Atmos. Chem. Phys., 9, 8413–8430, 2009
www.atmos-chem-phys.net/9/8413/2009/

Atmospheric Chemistry and Physics

Origin of aerosol particles in the mid-latitude and subtropical upper troposphere and lowermost stratosphere from cluster analysis of CARIBIC data

M. Köppe[1], M. Hermann[1], C. A. M. Brenninkmeijer[2], J. Heintzenberg[1], H. Schlager[3], T. Schuck[2], F. Slemr[2], D. Sprung[4,2], P. F. J. van Velthoven[5], A. Wiedensohler[1], A. Zahn[4], and H. Ziereis[3]

[1]Leibniz Institute for Tropospheric Research, Permoserstr. 15, 04318 Leipzig, Germany
[2]Atmospheric Chemistry Division, Max Planck Institute for Chemistry, P.O. Box 3060, 55020 Mainz, Germany
[3]Institute of Atmospheric Physics, German Aerospace Center, Oberpfaffenhofen, 82230 Wessling, Germany
[4]Institute for Meteorology and Climate Research, Karlsruhe Institute of Technology, P.O. Box 3640, 76021 Karlsruhe, Germany
[5]Royal Netherlands Meteorological Institute, P.O. Box 201, 3730 AE de Bilt, The Netherlands

Received: 25 May 2009 – Published in Atmos. Chem. Phys. Discuss.: 18 June 2009
Revised: 25 September 2009 – Accepted: 7 October 2009 – Published: 5 November 2009

Abstract. The origin of aerosol particles in the upper troposphere and lowermost stratosphere over the Eurasian continent was investigated by applying cluster analysis methods to in situ measured data. Number concentrations of submicrometer aerosol particles and trace gas mixing ratios derived by the CARIBIC (Civil Aircraft for Regular Investigation of the Atmosphere Based on an Instrument Container) measurement system on flights between Germany and South-East Asia were used for this analysis. Four cluster analysis methods were applied to a test data set and their capability of separating the data points into scientifically reasonable clusters was assessed. The best method was applied to seasonal data subsets for summer and winter resulting in five cluster or air mass types: stratosphere, tropopause, free troposphere, high clouds, and boundary layer influenced. Other source clusters, like aircraft emissions could not be resolved in the present data set with the used methods. While the cluster separation works satisfactory well for the summer data, in winter interpretation is more difficult, which is attributed to either different vertical transport pathways or different chemical lifetimes in both seasons. The geographical distribution of the clusters together with histograms for nucleation and Aitken mode particles within each cluster are presented. Aitken mode particle number concentrations show a clear vertical gradient with the lowest values in the lowermost stratosphere (750–2820 particles/cm^3 STP, minimum of the two 25% – and maximum of the two 75%-percentiles of both seasons) and the highest values for the boundary-layer-influenced air (4290–22 760 particles/cm^3 STP). Nucleation mode particles are also highest in the boundary-layer-influenced air (1260–29 500 particles/cm^3 STP), but are lowest in the free troposphere (0–450 particles/cm^3 STP). The given submicrometer particle number concentrations represent the first large-scale seasonal data sets for the upper troposphere and lowermost stratosphere over the Eurasian continent.

1 Introduction

The implementation of aerosols in global atmospheric models still represents one of the largest uncertainties in assessments of climate change (IPCC, 2007). The complexity of realistically simulating aerosol is obviously due to the large number of relevant physical and chemical particle properties, the great variety of aerosol processes, and the variability of the spatiotemporal distribution of atmospheric aerosol particles (Textor et al., 2007). The need to improve knowledge of atmospheric aerosol properties pertains to their impact on atmospheric chemistry (Seinfeld and Pandis, 2006) and on radiative forcing (Lohmann and Feichter, 2005; IPCC, 2007; Heintzenberg and Charlson, 2009). This holds true in particular for aerosol particles in the upper troposphere/lowermost stratosphere (UT/LMS) region, where they influence cirrus

Correspondence to: M. Hermann
(hermann@tropos.de)

clouds and hence the radiation budget of the atmosphere (Kärcher, 2003; Heintzenberg and Charlson, 2009). Furthermore these particles change the chemical composition of the UT/LMS by heterogeneous chemical processes (e.g. Bell et al., 2005; Søvde et al., 2007). Progress in understanding aerosol properties throughout the troposphere is strongly hindered by the technical challenges of airborne measurements and the associated high costs, both leading to a lack of reliable data sets. Fortunately new active satellite instruments, like CALIPSO (http://www-calipso.larc.nasa.gov), close this gap partly, by providing global maps of aerosol properties (e.g. Liu et al., 2008). However, these measurements are always restricted to the optically active particles (larger than ~100 nm) and a source attribution is not always possible because co-located trace gas information is lacking.

For a better representation of atmospheric aerosol particles (or trace gases), e.g. in three-dimensional atmospheric models, in situ long-term measurements in the free troposphere are essential (Kunz et al., 2008). Within the CARIBIC project (Civil Aircraft for Regular Investigation of the Atmosphere Based on an Instrument Container; Brenninkmeijer et al., 1999, 2007; http://www.caribic-atmospheric.com), long-term measurements of aerosol particle and trace gas concentrations in the UT/LMS were realized on a monthly basis since 1997. Using an instrumented container on a passenger aircraft (first a Boeing B767 of LTU International Airways and later an Airbus A340-600 of Lufthansa AG) over 220 successful measurement flights were conducted until December 2008. The geographic region covered on mainly five flight routes spans from northern mid-latitudes to southern subtropics (http://www.caribic-atmospheric.com). Accordingly, the CARIBIC project helps to close some of the observational gaps by yielding a large-area data set, which can help to validate atmospheric models (Kim et al., 2008) or to verify satellite based remote sensing instruments (Peylin et al., 2007).

In this study we present and analyze CARIBIC data from flights between Germany and South-East Asia. By using multivariate statistical analysis methods (mainly cluster analysis), we tried to identify types of air masses and hence the origins of sub-micrometer aerosol particles in the UT/LMS. This approach is highly experimental, because the sequence of analysis steps is not known a priori and the results can be hard to interpret (Backhaus et al., 2006). It was chosen, however, because of the large number of variables and data points in the CARIBIC data set. In addition to published methods of statistical analyses for similar data sets (Borchi and Marenco, 2002; Borchi et al., 2005), new combinations of multivariate analysis methods were applied and the results of the various methods are compared.

Section 2 begins with a short description of the aircraft measurements and the data set. The theory of univariate and multivariate statistical analyses and in particular the new method of analysis is presented in Sect. 3. In Sect. 4 the results are discussed and they are summarized in Sect. 5.

2 Experimental and data set

The core of the CARIBIC project is the monthly operation of a modified airfreight-container equipped with fully automated in situ measuring instruments and sampling devices for trace gases and aerosol particles. For detailed information about CARIBIC, such as the inlet and the measurement system, we refer to Brenninkmeijer et al. (1999, 2007), references therein, and the CARIBIC website http://www.caribic-atmospheric.com. The aerosol measurement techniques are explained in more detail by Hermann and Wiedensohler (2001), Papaspiropoulos et al. (2002), Hermann et al. (2003), and Nguyen et al. (2006, 2007).

For analysis, the in situ measured, quality-checked data were combined with aircraft flight parameters and meteorological information (provided by the Koninklijk Nederlands Meteorologisch Instituut, KNMI, de Bilt, the Netherlands, http://www.knmi.nl/samenw/campaign_support/CARIBIC) to so-called "meteo-merged-files". Exemplary parameters in these files are: Trace gas mixing ratios of carbon monoxide (CO), ozone (O_3), and acetonitrile (CH_3CN); number concentrations of aerosol particles between 4 and 12 nm diameter (nucleation mode, N_{4-12}) as well as particles larger than 12 nm diameter (Aitken mode, N_{12}); aircraft pressure-altitude and position; as well as model-derived potential vorticity (PV). Table 1 gives a list of all 52 parameters. In composing the meteo-merged-files, all parameters were averaged over 10 s intervals, centered at coordinated universal times (UTC) ending with five seconds, i.e. "hh:mm:s5". In order to have more complete data points, for this study the mixing ratios of the two trace gases acetone (CH_3COCH_3) and acetonitrile, which were obtained with a time resolution of 60 s, were linearly interpolated when the gap between following data points was smaller than 121 s.

Unfortunately, in the data set there are many missing values for individual parameters, caused by instrument malfunctioning or calibration periods. Eventually the data set was reduced strongly by having to eliminate all data points from multivariate statistical analyses affected by missing values for individual parameters. For summer and winter seasons, defined below, only 26% and 38%, respectively, of the original data points could be used. The missing values occur mainly in the data of the three trace gases CH_3CN, CH_3COCH_3, and NO_y, i.e. in the data of two instruments. For these instruments between 40% and 70% of the missing values are caused by regular instrument calibration, which should not influence the results of the data analysis. Instrument malfunctioning is responsible for the rest of the missing values of the two instruments.

In order to be sure that the data reduction does not influence the following multivariate analysis, we compared histograms of the original data set and the reduced data set with respect to e.g. the geographical distribution, altitude, or the local time of day. Besides the geographical distribution of data points for summer, which shows a larger fraction

Table 1. Parameters in the CARIBIC meteo-merged-files. "mr" stands for mixing ratio and the ^{md} for model-derived parameters. Bold parameters were used for the multivariate statistical analyses methods in this study.

Parameter	Unit	Parameter	Unit
PCUTC (container master computer)	days since 1 January 1900 and time in fraction of day	**Liquid water** ($\mathbf{H_2O_{cloud}}$) mr	ppmv
UTC ARINC Julian time (aircraft)	days since 1 January 1900 and time in fraction of day	**Ozone** ($\mathbf{O_3}$) mr	ppbv
Flight phase	0–10, e.g., 4= take off	Nitric oxide (NO) mr	ppbv
Position latitude	°	**Reactive nitrogen** (NO_y) mr	ppbv
Position longitude	°	CARIBIC Master PC time	s
True Heading	°	**Carbon monoxide** (CO) mr	ppbv
Wind speed	kn	**Concentration of particles between 4 and 12 nm** ($\mathbf{N}_{\mathbf{4}-12}$)	particles/cm^3 STP
Wind direction	°	**Concentration of particles larger than 12 nm** ($N_{\mathbf{12}}$)	particles/cm^3 STP
Pitch angle	°	**Concentration of particles larger than 18 nm** ($\mathbf{N_{18}}$)	particles/cm^3 STP
Roll angle	°	**Acetonitrile** ($\mathbf{CH_3C}N$) mr	pptv
Altitude	feet	**Acetone** ($\mathbf{CH_3COCH_3}$) mr	pptv
Barro-corrected altitude	feet	Static temperature ^{md}	K
True airspeed	kn	**Potential vorticity (PV)**^{md}	PVU
Total air temperature	°C	Potential temperature^{md}	K
Altitude rate	feet/min	**Equivalent potential temperature**^{md}	K
Static air temperature	°C	Amount of water vapor^{md}	kg/kg
Corrected angle of attack	°	Eastward wind^{md}	m/s
Total pressure	mbar	Northward wind^{md}	m/s
Vertical velocity	feet/min	**Vertical wind speed**^{md}	Pa/s
Ground speed	kn	Wind speed^{md}	m/s
Local time	Julian time	Wind direction^{md}	°
Static pressure	mbar	Total amount of water^{md}	ppmv
Potential temperature	K	Relative humidity^{md}	%
Altitude	m	**Cell cloud cover**^{md}	0 to 1
Status	sensor work, 0–3	Cloud water content^{md}	kg/m^3
Gaseous water ($\mathbf{H_2O}_{gas}$) mr	ppmv	Cloud ice content^{md}	kg/m^3

of data points for the longitudes 20–50° E and 100–110° E, and a smaller fraction in the region 60–80° E compared to the original data set, all compared histograms look similar, which gives some confidence that the reduced data set is not strongly biased.

Before applying the multivariate statistical analysis methods to the meteo-merged-files, the data set was reduced in order to keep computing burden within reasonable limits. In this reduction step those variables that are a priori irrelevant for determining the particle origin, such as the aircraft flight parameters "pitch angle" or "roll angle", were eliminated. Consequently, the CARIBIC data subset used in this study consists of 25 essential parameters, which are called "original variables" hereafter (printed bold in Table 1).

