represent the 10th and 90th percentiles of each bin.

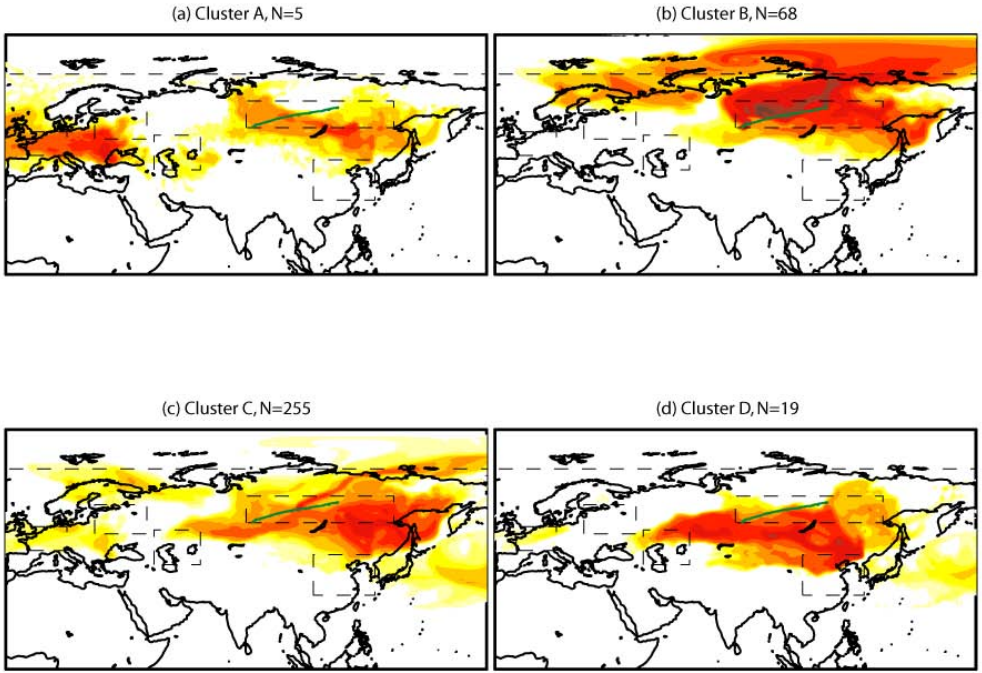


**Fig. 5.** Average footprint maps for data belonging to clusters A–D throughout Flight 1 (11 April 2006). Logarithmic color scale gives the mean particles PES<300 m. The number of elements N in a particular cluster is given above each map.



