Towards an online-coupled chemistry-climate model: evaluation of trace gases and aerosols in COSMO-ART


1 Laboratory for Air Pollution/Env. Technology, Empa Materials and Science, 8600 Duebendorf, Switzerland
2 C2SM Center for Climate Systems Modeling, ETH, Zurich, Switzerland
3 Institute for Meteorology and Climate Research, Karlsruhe Institute of Technology, Karlsruhe, Germany
4 School of Earth Atmospheric, and Environmental Sciences, National Centre for Atmospheric Science, University of Manchester, Manchester, UK
5 Department of Physics, University of Helsinki, Helsinki, Finland
6 Air Quality Research, Finnish Meteorological Institute, Helsinki, Finland
7 TNO Princetonlaan 6, 3584 CB Utrecht, The Netherlands
8 CIRES and Dept. of Chemistry and Biochemistry, Univ. of Colorado, Boulder, CO, USA
9 Institut IEK-8, Troposphäre, Forschungszentrum Jülich, Jülich, Germany
10 Laboratory of Atmospheric Chemistry, Paul Scherrer Institute, Villigen, Switzerland
11 Leibniz Institute for Tropospheric Research, Leipzig, Germany
12 Division of Nuclear Physics, Department of Physics, Lund University, Lund, Sweden

Received: 9 July 2011 – Published in Geosci. Model Dev. Discuss.: 4 August 2011
Revised: 23 November 2011 – Accepted: 25 November 2011 – Published: 2 December 2011

Abstract. The online-coupled, regional chemistry transport model COSMO-ART is evaluated for periods in all seasons against several measurement datasets to assess its ability to represent gaseous pollutants and ambient aerosol characteristics over the European domain. Measurements used in the comparison include long-term station observations, satellite and ground-based remote sensing products, and complex datasets of aerosol chemical composition and number size distribution from recent field campaigns. This is the first time these comprehensive measurements of aerosol characteristics in Europe are used to evaluate a regional chemistry transport model. We show a detailed analysis of the simulated size-resolved chemical composition under different meteorological conditions. Mean, variability and spatial distribution of the concentrations of O\textsubscript{3} and NO\textsubscript{x} are well reproduced. SO\textsubscript{2} is found to be overestimated, simulated PM\textsubscript{2.5} and PM\textsubscript{10} levels are on average underestimated, as is AOD. We find indications of an overestimation of ship emissions. Time evolution of aerosol chemical composition is captured, although some biases are found in relative composition. Nitrate aerosol components are on average overestimated, and sulfates underestimated. The accuracy of simulated organics depends strongly on season and location. While strongly underestimated during summer, organic mass is comparable in spring and autumn. We see indications for an overestimated fractional contribution of primary organic matter in urban areas and an underestimation of SOA at many locations. Aerosol number concentrations compare well with measurements for larger size ranges, but overestimations of particle number concentration with factors of 2–5 are found for particles smaller than 50 nm. Size distribution characteristics are often close to measurements, but show discrepancies at polluted sites. Suggestions for further improvement of the modeling system consist of the inclusion of a revised secondary organic aerosols scheme, aqueous-phase chemistry and improved aerosol boundary conditions. Our work sets the basis for subsequent studies of aerosol characteristics and climate impacts with COSMO-ART, and highlights areas where improvements are necessary for current regional modeling systems in general.

Correspondence to: D. Brunner (dominik.brunner@empa.ch)
1 Introduction

Aerosols affect climate through changes in the radiation budget (direct effect), the subsequent changes in atmospheric stratification (semi-direct effect, Haywood and Boucher, 2000) and through changes in cloud development and lifetime due to the differences in available cloud condensation/ice nuclei (indirect effects, Lohmann and Feichter, 2005). Aerosols also constitute a health concern if they are small enough to traverse the human respiratory tract (Laden et al., 2006; Dockery et al., 1996). Once in the lungs their toxicity depends on size (Donaldson et al., 2000) and chemical composition (Aktories et al., 2009; Hoek et al., 2002). Within the climate system, their influence on the radiation budget depends on their optical properties, and how they affect clouds is a function of size and hygroscopicity. Size, chemical composition, and optical properties are therefore indispensable parameters that need to be well represented if any study of aerosol effects should be accurate.

