

Meteorological and trace gas factors affecting the number concentration of atmospheric Aitken ($D_p = 50$ nm) particles in the continental boundary layer: parameterization using a multivariate mixed effects model

S. Mikkonen¹, H. Korhonen^{1,2}, S. Romakkaniemi¹, J. N. Smith^{1,2,3}, J. Joutsensaari¹, K. E. J. Lehtinen^{1,2}, A. Hamed¹, T. J. Breider⁴, W. Birmili⁵, G. Spindler⁵, C. Plass-Duelmer⁶, M. C. Facchini⁷, and A. Laaksonen^{1,8}

¹Department of Applied Physics, University of Eastern Finland, P.O.Box 1627, 70211 Kuopio, Finland

²Finnish Meteorological Institute, Kuopio unit, P.O.Box 1627, 70211 Kuopio, Finland

³Atmospheric Chemistry Division, National Center for Atmospheric Research, P.O.Box 3000, Boulder, Colorado 80307, USA

⁴School of Earth and Environment, University of Leeds, LS2 9JT Leeds, UK

⁵Leibniz Institute for Tropospheric Research, Permoserstrasse 15, 04318 Leipzig, Germany

⁶Hohenpeissenberg Meteorological Observatory, Hohenpeissenberg, Germany

⁷Institute of Atmospheric Sciences and Climate, ISAC – CNR, Consiglio Nazionale delle Ricerche, Bologna, Italy

⁸Finnish Meteorological Institute, P.O.Box 503, 00101 Helsinki, Finland

Received: 1 July 2010 – Published in Geosci. Model Dev. Discuss.: 13 August 2010

Revised: 19 November 2010 – Accepted: 16 December 2010 – Published: 11 January 2011

Abstract. Measurements of aerosol size distribution and different gas and meteorological parameters, made in three polluted sites in Central and Southern Europe: Po Valley, Italy, Melpitz and Hohenpeissenberg in Germany, were analysed for this study to examine which of the meteorological and trace gas variables affect the number concentration of Aitken ($D_p = 50$ nm) particles. The aim of our study was to predict the number concentration of 50 nm particles by a combination of in-situ meteorological and gas phase parameters. The statistical model needs to describe, amongst others, the factors affecting the growth of newly formed aerosol particles (below 10 nm) to 50 nm size, but also sources of direct particle emissions in that size range. As the analysis method we used multivariate nonlinear mixed effects model. Hourly averages of gas and meteorological parameters measured at the stations were used as predictor variables; the best predictive model was attained with a combination of relative humidity, new particle formation event probability, temperature, condensation sink and concentrations of SO_2 , NO_2 and ozone. The

seasonal variation was also taken into account in the mixed model structure. Model simulations with the Global Model of Aerosol Processes (GLOMAP) indicate that the parameterization can be used as a part of a larger atmospheric model to predict the concentration of climatically active particles. As an additional benefit, the introduced model framework is, in theory, applicable for any kind of measured aerosol parameter.

1 Introduction

It is commonly known that atmospheric aerosols have a great effect on the radiation budget, formation of clouds, climate change and human health (IPCC, 2007; Kerminen et al., 2005; Pope and Dockery, 2006). Acquiring a quantitative understanding of the Cloud Condensation Nucleus (CCN) production from different natural and anthropogenic sources is one of the key topics in aerosol research (Wiedensohler et al., 2009). Theoretical frameworks have been derived to investigate the efficacy by which nucleated particles produce CCN in the atmosphere (e.g. Pierce and Adams, 2007; Kuang et al., 2008) and recent atmospheric measurements and modelling studies have shown that newly formed particles



Correspondence to: S. Mikkonen
(santtu.mikkonen@uef.fi)

can affect concentrations of CCN (Ghan et al., 2001; Lihavainen et al., 2003; Kerminen et al., 2005). Laaksonen et al. (2005) suggested that new particle formation (NPF) might be an important source of CCN even in polluted environments with strong primary particle emissions; the growth from nucleation to CCN-size may take only few hours if enough condensable vapours are available.

In order to obtain better predictions with global or regional climate models an adequate description of aerosol dynamics is needed. In global scale atmospheric models, modelling aerosol processes is necessarily a compromise between the accurate description of microphysical processes on the one hand, and computational efficiency – requiring simplified parameterisations of these processes, on the other hand. Simplified parameterisations of the size distribution evolution as well as the chemical composition of aerosol populations could help to reduce computational time efficiently (Dusek et al., 2006; Kokkola et al., 2009). Typically the aerosol particle size distribution is modelled using either a modal approach (e.g. Whitby and McMurry, 1997) or a sectional approach employing relatively few size bins. These methods are not very well suited for predicting new particle formation and growth to CCN sizes (Korhonen et al., 2003), and methods that can directly predict particle concentrations at climate relevant sizes would be of great advantage. Aerosols have two principal impacts on climate: particles larger than 100 nm reflect and absorb radiation (direct effect) and particles of sizes starting from 50 nm can act as CCN (indirect effect). Particles smaller than 50 nm do not have significant effect on climate; a detailed description of their dynamics in large scale climate models would be a waste of computational resources. In addition, there has been no uniform theoretical description of the atmospheric particle nucleation and growth process, so a parameterisation of the concentrations of 50 nm particles could circumvent some of the uncertainties associated with that lack of knowledge. A prime motivation for the current study was to derive a statistical relationship between meteorological, trace gas, and aerosol properties and 50 nm particle concentrations. The resulting parameterisation could then be used in large scale models so that no computing resources go into aerosol dynamics below 50 nm. We do not suggest that this population of 50 nm particles are themselves climatically important: this depends on factors such as temperature and water vapour supersaturation as well as aerosol physico-chemical properties. These factors, combined with aerosol dynamical processes such as coagulation and evaporation/condensation, ultimately determine the climatic impact of this population and are expected to be calculated in the model. Thus, by incorporating our parameterization modelers have an opportunity to lower the computational cost of their calculation without sacrificing accuracy; environmental conditions and aerosol chemical properties from the model will ultimately determine the direct and indirect impacts of these particles.

The use of advanced statistical tools has been rare in the analysis of the formation and growth of atmospheric aerosols. Kulmala et al. (2004) studied several physical and chemical properties affecting particle growth, and statistical properties of NPF events have been studied in Hyvönen et al. (2005) and in Mikkonen et al. (2006), but factors affecting the growth of newly formed particles have not undergone careful statistical studies due in large part to the fact that the exact processes by which this growth occurs are poorly understood. Since simple yet comprehensive mechanisms are currently not available for these processes, statistical analyses such as ours can be used to develop effective representations of the causalities and interdependencies of the gases and particles in the atmosphere.

The problem in using general data analysis methods is that the measurement data is not normally distributed and typically contains different autocorrelation structures. In this manuscript, we describe an advanced statistical data analysis method that takes into account the structure of the data and uses it to find predictors and indicators for the number concentration of particles at a selected size. The study was carried out in an explorative manner, i.e., we did not make limiting preconceived assumptions about which variables should be included in the analysis but we used all the variables measured in our three measurement sites. This ensures that no significant variables were left out from the analysis due to any limitations. The method presented here is applicable for any other dataset from different sites.

The main objectives of the study are to find the factors affecting the growth to, and primary production of, particles that can be considered the minimum potential CCN size and to find a parameterization which can be used as a part of a larger atmospheric model to predict the concentration of climatically active particles. In this article we present a multiple-site, and multi-annual statistical analysis that led to the identification of important parameters the size distribution function at 50 nm, $dN/d\log D_p|_{50\text{ nm}}$. Following this we present the optimized parameterization that incorporates the observations from all three focus sites. The results of the multivariate mixed model are compared to data from these sites to quantify the predictive ability of the model. In a final step, we investigate the usefulness of the statistical model by incorporating the developed statistical parameterization into the Global Model of Aerosol Processes (GLOMAP).