ACPD
9, 6207–6245, 2009

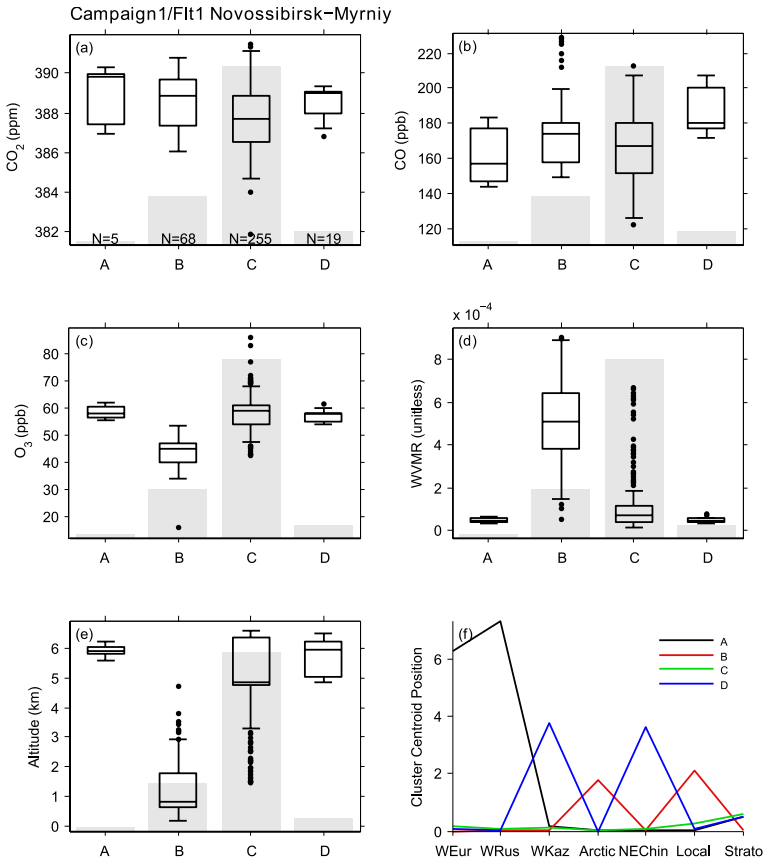


**Fig. 6.** Box plots of median and interquartile range (IQR) for $CO_2$ concentrations measured during Flight 1 in each April 2006 cluster (clusters A–D in abscissa, same denomination as in Fig. 5) **(a)**. Outliers within 1 IQD are within the whiskers, outliers beyond 1 IQR are shown as points. **(b)** Same for CO. **(c)** Same for $O_3$. **(d)** Same for water vapour mixing ratio. **(e)** Same for altitude. **(f)** Cluster centroid position vector for each cluster.