Up to now, climate modeling studies including aerosols often lack a comprehensive description of aerosol characteristics, due to the high computational demand of such a complex effort. Approaches range from simple bulk mass aerosol schemes with only externally mixed aerosols, up to multi-component, size-resolving aerosol modules including explicit aging of aerosols and interactions with radiation and clouds. Often these modules lack parts (or all) of the interaction between gas- and aerosol-phase. Nucleation of ammonium-sulfate particles is represented in most models, and also the condensation of organics onto particles is included in some. Nitrates, which can represent up to 50% of ambient aerosol mass in polluted regions (Putaud et al., 2004), were missing for example in all but two models participating in the Fourth Assessment Report of the Intergovernmental Panel on Climate Change (IPCC, Meehl et al., 2007). This was probably due to the lack of the necessary, but computationally expensive, gas-phase chemistry leading to nitrate formation.

Current efforts try to bridge the gap between accurate representation of all aerosol components while retaining the ability to model climatic timescales. To reach this goal it is necessary to couple climate and air quality models. One such modeling system which focuses on the regional scale combines the numerical weather prediction model of the Consortium for Small Scale Modeling (COSMO, Baldauf et al., 2011) with an extension for Aerosols and Reactive Trace gases: COSMO-ART (Vogel et al., 2009). It is based on state-of-the-art components for the description of meteorology, chemistry and aerosols and features an integrated approach to couple them. Such an “online”-coupling allows for consistent treatment of all components by the same parameterization (e.g. advection, diffusion, convection) and avoids unnecessary interpolation steps. Additionally, simulation of feedbacks between chemistry, aerosols and meteorology becomes possible. Grell and Baklanov (2011) showed the importance of this approach and its benefits compared to traditional “offline” models, and Zhang (2008) gave a comprehensive overview of the available modeling systems. COSMO-ART is in its composition very similar to the Weather Research and Forecasting model (WRF) extended by chemistry and aerosols: WRF/chem. Grell et al. (2005) presented a comprehensive evaluation for this modeling system. Most of the components of COSMO-ART are well known and tested. However, their interplay and integration into the modeling system lacks a thorough evaluation.

In this work we analyse COSMO-ART regarding its ability to represent ambient concentrations of gaseous and particulate matter constituents over Europe under different meteorological conditions. Through a detailed analysis of aerosol size distributions and chemical composition we set the basis for subsequent analyses of aerosol-climate interactions in COSMO-ART. We have collected an extensive evaluation dataset of satellite-derived NO\textsubscript{2} and aerosol optical depth (AOD), long-term station measurements for gas-phase tracers, bulk aerosol mass and optical properties, as well as aerosol mass spectrometer (AMS) measurements of aerosol chemical composition and measurements of aerosol size distribution. The comprehensive datasets of aerosol characteristics have been created during recent field campaigns of the European integrated Project on aerosol cloud climate air quality interactions (EUCAARI, Kulmala et al., 2009), during intensive measurement campaigns of the European Monitoring and Evaluation Programme (EMEP, http://www.epmep.int) and in coordinated measurements of the European Supersites for Atmospheric Aerosol Research (EUSAAR, http://www.eusaar.net) and the German Ultrasfine Aerosol Network (GUAN, Birmili et al., 2009).

Our simulations employ full gas-phase chemistry and aerosol dynamics. Spatial and temporal resolution of input data (meteorology, anthropogenic emissions) and model setup is on the top end of currently possible simulations. While the modeling system is currently still too expensive to be used for climate simulations, the results of our evaluation efforts can be seen as a benchmark for what degree of accuracy in simulation gas and aerosol characteristics can be expected in future fully-coupled regional chemistry-climate models, and identify model deficiencies which would need to be remedied before such simulations can be made.