2 Methods

2.1 Atmospheric particle data

The present paper concentrates on the description of Aitken particles in the troposphere of Central Europe and Northern Italy. As a measure for the concentration of Aitken mode particles we use the particle number density

of these distributions at $D_p = 50 \text{ nm}$ $dN/d\log D_p|_{50\text{nm}}$, denoted N_{50} hereafter. Particle number size distributions were collected at the measurement sites of San Pietro Capofiume (SPC; Po Valley, Italy), Melpitz (East Germany), and Hohenpeissenberg (South Germany) using Differential Mobility Particle Sizer (DMPS) instruments. DMPS systems were set to measure the dry mobility diameter of the particles. The mentioned regions are characterised by a population density of 50–200 inhabitants km^{-2} , and feature substantial anthropogenic gaseous and particulate emissions from diffuse sources such as industry, domestic heating and traffic.

In the continental troposphere, Aitken particles often occur as a distinct particle mode, with mean diameters between 45 and 90 nm (Birmili et al., 2001). Aitken mode particles are, in the most general picture, a mixture between aged secondary particles originating from gas-to-particle conversion, or new particle formation (NPF), and “primary” particles emitted directly from anthropogenic sources. NPF, which is ultimately the result of photochemical processes, has been observed at all three sites under study (Jaatinen et al., 2009; Paasonen et al., 2009), and is expected to influence the concentrations of Aitken particles via particle growth induced by condensation and coagulation. A detailed principal component analysis of particle number size distributions in the Leipzig/Melpitz area yielded at least four statistically independent sources of Aitken particles at the rural observation site Melpitz, including aged nucleation mode particles from regional-scale secondary formation, two types of particles originating from anthropogenic, mainly urban particle emissions, and a fourth (though less significant) type deriving from long-range transport (Costabile et al., 2009). A one-year characterisation of the hygroscopic properties of atmospheric particles at Melpitz confirmed the presence of different particle types at $D_p = 50 \text{ nm}$, exhibited by their different hygroscopic properties (Kinder, 2010). Three hygroscopicity classes were identified: hydrophobic particles, associated with fresh direct anthropogenic emissions, accounted for 7–35% of the particle number at 50 nm, depending on season and the large-scale weather situation. Less hygroscopic and more hygroscopic particles, which are broadly associated with secondary particles at different stages of the atmospheric ageing process, accounted for 12–54% and 26–68% of the particle number at 50 nm, respectively. As a summary, it is necessary to consider tropospheric Aitken particle ($D_p = 50 \text{ nm}$) concentrations as being influenced by a variety of source processes.

We classified NPF events into three classes using the visual methods described in Hamed et al. (2007). A day is considered an event day if the formation of new aerosol particles starts in the nucleation mode size range and the mode is observed over a period of several hours showing signs of growth. If no NPF is observed, the day is classified as a non-event day (NE). A large number of days did not

fulfil the criteria to be classified either clear event or NE day and they are considered as unclassified days (UC). In each data set, there are periods of missing data, as well as periods of low quality data, i.e., one or more variables did not have measured value. To avoid the eventual biasing of our results, the number of observations used in the analysis was decreased so that slightly more than half of the observations had measured values for all of the variables used in the final statistical model.

SPC is a low-land observation site in the Po Valley (North Italy). The particle size distribution measurements were made between 24 March 2002 and 30 April 2005. The DMPS system was operational on 814 days during the time period, which included 293 NPF event days and 270 non-event days. 251 days could not be classified. A description of the data set and the analyses performed is given in Hamed et al. (2007).

Melpitz is a low-land (86 m a.s.l.) research station in Eastern Germany, surrounded by flat terrain consisting mainly of agricultural soil, pastures, and forests. Atmospheric particle size distributions and NPF have been analysed at Melpitz since 1996 (Birmili and Wiedensohler, 2000; Engler et al., 2007). The particle size distribution measurements used in this paper lasted between 1 July 2003 and 30 June 2006. This particular time period included 270 event days, 414 non-event days and 130 unclassified days (Hamed et al., 2010).

Hohenpeissenberg is a mid-level mountain site (980 m a.s.l.) in Southern Germany, located about 30 km north of the Alpine mountain ridge. In Hohenpeissenberg the DMPS measurements lasted between 1 April 1998 and 3 August 2000, which included 85 event days, 220 non-event days and 40 unclassified days. For details of these measurements and a previous analysis of the connection between NPF and gaseous sulphuric acid, see Birmili et al. (2003).

2.2 Data selection and pre-processing

The aim of our study was to find the factors affecting the growth to, and primary production of, particles that can be considered the minimum potential CCN size (here chosen to be 50 nm in diameter). For this purpose, several different statistical models were tested in their ability to predict N_{50} as a function of various factors. Most of our statistical models included combinations of hourly averages of gas and meteorological parameters measured at all of the three stations, including temperature, relative humidity, radiation, O_3 , SO_2 , NO_x , condensation sink, wind speed and direction and many other parameters which were either not present in all datasets or did not have any relevance to the fine particle concentrations. We also used the probability that the day is a non-event day (PrNE), i.e., a day when NPF is not observed (Hamed et al., 2007). PrNE was calculated with discriminant analysis according to Mikkonen et al. (2006). The calculated non-event probability was used instead of

Table 1. Results from Discriminant analysis.

Station	Predictors	Classification error	Missed events	Predicting non-event to event
SPC (Mikkonen et al., 2006)	RH, O ₃ , Radiation	13.4%	1.9%	4.9%
Melpitz	RH, O ₃ , SO ₂ , Radiation, Temperature	6.7%	7.4%	0.7%
Hohenpeissenberg	RH, O ₃ , SO ₂ , Radiation, NO ₂	6.1%	11.7%	2.7%
Combined data	RH, O ₃ , SO ₂ , Radiation, NO ₂ , Temperature	7.7%	8.4%	1.5%

observed event classification because otherwise we would have had to exclude the unclassified days, which would have subsidised the data drastically. In addition, the probability of a non-event day can be estimated also for those days where the visual event classification has not been made at all, which enables the use of the statistical model in predictive purposes.

Since the condensation sink (CS) is computed from the size-distribution of the particles by the method described by Pirjola et al. (1998) and Kulmala et al. (2001), there is a risk of circular argumentation when using it in our statistical model. Even if the contribution of the smallest particles to the total value of CS is small it may still cause bias to the estimation if we first use the number of small particles to calculate the CS and then use the CS to predict the growth of the same particles. That is why we used only the number of particles larger than 50 nm in the calculation of the condensation sink, acknowledging that the contribution of sub 50 nm particles to CS amounts to a few percent.

2.3 Computing event probabilities with discriminant analysis

Probabilities for event and non-event days were computed with discriminant analysis method described in Mikkonen et al. (2006). Discriminant analysis is a multivariate statistical analysis method which is commonly used to classify observations into different groups. When the distribution in each group cannot be assumed multivariate normal, which is the case with atmospheric aerosol measurements, non-parametric discriminant methods must be used. Non-parametric methods are based on group-specific probability densities and they are used to produce a classification criterion based on those probabilities. In addition, the non-parametric method is more robust for multicollinearity (i.e. when some variables measure partly the same effect), which might occur in the analysis of atmospheric data.

The best classification result in SPC was reached with the combination of daily averages of relative humidity, ozone concentration and global radiation (Mikkonen et al., 2006) and these variables were used in the computation of the event probability. A similar analysis was made also for the two other datasets and the predictors were slightly site dependent.

The best predictor models for each site are shown in Table 1 with prediction errors. Finally, we combined all the data from different sites to find the best discriminates for the NPF events for the full dataset. We found out that the best predictor sets for the individual sites are subgroups of the best predictors of the combined data, i.e., in specific sites it is possible to get equally good classification with fewer parameters.

2.4 Predicting the size-distribution

Because of the complexity of processes affecting the concentration of small particles in the atmosphere, we chose to use generalized linear models in the analysis. The basic form of a generalized linear model is given by

$$g(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}. \quad (1)$$

On the left side of Eq. (1), \mathbf{y} is the vector of measurements of the studied variable (in our case N_{50}) and $g(\cdot)$ refers to so-called link function, which relates the linear predictors (e.g., measured temperature or SO₂ concentration) to the expected value of \mathbf{y} (McCullagh and Nelder, 1989). In our analysis, the natural logarithm turned out to be a suitable link function, since the density function of measured particle concentration ($dN/d\log D_p$) followed the Gamma distribution and the natural logarithm of gamma-distributed data follows normal distribution. Using a model based on gamma-distribution was also an option but we chose a log-linear model since estimating and interpreting it is somewhat easier. On the right side of Eq. (1), $\mathbf{X}\boldsymbol{\beta}$ is the fixed part of the model (as in the case of standard linear models) so that \mathbf{X} denotes the ($n \times p$) observation matrix (e.g., measured temperature or SO₂ concentration) and $\boldsymbol{\beta}$ denotes the unknown ($p \times 1$) vector of fixed intercept and slope effects of the model. The remaining term, $\boldsymbol{\varepsilon}$, is the vector of the residuals of the model.