For the air-mass-based assessment of submicrometer aerosol particle types in the UT/LMS, we primarily use specific trace gases as indicators (tracers) for the origin of respective air masses. In particular, O_3, CO, and CH_3CN are relevant and can be used as indicators of stratospheric air (e.g. Zahn and Brenninkmeijer, 2003), boundary layer air (e.g. Zahn et al., 2002; Hoor et al., 2005), and biomass-burning-influenced air (de Gouw et al., 2003, 2006), respectively.

Some of these essential tracers were available in sufficient degree only for the CARIBIC flights since 2006, which predominantly cover the South-East Asia route. Consequently, we exclusively used the CARIBIC data obtained on long-distance flights from Frankfurt (Germany) to Manila (Philippines) and back. Each flight consisted of two legs with a brief stopover in Guangzhou (China). The actual flight tracks of the flights used in this study are displayed in Fig. 1. They cover an area from Eastern Europe (mid-latitudes, ~50° N,

Table 2. Seasonal assignment of the flights to the summer-, winter-, and test data set. Each flight leg of the monthly flights obtained its individual number. Missing numbers indicate flights conducted on alternative routes. Flights assigned to the summer data set are highlighted by red, flights of winter data set by blue, and flights not assigned to either one of them by grey color. Flights of the test data set are indicated by a "T", flights with no data (e.g. no power to the container) by "n.a.". Airports codes are Frankfurt (FRA), Guangzhou (CAN), and Manila (MNL).

Number	Direction	Date	Note	Number	Direction	Date	Note
146	FRA-CAN	27 Apr 2006	T	184	MNL-CAN	7 Mar 2007	
147	CAN-MNL	28 Apr 2006	T	185	CAN-FRA	7 Mar 2007	
148	MNL-CAN	28 Apr 2006	T	186	FRA-CAN	18 Apr 2007	
149	CAN-FRA	28 Apr 2006	T	187	CAN-MNL	19 Apr 2007	
150	FRA-CAN	29 May 2006	T	188	MNL-CAN	19 Apr 2007	
151	CAN-MNL	30 May 2006	T	189	CAN-FRA	19 Apr 2007	
152	MNL-CAN	30 May 2006	T	190	FRA-CAN	22 May 2007	
153	CAN-FRA	30 May 2006	T	191	CAN-MNL	23 May 2007	
154	FRA-CAN	5 Jul 2006	T	192	MNL-CAN	23 May 2007	
155	CAN-MNL	6 Jul 2006	T	193	CAN-FRA	23 May 2007	
156	MNL-CAN	6 Jul 2006	T	194	FRA-CAN	21 Jun 2007	
157	CAN-FRA	6 Jul 2006	T	195	CAN-MNL	22 Jun 2007	
158	FRA-CAN	31 Jul 2006	T	196	MNL-CAN	22 Jun 2007	
159	CAN-MNL	1 Aug 2006	T	197	CAN-FRA	22 Jun 2007	
160	MNL-CAN	1 Aug 2006	T	198	FRA-CAN	17 Jul 2007	
161	CAN-FRA	1 Aug 2006	T	199	CAN-MNL	18 Jul 2007	
162	FRA-CAN	7 Sep 2006	T	200	MNL-CAN	18 Jul 2007	n.a.
163	CAN-MNL	8 Sep 2006	T	201	CAN-FRA	18 Jul 2007	n.a.
164	MNL-CAN	8 Sep 2006	T	202	FRA-CAN	14 Aug 2007	
165	CAN-FRA	8 Sep 2006	T	203	CAN-MNL	15 Aug 2007	
166	FRA-CAN	19 Oct 2006		204	MNL-CAN	15 Aug 2007	
167	CAN-MNL	20 Oct 2006		205	CAN-FRA	15 Aug 2007	
168	MNL-CAN	20 Oct 2006		210	FRA-CAN	24 Oct 2007	
169	CAN-FRA	20 Oct 2006		211	CAN-MNL	25 Oct 2007	
170	FRA-CAN	14 Nov 2006		212	MNL-CAN	25 Oct 2007	
171	CAN-MNL	15 Nov 2006		213	CAN-FRA	25 Oct 2007	
172	MNL-CAN	15 Nov 2006		214	FRA-CAN	13 Nov 2007	
173	CAN-FRA	15 Nov 2006		215	CAN-MNL	14 Nov 2007	
174	FRA-CAN	13 Dec 2006		216	MNL-CAN	14 Nov 2007	
175	CAN-MNL	14 Dec 2006		217	CAN-FRA	14 Nov 2007	n.a.
176	MNL-CAN	14 Dec 2006		220	FRA-CAN	25 Feb 2008	
177	CAN-FRA	14 Dec 2006		221	CAN-MNL	26 Feb 2007	
178	FRA-CAN	5 Feb 2007		222	MNL-CAN	26 Feb 2008	
179	CAN-MNL	6 Feb 2007		223	CAN-FRA	26 Feb 2008	
180	MNL-CAN	6 Feb 2007		224	FRA-CAN	26 Mar 2008	
181	CAN-FRA	6 Feb 2007		225	CAN-MNL	27 Mar 2008	
182	FRA-CAN	6 Mar 2007		226	MNL-CAN	27 Mar 2008	
183	CAN-MNL	7 Mar 2007		227	CAN-FRA	27 Mar 2008	

~10 ° E) via Central and East Asia to South-East-Asia (tropics, ~15° N, ~120° E). On this flight route, the actual flight tracks were mainly determined by geography and at least for the Asian part mostly fall into a relative narrow flight corridor (in contrast to e.g. the North-America flight route). Hence we assume no strong bias towards specific synoptic conditions, as it might be expected for commercial aircraft. The final data set consists of 36 long-distance flights, conducted between April 2006 and March 2008 (cf. Table 2). Finally, data are restricted to cruise altitudes between ~8 and ~12 km altitude (~335 to ~200 hPa pressure).

Previous studies have already shown that the concentrations of submicrometer aerosol particles undergo strong geographical and seasonal variations (e.g. Hermann et al., 2003). Therefore, a summer (15 flights) and a winter data set (17 flights) were extracted from the available data set. Because of the limited number of flights, each corresponding to only one specific meteorological situation in an overflown region, a finer division, e.g., into four seasons was statistically not reasonable. For the division not only the time of the year, but also the synoptic weather situation in the sampling area was considered. Firstly, flights in June, July, and August were assigned to boreal summer and those in December, January, and February to boreal winter. In order to get better statistics, we expand the summer and winter seasons by adding flights from spring and autumn, if the synoptic weather conditions for the flight date showed similar patterns like in summer or winter, respectively. Therefore we checked the synoptic meteorological situation indicated by the positions of controlling anticyclones, cyclones, and frontal zones of general flow patterns (Liljequist and Cehak, 1994; Gerstengarbe et al., 1999a; Malberg, 1997), by pressure level maps (http://www.wetterzentrale.de; http://www.arl.noaa.gov), and by model-derived vertical cross sections of the atmosphere (http://www.knmi.nl/samenw/campaign_support/CARIBIC). The four April flights were not assigned, because the synoptic meteorological situation was ambiguous. Most of the data points are located within a corridor of ∼1500 km width (Fig. 1), within the scale of synoptic-scale systems. Therefore existing differences in flight routes between summer and winter should play a minor role compared to the shift in atmospheric circulation associated with the seasons.

Interpreting the present data sets, one should be aware that because of the fixed flight schedule of commercial aircraft and the mainly east-west flight route the data sets are biased towards certain local times of day (LT). In the winter data set, where this effect is stronger compared to the summer data set, 57% of the data points were measured in the six hours between 00:00 and 06:00 LT. On the other hand there is a minimum between 15:00 and 21:00 LT, representing only 6% of the data points. Measurements at mid-latitudes, between 40° N and 55° N, are about 80% night time measurements. The same holds true for data points below 23° N, which are about 70% night time measurements.

For testing and verification of the multivariate statistical methodology (cf. Sect. 3) with reduced computing burden, first a smaller test data set was generated from the summer data set (cf. Table 2).

3 Statistical methods

In general, statistical analysis methods can be divided into univariate and multivariate methods by the number of simultaneously investigated variables. While univariate methods readily yield results and hence are suited to give an overview of the data set, multivariate methods are more complex, but also provide deeper insight. Multivariate statistical analyses have been used in atmospheric research for example to investigate the relation between synoptic weather patterns and precipitation chemistry (Dorling and Davies, 1995), the classification of European climates (Gerstengarbe et al., 1999b), the discrimination of air masses near the extra tropical tropopause (Borchi and Marenco, 2002; Borchi et al., 2005), or aircraft-derived profiles of trace gas and particulate pollution in the lower troposphere (Taubman et al., 2006; Hains et al., 2008).

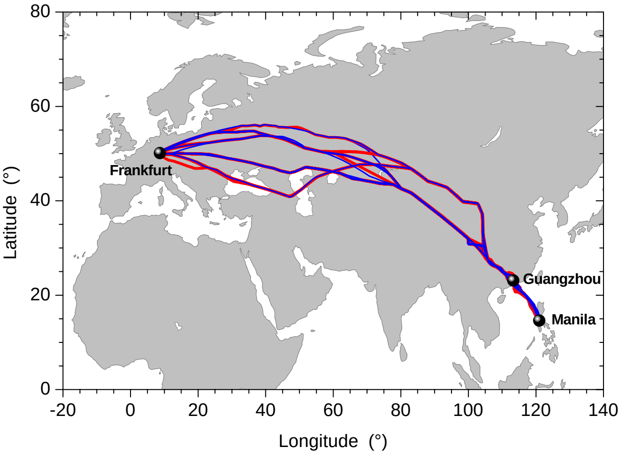

Fig. 1. Geographic sampling area and individual CARIBIC flight tracks on the South-East Asia route, which were used for the cluster analysis in this study. Red color indicates summer flights (15), blue winter flights (17). Each outward and return flight consists of two flight legs with a brief stopover in Guangzhou (China).

The choice of a statistical analysis method for a given problem is not straight forward, as different methods partially overlap in the resulting information content (Backhaus et al., 2006). Therefore we experimented with different combinations of statistical analysis methods. In the following sections, the univariate and multivariate statistical analysis methods are briefly described. Details, in particular on mathematical background, are given for instance in Backhaus et al. (2006). All statistical analyses were performed with the commercial software code SPSS® (statistical computer software SPSS, version 11.0.1, SPSS Inc., Chicago, Illinois, USA; http://www.spss.com).

3.1 Univariate statistical methods

Explorative data analysis methods are generally used to get an overview of the data set and to detect suspicious data points (outliers) and errors. However, there are no fixed prescriptions as to how an explorative data analysis should be conducted. The choice of methods clearly depends on the respective data sets and on the subsequent multivariate statistical methods (Brosius, 2006; Backhaus et al., 2006; Leyer and Wesche, 2007).

If an outlier in one variable gets identified, it is most often useful to exclude the respective data point from analysis, because outliers can affect results strongly and complicate conclusions (Brosius, 2006; Backhaus et al., 2006). The present CARIBIC data were examined by using the SPSS®

functions *extreme values*, *percentiles*, *M-estimators*, and (the multivariate) *hierarchical cluster analyses* (here the single linkage algorithm, cf. Sect. 3.2). The SPSS® sub function *extreme values* identifies the five highest and lowest values of a variable together with the corresponding row numbers of the data matrix. By using this function, erroneous data points could be eliminated. In statistical analyses percentiles are often used to visualize the distribution of values. Thereby, extremes that show values for instance above the 95% percentile can be eliminated from data sets. Here, we refrained from this relatively crude elimination by percentiles, because extreme values of tracers are likely to help to reveal the origin of an air mass. Nevertheless, we did use percentiles in this study to verify the elimination of outliers process resulting by the above listed techniques. M-estimators are statistical indices derived by combining all values of one investigated variable to a mean value. To identify the influence of extreme values so-called maximum likelihood estimators can be used. Herein individual values get different weights according to their distance to the remaining values of the respective variable. For that purpose SPSS® offers four different indices. If the distribution of values is not symmetric, the four M-estimators differ significantly from the arithmetic mean. In this work, M-estimators were recalculated after an elimination step to examine the effect. Finally, the multivariate single linkage algorithm was used, which is highly suited to find outliers in data sets by separating clusters of highly different numbers of data points (Brosius, 2006; Backhaus et al., 2006) (cf. Sect. 3.2). For both seasonal data sets, the fraction of outliers was below 0.4%.