We begin with a description of the system, its setup and the measurement datasets used in evaluation. The second chapter describes the findings of our evaluation against the different datasets and discusses the results. The last chapter provides a more in-depth discussion of simulated aerosol characteristics. We conclude with implications for future studies and give directions for further developments of the modeling system.
2 Methods

2.1 Modeling system

COSMO-ART is a regional chemistry transport model, online-coupled to the COSMO regional numerical weather prediction and climate model (Baldauf et al., 2011). COSMO is operationally used for numerical weather prediction (NWP) purposes by several European national meteorological services and research institutes. In its climate version (Rockel et al., 2008) it has been used in several studies of regional climate impact assessment (e.g. Jaeger and Seneviratne, 2010; Suklitsch et al., 2008; Hohenegger et al., 2008) and participated in the IPCC fourth assessment report modeling ensemble (Christensen et al., 2007). The extension for Aerosols and Reactive Trace gases (ART) contains a modified version of the Regional Acid Deposition Model, Version 2 (RADM2) gas-phase chemistry mechanism (Stockwell et al., 1990). It has been extended by a more sophisticated isoprene scheme of Geiger et al. (2003) for a better description of biogenic volatile organic compounds (VOC), but does not include recent findings regarding formation of secondary organic aerosols and OH recycling due to isoprene chemistry (e.g. Paulot et al., 2009). Aerosols are represented by the modal aerosol module MADE (Modal Aerosol Dynamics Model for Europe, Ackermann et al., 1998), improved by explicit treatment of soot aging through condensation of inorganic salts (Riems et al., 2003) and additional modes for mineral dust (Stanelle et al., 2010) and sea salt. Nucleation of new particles is formulated according to Kerminen and Wexler (1994) allowing for binary homogeneous nucleation of sulfuric acid. The condensation of vapours from biogenic and anthropogenic VOCs is parametrized with the Secondary Organic Aerosol Model (SORGAM) of Schell et al. (2001). This is still a commonly used module, although Fast et al. (2009) showed that this scheme underpredicts SOA concentrations by up to a factor of 10 in very polluted regions. Biogenic VOC emission fluxes, considering isoprene, α-pinene, other monoterpenes and a class of unidentified compounds, are calculated online with a Guenther-type model presented in Vogel et al. (1995), using land use data from the Global Land Cover 2000 (GLC2000) dataset (Bartholomé and Belward, 2005). Seasalt emissions follow Lundgren (2006), and mineral dust is parameterized as described in Vogel et al. (2006). Dry deposition is modeled by a resistance approach (Baer and Nester, 1992). Washout of aerosols is included by a parameterization of Rinke (2008). Wet removal of gases and aqueous-phase chemistry are currently not considered. COSMO-ART is fully online-coupled, and currently allows for feedbacks of aerosols on radiation (direct/semi-indirect effects). Cloud feedbacks (indirect effects) have been included in a research version (Bungert et al., 2011) but were not used in this work. A complete description of the modeling system can be found in Vogel et al. (2009) and references therein. In our study, COSMO-ART based on COSMO version 4.17 is used.

For meteorology we used initial and boundary conditions from the European Centre for Medium-Range Weather Forecasts (ECMWF) Integrated Forecast System (IFS) model, with an update frequency of 3h. For runs on climatic timescales boundary data could e.g. be provided by the ECHAM-HAM (Stier et al., 2005) or (for past episodes) by ERA-40 (Uppala et al., 2005)/ERA-Interim (Simmons et al., 2007) reanalyses, which are all based on the IFS and would therefore deliver comparable meteorology. Boundary data for gas-phase species, including most of the lumped NMVOC compounds, were provided through simulations of the Model for Ozone and Related Chemical Tracers (MOZART) driven by meteorological data from the National Center for Environmental Prediction (NCEP) presented in Emmons et al. (2010), with an update frequency of 6h. No boundary data for aerosol components were available from MOZART or other models that matched our aerosol mechanism. Therefore, we took the output of a previous (otherwise identical) simulation of COSMO-ART and chose one point in the Northern Atlantic (8.7° W, 47.4° N, see Fig. 1). We averaged the simulated aerosol characteristics over the complete simulation period, and used this vertical column as lateral boundary conditions for all aerosol variables. While this gives more realistic aerosol concentrations at the boundaries, the total inflow will still be underestimated. In this work we will show that simulated particulate matter concentrations are often underestimated, which will also be the case for boundary conditions based on such a simulation.