However, the basic form of the generalized linear model, Eq. (1), is ill-suited for aerosol measurement data in which standard independency and homogeneity assumptions are not met. Therefore, we chose to use a mixed model structure in which a random component (denoted $\mathbf{Z}\mathbf{u}$) is added to Eq. (1). The main idea of a mixed model is to estimate not only the mean of the measured response variable \mathbf{y} , but

also the variance-covariance structure of the data. Modelling the (co)variances of the variables reduces the bias of the estimates and prevents autocorrelation of the residuals.

Using matrix notation, a linear mixed model can be written as follows (McCulloch and Searle, 2001):

$$g(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\varepsilon}. \quad (2)$$

Here $\mathbf{Z}\mathbf{u} + \boldsymbol{\varepsilon}$ is the random part of the model. \mathbf{u} is a $(q \times 1)$ vector of random effects with a q -dimensional normal distribution with zero expectation and $(q \times q)$ covariance matrix denoted by \mathbf{G} . Note that the structure of the covariance matrix \mathbf{G} is not defined in advance. On the other hand, \mathbf{Z} is the $(n \times q)$ design matrix of the random effects vector, \mathbf{u} . With adequate choices of the matrix \mathbf{Z} , different covariance structures $\text{Cov}(\mathbf{u})$ and $\text{Cov}(\boldsymbol{\varepsilon})$ can be defined and fitted. Successful modelling of variances and covariances of the observations provides valid statistical inference for the fixed effects $\boldsymbol{\beta}$ of the mixed model. In contrast to general linear models, the error terms $\boldsymbol{\varepsilon}$ can be correlated and the covariance matrix of the residuals is denoted by \mathbf{R} . From this it follows that the distribution of observations can be postulated as a normal distribution with the expectation of $\mathbf{X}\boldsymbol{\beta}$ and covariance matrix \mathbf{V} , which is given by $\mathbf{V} = \mathbf{Z}\mathbf{G}\mathbf{Z} + \mathbf{R}$.

One of the greatest advantages of a multivariate model is that when all parameters are in the same model the interpretation of estimates of single parameters is easy and the results are more valid than in single variable analysis. For example, bias caused by yearly variation is cleaned from the other variables in the model. We made the analysis in a stepwise manner, i.e., the parameters were added and removed to and from the model according to their statistical significance and the total increase of the explanation capability of the model with the all other parameters included.

The final model, used in all datasets is given by

$$\begin{aligned} \ln(dN/d\log D_p|_{50}) &= (\beta_0 + u_m) + (\beta_1 + v_{1m}) \cdot \text{RH} + (\beta_2 + v_{2m}) \cdot \text{SO}_2 \\ &+ (\beta_3 + v_{3m}) \cdot \text{NO}_2 + (\beta_4 + v_{4h}) \cdot \text{O}_3 + \beta_5 \cdot \text{PrNE} \\ &+ \beta_6 \cdot \text{Temp} + v_5 \cdot \text{CS}^{-1} + \varepsilon_{jt} \end{aligned} \quad (3)$$

where β_0 is the fixed intercept term, β_i are the fixed slopes, u_m are month-specific random intercepts and $v_1 - v_4$ are the random month-specific (m) or hour-specific (h) slopes. v_5 is location specific random effect for taking account the condensation sink in the estimation of other parameters but it is set to zero (or constant) when the model is used in prediction. The other variables used in the model are: relative humidity (RH, %), concentrations of SO_2 ($\mu\text{g m}^{-3}$), NO_2 ($\mu\text{g m}^{-3}$) and ozone ($\mu\text{g m}^{-3}$), probability that the day is not a NPF event day (PrNE) and air temperature (K).

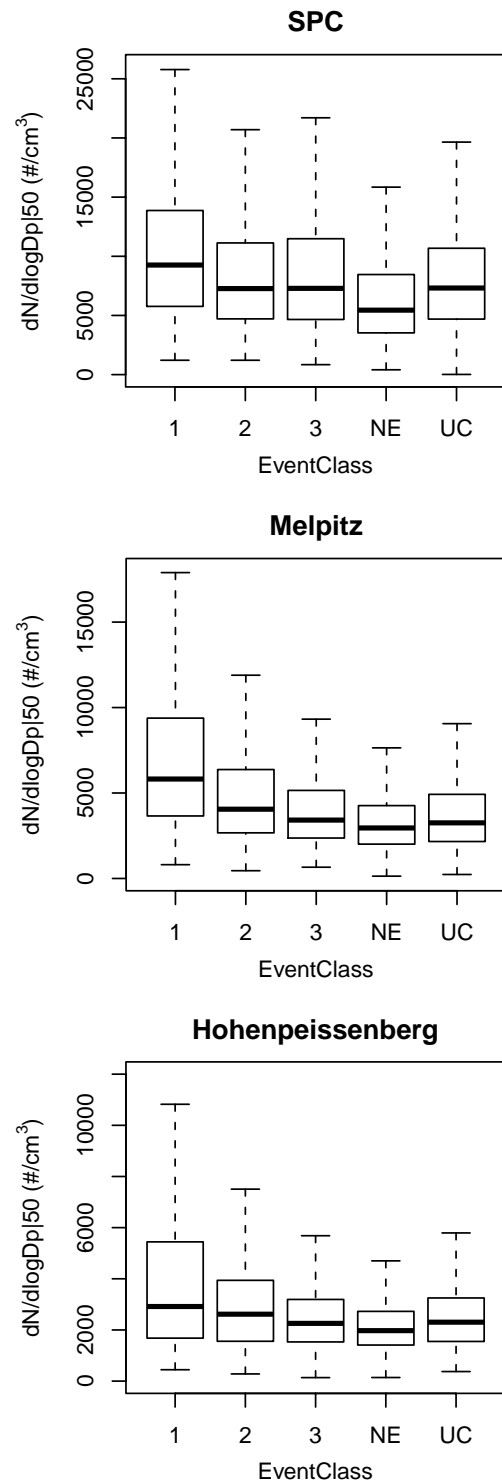


Fig. 1. Experimental particle concentrations at 50 nm ($dN/d\log D_p$) for different event classes. The box plots indicate the median (bar), 25th and 75th percentiles (box), and the minimum and maximum values (whiskers) that were not considered outliers. “1”, “2”, and “3” indicate NPF event classes of decreasing intensity (see text). NE: non-events, and UC: unclassified (i.e. ambiguous) days.

Table 2. Signs and magnitudes of the parameters of the model ($\beta_i + v_i$). Minus sign denotes a negative regression effect on N_{50} and plus sign denotes a positive effect and the number of signs describes the magnitude of the effect.

Predictor		Sign of Effect			
		SPC	Melpitz	Hohenpeissenberg	all
RH	Winter	--	--	-	--
	Rest of the year	-	-	-	-
SO ₂	Summer	++	++	++	++
	Rest of the year	-	-/0	+	+/0
NO ₂	Summer	Not significant	++	+	++
	Rest of the year	Not significant	+	+	+/0
O ₃	Daytime	-	-	-	-
	Night/early morning	--	--	--	--
Condensation Sink		++	++	++	++
Temperature		-	-	Not significant	-
PrNE		--	--	--	--

3 Results

3.1 Effect of new particle formation event

As described in Sect. 2, all measurement days were classified as a NPF event, non-event or unclassified day. NPF events were further classified to three classes, according to the intensity of the particle formation event. The experimental observations show that N_{50} was significantly higher on class 1 event days compared to the rest of the data (Fig. 1). On the other hand, on non-event days the concentration of N_{50} was the lowest. This observation confirms that the number of Aitken particles (50 nm) is influenced by NPF events, which are characterised by the evolution of nucleation mode particles visible at an initial diameter around 3 nm. The impact can be twofold. The first option is that freshly nucleated particles grow to 50 nm on the very same day. This requires considerable particle growth and thus amounts of condensable vapours. At Melpitz, the nucleation mode has been shown to occasionally reach mode diameters between 50 and 80 nm on the very same day when NPF happened (Wehner et al., 2005). A second option is that particles from the previous day, on which NPF might have happened, grew into the Aitken particle size range. This second option bears some relevance, since NPF events tend to cluster in series of subsequent days that are characterised by synoptic-scale weather conditions (high solar radiation, intense vertical mixing) that are favourable to particle nucleation.