Before starting the multivariate analysis, two data processing steps were carried out. Although not necessarily required, normalization of the data might help data analysis (Leyer and Wesche, 2007). For that reason the original variables of the CARIBIC data subsets were examined for normal distribution by using histograms and significance tests (e.g. the Kolmogorov-Smirnov test). If the applied tests revealed a non-normal distribution, the corresponding original variable was transformed by applying either the logarithm, the square, or the radical function in order to bring it in a nearly normally distributed shape (Leyer and Wesche, 2007). After having evaluated each variable for normal distribution, the original variables were processed by centering and standardization. This step is recommended since measured variables often have highly different value ranges (Brosius, 2006; Backhaus et al., 2006; Leyer and Wesche, 2007). Centering and standardization leads to comparability of respective variables. For centering purposes the difference between the measured value and the arithmetic mean was calculated. Afterwards, within the standardization, this difference was divided by the standard deviation of the variable distribution. Accordingly, these processed variables (hereafter called z-parameters) have a mean of zero and a standard deviation of one.

3.2 Multivariate statistical methods

Atmospheric science often deals with complex relationships involving more than one variable. One of the main problems is to identify the controlling parameters among a large number of variables that subsequently can be used as input for multivariate analysis (Brosius, 2006; Backhaus et al., 2006; Leyer and Wesche, 2007). Therefore we occasionally used the principal component analysis (PCA) before applying the cluster analysis (CA).

PCA is one possible method to reduce the complexity of a data set by attributing as much as possible of the total variance of the original data to fewer newly defined variables (Brosius, 2006; Backhaus et al., 2006). This data reduction is accomplished by finding linear combinations of original variables, also referred to as factors or principal components (PCs). The resulting linear combinations are chosen to be uncorrelated with each other and account for successively smaller fractions of the total variance.

The data reduction in the course of the PCA can also be used to successive eliminate "ineffective" variables of a data set, e.g. by applying the MSA (Measure of Sampling Adequacy) criterion (Brosius, 2006; Backhaus et al., 2006). The MSA criterion enables to investigate whether or not a variable contributes significantly to the variance of the data set. It compares the single and partial correlations between variables. High MSA values (close to one) generally indicate that a particular variable is "useful" and should be kept in the data set. In two of the analysis procedures we applied the MSA criterion to eliminate variables and directly afterwards PCA was aborted (cf. Fig. 2). Consequently, for these two procedures no PCs were extracted but instead the z-parameters, identified as relevant, were used for the subsequent analysis methods.

The CA is a technique to group similar observations or data points into classes called clusters (Brosius, 2006; Backhaus et al., 2006). In our case, these clusters helped to discriminate air masses and hence to identify particle origins. Two kinds of CAs, referred to as the hierarchical cluster analysis (HCA) and K-means cluster analysis (KCA) were applied. As HCA algorithms, two common types, the single linkage and Ward algorithms, were used. The single linkage algorithm is particularly capable of separating clusters with highly different numbers of data points. Often it separates a few large and many small clusters, with the small clusters allowing the identification of outliers in data sets (Brosius, 2006; Backhaus et al., 2006). This valuable characteristic was used in addition to the univariate methods listed in Sect. 3.1. The *Ward* algorithm uses the heterogeneity of the cluster as criterion for clustering data points, whereby the variance within clusters should become minimal. Since homogeneity inside clusters is the goal of the *Ward* algorithm, it is the preferred algorithm in many applications of CAs (Leyer and Wesche, 2007). The *Ward* algorithm can be regarded as practical classification algorithm for

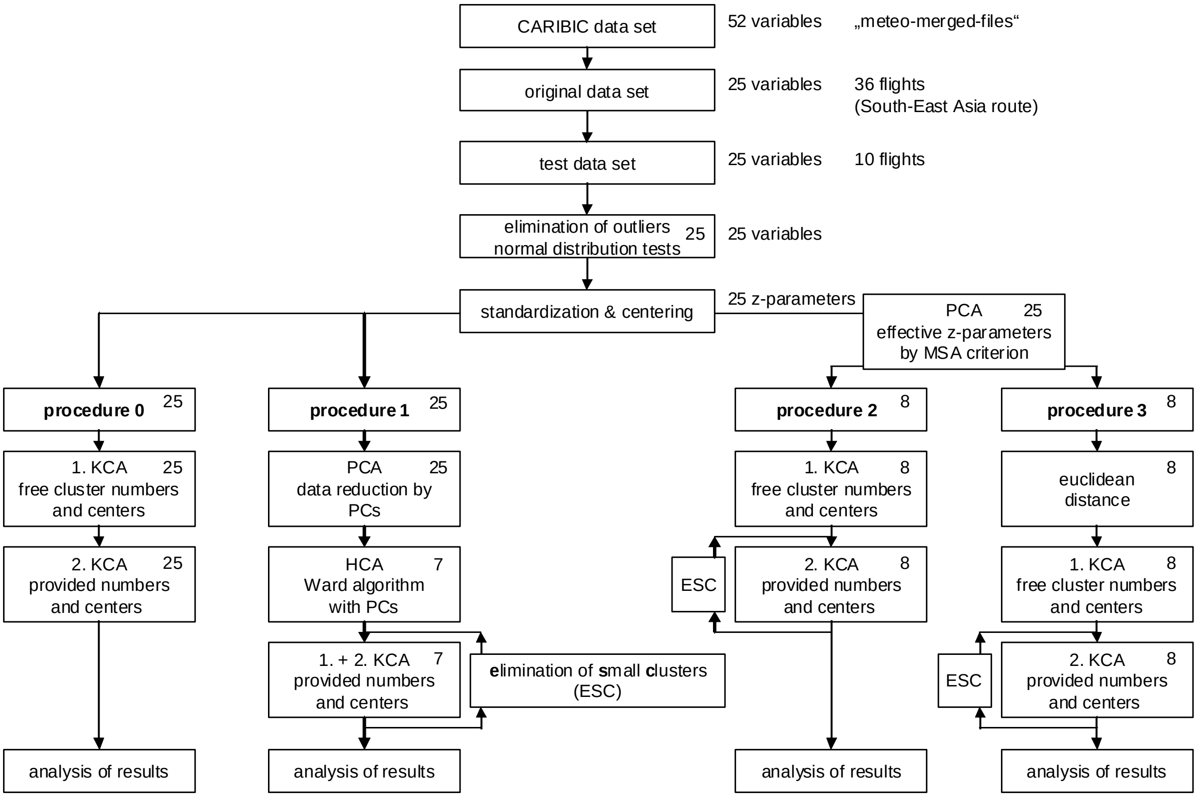

Fig. 2. Flow diagram for composition and implementation of the statistical analysis methods of the original - and test data sets. Numbers in the upper right corner of the boxes give the number of investigated z-parameters or principal components (in procedure 1), respectively.

the objective of this study, because it mostly finds uniform and compact clusters in data sets and even promotes their formation (Backhaus et al., 2006; Leyer and Wesche, 2007). Unfortunately, HCA requires a large computer capacity and has the drawback that once a data point is attributed to a particular cluster, a reassignment to a more appropriate cluster in a later step of the HCA is not feasible. Consequently, non-hierarchical techniques like KCA are commonly preferred. For the KCA a number of clusters and cluster centers have to be provided by the user and then the KCA assigns each data point to that cluster that has the nearest centroid (arithmetic mean). The clustering itself is performed in a way that the sum of distances within a cluster is minimized. Assignment or permutation of data points is not concluded before the “optimal” clustering is found.

As briefly explained, the different CA techniques have their advantages and disadvantages and can provide dissimilar results. While the entire purpose is to minimize the differences between the data points within clusters (homogeneity) and maximize the differences between clusters (heterogeneity), the essential challenge is to find the “correct” clustering. This question clearly depends on the nature of the data set and the problem in question. Therefore we tested and compared different techniques of both HCA and KCA.

3.3 “Optimal” multivariate statistical method for CARIBIC data

With increasing complexity of statistical analysis computing time is rising. Thus a first issue is whether simple and quick statistical methods can provide similar results as complex and more elaborate methods do. To this end we tested four different procedures (cf. Fig. 2).

Procedure 0 is the simplest but crudest of all, clustering z-parameters (cf. Sect. 3.1) of particular CARIBIC data subsets with a KCA only. From the outset it was not known how many clusters would be required for an optimal clustering of CARIBIC data targeting the meaningful separation of air masses. Consequently, all clustering methods described here were tested for the range of four to seven clusters. Procedure 0 is the quickest, because it does not need further data preprocessing before the CA. To check the quality of the resulting clustering, various scatter plots, each involving two tracers, were investigated. As an example, Fig. 3 shows the scatter plot of O_3 vs. CO for the six-cluster separation process, in order to check the homogeneity of resulting stratosphere clusters. A comparison based on a cluster separation with five clusters yielded similar results. In all four graphs (a–d) the well-known negative correlation between O_3 and CO appears which marks

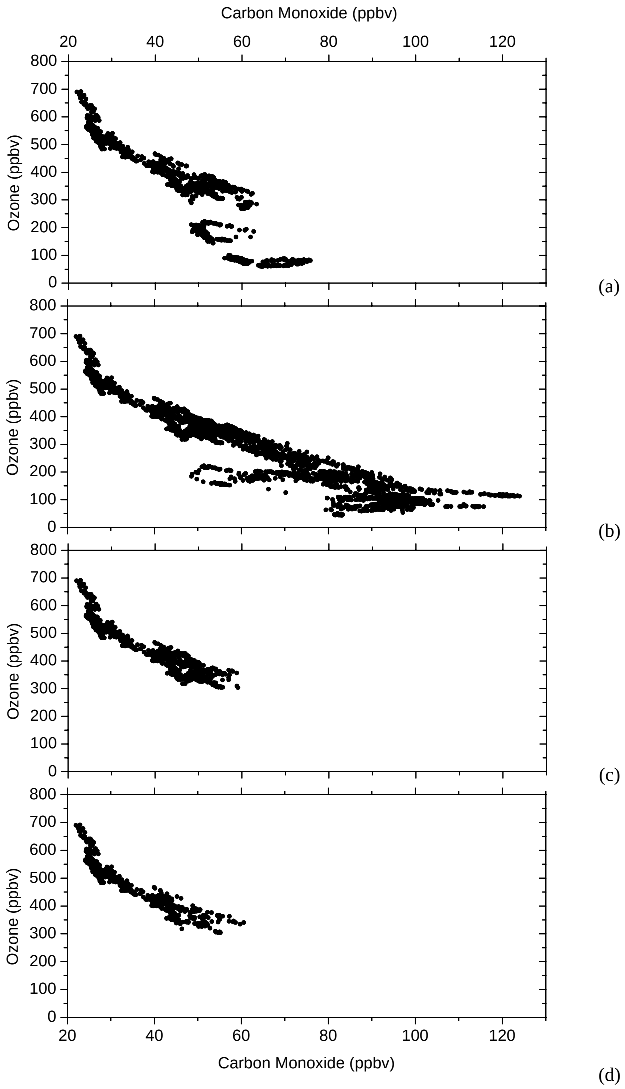

Fig. 3. O_3-CO-scatter plots for the stratosphere clusters of the test data set (six-cluster separation). **(a)** procedure 0, **(b)** procedure 1, **(c)** procedure 2, and **(d)** procedure 3_50.

the extra-tropical tropopause transition layer sandwiched between the tropopause and the unperturbed lowermost stratosphere (Hoor et al., 2002). Figure 3 shows that the stratosphere cluster of procedure 0 (Fig. 3a) contains data points with O_3 values from nearly 50 ppbv up to 700 ppbv whilst CO values are always below 80 ppbv. Since O_3 values below 100 ppbv describe almost always tropospheric air (Zahn and Brenninkmeijer, 2003) and procedure 0 grouped data points with upper tropospheric up to stratospheric values into the same cluster, this clustering is certainly not satisfactory.