The emission inventory for Europe developed by TNO (Netherlands) within the Monitoring Atmospheric Composition and Climate (MACC) project (TNO/MACC, Kuenen et al., 2011; Denier van der Gon et al., 2010) provides anthropogenic emissions. This is a follow-up and improvement of the earlier TNO-GEMS emission database (Visschedijk et al., 2007). Therein, emissions from 10 different SNAP (Selected Nomenclature for sources of Air Pollution) source categories are represented by a spatial pattern of annual emission totals for the years 2003–2007, and statistical time functions for species, country and source category dependent monthly, weekly and daily cycles. Our speciation of non-methane volatile organic compounds (NMVOC) mass totals is done using composition information from Pasquant (2002) and a translation matrix to RADM2 (J. Keller, PSI, Switzerland, personal communication, 2009). Aerosol emissions are provided as mass totals of particulate matter below 10µm (PM_{10}) and below 2.5µm (PM_{2.5}) in diameter. We distribute them onto the different MADE modes following Elleman and Covert (2010), with a disaggregation into chemical components using a split table from TNO (Table 1). Emission country totals per SNAP category from the International Institute for Applied Systems Analysis (IIASA, PRIMES09 scenario) serve to extrapolate TNO/MACC emissions to years after 2007. With its spatial resolution of about
Fig. 1. Model domain and measurement station positions. AIRBASE/EMEP: gas-phase and bulk aerosol mass, AMS: aerosol chemical composition (EIMP/EUCAARI/EUSAAR), DMPS/SMPS: aerosol size distribution (EUSAAR/GUAN), AERONET: aerosol optical depth. The position of the column used for aerosol boundary conditions is marked with a red cross (aerosol IC/BC). Stations with AMS and/or DMPS/SMPS measurements: (a) Payerne (CH), (b) Aucencorth/Bush (UK), (c) Mel-pitz (DE), (d) Vavihill (SE), (e) Hyytiäliä (FI), (f) K-Puszta (HU), (g) Cabauw (NL), (h) Helsinki (FI), (i) Barcelona (ES), (j) Montseny (ES), (k) Kosetice (CZ), (m) Aspvreten (SE), (n) Birkenes (NO), (o) Mace Head (IE), (p) Ispra (IT), (q) Wald-hof (DE). Several named stations measured both chemical composition and size distribution.

8 km (0.125 × 0.0625 °), the description of the time evolution of emissions and the comprehensive set of emitted species this dataset is one of the most detailed currently available emission inventories for Europe. Preparation of all input datasets for COSMO-ART is done using INT2COSMO-ART (Appendix A).

Our modeling domain (Fig. 1) covers the greater European region, with a horizontal resolution of 0.17° and a grid of 200 × 190 points. Vertically, the model is discretized into 40 terrain-following hybrid sigma levels, with the lowest level at 10 m (layer thickness: 20 m) and ranging up to approx. 24,000 m (20 hPa). A Runge-Kutta time integration scheme is employed with time steps of 40 s. Tracers are advected horizontally via a semi-lagrangian method conserving mass over the total domain ("globally mass-conserving"). The overall model configuration closely follows the current operational setup of COSMO-EU of the German Meteorological Service (DWD).