Since all three stations studied are located in anthropogenically influenced areas and encounter pollution episodes, several of the non-event days also showed high particle concentrations. In SPC the concentrations in weak event classes 2 and 3 do not differ significantly from each other or

from the unclassified days. At the German stations enhanced concentrations also occurred on class 2 event days, more pronounced in Melpitz.

3.2 Estimated parameters of the statistical model

In order to find the best predictive statistical model for the number concentration of 50 nm particles we first performed tests for the data from each of the three measurement sites separately. Owing to a scarcity of observations, we needed to exclude the data for September in SPC and for February in Hohenpeissenberg.

Table 2 shows the signs and strengths of parameters ($\beta_i + v_i$) in Eq. (3); the magnitudes of the parameters are listed in the Appendix A. Since the parameters for the three datasets were very close to each other, we could merge the data and find the best predictors for the combined dataset. All months were also given individual intercept terms, which would give information about the monthly variation if the monthly slopes were constant, but as the slopes differ too, the intercepts cannot be fully interpreted. We found that RH and the concentrations of SO₂ and NO₂ had significant additional variance components for different times of the year in all sites, whereas other parameters showed no significant seasonal variation.

When the additional variance is taken into account, the statistical model suggests that the regression effect of RH is negative in all sites, i.e., when the RH is high, N_{50} is low. High atmospheric relative humidity has proved to be a factor disfavoured NPF (e.g. Birmili and Wiedensohler, 2000; Boy and Kulmala, 2002; Mikkonen et al., 2006; Hamed et al., 2007), so the observed inverse relationship between RH and N_{50} is in line with the inverse relationship between RH and NPF. The overall regression effect of RH was found

to be negative but it varies between months (Fig. 2a). The adverse effect of RH was most intense in winter (January–December), when the measured relative humidity reaches its highest overall levels. This is not surprising because the highest RH values are also associated with more clouds and precipitation and, thus, wet particle deposition.

SO₂ concentration had a significant positive regression effect in the summer at all sites but the effect was negative or zero in wintertime (Fig. 2b), especially in SPC and Melpitz. Sulphuric acid has been shown to be involved in nucleation and growth of newly formed particles (e.g. Kulmala et al., 2006; Laaksonen et al., 2008a). A product of SO₂ and radiation divided by CS has been used as a proxy for H₂SO₄ formation (Hamed et al., 2010). The positive effect of SO₂ concentration in summer months suggests that increased amount of radiation together with lower CS increases the concentration of H₂SO₄ and thus gives a significant contribution to the particle formation and growth.

NO_x has been suggested to have a positive influence on particle nucleation (Laaksonen et al., 2008a) and thus affect the number of Aitken particles. NO₂ was found to be a significant predictor in our model for Melpitz and Hohenpeissenberg and had a positive effect on *N*₅₀. In SPC the effect of NO₂ seems insignificant, but this may be due to poor quality of the NO₂ data. The effect of the NO₂ concentration in the combined dataset is at its highest in summer and in early autumn (Fig. 2c). The effect of NO₂ being a significant predictor for *N*₅₀ is interesting and we do not completely understand it but more thorough settling of it is out of the scope of the current study.

The aerosol condensation sink determines how rapidly molecules will condense onto pre-existing aerosols (Kulmala et al., 2005). The regression effect of condensation sink was found to be positive in all datasets, i.e., high CS was predicted to favour high particle concentrations at 50 nm. This was somewhat unexpected since previous studies have found low CS to favour new particle formation (Vehkamäki et al., 2004; Hamed et al., 2007; Jaatinen et al., 2009). It is possible that CS acts as an indicator of the growth of the smaller particles to the size of 50 nm. An alternative explanation is the contribution of direct anthropogenic particle emissions to both, *N*₅₀ and CS. The relevance of anthropogenic particle sources (traffic, industry), which are concentrated in urban areas but are also spread in a diffuse distribution in rural areas, on the observations at Melpitz was pointed out by Costabile et al. (2009). The effect of CS was used in the model as a random, site specific effect, to take account the random variation in the 50 nm particle concentration caused by the larger particles existing in the air, but to avoid circularity problems it is not used in prediction or in the GLOMAP modelling study.

Temperature had a small negative effect in SPC and Melpitz but the effect was not significant in Hohenpeissenberg. The negative regression effect of temperature in SPC and in

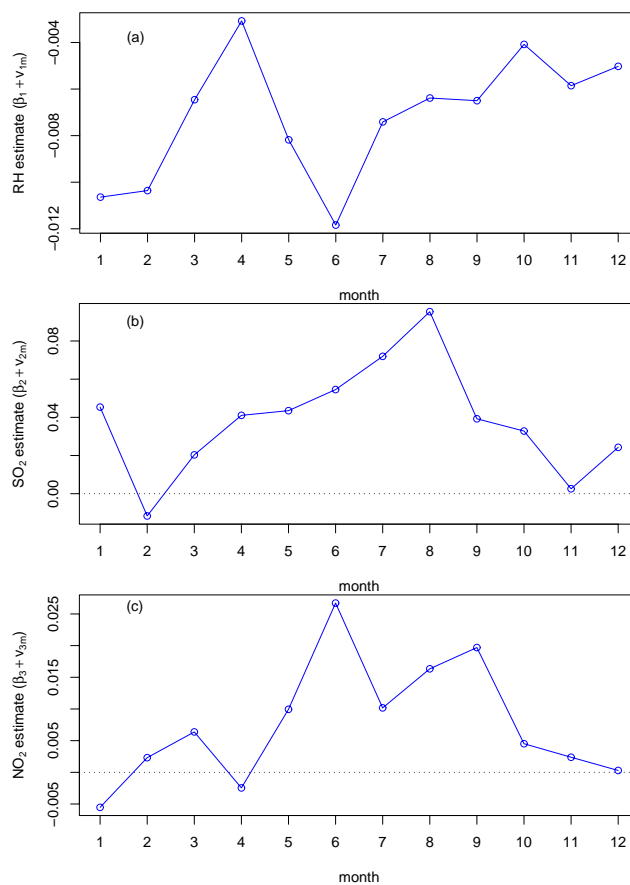


Fig. 2. Seasonal estimates of (a) RH, (b) SO₂, and (c) NO₂ for the combined dataset. Note that the estimates of different parameters are not comparable, since the values of the variables are not standardised.

Melpitz is in line with the suggestions of Birmili et al. (2003) and Hamed et al. (2007) that the high temperatures suppress new particle formation in polluted areas, i.e., there is less NPF on high temperature days, and through this affect also to the number concentration of small particles. In addition, high temperatures are generally associated with taller atmospheric mixed layer heights. In a taller mixed layer, primary particle emissions contributing to the Aitken mode will dilute in a bigger volume of air and thus lead to lower concentrations of such particles. Not significant effect in Hohenpeissenberg may be caused by the fact that the measurement station is on top of a small mountain (980 m a.s.l.), where it is generally above the nocturnal boundary layer and in winter time also occasionally above the daytime boundary layer.

The effect of PrNE was significant and negative which was expected as the number concentration of 50 nm particles typically increases on NPF days (i.e., when PrNE is zero or low). PrNE was calculated with method described in Mikkonen et al. (2006) by using daily averages of predictive parameters described in Sect. 2.3. There are several other

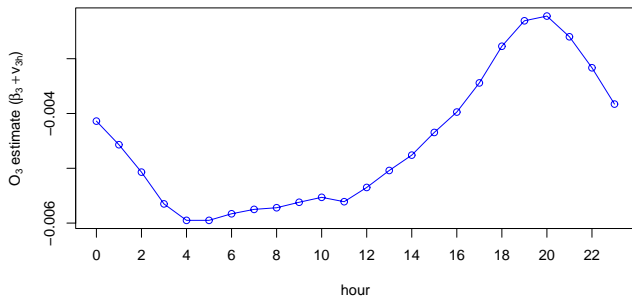


Fig. 3. Variation of the random effect of ozone concentration within a day.

possible methods to compute the probability of NPF (or nonevent), e.g., Hyvönen et al. (2005) introduced a method for boreal forest areas and some cases even local proportion of nonevent days could be valid approximation, so method for determining the PrNE does not limit the use of the parameterization for N_{50} .