In procedure 1 the numbers of z-parameters were first reduced by applying a PCA before the CA. With the extracted PCs an HCA and afterwards a KCA was carried out. This procedure follows the methodology developed in Borchi and Marenco (2002) and Borchi et al. (2005), who applied it to a similar data set, however with a smaller number of parameters. In the course of KCA another modification in comparison to procedure 0 was made. Very small clusters, representing only very small fractions of all data points (<1%), characterize individual events, like crossing an individual cumulonimbus. They are not representative for the intended large-scale air mass-based characterization of submicrometer aerosol particles in the UT/LMS. Thus they were eliminated and the KCA was reapplied.

Correlation analyses in the course of the PCA exposed the very low load of some essential z-parameters – like the particle number concentrations – on the extracted PCs. This results in a too large information loss concerning the variance of the particle number concentrations, which is not desired for the overall purpose of this study. Concerning its separation capability, procedure 1 shows similar results as procedure 0 (cf. Fig. 3b). O_3 values lie again between 50 and 700 ppbv, but this time CO values of even more than 120 ppbv are reached, clearly indicating tropospheric air. For this reason, also procedure 1 was excluded from the main statistical analyses of the CARIBIC data subsets.

As procedures 0 and 1 turned out to be inadequate for the purpose of this study new procedures were developed. Procedure 2 also starts with a PCA of z-parameters. After a correlation analysis concerning adequacy for subsequent multivariate statistical analysis, the most effective and relevant z-parameters, not PCs, were identified by using the MSA criterion in the already described manner. The method's suggestion to eliminate all particle number concentrations was however disregarded, because this would of course not allow the identification of aerosol origins. Only the particle number concentration for particles larger than 18 nm diameter (N_{18}) was excluded because it is highly correlated with N_{12}. The following eight remaining z-parameters finally entered a KCA: O_3 (ozone), CO (carbon monoxide), NO_y (reactive odd nitrogen), H_2O_{gas} (water vapor), CH_3CN (acetonitrile), CH_3COCH_3 (acetone), N_{4-12} (nucleation mode particles), and N_{12} (Aitken mode particles). Like before (procedure 1) too small clusters were eliminated and the KCA reapplied. Procedure 2 yields better results compared to procedures 0 and 1. For instance, data points of stratospheric and tropospheric air masses were separated in clearly different clusters as can be seen in Fig. 3c.

Like procedure 2, procedure 3 starts with a PCA, but afterwards an additional step was performed. The idea was that generally those air masses that have been recently influenced by a particle or trace gas source or sink show a clear signature in the respective tracer. The more the air mass ages or mixes with other air masses, the more its tracer signature wanes, while being transformed into background air. In terms of an n-dimensional data set in the z-space, aged or well-mixed air masses should be located close to the origin of the coordinate system. Air masses recently influenced by a source or sink process should show a greater distance to the origin. The idea behind procedure 3 was to eliminate this atmospheric background enabling a clearer clustering of the remaining data points. Therefore the Euclidean distance from the origin of the coordinate system was calculated for each data point. Hereafter 25% and 50% of data points, respectively, with the smallest distance to the origin were eliminated. The remaining data points were then clustered according to procedure 2. However, procedure 3 did not significantly improve the separation of resulting clusters in comparison to procedure 2 (cf. Fig. 3d). Moreover, procedure 3 always showed the same results for the two seasonal data subsets as procedure 2. The likely reason for this similarity is that procedure 2 generates a background air cluster close to the origin by its own, which has the same effect as eliminating the data points close to the origin. Consequently, only results of procedure 2 will be discussed below.

4 Results and discussion

The results of the cluster analysis for the summer and winter data sets are presented in Sect. 4.1 and 4.2, respectively. At the outset, it is conspicuous that there is a significant difference in the total number of data points between summer (15 609) and winter (25 111), compared to the only slightly different number of flights (15 and 17, respectively). This difference is due to missing values of individual parameters in the meteo-merged files (cf. Sect. 2) and can be explained by an increasing instrument reliability with time.

In a first step, cluster separations with four through seven clusters were tested and compared for each season separately to find the optimal cluster separation, as was done with the test data set (Sect. 3.3). On the basis of whether the particular clusters were clearly separated and assignable to a source signature, and how data points were reassigned between the different cluster separations, the optimal cluster separation was chosen. The investigation of the test data set had already pointed out that the optimal number of clusters for the present data set lies at five or six clusters. Separations with four clusters or less are not satisfactory as not all identifiable types of air masses get their own cluster. On the other hand having a too large number of clusters leads to a splitting of existing clusters into two parts, e.g. the stratosphere cluster splits into two stratosphere clusters. Again results of procedure 3 were used to confirm the optimal number of clusters obtained for procedure 2. Eventually the optimal cluster separation for both seasonal data sets was achieved with five clusters. The identified air masse types or clusters are:

1. Lowermost stratospheric air
(found in summer and winter, “stratosphere cluster” (LMS), in the following graphs indicated by yellow symbols)

2. Mixed air of the tropopause region
(summer and winter, “tropopause cluster” (TP), orange symbols)

3. Air masses influenced by high clouds, probably mostly deep convective
(summer only, “high clouds cluster” (HC), blue symbols)

4. Air of the free tropospheric background
(summer and winter, “free troposphere cluster” (FT), green symbols)

5. Air masses which had contact with the boundary layer
(summer and winter, “boundary layer cluster” (BL), red symbols).

Further clusters, such as for aircraft emissions, possibly based on elevated NO_y and particle number concentration values (Hermann et al., 2008), were not found with the described statistical analysis methods in the present data set. The most likely reason for not finding such a cluster is that the aircraft density on our specific South-East Asia route, although growing rapidly in recent years, is still lower by at least a factor of two, compared to the highly dense North-Atlantic flight route (cf. e.g. the Global Aviation Emissions Inventory of the FAA, SAGE, Fig. 11, http://www.faa.gov/about/office_org/headquarters_offices/aep/models/sage/). But even there only 5–20% of the high particle number concentration peaks could be attributed to aircraft emission plumes (Hermann et al., 2008). Secondly, a minor reason, clear aircraft emission peaks in the aerosol number concentration are usually short, on the order of a few seconds only. Hence, due to the 10 s averaging in the meteo-merged files, clear and large aircraft signals might be diminished. These reasons do not exclude that on other flight routes an aircraft emission cluster might be found with the methodology described here.

4.1 Results for boreal summer

Figure 4 shows the distribution of the five clusters obtained with procedure 2 applied to the summer data set. Data points are displayed with respect to the three original variables O_3, CO, and N_{12} and in the respective cluster color. Additionally, Table 3 lists the results of procedure 2 in terms of arithmetic means, medians, and relative standard deviations for the original variables. In Table 3 (and also in Table 4) the order of clusters is arranged according to the number of data points. In summer, there is one bigger cluster, the tropopause cluster, and four approximately evenly divided medium-sized clusters.

The summer stratosphere cluster is characterized by lowest CO (32–57 ppbv) and highest O_3 values (287–483 ppbv). Here and in the following mixing ratios and concentration ranges are indicated by the 25%- and 75%-percentiles.

Table 3. Arithmetic means, medians, and standard derivations for the five-cluster separation of procedure 2 of the summer data set. The circular parameters wind direction an local time are not displayed because the respective averages are meaningless. Abbreviations for the clusters are: lowermost stratosphere = LMS, tropopause = TP, high clouds = HC, free troposphere = FT, and boundary layer = BL. The total number of data points was 15 609.

Summer			Mean					Median					Standard Deviation(%)		
Cluster name	TP	HC	LMS	BL	FT	TP	HC	LMS	BL	FT	TP	HC	LMS	BL	FT
Parameter/% of Data	30.7	19.9	17.9	16.7	14.8										
Altitude (feet)	36 316	32 298	37 124	36 179	33 906	36 457	31 528	37 452	37 038	33 538	5.1	12	3.8	6.5	7.8
Julian Day (0–365)	185	191	176	175	187	174	188	173	173	147	18	16	16	17	16
Latitude (°)	41	34	49	35	46	43	33	50	33	47	19	30	10	32	11
Longitude (°)	76	92	52	84	56	84	104	45	104	54	36	32	47	40	49
Wind Speed (kn)	52	29	47	32	28	50	27	47	26	24	54	51	46	73	60
Static Temperature (K)	227	236	225	226	227	228	240	225	227	227	2.9	4.4	1.9	3.2	3.4
Vertical Velocity (feet/min)	11.1	11.4	2.1	9.7	10.9	2.3	−1.0	0.4	2.2	−0.4	1158	954	5310	1246	1022
Static Pressure (mbar)	225	274	216	227	253	222	281	212	216	256	9.5	17	7.0	12	13
Potential Temperature (K)	348	343	349	346	336	352	347	349	351	335	2.8	3.0	2.5	3.1	1.9
H_2O_{gas} mr (ppmv)	34	688	21	215	142	28	446	17	117	88	73	102	72	175	127
H_2O_{cloud} mr (ppmv)	0	20	0	12	21	0	0	0	0	0	1478	337	2911	644	715
O_3 mr (ppbv)	134	65	378	90	130	120	67	361	78	114	43	28	29	41	32
NO_y mr (ppbv)	0.96	0.48	1.98	1.60	0.94	0.44	2.04	0.84	1.52	41	41	53	23	57	35
CO mr (ppbv)	72	92	45	100	96	70	89	48	95	95	16	16	30	23	15
N_{4-12} (particles/cm^3 STP)	783	1464	1300	6710	363	506	465	861	2825	157	145	231	161	152	163
N_{12} (particles/cm^3 STP)	2499	2028	2300	11 664	8274	2125	1687	1864	7215	4777	54	72	157	114	139
N_{18} (particles/cm^3 STP)	2136	1628	1940	9036	6958	1813	1378	1642	5729	4014	53	73	135	114	139
Potential Vorticity (PVU)	2.6	0.4	6.9	1.7	2.5	2.5	0.3	7.0	0.4	1.3	63	81	24	121	79
Eq. Pot. Temperature (K)	347	343	347	346	336	351	348	348	351	334	2.7	3.1	2.4	3.0	1.9
Vertical Wind Speed (Pa/s)	−0.4	−0.42	0.09	−0.25	0.21	0.01	−0.11	0.14	−0.05	0.19	1419	378	548	397	336
Cloud Cover (0–1)	0.02	0.24	0.01	0.17	0.05	0.00	0.12	0.00	0.03	0.00	312	114	516	146	280
CH_3CN mr (pptv)	127	112	114	144	122	125	112	115	142	121	18	19	18	21	18
CH_3COCH_3 mr (pptv)	509	776	269	940	1233	493	732	253	917	1173	41	33	65	30	31

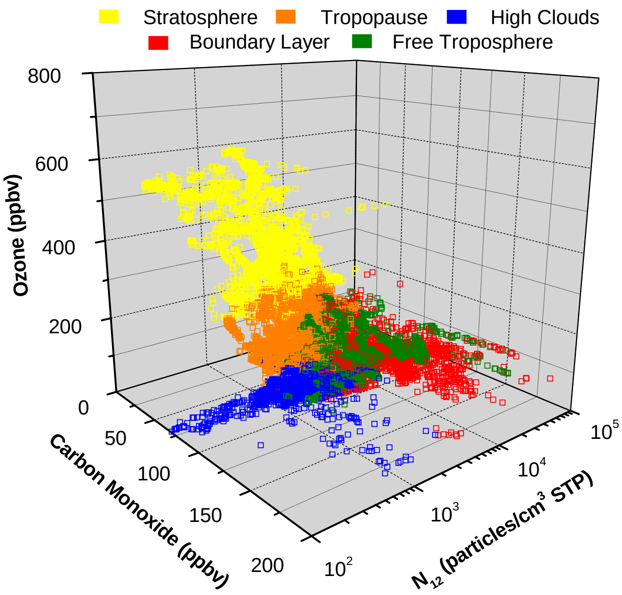

Fig. 4. O_3-CO-N_{12} scatter plot for the five-cluster separation of the summer data set (procedure 2).