ACPD
9, 6207–6245, 2009



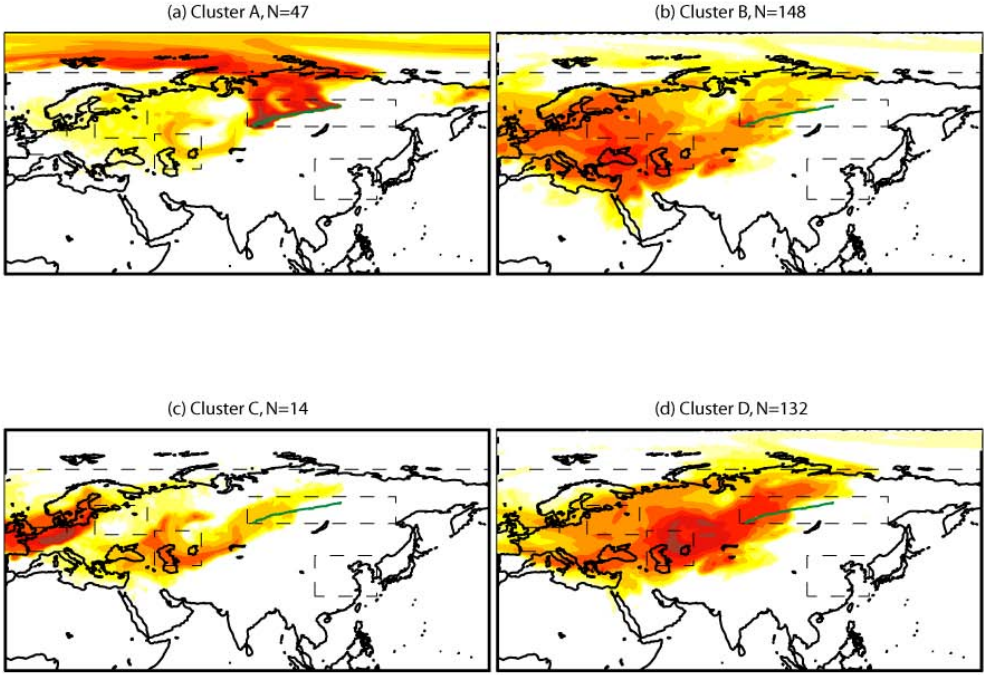


**Fig. 7.** Same as Fig. 5 for Flight 5 in September 2006.



ACPD
9, 6207–6245, 2009

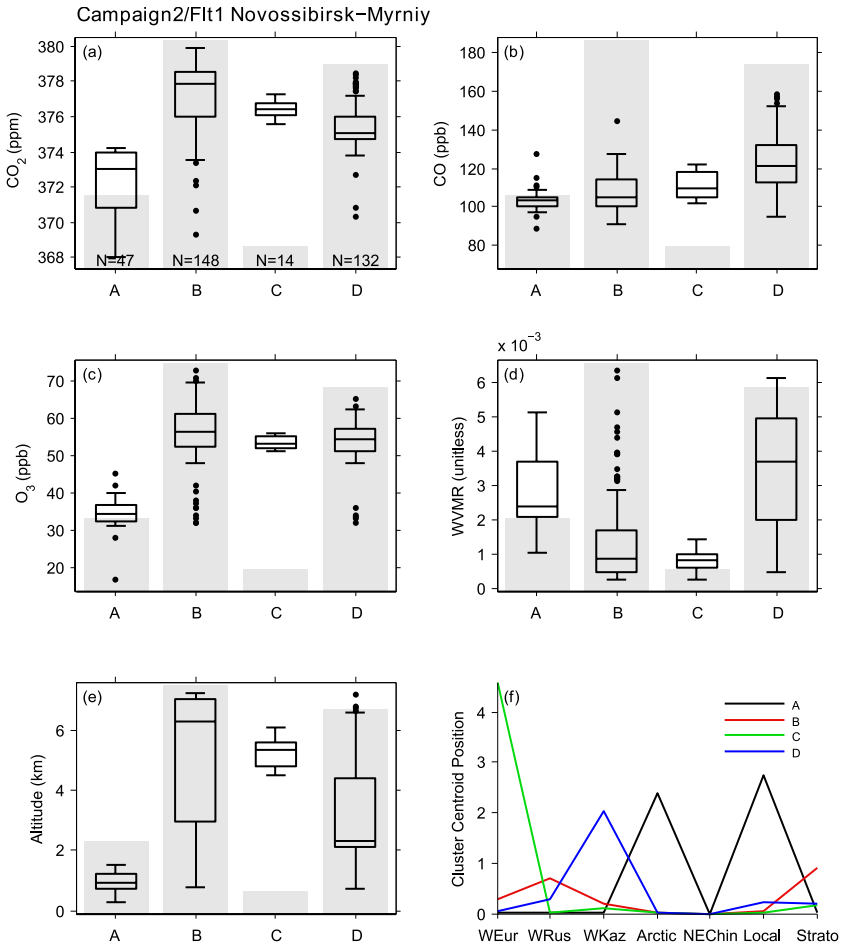


**Fig. 8.** Same as Fig. 6 for Flight 5 in September 2006.



ACPD
9, 6207–6245, 2009



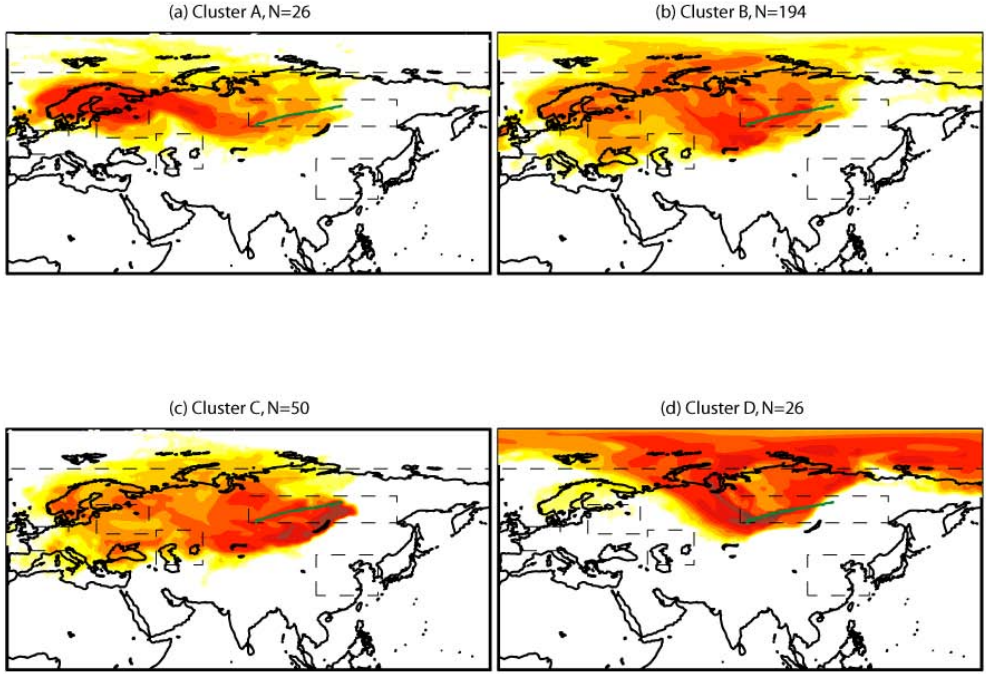