Oxidation of organics is known to be a significant factor in the growth of the particles (Laaksonen et al., 2008b). Ozone concentration has been found to affect the oxidation of organic species and thus affect the particle formation and growth (e.g., Joutsensaari et al., 2005; Vaattovaara et al., 2006). Ozone concentration was the only variable that had significant diurnal variation in the estimate. Figure 3 shows that while the overall effect is negative, it has its lowest values just before sunrise and then rises until sunset, which indicates that it is probably acting as an (inverse) tracer of some pollutants which are sinks for small particles. For other variables, diurnally varying estimates did not increase the predicting ability of the model so much that it would either have statistical significance or compensate for the increased computational cost.

3.3 Statistical model predictions

Figure 4 illustrates how well the statistical model for the combined data predicts the observations at the three stations in randomly selected periods. The figure shows that the predicted values follow the observations fairly well in all stations. Overall the statistical model finds the peaks of the number concentration but slightly underestimates the highest peaks and the fastest fluctuations.

Figure 5 presents the scatter plots of the natural logarithms of observed and predicted N_{50} for event and non-event days. It shows that when the great number of data points is taken into account, the predicted values are quite well in line with the observations. Only an insignificant number of the highest observations is underestimated and some of the lowest values are overestimated. Together with Fig. 4 it confirms that the prediction ability of the statistical model is adequate in all of the stations and for both event and non-event days.

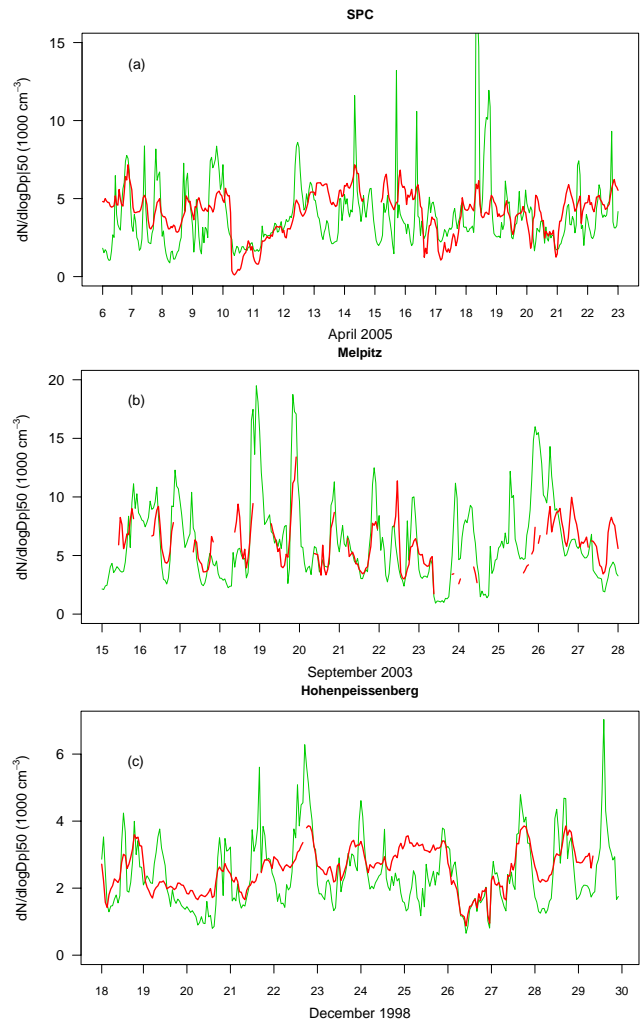


Fig. 4. Observed (green line) and predicted (red line) time series from illustrative example periods in all stations, when the combined data is used in the parameter estimation. Gaps in red line are due to missing datapoints.

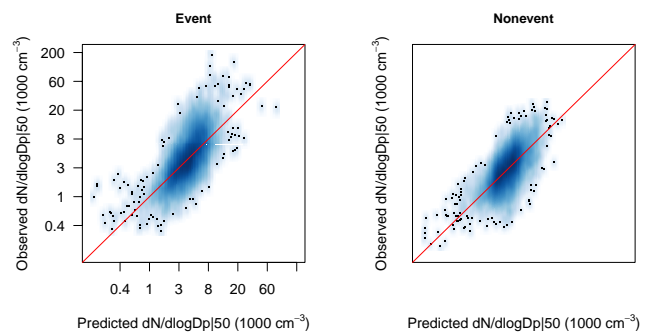


Fig. 5. Observed vs. predicted $dN/d\log D_p |_{50\text{ nm}}$ (N_{50}) for event and non-event days in the combined dataset. Diagonal line represents the perfect fit and the tone of the colour shows the density of the data points.

We computed coefficients of determination (R^2) for each dataset separately and for the combined dataset. R^2 indicates how well the statistical model predicts the total variation of the dependent variable (here N_{50}) and is commonly used to give information about the goodness of a fit. The best prediction ability was found in data from SPC where the R^2 was more than 0.6, which indicates that the statistical model explains more than 60% of the total variation of the particle concentration. At the German stations R^2 values were approximately 0.5. For the combined dataset the model could explain more than 50% of the total variation of the number concentration of 50 nm particles, which can be considered as fairly good result for this kind of data. The biggest single factor decreasing the R^2 values is the highest peaks on the number concentration which the model does not capture very well. However, these peaks are often caused by local pollution events and thus are almost impossible to predict. The current statistical model does not have specific description for local transport trajectories, or air mass origin but these effects are partly described by the gas phase data.

3.4 Test of statistical model in a global aerosol model

3.4.1 Model set-up

The derived parameterisation for aerosol size distribution at 50 nm (Eq. 4) was tested in a global scale aerosol model GLOMAP to predict the concentration of climate active particles over Europe. The model is an extension to the TOMCAT 3-D chemical transport model (Chipperfield, 2006; Stockwell and Chipperfield, 1999), and its detailed description can be found in Spracklen et al. (2005). Here, we used a sectional moving centre scheme with 20 size sections to cover the aerosol particle size range of 3 nm to 25 μm . We performed two simulations for April 2000: (1) a baseline simulation using the GLOMAP standard set-up, and (2) a test simulation using the statistical model developed in this study.

The standard set-up of GLOMAP explicitly simulates particle formation via binary H_2SO_4 - H_2O nucleation in the free troposphere (Kulmala et al., 1998), activation nucleation in the boundary layer (Kulmala et al., 2006), and primary particle emissions according to the AEROCOM emission data base (<http://nansen.ipsl.jussieu.fr/AEROCOM>). Particles formed or emitted at sizes below 50 nm grew to this size by condensation of sulphuric acid and oxidation products of monoterpenes and, to a lesser extent, by coagulation. In the test run, boundary layer nucleation and primary emissions at particle sizes $D_p < 50$ nm were omitted and replaced with the parameterization developed in this study. The parameterization was used to calculate N_{50} at each aerosol model time step (15 min). Input for the statistical model was taken from the model predicted SO_2 field, offline NO_2 and O_3 fields predicted with a coupled

chemistry aerosol model, and European Centre for Medium-range Weather Forecasts (ECMWF) temperature and relative humidity fields.

3.4.2 Results

Figure 7 compares the model-predicted potential CCN concentration ($D_p > 50$ nm) against measured April mean data at Melpitz, SPC, Hohenpeissenberg, and 15 other sites. For the three sites analysed in this study, we present averages of several years of measurement data (i.e., multi-annual averages), since the analysed measurement periods at the three sites did not overlap (see measurement periods in Sect. 2.1). The data for the other 15 sites is from the European Integrated project on Aerosol Cloud Climate and Air Quality Interactions (EUCAARI) during April 2008 and/or 2009, and were chosen because of their comprehensiveness. Note that since the model was run only for April 2000 (due to computational expense of a global model) one can expect only a rough agreement between the model and measurements. However, we wanted to include the EUCAARI data in order to demonstrate that the statistical parameterization gives reasonable results also outside the geographical domain for which it was derived.