Aitken mode particle number concentrations (N_{12}) are low to intermediate (1110–2820 particles/cm^3 STP). Additional indicators for stratospheric air masses are the highest averages for NO_y and PV (cf. Table 3). Furthermore, this cluster combines the data points with the lowest averages for the trace gases H_2O, CO, and CH_3COCH_3, which have their sources almost exclusively near the ground. The stratosphere cluster is the only summer cluster, which is strongly dominated by individual flight sequences, here namely the CARIBIC flights 150–153 and 194–197, which contribute each about one third of all data points of this cluster. Consequently the cluster values should be regarded typical for early summer conditions. All other individual contributions of flight sequences to a summer cluster where below 30%.

Because of the typical aircraft cruise altitudes, the tropopause cluster is the largest summer cluster and contains almost 31% of all data points (Table 3). Its O_3 values lie mainly in the range of 87–177 ppbv, CO values in the range of 63–79 ppbv. The tropopause cluster has the second highest averages for O_3 and PV, and the second lowest averages for H_2O, CO, and CH_3COCH_3 (Table 3). N_{12} values lie in the intermediate range of 1650–3140 particles/cm^3 STP. An analysis of static pressure, altitude (cf. Table 3), and vertical cross sections for the potential vorticity (http://www.knmi.

Table 4. Arithmetic means, medians, and standard derivations for the five-cluster separation of procedure 2 of the winter data set. The circular parameters wind direction an local time are not displayed because the respective averages are meaningless. Abbreviations for the clusters are: lowermost stratosphere = LMS, tropopause = TP, free troposphere = FT, and boundary layer = BL. The total number of data points was 25 111.

Winter	Mean					Median					Standard Deviation (%)				
Cluster Name	FT-1	TP	LMS	FT-2	BL	FT-1	TP	LMS	FT-2	BL	FT-1	TP	LMS	FT-2	BL
Parameter/% of Data	32.4	29.1	16.1	14.8	7.6										
Altitude (feet)	34 923	35 460	36 088	33 514	36 091	35 043	36 440	37 440	33 146	37 033	7.4	6.1	6.3	7.6	4.5
Julian Day (0–365)	357	5	39	41	317	349	349	38	67	293	54	55	48	58	48
Latitude (°)	40	44	48	37	36	43	46	49	37	34	29	17	13	25	36
Longitude (°)	73	69	52	81	83	79	69	47	92	97	44	36	48	37	43
Wind Speed (Kn)	54	52	48	61	49	49	47	47	53	47	55	55	43	59	50
Static Temperature (K)	219	217	222	223	220	218	216	222	222	220	3.6	2.8	2.5	4.0	2.5
Vertical Velocity (feet/min)	3.1	12.5	18.4	8.5	6.5	−0.2	0.5	1.5	−1.9	2.7	4377	851	741	1184	1179
Static Pressure (mbar)	241	234	227	257	227	238	223	212	260	216	13	11	11	12	8.1
Potential Temperature (K)	329	329	339	330	336	325	328	341	324	339	3.7	2.8	3.2	3.5	3.1
H_2O_{gas} mr (ppmv)	46	21	8	69	75	34	20	7	63	58	87	40	39	60	80
H_2O_{cloud} mr (ppmv)	2	0	0	2	3	0	0	0	0	0	384	1644	3481	432	400
O_3 mr (ppbv)	57	135	356	67	59	54	128	335	60	55	34	32	38	38	31
NO_y mr (ppbv)	0.33	0.72	1.23	0.36	1.11	0.31	0.73	1.11	0.31	0.93	60	34	38	55	71
CO mr (ppbv)	87	68	40	92	106	85	67	38	91	92	16	17	22	15	27
N_{4-12} (particles/cm^3 STP)	2571	665	883	239	25 152	1159	443	303	158	15 618	127	126	494	140	113
N_{12} (particles/cm^3 STP)	3671	1825	1338	2395	20 006	3083	1680	1014	1642	12 019	66	49	138	98	111
N_{18} (particles/cm^3 STP)	3018	1589	1116	2101	16 065	2627	1465	863	1454	9905	62	48	122	99	109
Potential Vorticity (PVU)	1.3	3.8	6.5	0.8	0.6	1.3	3.9	6.6	0.5	0.3	78	36	16	98	143
Eq. Pot. Temperature (K)	329	328	338	329	336	325	327	339	324	339	3.6	2.8	3.2	3.4	3.1
Vertical Wind Speed (Pa/s)	0.08	0.04	−0.01	0.08	0.10	0.14	0.03	0.03	0.13	0.17	823	1476	4014	1068	486
Cloud Cover (0–1)	0.07	0.01	0.00	0.10	0.07	0.00	0.00	0.00	0.00	0.01	225	509	995	200	186
CH_3CN mr (pptv)	89	107	112	121	126	88	104	110	117	119	23	21	19	25	39
CH_3COCH_3 mr (pptv)	446	330	147	558	698	440	320	140	518	653	32	35	50	33	35

nl/samenw/campaign_support/CARIBIC) gives further confidence that data points of this cluster were measured almost exclusively in the tropopause region.

The high clouds cluster contains the data points with the lowest O_3 values (55–75 ppbv) and medium-range CO values (84–97 ppbv). It is characterized by the highest H_2O_{gas} and ECMWF-derived vertical wind speed values (Table 3, note: negative values of the vertical wind speed indicate updraft). Furthermore, the NO_y values are the lowest, probably due to wet scavenging in deep convective clouds (e.g. Garrett et al., 2006). Interestingly, this high clouds cluster seems to have two orthogonal branches, one dominant, with low CO, CH_3COCH_3 (not shown), and N_{12} values, the other one with high CO, CH_3COCH_3 (not shown), and N_{12} values. These two branches probably represent the transport of relatively clean air and of polluted air, respectively.

The boundary layer cluster is characterized by the highest CO (87–106 ppbv) and N_{12} values (4290–13 400 particles/cm^3 STP). More importantly, it has also the highest averages for N_{4-12} and CH_3CN (Table 3), the latter being one of the best indicators for biomass burning (de Gouw et al., 2003, 2006). Admittedly, there are data points in the boundary layer cluster, which could also be attributed to the high clouds cluster, or vice versa, as for instance deep convective clouds transport boundary layer air into the UT. Nevertheless, there are clear differences in trace gas and particle concentrations, which led to the separation of the two clusters.

The free troposphere cluster comprises, as expected, the data points around the center of the point cloud in Fig. 4. It is characterized mostly by intermediate parameter values, but stands out with the second highest N_{12} and the highest CH_3COCH_3 values (Table 3).

Figure 4 displays the general problem of cluster analysis: data points are only assignable to dissimilar clusters, if they differ significantly at least in one variable. If the air masses are old and well mixed characteristic fingerprints get lost. Considering this principal problem, Fig. 4 shows a surprisingly successful separation of air masses. As additional check for the above cluster attribution, we display the geographic distribution of the clusters along the South-East Asia route in Fig. 5. Data points of the stratosphere are located north of 40° latitude, with a concentration in the region between Germany and the Ural mountains. This concentration is also indicated by the median values of 50° latitude and 45° longitude in Table 3. The tropopause cluster comprises data points from the whole sampling area, which is again indicated by the median values

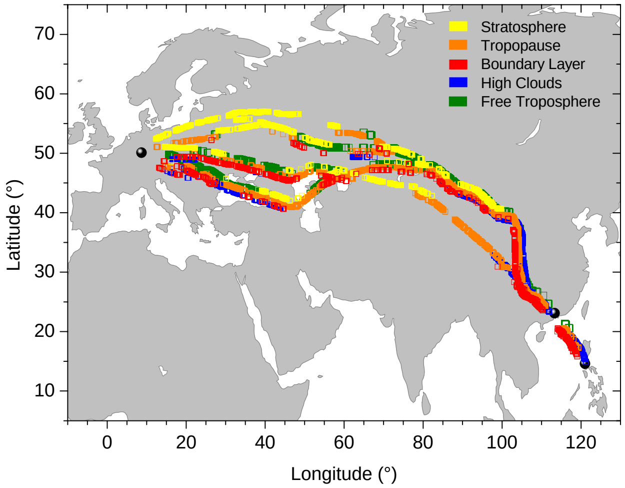

Fig. 5. Regional distribution of data points for the five-cluster separation of the summer data set (procedure 2). Colors give the attribution of the data points to its cluster. For the sake of clarity, all data points of a cluster were shifted by up to ±1° in longitude and/or latitude with respect to the flight tracks.

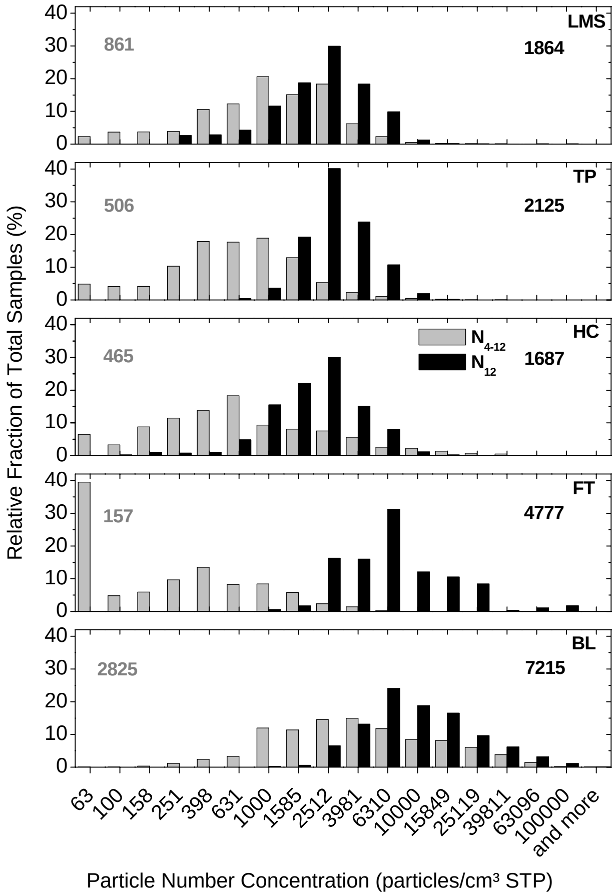

Fig. 6. N_{4-12} and N_{12} histograms for the five-cluster separation for the summer data set (procedure 2). Columns are representative for the concentration window between the particle number concentrations given below the columns and the value at the next lower tick. Numbers indicate the respective medians. Abbreviations used for the clusters are: lowermost stratosphere = LMS, tropopause = TP, high clouds = HC, free troposphere = FT, and boundary layer = BL.

of 43° latitude and 84° longitude. Indeed, vertical cross sections for the potential vorticity (http://www.knmi.nl/samenw/campaign_support/CARIBIC) show the flight track to be close to the tropopause for each summer flight, with the exception of one to two hours in the vicinity of Guangzhou. For the short flight legs between Guangzhou and Manila (high tropopause levels) the attribution of tropopause air is not as clear as for the longer flight legs. However, these tropopause data points show on average the greatest distance to the cluster centroid, and hence could be possibly attributed to another cluster when using a different clustering method. Furthermore they account only for less than 3% of the tropopause cluster.

The data points of the high clouds cluster are mainly found over China and South-East Asia (cf. latitude and longitude medians in Table 3), but also over Eastern Europe. This is in agreement with satellite data from the International Satellite Cloud Climatology Project (ISCCP), which show high frequencies particularly for deep convective clouds in summer for both regions (cf. http://isccp.giss.nasa.gov/products/browsed2.html). Likewise, the boundary layer cluster is found mainly over China plus South-East Asia and Eastern Europe. A remarkable difference is a concentration of boundary layer cluster points (∼22% of the points) over the Ukraine and southern Russia. Considering the elevated level of CH_3CN for this cluster, these signals might arise from cropland burning, which is an agricultural practice in these regions. According to Korontzi et al. (2006) there is a maximum in agriculture fire activity in the Ukraine in July/August, by field burning after the harvesting of winter and spring wheat. Indeed, MODIS satellite pictures of fire activity (http://rapidfire.sci.gsfc.nasa.gov/firemaps/) indicate such a fire maximum for July/August in the region of the boundary layer cluster points also for the years 2006 and 2007. The data points of the free troposphere cluster in summer are mainly located over Eastern Europe and Central Asia.