**Fig. 9.** Same as Fig. 5 for Flight 9 in August 2007.



ACPD
9, 6207–6245, 2009

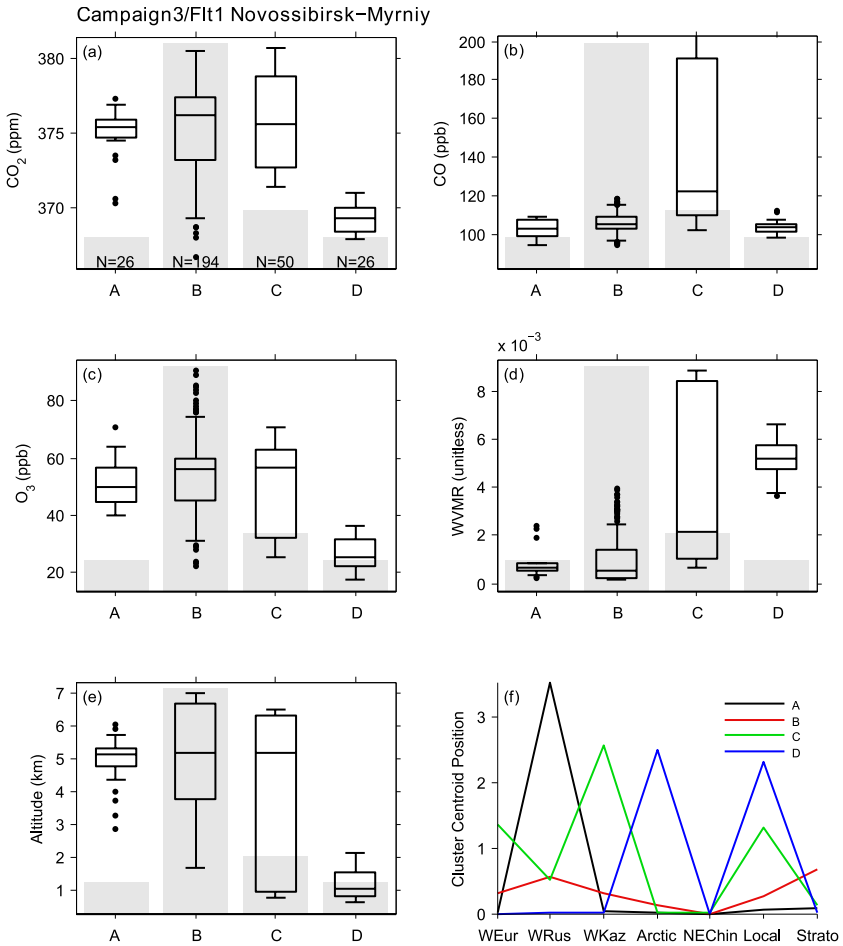


**Fig. 10.** Same as Fig. 6 for Flight 9 in August 2007.



ACPD
9, 6207–6245, 2009

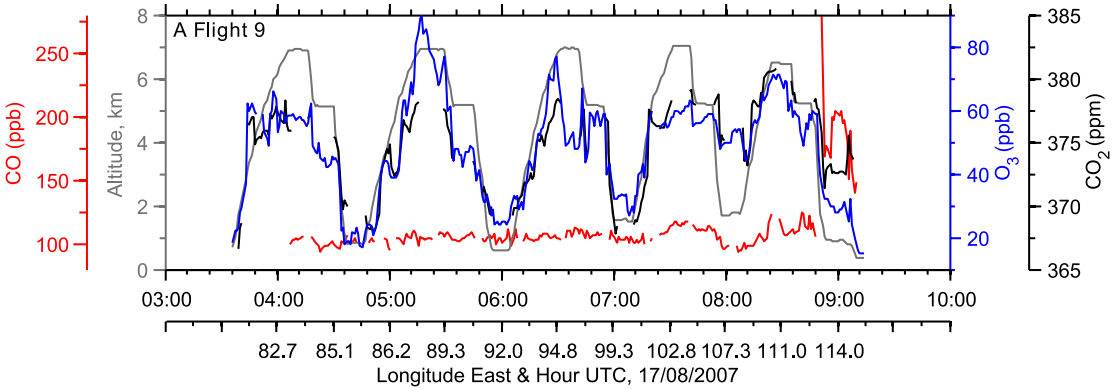


**Fig. 11.** Time series of measured $CO_2$ (black), CO (red) and $O_3$ (blue) concentrations as well as altitude (grey). Longitude is also given every half hour.



ACPD
9, 6207–6245, 2009

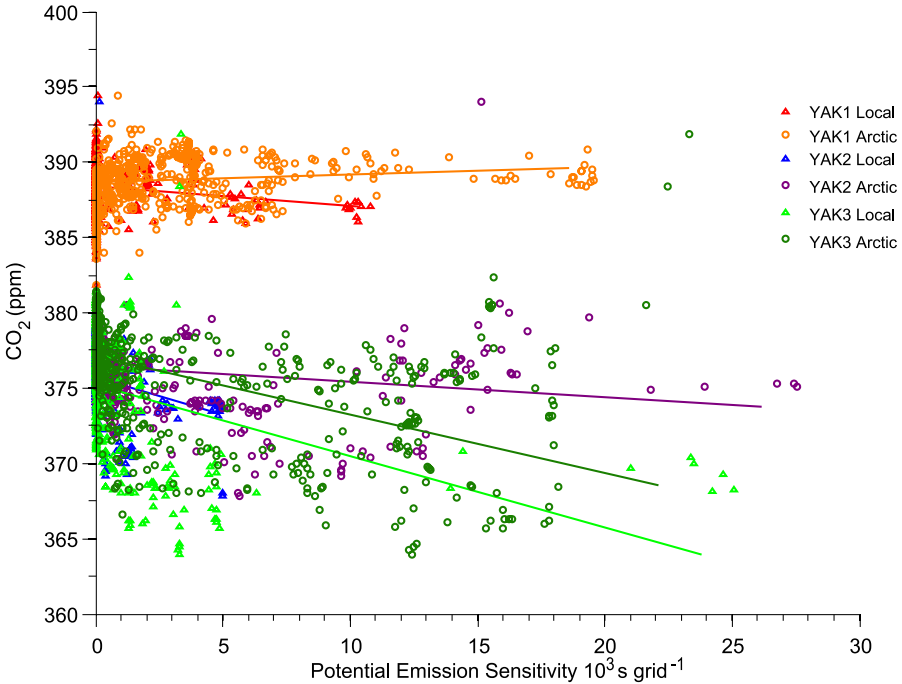


**Fig. 12.** $CO_2$ measurements from the three campaigns plotted against PES in the Arctic and Siberia regions, with robust linear regression fit.



ACPD
9, 6207–6245, 2009