The parameterization (test run) brings the model results significantly closer to observations at Melpitz and Hohenpeissenberg: the baseline run predicts 110% and 52% higher than observed potential CCN concentration, respectively. Using the statistical model, the overprediction is reduced to 34% and 22%, respectively. However, at SPC the agreement with the observations weakens in the test run compared to the baseline run (underpredictions of 42% and 18%, respectively). This can be because the modelled gas concentrations, which are used as input for the statistical model, are in poor agreement with the measurements. For example, the modelled O_3 concentration at SPC is a factor of ~ 2 higher and the NO_x concentration a factor of ~ 5 lower than the measured mean values for April. As a whole, SPC is a challenging site for large scale aerosol models such as GLOMAP that have a coarse spatial resolution. The site is located in a valley where pollution frequently builds up or clears out. Since GLOMAP has a spatial resolution of $2.8^\circ \times 2.8^\circ$, it is completely unable to resolve the local topological features around the site.

Of the 15 EUCAARI sites, using the parameterisation instead of the baseline model set-up brings the predicted CCN much closer to observations at 7 locations and deteriorates the agreement clearly at 4 locations. At 3 sites (K-Puzsta, Kosetice and Finokalia) there is relatively small difference between the baseline and the parameterized runs. At Waldhof the observed CCN is approximately halfway between the predicted values from the baseline and parameterized runs.

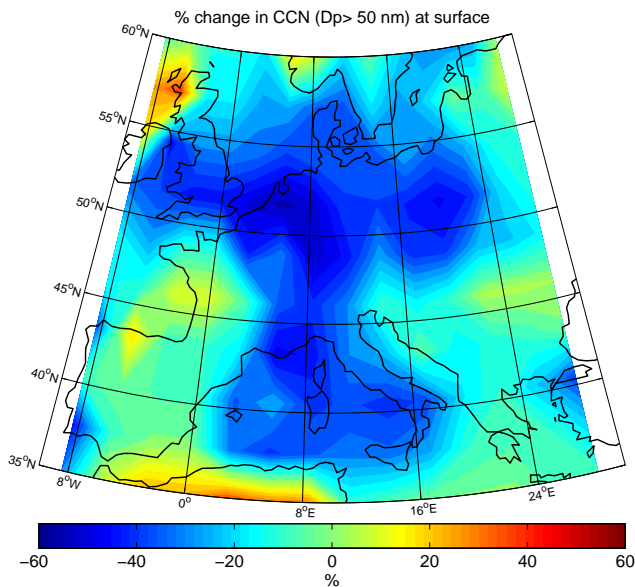


Fig. 6. Relative change in model predicted potential CCN concentrations when the standard aerosol model set-up is replaced with the test model set-up.

While this preliminary test of the parameterisation against observations is incomplete in that it does not simulate the exact years of the observations, it does indicate that the derived parameterisation has potential to describe CCN formation at very different environments from Arctic to polluted rural. These results give confidence to apply the statistical framework also to measurements from other sites in order to further improve the derived parameterisation.

As indicated in Fig. 7, the parameterization has a tendency to predict lower simulated potential CCN concentrations than the standard model version over Central Europe (Fig. 6). Reductions of the order of 30–60% are seen over northern parts of Central and Western Europe (Germany, Poland, Benelux, England) and Italy. Over the rest of Europe (apart from a small region in Northern Scotland) the baseline and test runs agree within 20%. Note, however, that since the parameterisation was developed based on aerosol data from the Central Europe, it is most likely not valid outside this region (e.g. over the oceans were totally different processes determine the aerosol and CCN concentrations).

4 Conclusions

An advanced statistical model structure was introduced and found to be an adequate tool to analyse tropospheric Aitken particle ($D_p = 50$ nm) concentrations and for making predictions based on in-situ meteorological (temperature, RH) and gas phase parameters (SO_2 , NO_2 , O_3). The statistical model can also be used for forecasting the particle concentration with the estimated regression coefficients. A key result was that some of those variables which control the occurrence of new particle formation events also influence

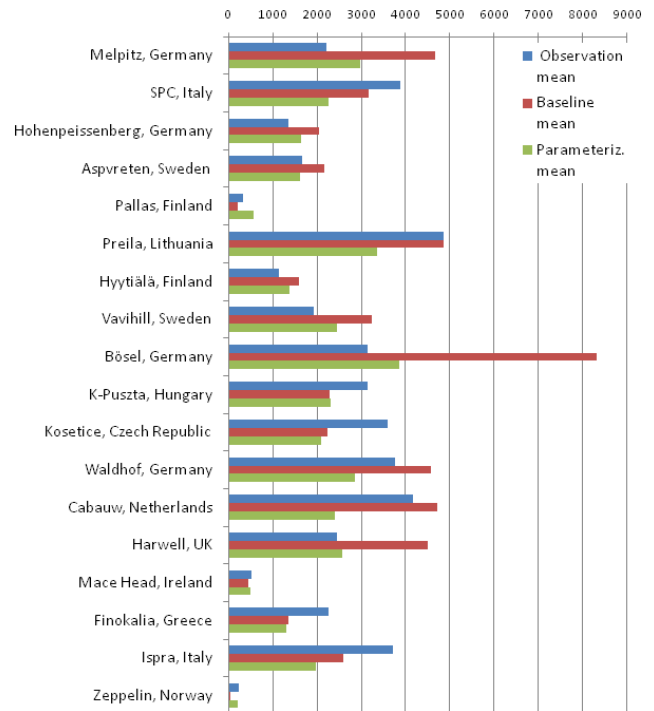


Fig. 7. Mean observed and predicted potential CCN ($D_p > 50$ nm) concentrations at the 3 measurement sites (measurement data from the time periods indicated in the text) and at 15 EUCAARI sites (measurement data for April 2008 and 2009). Note that the model is run for April 2000. Units are cm^{-3} .

the number concentration of 50 nm particles. This is explained by a significant transfer rate of newly formed particles into the bigger size ranges by condensation and/or coagulation. A notable exception was the condensation sink, which was found to be a factor disfavouring NPF but had a significant positive correlation with the number concentration of 50 nm particles.

The same statistical model framework could be used for any other particle sizes and in other locations. The same parameterization can be used at least in areas with similar concentrations of particles and pollutants but extrapolation of the results to clean environments, like boreal forests, needs to be confirmed before use. Our preliminary tests with the global scale aerosol model GLOMAP indicate that the parameterization can be used as a part of a larger atmospheric model to predict the concentration of climatically active particles. The statistical model for the prediction of 50 nm particles could be a significant step towards shorter computation times in global climate models; it seems to work adequately in boundary layer but it still does not solve the computational efficiency problems in free troposphere. Equally, the use of the statistical model for N_{50} could bypass some of the current uncertainties in the theoretical description of the nucleation and growth process, particularly when predicting potential CCN concentration.

Appendix A

Function coefficients

Table A1. Coefficients for the function (1).

Month	Intercept ($\beta_0 + u$)	RH ($\beta_1 + v_1$)	SO ₂ ($\beta_2 + v_2$)	NO ₂ ($\beta_3 + v_3$)
1	11.1982	-0.01064	0.045368	-0.005527
2	11.362	-0.01036	-0.01166	0.002313
3	10.833	-0.006458	0.02034	0.006373
4	10.8089	-0.003074	0.041041	-0.002457
5	11.1658	-0.00818	0.043503	0.00995
6	11.4385	-0.01184	0.05454	0.026713
7	11.2016	-0.00741	0.07196	0.010171
8	10.97091	-0.006385	0.09535	0.016337
9	10.97628	-0.006499	0.039237	0.019693
10	10.8471	-0.004082	0.03283	0.004513
11	11.02423	-0.005856	0.00258	0.002383
12	10.8648	-0.005025	0.02428	0.000293

Table A2. Coefficients for the function (1).

Hour	O ₃ ($\beta_4 + v_4$)
0	-0.004139
1	-0.00457
2	-0.00507
3	-0.00565
4	-0.00595
5	-0.00595
6	-0.00583
7	-0.00575
8	-0.00572
9	-0.00562
10	-0.00553
11	-0.00561
12	-0.00535
13	-0.00504
14	-0.00476
15	-0.004346
16	-0.003974
17	-0.003442
18	-0.002773
19	-0.002308
20	-0.002222
21	-0.0026
22	-0.003166
23	-0.003828

Table A3. Coefficients for the function (1).

prNE (β_5)	T (β_6)
-0.2854	-0.00645

Table A4. Coefficients not used in prediction.