Finally, Fig. 6 compares the histograms of nucleation (N_{4-12}) and Aitken mode (N_{12}) particle number concentrations for the five summer clusters. These histograms and the respective particles number concentrations can be regarded as typical for the specific air masse type in the northern hemispheric mid-latitudes and subtropics over the Eurasian continent in summer. As Fig. 6 shows, N_{12} increases when going from the stratosphere (1110–2820 particles/cm^3 STP) downward into the free troposphere (3260–8270 particles/cm^3 STP). The highest concentrations in this study are found for the boundary layer cluster (4290–13 390 particles/cm^3 STP). Similar Aitken mode UT/LMS concentrations were not only found by previous CARIBIC studies for other flight routes (Hermann et al., 2003, 2008), but also by other research groups. For the LMS, Aitken mode particle number concentrations are reported in the range of 100–2500 particles/cm^3 STP (de Reus et al., 1998, 1999). For UT, over Western Europe, Schröder et al. (2002) measured somewhat lower concentrations in the range of 300–2500 particles/cm^3 STP. Singh et al. (2002) present concentrations in the range of 1000–8000 particles/cm^3 STP over the North Atlantic in fall. Similar numbers (300–10 000 particles/cm^3 STP) were observed by Minikin et al. (2003) over Western Europe.

The nucleation mode particles, N_{4-12}, can be regarded as an indicator for recent new particle formation events (cf. e.g. Hermann et al., 2003), as these particles have UT lifetimes of hours only (Williams et al., 2002). In summer, the stratosphere and the tropopause clusters show a similar range of N_{4-12} values of 260–1700 particles/cm^3 STP, whereby the tropopause cluster distribution is shifted a little bit to lower concentrations (cf. averages in Table 3). Hence particle formation takes place in these regions. The high clouds cluster yields the broadest N_{4-12} distribution, indicative for clouds acting both as particle sink and as particle source (cf. Weigelt et al., 2009). The highest N_{4-12} values are found in the boundary layer cluster (1260–7530 particles/cm^3 STP), a clear indication for particle formation in such air masses. As transport into the UT is mostly associated with clouds, and nucleation mode particles do not survive this transport in clouds, because either transport time is too long (e.g. in warm conveyor belts) or they are effectively scavenged by cloud droplets (e.g. in deep convective clouds; Ekman et al., 2006), it is very likely that the boundary layer provides only the precursor gases but particle formation takes place in the UT. On the other hand, the free troposphere cluster shows the lowest N_{4-12} values (0–450 particles/cm^3 STP). This is in agreement with vertical profiles of nucleation mode particles, which often show a C-shaped profile with high concentrations in the boundary layer and at the tropopause, but lower values inbetween (Schröder et al., 2002; Singh et al., 2002; Krejci et al., 2003). Nucleation mode particle number concentrations for the mid-latitude UT reported by Schröder et al. (2002) (50–8000 particles/cm^3 STP) and Singh et al. (2002) (1000–10 000 particles/cm^3 STP) fall into the same range as observed in this study.

4.2 Results for boreal winter

Figure 7 illustrates the results of the five cluster separation obtained with procedure 2 for the winter data set. The respective cluster means, medians, and relative standard deviations for the original variables are listed in Table 4. In winter, there are two bigger clusters, two medium-sized clusters, and one small cluster.

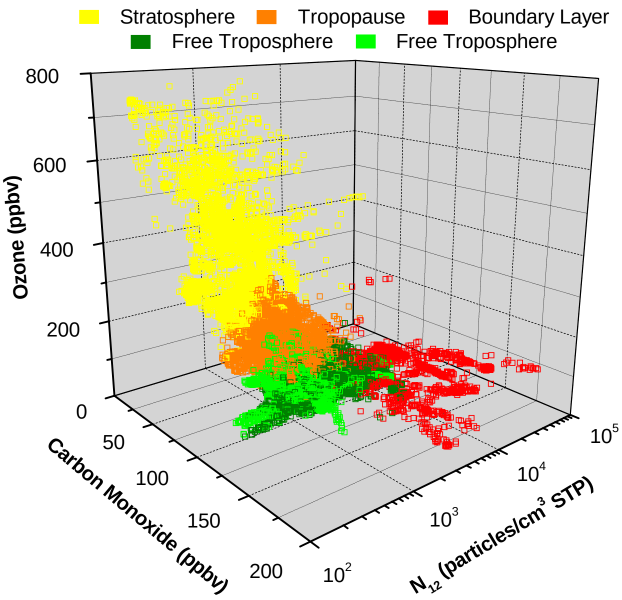

Fig. 7. O_3-CO-N_{12} scatter plot for the five-cluster separation of the winter data set (procedure 2).

The winter stratosphere cluster is of medium size with a smaller fraction of data points compared to summer. This is not expected, because having the same flight levels and flight routes as well as a lower local winter tropopause more data points should lie in the stratosphere. Apparently, in winter some of the data points with stratospheric influence were attributed to the tropopause cluster, as indicated by the higher PV averages in winter compared to summer. The stratosphere cluster combines again the highest O_3 values (249–434 ppbv) with the greatest absolute O_3 variance and the lowest CO values (34–46 ppbv). N_{12} values are low with 720–1450 particles/cm^3 STP. Again, the stratospheric cluster is characterized by the highest averages of NO_y and PV and the lowest averages of H_2O, CO, and CH_3COCH_3 (Table 4). As in the summer data, the tropopause cluster is large, has the second highest averages for O_3 and PV, and the second lowest averages for H_2O, CO, and CH_3COCH_3 (Table 4). N_{12} values lie in the intermediate range of 1250–2130 particles/cm^3 STP.

The winter boundary layer cluster is again not only characterized by the overall highest CO (84–126 ppbv) and N_{12} (7220–22 760 particles/cm^3 STP) values, but also by the highest seasonal averages in CH_3COCH_3, CH_3CN, and N_{4-12}. This cluster is the smallest winter cluster. No high clouds cluster was found for the winter data set, which can not be attributed to the frequency of cloud encounters, which is approximately equal for the summer and winter data set. The difference between the two seasons is rather caused by different cloud types and cloud properties. In summer, a larger fraction of the clouds over the continent are deep convective clouds, which change the UT air more strongly

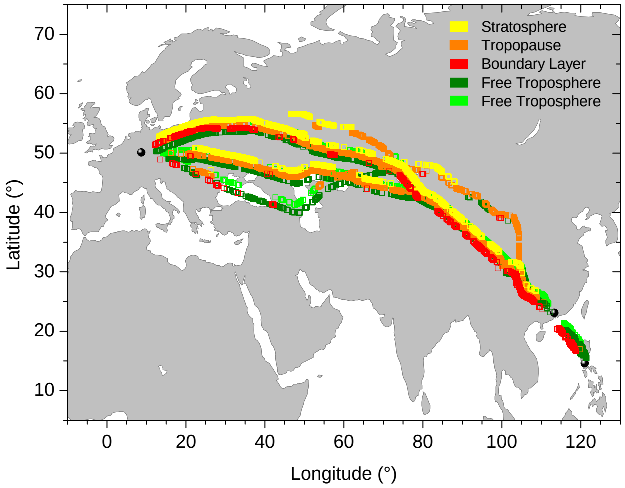

Fig. 8. Regional distribution of data points for the five-cluster separation of the winter data set (procedure 2). Same representation as in Fig. 5.

than other cloud types do. In winter, the convective activity over the Eurasian continent is smaller (cf. ISCCP; http://isccp.giss.nasa.gov/products/browsed2.html).

The remaining data points of the winter data set are distributed between a large (FT-1, dark green) and a medium-sized (FT-2, light green) free troposphere cluster. In the O_3-CO-N_{12} scatter plot, they are hardly to distinguish, which is also confirmed by similar parameter averages in Table 4. Interestingly, the majority of FT-2 data points were obtained in February/March (73%), while for FT-1 most of the data points belong to the period October–December (65%). The FT-2 cluster is dominated by one flight sequence, 224–227, which contributes about 45% of all data points to this cluster, while for the FT-1 cluster (as well as for the stratosphere and the tropopause cluster) the individual contribution of flight sequences is again below 30%.

The geographic distribution of the winter clusters on the South-East Asia route is presented in Fig. 8. Data points of the stratosphere cluster again form a data point cloud for the section between Germany and the Ural Mountains (cf. latitude and longitude medians in Table 4). But in contrast to summer, there are also stratospheric data points below 40° N, between Kazakhstan and China. This change is caused by the seasonal variation of the tropopause height. The tropopause cluster spreads again over the whole flight route, but this time there is only one data point on the short flight leg between Guangzhou and Manila. Data points of the boundary layer cluster are mainly found over China and South-East Asia (cf. latitude and longitude medians in Table 4), but there is a second concentration of data points over Eastern Europe above 52° latitude. This cluster and particular the data point cloud above 52° are strongly dominated by data from the CARIBIC flight 166 (Frankfurt–Guangzhou, 19 October 2006), which contributes about 55% of the cluster data points. This flight is characterized by strong pollution plumes and was already discussed by Slemr et al. (2009). The respective air masses over Eastern Europe probably have their origin over the Iberian Peninsula, where a low pressure system over the western Atlantic Ocean together with convective activity over the Iberian Peninsula let to a transport of boundary-layer-influenced air into the upper troposphere (http://www.knmi.nl/samenw/campaign_support/CARIBIC; http://www.wetterzentrale.de). The light and the dark green free troposphere clusters describe almost 60% of all data points of the winter data set. They are scattered over the whole sampling area, whereby the medium sized-cluster (FT-2, light green) is concentrated more over Central Asia, China, and South-East Asia. This is indicated also by the lower latitude and higher longitude medians in Table 4 compared to the dark green cluster (FT-1). Furthermore, FT-2 shows the higher H_2O, CH_3COCH_3, and CH_3CN mixing ratios of both clusters, but the lower N_{4-12} and N_{12} values, while the other trace gases are nearly similar. Evidently, the cluster separation for winter is less satisfactory than for summer, because a clear attribution to a "source" is not easily done. There are two possible reasons for this difference. First, stronger vertical transport over the continent in summer due to high-reaching deep convective clouds could lead to younger air masses and hence clearer air mass characteristics. Secondly, the lower chemical reactivity of most trace gases in winter leads to a more homogeneous distribution and hence less clear air mass characteristics.

Finally, in Fig. 9 the histograms of N_{4-12} and N_{12} for the five winter clusters are displayed. Similar to summer, the N_{12} values increase from the stratosphere into the troposphere, however concentrations are generally lower. In the stratosphere only 720–1450 particles/cm^3 STP are found, in the free troposphere 1100–4560 particles/cm^3 STP. The boundary layer cluster has the highest N_{12} values with 7220–22 760 particles/cm^3 STP. This cluster comprises also the highest N_{4-12} values with 6240–29 500 particles/cm^3 STP. In the stratosphere, at the tropopause, and also partly in the free troposphere particle formation appears to be less frequent, given that N_{4-12} values are mostly below 850 particles/cm^3 STP (75%-percentile). Only in the larger free troposphere cluster (FT-1, dark green) nucleation events are more frequent, indicated by concentrations typically in the range of 530–3930 particles/cm^3 STP. Again, particle number concentrations are in agreement with previous CARIBIC studies and literature values (e.g. Young et al., 2007).