2.2 Measurement data

Meteorological parameters have been taken from the operational surface synoptic observations (SYNOP) network, providing measurements for temperature, dew point temperature, wind speed and direction at or in the vicinity of most measurement points of chemical composition. In the EMEP programme a number of stations throughout Europe report quality-controlled, long-term measurements of gaseous precursor substances and aerosol variables. AIRBASE (European AIR quality dataBASE, http://airbase.eionet.europa.eu/) provides measurements at a much larger number of stations, but with heterogeneous quality and mostly at rather polluted locations not representative for the model grid size of 0.17° (approx. 19 km at model domain center). While AIRBASE, in its recently published version 5, provides data up to the end of 2009, EMEP data were only available until 2008. As one of our simulation periods is in 2009, we settled on the following method to provide a homogeneous dataset of measurements for gas-phase species and aerosol mass for all periods: We retrieved data from AIRBASE, but restricted the stations used to those which also report to EMEP. As discrepancies between modelled and measured values might be related to the type and location of a measurement station, we have additionally disaggregated the selected stations into categories based on the representativeness study done by Henne et al. (2010), which includes a more comprehensive analysis of the surroundings of each station. Therein, stations are classified regarding their pollution burden and usability in a model evaluation. We have used the “alternative classification” described in the Supplement S3 in Henne et al. (2010), which gives classes ranging from very clean stations ("rural") via stations with very variable pollution levels ("rural/coastal") and stations representative for a larger area ("rural"), up to stations with a strong influence of large urban areas in their vicinity ("suburban/urban"). Most EMEP stations are found in the “rural” and “rural/coastal” classes, and are seen as the most representative when evaluating model results.

The Aerosol Robotic Network (AERONET) (Holben et al., 1998) provides measurements of aerosol optical depth (AOD) for analysis of the optical properties. Aerosol mass spectrometer (AMS) measurements give quantitative measurements of the chemical composition of submicron non-refractory aerosol mass (NR−PM1) with high temporal resolution (Canagaratna et al., 2007). AMS data collected at several sites throughout Europe during measurement campaigns of the EMEP/EUCAARI project in October 2008 and March 2009 were used, as well as from an EMEP intensive campaign in June 2006. No evaluation of elemental carbon has been made, as the different measurement techniques used make even inter-station comparison difficult (Andreae and Gelencsér, 2006), and devising a homogenized dataset was out of scope for this work. Homogenized measurements of aerosol size distribution from scanning mobility particle
Table 1. Contributions (in % mass) to PM$_{2.5}$ emissions as used in the TNO/GEMS emission inventory. Sodium (Na) is not used directly in the simulations, but added to the “other primary” category, representing the remaining, non-carbonaceous primary PM$_{2.5}$ part (including e.g. minerals, metal oxides, product emissions). Sulfate contributions have been calculated assuming 2% of total emitted SO$_2$ mass (IIASA RAINS emissions for 2000) is H$_2$SO$_4$ for all SNAP categories except SNAP 1 and 3. There, measured compositions of coal fly ash (as dominant contributor to source category) as reported by Lipsky et al. (2002) and Senior et al. (2000) are used as basis. OC depicts organic carbon, a ratio of 1.3 has been used to convert OC to organic aerosol (OA).

<table>
<thead>
<tr>
<th>SNAP</th>
<th>Description</th>
<th>SO$_2^2$</th>
<th>OC</th>
<th>BC</th>
<th>Na</th>
<th>other primary</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Energy transformation</td>
<td>15.0</td>
<td>2.0</td>
<td>1.0</td>
<td>1.5</td>
<td>80.0</td>
</tr>
<tr>
<td>2</td>
<td>Small combustion sources</td>
<td>2.0</td>
<td>35.0</td>
<td>18.0</td>
<td>1.0</td>
<td>44.0</td>
</tr>
<tr>
<td>3</td>
<td>Industrial combustion</td>
<td>10.0</td>
<td>3.0</td>
<td>1.0</td>
<td>1.0</td>
<td>84.0</td>
</tr>
<tr>
<td>4</td>
<td>Industrial process emissions</td>
<td>3.0</td>
<td>3.0</td>
<td>5.0</td>
<td>1.5</td>
<td>88.0</td>
</tr>
<tr>
<td>5</td>
<td>Extraction of fossil fuels</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.5</td>
<td>99.5</td>
</tr>
<tr>
<td>6</td>
<td>Solvent and product use</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>7</td>
<td>Road transport</td>
<td>1.0</td>
<td>32.0</td>
<td>49.0</td>
<td>0.1</td>
<td>17.0</td>
</tr>
<tr>
<td>8</td>
<td>Non road transport (e.g. international shipping)</td>
<td>3.0</td>
<td>31.0</td>
<td>41.0</td>
<td>0.2</td>
<td>26.0</td>
</tr>
<tr>
<td>9</td>
<td>Waste handling and disposal</td>
<td>0.0</td>
<td>31.0</td>
<td>20.0</td>
<td>0.0</td>
<td>49.0</td>
</tr>
<tr>
<td>10</td>
<td>Agriculture</td>
<td>0.0</td>
<td>48.0</td>
<td>15.0</td>
<td>0.0</td>
<td>37.0</td>
</tr>
</tbody>
</table>