Location	1/CS (v_5)
HOH	-0.00063
MEL	-0.00149
SPC	-0.00189

Acknowledgements. This work was supported by Graduate school in Physics, Chemistry, Biology and Meteorology of Atmospheric composition and climate change and by the Alfred Kordelin Foundation. The Academy of Finland Center of Excellence program (project number 1118615) is also acknowledged. JS acknowledges the financial support of the Saastamoinen Foundation and the US Department of Energy (grant DE-FG-02-05ER63997). The National Center for Atmospheric Research is sponsored by the U. S. National Science Foundation. SR acknowledges the financial support of The Academy of Finland (project number 123466). AH acknowledges the financial support of Finnish Cultural Foundation. We would also like to thank Ari Asmi for providing the EUCAARI (contract no. 036833-2) data.

Edited by: I. Rutt

References

- Birmili, W. and Wiedensohler, A.: New particle formation in the continental boundary layer: Meteorological and gas phase parameter influence, *Geophys. Res. Lett.*, 27, 3325–3328, 2000.
- Birmili, W., Wiedensohler, A., Heintzenberg, J., and Lehmann, K.: Atmospheric particle number size distribution in Central Europe: statistical relations to air masses and meteorology, *J. Geophys. Res.*, 106, 32005–32018, doi:10.1029/2000JD000220, 2001.
- Birmili, W., Berresheim, H., Plass-Dülmer, C., Elste, T., Gilge, S., Wiedensohler, A., and Uhrner, U.: The Hohenpeissenberg aerosol formation experiment (HAFEX): a long-term study including size-resolved aerosol, H₂SO₄, OH, and monoterpenes measurements, *Atmos. Chem. Phys.*, 3, 361–376, doi:10.5194/acp-3-361-2003, 2003.
- Boy, M. and Kulmala, M.: Nucleation events in the continental boundary layer: Influence of physical and meteorological parameters, *Atmos. Chem. Phys.*, 2, 1–16, doi:10.5194/acp-2-1-2002, 2002.
- Chipperfield, M. P.: New version of the TOMCAT/SLIMCAT off-line chemical transport model: Intercomparison of stratospheric tracer experiments, *Q. J. R. Meteorol. Soc.*, 132, 1179–1203, doi:10.1256/qj.05.51, 2006.
- Costabile, F., Birmili, W., Klose, S., Tuch, T., Wehner, B., Wiedensohler, A., Franck, U., König, K., and Sonntag, A.: Spatio-temporal variability and principal components of the

- particle number size distribution in an urban atmosphere, *Atmos. Chem. Phys.*, 9, 3163–3195, doi:10.5194/acp-9-3163-2009, 2009.
- Dusek, U., Frank, G. P., Hildebrandt, L., Curtius, J., Schneider, J., Walter, S., Chand, D., Drewnick, F., Hings, S., Jung, D., Borrmann, S., and Andreae, M. O.: Size matters more than chemistry for cloud-nucleating ability of aerosol particles, *Science*, 312, 1375–1378, doi:10.1126/science.1125261, 2006.
- Engler, C., Rose, D., Wehner, B., Wiedensohler, A., Brüggemann, E., Gnauk, T., Spindler, G., Tuch, T., and Birmili, W.: Size distributions of non-volatile particle residuals ($D_p < 800$ nm) at a rural site in Germany and relation to air mass origin, *Atmos. Chem. Phys.*, 7, 5785–5802, doi:10.5194/acp-7-5785-2007, 2007.
- Ghan, S. J., Easter, R. C., Chapman, E. G., Abdul-Razzak, H., Zhang, Y., Leung, L. R., Laulainen, N. S., Saylor, R. D., and Zaveri, R. A.: A physically based estimate of radiative forcing by anthropogenic sulfate aerosol, *J. Geophys. Res.*, 106, 5279–5293, doi:10.1029/2000JD900503, 2001.
- Hamed, A., Joutsensaari, J., Mikkonen, S., Sogacheva, L., Dal Maso, M., Kulmala, M., Cavalli, F., Fuzzi, S., Facchini, M. C., Decesari, S., Mircea, M., Lehtinen, K. E. J., and Laaksonen, A.: Nucleation and growth of new particles in Po Valley, Italy, *Atmos. Chem. Phys.*, 7, 355–376, doi:10.5194/acp-7-355-2007, 2007.
- Hamed, A., Birmili, W., Joutsensaari, J., Mikkonen, S., Asmi, A., Wehner, B., Spindler, G., Jaatinen, A., Wiedensohler, A., Korhonen, H., Lehtinen, K. E. J., and Laaksonen, A.: Changes in the production rate of secondary aerosol particles in Central Europe in view of decreasing SO_2 emissions between 1996 and 2006, *Atmos. Chem. Phys.*, 10, 1071–1091, doi:10.5194/acp-10-1071-2010, 2010.
- Hyvönen, S., Junninen, H., Laakso, L., Dal Maso, M., Grönholm, T., Bonn, B., Keronen, P., Aalto, P., Hiltunen, V., Pohja, T., Launiainen, S., Hari, P., Mannila, H., and Kulmala, M.: A look at aerosol formation using data mining techniques, *Atmos. Chem. Phys.*, 5, 3345–3356, doi:10.5194/acp-5-3345-2005, 2005.
- IPCC: Climate Change 2007: The Physical Science Basis, in: Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, edited by: Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K. B., Tignor, M., and Miller, H. L., Cambridge University Press, Cambridge, UK and New York, NY, USA, 2007.
- Jaatinen, A., Hamed, A., Joutsensaari, J., Mikkonen, S., Birmili, W., Wehner, B., Spindler, G., Wiedensohler, A., Decesari, S., Mircea, M., Facchini, M. C., Junninen, H., Kulmala, M., Lehtinen, K. E. J., and Laaksonen, A.: A comparison of new particle formation events in the boundary layer at three different sites in Europe, *Boreal Environ. Res.*, 14, 481–498, 2009.
- Joutsensaari, J., Loivamäki, M., Vuorinen, T., Miettinen, P., Nerg, A.-M., Holopainen, J. K., and Laaksonen, A.: Nanoparticle formation by ozonolysis of inducible plant volatiles, *Atmos. Chem. Phys.*, 5, 1489–1495, doi:10.5194/acp-5-1489-2005, 2005.
- Kerminen, V. M., Lihavainen, H., Komppula, M., Viisanen, Y., and Kulmala, M.: Direct observational evidence linking atmospheric aerosol formation and cloud droplet activation, *Geophys. Res. Lett.*, 32, L14803, doi:10.1029/2005GL023130, 2005.
- Kinder, F.: Hygroscopic Properties of Atmospheric Aerosol Particles in Central Europe, Diploma thesis, Faculty of Physics and Earth Science, Institute for Meteorology, University of Leipzig, Germany, April 2010.
- Kokkola, H., Hommel, R., Kazil, J., Niemeier, U., Partanen, A.-I., Feichter, J., and Timmreck, C.: Aerosol microphysics modules in the framework of the ECHAM5 climate model intercomparison under stratospheric conditions, *Geosci. Model Dev.*, 2, 97–112, doi:10.5194/gmd-2-97-2009, 2009.
- Korhonen, H., Napari, I., Timmreck, C., Vehkamäki, H., Pirjola, L., Lehtinen, K. E. J., Lauri, A., and Kulmala, M.: Heterogeneous nucleation as a potential sulphate-coating mechanism of atmospheric mineral dust particles and implications of coated dust on new particle formation, *J. Geophys. Res.*, 108(D17), 4546, doi:10.1029/2003JD003553, 2003.
- Kuang, C., McMurry, P. H., and McCormick, A. V.: Determination of cloud condensation nuclei production from measured new particle formation events, *Geophys. Res. Lett.*, 36, L09822, doi:10.1029/2009GL037584, 2009.
- Kulmala, M., Laaksonen, A., and Pirjola, L.: Parameterizations for sulfuric acid/water nucleation rates, *J. Geophys. Res.*, 103, 8301–8307, 1998.
- Kulmala, M., Hämeri, K., Aalto, P. P., Mäkelä, J. M., Pirjola, L., Nilsson, E. D., Buzorius, G., Rannik, U., Dal Maso, M., Seidl, W., Hoffmann, T., Jansson, R., Hansson, H.-C., Viisanen, Y., Laaksonen, A., and O’Dowd, C. D.: Overview of the international project on biogenic aerosol formation in the boreal forest (BIOFOR), *Tellus B*, 53, 324–343, 2001.
- Kulmala, M., Vehkamäki, H., Petäjä, T., Dal Maso, M., Lauri, A., Kerminen, V.-M., Birmili, W., and McMurry, P.: Formation and growth rates of ultrafine atmospheric particles: a review of observations, *J. Aerosol Sci.*, 35, 143–176, doi:10.1016/j.jaerosci.2003.10.003, 2004.
- Kulmala, M., Petäjä, T., Mönkkönen, P., Koponen, I. K., Dal Maso, M., Aalto, P. P., Lehtinen, K. E. J., and Kerminen, V.-M.: On the growth of nucleation mode particles: source rates of condensable vapor in polluted and clean environments, *Atmos. Chem. Phys.*, 5, 409–416, doi:10.5194/acp-5-409-2005, 2005.
- Kulmala, M., Lehtinen, K. E. J., and Laaksonen, A.: Cluster activation theory as an explanation of the linear dependence between formation rate of 3 nm particles and sulphuric acid concentration, *Atmos. Chem. Phys.*, 6, 787–793, doi:10.5194/acp-6-787-2006, 2006.
- Laaksonen, A., Hamed, A., Joutsensaari, J., Hiltunen, L., Cavalli, F., Junkermann, W., Asmi, A., Fuzzi, S., and Facchini, M. C.: Cloud condensation nucleus production from nucleation events at a highly polluted region, *Geophys. Res. Lett.*, 32, L06812, doi:10.1029/2004GL022092, 2005.
- Laaksonen, A., Kulmala, M., Berndt, T., Stratmann, F., Mikkonen, S., Ruuskanen, A., Lehtinen, K. E. J., Dal Maso, M., Aalto, P., Petäjä, T., Riipinen, I., Sihto, S.-L., Janson, R., Arnold, F., Hanke, M., Ücker, J., Umann, B., Sellegri, K., O’Dowd, C. D., and Viisanen, Y.: SO_2 oxidation products other than H_2SO_4 as a trigger of new particle formation. Part 2: Comparison of ambient and laboratory measurements, and atmospheric implications, *Atmos. Chem. Phys.*, 8, 7255–7264, doi:10.5194/acp-8-7255-2008, 2008a.