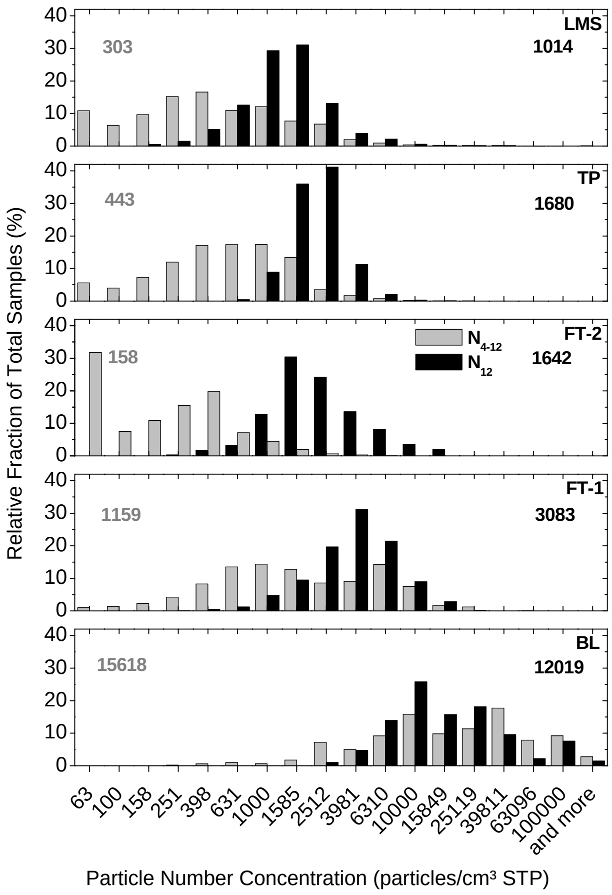

Fig. 9. N_{4-12} and N_{12} histograms for the five cluster separation of the winter data set (procedure 2). Same representation as in Fig. 6.

5 Summary

The origin of nucleation mode and Aitken mode particles in the upper troposphere and lowermost stratosphere (UT/LMS) at northern mid-latitudes and subtropics was investigated using cluster analysis methods. The analyses are based on in situ measurements of trace gases and aerosol particles conducted during flights between Germany and South-East Asia with the CARIBIC (Civil Aircraft for Regular Investigation of the Atmosphere Based on an Instrument Container) measurement system. Four cluster analysis methods were applied to the data set and were tested for their capability to separate the data points into meaningful clusters. With the preferred method applied to seasonal data subsets for summer and winter five cluster or air mass types were identified: stratosphere, tropopause, high clouds, free troposphere, and boundary layer influenced. Other source clusters, like aircraft emissions, were not identified in the present data set with the used methods. While the cluster separation for summer works satisfactory well, the results for winter are less clear and interpretation is more difficult. This difference is either caused by the different vertical transport pathways or the different chemical lifetimes of trace gases in the two seasons. Figure 10 summarizes the results of the cluster analysis by showing the relative frequency of cluster or air mass types in 8–12 km altitude along the South-East Asia route (Fig. 1). The data represent the two years period early summer 2006 to late winter 2008, and might look slightly differently for other years. As the diagram represents also a north south distribution, because of the flight route, data points from the stratosphere and the tropopause region are concentrated more to the left and high clouds data points more to the right.

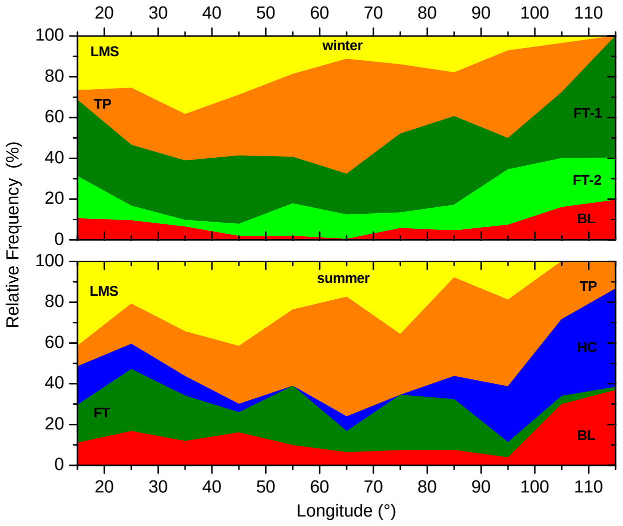

Fig. 10. Relative frequency of cluster types along the South-East Asia route. Data points were partitioned into 10° longitude bins. Cluster abbreviations are: lowermost stratosphere = LMS, tropopause = TP, high clouds = HC, free troposphere = FT, and boundary layer = BL.

This study presents the first seasonal submicrometer particle number concentration data sets in the UT/LMS over the Eurasian continent (Eastern Europe, Central Asia, East Asia, and South-East Asia) and analyses the origin of these particles. Aitken mode particle number concentrations show a clear vertical gradient with the lowest values in the lowermost stratosphere (750–2820 particles/cm^3 STP, minimum of the two 25%- and maximum of the two 75%-percentiles of both seasons) and the highest values for the boundary-layer-influenced air (4290–22 760 particles/cm^3 STP). Nucleation mode particles are also highest in the boundary-layer-influenced air (1260–29 500 particles/cm^3 STP, but are lowest in the free troposphere (0–450 particles/cm^3 STP), in agreement with previous studies. The obtained particle number concentrations for air mass types can be useful as input parameters for and for validating the results of global atmospheric aerosol models.

This study is the first step in the effort to identify the origin of aerosol particles in the UT/LMS by using multivariate statistical analysis methods. In the future, the results could be improved by enlarging the analyzed data set using data from other CARIBIC routes. By including trajectory information as well as hydrocarbon and halocarbon trace gas mixing ratios obtained from the CARIBIC whole air sampler, additional information on the particle origin might be derived. However, as the merging of a low-resolution, high-information-content datasets with the high-resolution datasets used here, increases the amount of data handling and computational burden enormously, this additional step was beyond the scope of this work.

Appendix A

List of abbreviations

CA = cluster analysis
CARIBIC = Civil Aircraft for Regular Investigation of the Atmosphere based on an Instrument Container
ECMWF = European Centre for Medium-Range Weather Forecasts
HCA = hierarchical cluster analysis
KCA = K-means cluster analysis
LMS = lowermost stratosphere
mr = mixing ratio
MSA = measure of sampling adequacy
PC = principle component
PCA = principle component analysis
ppbv = parts per billion by volume (10^{-9})
ppmv = parts per million by volume (10^{-6})
pptv = parts per trillion by volume (10^{-12})
PV = potential vorticity
PVU = potential vorticity unit (10^{-6} K m^2 kg^{-1} s^{-1})
SPSS = Statistical Package for the Social Sciences
STP = standard temperature and pressure (273.15 K, 1013.25 hPa)
UT = upper troposphere

Acknowledgements. We are deeply grateful to Andreas Waibel and Thomas Dauer (Lufthansa Passage and Lufthansa Technik) for their support of CARIBIC and to Dieter Scharffe and Claus Koeppel (Max Planck Institute for Chemistry) for the equipment support and container operation. The development of the CARIBIC system was financially supported by the German Ministry of Education and Science (AFO 2000 program) and its operation benefits from the European Commission's DGXII Environment RTD 4th, 5th, 6th, and 7th Framework programs.

Edited by: M. Petters

References

Backhaus, K., Erichson, B., Plinke, W. and Weiber, R.: Multivariate Analysemethoden, 11th edition, Springer, Berlin, 830 pp., 2006.

Bell, N., Koch, D., and Shindell, D. T.: Impacts of chemical-aerosol coupling on tropospheric ozone and sulfate simulations in a general circulation model, J. Geophys. Res., 110, D14305, doi:10.1029/2004JD005538, 2005.

Borchi, F. and Marenco, A.: Discrimination of air masses near the extratropical tropopause by multivariate analyses from MOZAIC data, Atmos. Environ., 36, 1123–1135, 2002.

Borchi, F., Oikonomou, E., and Marenco, A.: Extratropical case study of stratosphere-troposphere exchange using multivariate analyses from MOZAIC aircraft data, Atmos. Environ., 39, 6537–6549, 2005.

Brenninkmeijer, C. A. M., Crutzen, P. J. , Fischer, H., Güsten, H., Hans, W., Heinrich, G., Heintzenberg, J., Hermann, M., Immelmann, T., Kersting, D., Maiss, M., Nolle, M., Pitscheider, A., Pohlkamp, H., Scharffe, D., Specht, K., and Wiedensohler, A.: CARIBIC civil aircraft for global measurement of trace gases and aerosols in the tropopause region, J. Atmos. Ocean. Tech., 16, 1373–1383, 1999.

Brenninkmeijer, C. A. M., Crutzen, P., Boumard, F., Dauer, T., Dix, B., Ebinghaus, R., Filippi, D., Fischer, H., Franke, H., Frieß, U., Heintzenberg, J., Helleis, F., Hermann, M., Kock, H. H., Koeppel, C., Lelieveld, J., Leuenberger, M., Martinsson, B. G., Miemczyk, S., Moret, H. P., Nguyen, H. N., Nyfeler, P., Oram, D., O'Sullivan, D., Penkett, S., Platt, U., Pupek, M., Ramonet, M., Randa, B., Reichelt, M., Rhee, T. S., Rohwer, J., Rosenfeld, K., Scharffe, D., Schlager, H., Schumann, U., Slemr, F., Sprung, D., Stock, P., Thaler, R., Valentino, F., van Velthoven, P., Waibel, A., Wandel, A., Waschitschek, K., Wiedensohler, A., Xueref-Remy, I., Zahn, A., Zech, U., and Ziereis, H.: Civil Aircraft for the regular investigation of the atmosphere based on an instrumented container: The new CARIBIC system, Atmos. Chem. Phys., 7, 4953-4976, 2007,
http://www.atmos-chem-phys.net/7/4953/2007/.

Brenninkmeijer, C. A. M.: http://www.caribic-atmospheric.com, last access: April 2009.

Brosius, F.: SPSS 14, Das mitp-Standardwerk, 1st edition, Mitp-Verlag, Bonn, 1056 pp., 2006.

de Gouw, J. A., Warneke, C., Parrish, D. D., Holloway, J. S., Trainer, M., and Fehsenfeld, F. C.: Emission sources and ocean uptake of acetonitrile (CH_3CN) in the atmosphere, J. Geophys. Res., 108, 4329, doi:10.1029/2002JD002897, 2003.

de Gouw, J. A., Warneke, C., Stohl, A., Wollny, A. G., Brock, C. A., Cooper, O. R., Holloway, J. S., Trainer, M., Fehsenfeld, F. C., Atlas, E. L., Donnelly, S. G., Stroud, V., and Lueb, A.: Volatile organic compounds composition of merged and aged forest fire plumes from Alaska and western Canada, J. Geophys. Res., 110, D10303, doi:10.1029/2005JD006175, 2006.

de Reus, M., Ström, J., Kulmala, M., Pirjola, L., Lelieveld, J., Schiller, C., and Zöger, M.: Airborne aerosol measurements in the tropopause region and the dependence of new particle formation on preexisting particle number concentration, J. Geophys. Res., 103, 31255–31263, 1998.

de Reus, M., Ström, J., Hoor, P., Lelieveld, J., and Schiller, C.: Particle production in the lowermost stratosphere by convective lifting of the tropopause, J. Geophys. Res., 104, 23935–23940, 1999.

Dorling, S. R. and Davis, T. D.: Extending cluster analysis – synoptic meteorology links to characterise chemical climates at six northwest European monitoring stations, Atmos. Environ., 29(2), 145–167, 1995.

Ekman, A. M. L., Wang, C., Ström, J., and Krejci, R.: Explicit simulation of aerosol physics in a cloud-resolving model: Aerosol transport and processing in the free troposphere, J. Atmos. Sci., 63, 682–696, 2006.

Fine, S. S.: http://www.arl.noaa.gov, last access: February 2008.

Garrett, T. J., Avey, J., Palmer, P. I., Stohl, A., Neuman, J. A., Brock, C. A., Ryerson, T. B., and Holloway, J. S.: Quantifying wet scavenging processes in aircraft observations of nitric acid and cloud condensation nuclei, J. Geophys. Res., 111, D23S51, doi:10.1029/2006JD007416, 2006.

Gerstengarbe, F.-W., Werner, P. C., and Rüge, U.: Katalog der Großwetterlagen Europas (1981–1998), nach Paul Hess und Helmuth Brezowski, 5th edition, Berichte des Deutschen Wetterdienstes 113, Potsdam/Offenbach a. M., 138 pp., 1999a.