sizer (SMPS) and differential mobility particle sizer (DMPS) instruments were provided in Asmi et al. (2011) as a result of the EUSAAR project and data from the GUAN network (Birmili et al., 2009), with 24 measurement sites in Europe. Figure 1 shows the locations of ground-based stations used in our evaluation.

Finally, satellite-derived datasets provide a vertically integrated view on model performance. In our analysis, tropospheric columns of NO$_2$ from the Ozone Monitoring Instrument (OMI) were used for gas-phase comparison. The NO$_2$ columns are based on the Empa OMI NO$_2$ retrieval (EOMINO) which includes several improvements as compared to operational products in particular regarding a better representation of topography and surface reflectance using high-resolution data sets (Zhou et al., 2009, 2010). To estimate the accuracy of the spatial distribution of simulated aerosol loadings aerosol optical depth (AOD) retrieved from the Moderate Resolution Imaging Spectrometer (MODIS) (Levy et al., 2007, MOD04_L2 product) were used.

2.3 Investigation periods

The selection of the investigation periods was driven by two goals: to evaluate model performance under typical weather conditions and in all seasons, and to have AMS measurement data available for comparison. Apart from the campaign measurement data, AIRBASE and satellite data were available for all simulations. The following periods were chosen:

2.3.1 “Winter case”: 23 January–11 February 2006

A stable high pressure system with very low surface temperatures was present over Europe from 23 January onwards, with only minor disturbances on 5–7 February. Over Switzerland and Eastern Europe, this resulted in an episode with strong temperature inversions and exceptionally high particulate matter (PM) concentrations. The Swiss legislative limit for daily mean PM$_{10}$ (particulate matter below 10 µm in diameter) of 50 µg m$^{-3}$ was exceeded every day between 27 January and 5 February at several measurement stations. This episode represents a typical winter situation where high pollution levels are building up through strong inversions and local emissions are the strongest contributors to pollution levels (Holst et al., 2008).

2.3.2 “Summer case”: 10–29 June 2006

This episode was characterized by dry, sunny and warm conditions due to a stable high pressure system from 10–24 June, and a transient low pressure system with embedded thunderstorms on 25 to 29 June. Such a situation is associated with strong photochemistry and high O$_3$ levels, representing a typical “summersmog” episode. AMS instruments were deployed in Payerne (CH), Harwell and Auchencorth (UK) during this period in the context of an EMEP intensive measurement campaign. We used data from Payerne and Auchencorth in our analysis.

2.3.3 “Autumn case”: 1–20 October 2008

A low pressure system over Scandinavia brought polar air masses towards Europe at the beginning of the month. From 5–20 October generally mild and sunny conditions prevailed. On 16 October a low pressure disturbance passed, bringing rain to Central Europe. Frequent disturbances by mesoscale systems gradually change a summertime atmosphere towards a wintertime one in this simulation. During this period,
an EMEP/EUCAARI measurement campaign took place, from which we received AMS data for Payerne (CH), Melpitz (DE), Vaihills (SE), Hyytiälä (FI) and K-Puszta (HU). EUSAAR size distribution data were available for this period.

2.3.4 “Spring case”: 1–20 March 2009

A low pressure system originating over the North Atlantic brought cold weather on 1 and 2 March. It was followed by spring-like conditions from 13–18 March, and a cold surge from NE on 20 March. We regard this situation as typical of spring, with first warm days including the initial onset of BVOC emissions, intermitted by “cleansing” periods with clouds, precipitation and strong mesoscale forcing. Another EMEP/EUCAARI campaign took place during this period, from which we present data from AMS instruments deployed in Payerne (CH), Melpitz (DE), Vaihills (SE), Hyytiälä (FI), Cabauw (NL), Helsinki (FI), Barcelona (ES), and Montseny (ES). EUSAAR size distribution data were available for this period.