- Laaksonen, A., Kulmala, M., O'Dowd, C. D., Joutsensaari, J., Vaattovaara, P., Mikkonen, S., Lehtinen, K. E. J., Sogacheva, L., Dal Maso, M., Aalto, P., Petäjä, T., Sogachev, A., Yoon, Y. J., Lihavainen, H., Nilsson, D., Facchini, M. C., Cavalli, F., Fuzzi, S., Hoffmann, T., Arnold, F., Hanke, M., Sellegri, K., Umann, B., Junkermann, W., Coe, H., Allan, J. D., Alfarra, M. R., Worsnop, D. R., Riekkola, M.-L., Hyötyläinen, T., and Viisanen, Y.: The role of VOC oxidation products in continental new particle formation, *Atmos. Chem. Phys.*, 8, 2657–2665, doi:10.5194/acp-8-2657-2008, 2008b.
- Lihavainen, H., Kerminen, V.-M., Komppula, M., Hatakka, J., Aaltonen, V., Kulmala, M., and Viisanen, Y.: Production “potential” cloud condensation nuclei associated with atmospheric new-particle formation in northern Finland, *J. Geophys. Res.*, 108(D24), 4782, doi:10.1029/2003JD0038872, 2003.
- McCullach, P. and Nelder, J. A.: *Generalized Linear Models*, Chapman and Hall, London, 1989.
- McCulloch, C. E. and Searle, S. R.: *Generalized, Linear, and Mixed Models*, Wiley, New York, 2001.
- Mikkonen, S., Lehtinen, K. E. J., Hamed, A., Joutsensaari, J., Facchini, M. C., and Laaksonen, A.: Using discriminant analysis as a nucleation event classification method, *Atmos. Chem. Phys.*, 6, 5549–5557, doi:10.5194/acp-6-5549-2006, 2006.
- Paasonen, P., Sihto, S.-L., Nieminen, T., Vuollekoski, H., Riipinen, I., Plaß-Dülmer, C., Berresheim, H., Birmili, W., and Kulmala, M.: Connection between new particle formation and sulphuric acid at Hohenpeissenberg (Germany) including the influence of organic compounds, *Boreal Environ. Res.*, 14, 616–629, 2009.
- Pierce, J. R. and Adams, P. J.: Efficiency of cloud condensation nuclei formation from ultrafine particles, *Atmos. Chem. Phys.*, 7, 1367–1379, doi:10.5194/acp-7-1367-2007, 2007.
- Pirjola, L., Laaksonen, A., Aalto, P., and Kulmala, M.: Sulfate aerosol formation in the Arctic boundary layer, *J. Geophys. Res.*, 103, 8309–8322, 1998.
- Pope III, C. A. and Dockery, D. W.: 2006 critical review – health effects of fine particulate air pollution: lines that connect, *J. Air Waste Manage.*, 56, 709–742, 2006.
- Spracklen, D. V., Pringle, K. J., Carslaw, K. S., Chipperfield, M. P., and Mann, G. W.: A global off-line model of size-resolved aerosol microphysics: I. Model development and prediction of aerosol properties, *Atmos. Chem. Phys.*, 5, 2227–2252, doi:10.5194/acp-5-2227-2005, 2005.
- Stockwell, D. and Chipperfield, M.: A tropospheric chemical-transport model: Development and validation of the model transport schemes, *Q. J. R. Meteorol. Soc.*, 125, 1747–1783, doi:10.1002/qj.49712555714, 1999.
- Vaattovaara, P., Huttunen, P. E., Yoon, Y. J., Joutsensaari, J., Lehtinen, K. E. J., O'Dowd, C. D., and Laaksonen, A.: The composition of nucleation and Aitken modes particles during coastal nucleation events: evidence for marine secondary organic contribution, *Atmos. Chem. Phys.*, 6, 4601–4616, doi:10.5194/acp-6-4601-2006, 2006.
- Vehkamäki, H., Dal Maso, M., Hussein, T., Flanagan, R., Hyvärinen, A., Lauros, J., Merikanto, P., Mönkkönen, M., Pihlatie, K., Salminen, K., Sogacheva, L., Thum, T., Ruuskanen, T. M., Keronen, P., Aalto, P. P., Hari, P., Lehtinen, K. E. J., Rannik, Ü., and Kulmala, M.: Atmospheric particle formation events at Värriö measurement station in Finnish Lapland 1998–2002, *Atmos. Chem. Phys.*, 4, 2015–2023, doi:10.5194/acp-4-2015-2004, 2004.
- Wehner, B., Petäjä, T., Boy, M., Engler, C., Birmili, W., Tuch, T., Wiedensohler, A., and Kulmala, M.: The contribution of sulfuric acid and non-volatile compounds on the growth of freshly formed atmospheric aerosols, *Geophys. Res. Lett.* 32(17), L17810, doi:10.1029/2005GL023827, 2005.
- Whitby, E. R. and McMurry, P. H.: Modal aerosol dynamics modeling, *Aerosol Sci. Tech.*, 27, 673–688, 1997.
- Wiedensohler, A., Chen, Y. F., Nowak, A., Wehner, B., Achtert, P., Berghof, M., Birmili, W., Wu, Z. J., Hu, M., Zhu, T., Takegawa, N., Kita, K., Kondo, Y., Lou, S. R., Hofzumahaus, A., Holland, F., Wahner, A., Gunthe, S. S., Rose, D., Su, H., and Pöschl, U.: Rapid aerosol particle growth and increase of cloud condensation nucleus activity by secondary aerosol formation and condensation: A case study for regional air pollution in northeastern China, *J. Geophys. Res.*, 114, D00G08, doi:10.1029/2008JD010884, 2009.