Gerstengarbe, F.-W., Werner, P. C., and Friedrich, K.: Applying non-hierarchical cluster analysis algorithms to climate classification: some problems and their solution, Theor. Appl. Climatol., 64, 143–150, 1999b.

Hains, J. C., Taubman, B. F., Thompson, A. M., Stehr, J. W., Marufu, L. T., Doddridge, B. G., and Dickerson, R. R.: Origins of chemical pollution derived from Mid-Atlantic aircraft profiles using a clustering technique, Atmos. Environ., 42, 1727–1741, 2008.

Heintzenberg, J. and Charlson, R. J.: Editors, Clouds in the Perturbed Climate System Their Relationship to Energy Balance, Atmospheric Dynamics, and Precipitation, MIT Press, Cambridge, MA, 576 pp., 2009.

Hermann, M. and Wiedensohler, A.: Counting efficiency of condensation particle counters at low-pressures with illustrative data from the upper troposphere, J. Aerosol Sci., 32, 975–991, 2001.

Hermann, M., Heintzenberg, J., Wiedensohler, A., Brenninkmeijer, C. A. M., Heinrich, G., and Zahn, A.: Meridional distributions of aerosol particle number concentrations in the upper troposphere and lower stratosphere obtained by CARIBIC flights, J. Geophys. Res., 108, 4114, doi:10.1029/2001JD001077, 2003.

Hermann, M., Brenninkmeijer, C. A. M., Slemr, F., Heintzenberg, J., Martinsson, B. G., Schlager, H., van Velthoven, P. F. J., Wiedensohler, A., Zahn, A., and Ziereis, H.: Submicrometer aerosol particle distributions in the upper troposphere over the mid-latitude north Atlantic – results from the third route of CARIBIC, Tellus, 60(B), 106–117, 2008.

Hoor, P., Fischer, H., Lange, L., Lelieveld, J., and Brunner, D.: Seasonal variations of mixing layer in the lowermost stratosphere as identified by the CO-O_3 correlation form in situ measurements, J. Geophys. Res., 107, 4044, doi:10.1029/2000JD000289, 2002.

Hoor, P., Fischer, H., and Lelieveld, J.: Tropical an extratropical tropospheric air in the lowermost stratosphere over Europe: A CO-based budget, Geophys. Res. Lett., 32, L07802, doi:10.1029/2004GL022018, 2005.

IPCC, 2007: Climate Change 2007: Synthesis Report, Contribution of Working Groups I, II and III to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change (Core Writing Team, edited by: Pachauri, R. K. and Reisinger, A.), IPCC, Geneva, Switzerland, 104 pp., 2007.

Kärcher, B.: Simulating gas-aerosol-cirrus interactions: Process-oriented microphysical model and applications, Atmos. Chem. Phys., 3, 1645–1664, 2003, http://www.atmos-chem-phys.net/3/1645/2003/.

Kim, B.: http://www.faa.gov/about/office_org/headquarters_offices/aep/models/sage/, last access: August 2009.

Kim, D., Wang, C., Ekman, A. M. L., Barth, M. C., and Rasch, P. J.: Distribution and direct radiative forcing of carbonaceous and sulfate aerosols in an interactive size-resolving aerosol-climate model, J. Geophys. Res., 113, 16309, doi:10.1029/2007JD009756, 2008.

Korontzi, S., McCarty, J., Loboda, T., Kumar, S., and Justice, C.: Global distribution of agricultural fires in croplands from 3 years of Moderate Resolution Imaging Spectroradiometer (MODIS) data, Global Biogeochem. Cy., 20, GB2021, doi:10.1029/2005GB002529, 2006.

Krejci, R., Ström, J., de Reus, M., Hoor, P., Williams, J., Fischer, H., and Hansson, H.-C.: Evolution of aerosol properties over the rain forest in Surinam, South America, observed from aircraft during the LBA-CLAIRE 98 experiment, J. Geophys. Res., 108(D18), 4561, doi:10.1029/2001JD001375, 2003.

Kunz, A., Schiller, C., Rohrer, F., Smit, H. G. J., Nedelec, P., and Spelten, N.: Statistical analysis of water vapour and ozone in the UT/LS observed during SPURT and MOZAIC, Atmos. Chem. Phys., 8, 6603–6615, 2008, http://www.atmos-chem-phys.net/8/6603/2008/.

Leyer, I. and Wesche, K.: Multivariate Statistik in der Ökologie, 1st edition, Springer, Berlin, 221 pp., 2007.

Liljequist G. H. and Cehak, K.: Allgemeine Meteorologie, 3rd edition, vieweg, Braunschweig/Wiesbaden, 412 pp., 1994.

Liu, D., Wang, Z., Liu, Z., Winker, D., and Trepte, C.: A height resolved global view of dust aerosols from the first year CALIPSO lidar measurements, J. Geophys. Res., 113, D16214, doi:10.1029/2007JD009776, 2008.

Lohmann, U. and Feichter, J.: Global indirect aerosol effects: a review, Atmos. Chem. Phys., 5, 715–737, 2005, http://www.atmos-chem-phys.net/5/715/2005/.

Malberg, H.: Meteorologie und Klimatologie: Eine Einführung, 3rd edition, Springer, Berlin, 354 pp., 1997.

Masuoka, E.: http://rapidfire.sci.gsfc.nasa.gov/firemaps/, last access: April 2009.

Minikin, A., Petzold, A., Ström, J., Krejci, R., Seifert, M., van Velthoven, P., Schlager, H., and Schumann, U.: Aircraft observations of the upper tropospheric fine particle aerosol in the northern and southern hemisphere at midlatitudes, Geophys. Res. Lett., 30(10), 1503, doi:10.1029/2002GL016458, 2003.

Müller, G.: http://www.wetterzentrale.de, last access: February 2008.

Nie, N. H.: http://www.spss.com, last access: February 2008.

Nguyen, H. N., Gudmundsson, A., and Martinsson, B. G.: Design and calibration of a multi-channel aerosol sampler for studies of the tropopause region from the CARIBIC platform, Aerosol Sci. Tech., 40, 649–655, 2006.

Nguyen, H. N. and Martinsson, B. G.: Analysis of C, N, and O in aerosol collected on an organic backing using internal blank measurements and variable beam size, Nucl. Instrum. Methods Phys. Res., Sect. B, 264, 96-109, doi:10.1016/j.nimb.2007.08.001, 2007.

Papaspiropoulos, G., Martinsson, B. G., Zahn, A., Brenninkmeijer, C. A. M., Hermann, M., Heintzenberg, J., Fischer, H., and van Velthoven, P. F. J.: Aerosol elemental concentrations in the tropopause region from intercontinental flights with the Civil Aircraft for Regular Investigation of the Atmosphere Based on an Instrument Container (CARIBIC) platform, J. Geophys. Res., 107(D23), 4671, doi:10.1029/2002JD002344, 2002.

Peylin, P., Bréon, F. M., Serrar, S., Tiwari, Y., Chédin, A., Gloor, M., Machida, T., Brenninkmeijer, C. A. M., Zahn, A., and Ciais, P.: Evaluation of Television Infrared observation Satellite (TIROSN) Operational Vertical Sounder (TOVS) spaceborne O_2 estimates using model simulations and aircraft data, J. Geophys. Res., 112, D09313, doi:10.1029/2005JD007018, 2007.

Rossow, W. B.: http://isccp.giss.nasa.gov/products/browsed2.html, last access: April 2009.

Schröder, F., Kärcher, B., Fiebig, M., and Petzold, A.: Aerosol states in the free troposphere at northern midlatitudes, J. Geophys. Res., 107(21), 8126, doi:10.1029/2000JD000194, 2002.

Seinfeld, J. H. and Pandis, S. N.: Atmospheric Chemistry and Physics: From Air Pollution to Climate Change, 2nd edition, J. Wiley, New York, 1232 pp., 2006.

Singh, H. B., Anderson, B. E., Avery, M. A., Viezee, W., Chen, Y., Tabazadeh, A., Hamill, P., Pueschel, R., Fuelberg, H. E., and Hannan, J. R.: Global distribution and sources of volatile and nonvolatile aerosol in the remote troposphere, J. Geophys. Res., 107(11), 4121, doi:10.1029/2001JD000486, 2002.

Slemr, F., Ebinghaus, R., Brenninkmeijer, C. A. M., Hermann, M., Kock, H. H., Martinsson, B. G., Schuck, T., Sprung, D., van Velthoven, P., Zahn, A., and Ziereis, H.: Gaseous mercury distribution in the upper troposphere and lower stratosphere observed onboard the CARIBIC passenger aircraft, Atmos. Chem. Phys., 9, 1957–1969, 2009, http://www.atmos-chem-phys.net/9/1957/2009/.

Søvde, O. A., Gauss, M., Isaksen, I. S. A., Pitari, G., and Marizy, C.: Aircraft pollution - a futuristic view, Atmos. Chem. Phys., 7, 3621–3632, 2007, http://www.atmos-chem-phys.net/7/3621/2007/.

Taubman, B. F., Hains, J. C., Thompson, A. M., Marufu, L. T., Doddridge, B. G., Stehr, J. W., Piety, C. A., and Dickerson, R. R.: Aircraft vertical profiles of trace gas and aerosol pollution over the Mid-Atlantic United States: statistics and meteorological cluster analysis, J. Geophys. Res., 111(D10), D10S07, doi:10.1029/2005JD006196, 2006.

Textor, C., Schulz, M., Guibert, S., Kinne, S., Balkanski, Y., Bauer, S., Berntsen, T., Berglen, T., Boucher, O., Chin, M., Dentener, F., Diehl, T., Feichter, J., Fillmore, D., Ginoux, P., Gong, S., Grini, A., Hendricks, J., Horowitz, L., Huang, P., Isaksen, I. S. A., Iversen, T., Kloster, S., Koch, D., Kirkevåg, A., Kristjansson, J. E., Krol, M., Lauer, A., Lamarque, J. F., Liu, X., Montanaro, V., Myhre, G., Penner, J. E., Pitari, G., Reddy, M. S., Seland, Ø., Stier, P., Takemura, T., and Tie, X.: The effect of harmonized emissions on aerosol properties in global models - an AeroCom experiment, Atmos. Chem. Phys., 7, 4489–4501, 2007, http://www.atmos-chem-phys.net/7/4489/2007/.

Trepte, C. R.: http://www-calipso.larc.nasa.gov, last access: April 2009.

Weigelt, A., Hermann, M., van Velthoven, P. F. J., Brenninkmeijer, C. A. M., Schlaf, G., Zahn, A., and Wiedensohler, A.: Influence of clouds on aerosol particle number concentrations in the upper troposphere, J. Geophys. Res., 114, D01204, doi:10.1029/2008JD009805, 2009.

Williams, J., de Reus, M., Krejci, R., Fischer, H., and Ström, J.: Application of the variability-size relationship to atmospheric aerosol studies: estimating aerosol lifetimes and ages, Atmos. Chem. Phys., 2, 133–145, 2002, http://www.atmos-chem-phys.net/2/133/2002/.

Young, L.-H., Benson, D. R., Montanaro, W. M., Lee, S.-H., Pan, L. L., Rogers, D. C., Jensen, J., Stith, J. L., Davis, C. A., Campos, T. L., Bowman, K. P., Cooper, W. A., and Lait, L. R.: Enhanced new particle formation observed in the northern midlatitude tropopause region, J. Geophys. Res., 112, D10218, doi:10.1029/2006JD008109, 2007.

Zahn, A., Brenninkmeijer, C. A. M., Asman, W. A. H., Crutzen, P. J., Heinrich, G., Fischer, H., Cuijpers, J. W. M., and van Velthoven, P. F. J.: The budget of O_3 and CO in the upper troposphere: The CARIBIC passenger aircraft results 1997–2001, J. Geophys. Res., 107(D17), 4337, doi:10.1029/2001JD001529, 2002.

Zahn, A. and Brenninkmeijer, C. A. M.: New Directions: A new tropopause defined, Atmos. Environ., 37, 439–440, 2003.