3 Evaluation

The following section contains a description of the results of our evaluation efforts, starting with meteorology, then trace gases and finally aerosol characteristics. Each section is accompanied by a figure/table summarizing the results for the species discussed. Section 4 then further elaborates on the results for aerosol characteristics.

3.1 Meteorology

COSMO-ART is in its meteorological core code identical with the NWP model COSMO, and its performance is continuously verified by several European weather services and in more detail also within field campaigns like e.g. in Barthlott et al. (2011). Meteorological evaluation has therefore been limited in this work to surface parameters. In all periods, the comparison of simulated temperature, dew point temperature, wind direction and wind speed show very good agreement with SYNOP measurement data both in terms of temporal variability and average values (Fig. 2). Sometimes the (diurnal) variability is underestimated by the simulations (not shown), which is not unexpected for such coarse grid simulations due to the averaging onto a 0.17° grid box (e.g. Schlünzen and Katzfey, 2003; Heinemann and Kerschgens, 2005). The means of temperature, wind speed and direction are well reproduced (Table 2). Except for the summer 2006 period, where the model shows a negative bias (Fig. 2), also relative humidity is realistically represented. The negative bias in summer 2006 might be related to an unrealistic initialization of soil moisture. Further investigation is needed to remedy this deficiency.

IFS analysis data were used to initialize and force the model at the lateral boundaries. Within the model domain COSMO runs freely, creating its own dynamics. This is not the best possible setup. Constant data assimilation from observations like it is done for operational analysis (e.g. nudging), or a reinitialization of meteorology after one or two days could further improve meteorology. However, using simple comparison with SYNOP data we found no significant loss in accuracy of the simulation over the whole integration period when compared against several SYNOP stations, suggesting that the lateral forcing provides a sufficiently strong constraint for the meteorology within the model domain. Some of the underestimated (diurnal) variability found would likely be improved at increased resolution.

The modest deficiencies found such as an underestimated diurnal variability are well known to NWP modellers and represent problems such models are currently faced with in general (e.g. Schlünzen and Katzfey, 2003; Heinemann and Kerschgens, 2005). Mean wind speeds simulated by the model, for example, are below 5% biases at nearly all stations in all periods (Table 2), and temperatures show essentially no bias. Overall, meteorology is well represented and these findings set the basis for a successful air quality simulation. They also highlight one of the key benefits of this modeling system: its direct coupling to an operational weather prediction model.
Table 2. Normalized mean biases (%) in spring, summer, autumn and winter (SP, SU, AU, WI) for SYNOP and AIRBASE datasets, satellite comparisons and aerosol datasets. Satellite biases are the mean over all grid points, only land points and only sea points. AERONET comparison shows biases model – AERONET (AER), model – MODIS (MOD) at AERONET station, and MODIS – AERONET (DIF). AMS and EUSAAR biases are the mean over all stations presented in the evaluation, the number of stations in each period is given in the notes.

<table>
<thead>
<tr>
<th></th>
<th>SP</th>
<th>SU</th>
<th>AU</th>
<th>WI</th>
<th>SP</th>
<th>SU</th>
<th>AU</th>
<th>WI</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Meteorology (SYNOP)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$T_{2m}$</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>WS$_{10m}$</td>
<td>3</td>
<td>3</td>
<td>6</td>
<td>−1</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>WD$_{10m}$</td>
<td>6</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>6</td>
<td>−15</td>
<td>−4</td>
<td>10</td>
</tr>
<tr>
<td>RH</td>
<td>6</td>
<td>−15</td>
<td>−4</td>
<td>10</td>
<td>6</td>
<td>−15</td>
<td>−4</td>
<td>10</td>
</tr>
<tr>
<td>Gas–phase tracers and aerosol bulk mass (12:00–18:00 LT)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$O_3$</td>
<td>−12</td>
<td>−15</td>
<td>2</td>
<td>−22</td>
<td>rural</td>
<td>−16</td>
<td>−3</td>
<td>2</td>
</tr>
<tr>
<td>$NO_2$</td>
<td>−2</td>
<td>−10</td>
<td>18</td>
<td>11</td>
<td>rural</td>
<td>7</td>
<td>−39</td>
<td>44</td>
</tr>
<tr>
<td>NO</td>
<td>−2</td>
<td>−14</td>
<td>−21</td>
<td>−40</td>
<td>rural</td>
<td>69</td>
<td>19</td>
<td>75</td>
</tr>
<tr>
<td>SO$_2$</td>
<td>108</td>
<td>0</td>
<td>49</td>
<td>39</td>
<td>rural</td>
<td>60</td>
<td>72</td>
<td>324</td>
</tr>
<tr>
<td>PM$_{10}$</td>
<td>−35</td>
<td>−48</td>
<td>3</td>
<td>−43</td>
<td>rural</td>
<td>8</td>
<td>−39</td>
<td>15</td>
</tr>
<tr>
<td>PM$_{2.5}$</td>
<td>−2</td>
<td>−55</td>
<td>−11</td>
<td>1</td>
<td>rural</td>
<td>13</td>
<td>21</td>
<td>43</td>
</tr>
<tr>
<td><strong>Satellite observations</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NO$_2$</td>
<td>31</td>
<td>1</td>
<td>56</td>
<td>59</td>
<td>all</td>
<td>10</td>
<td>−11</td>
<td>38</td>
</tr>
<tr>
<td>(OMI)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AOD</td>
<td>−58</td>
<td>−56</td>
<td>−55</td>
<td>−49</td>
<td>all</td>
<td>−50</td>
<td>−58</td>
<td>−31</td>
</tr>
<tr>
<td>(MODIS)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AOD at stations (AERONET/MODIS)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AOD</td>
<td>−22</td>
<td>−61</td>
<td>−14</td>
<td>−46</td>
<td>AER</td>
<td>−30</td>
<td>−57</td>
<td>−27</td>
</tr>
<tr>
<td><strong>Aerosol chemical composition</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NO$_2^-$</td>
<td>39</td>
<td>−48</td>
<td>47</td>
<td>8,3,5,</td>
<td>−34</td>
<td>−33</td>
<td>−72</td>
<td>34</td>
</tr>
<tr>
<td>(NH$_4^+$)</td>
<td>−33</td>
<td>−72</td>
<td>−34</td>
<td>8,3,5,</td>
<td>−33</td>
<td>−72</td>
<td>−34</td>
<td>8,3,5,</td>
</tr>
<tr>
<td>SO$_4^{2-}$</td>
<td>−86</td>
<td>−84</td>
<td>−81</td>
<td>8,3,5,</td>
<td>−86</td>
<td>−84</td>
<td>−81</td>
<td>8,3,5,</td>
</tr>
<tr>
<td>OA</td>
<td>1</td>
<td>−61</td>
<td>−17</td>
<td>8,2,5,</td>
<td>1</td>
<td>−61</td>
<td>−17</td>
<td>8,2,5,</td>
</tr>
<tr>
<td><strong>Aerosol number concentrations</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$N_{30050}$</td>
<td>149</td>
<td>97</td>
<td>9,−10,</td>
<td>−</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$N_{50}$</td>
<td>71</td>
<td>53</td>
<td>9,−10,</td>
<td>−</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$N_{100}$</td>
<td>64</td>
<td>76</td>
<td>9,−10,</td>
<td>−</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$N_{250}$</td>
<td>139</td>
<td>257</td>
<td>9,−10,</td>
<td>−</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

3.2 Gas-phase

3.2.1 Mean concentrations

We have calculated the distribution of median pollutant concentrations at all stations in the model domain over each simulation period. Shown in Fig. 3 are the distributions of $O_3$, NO$_2$, NO, SO$_2$, PM$_{10}$ and PM$_{2.5}$ for different station classes. They are presented by boxplots of the distribution of measured and modelled median values during each season (afternoon values of hours 12:00–18:00 local time) and allow to evaluate accuracy and potentially existing biases in our simulations. Table 2 gives a summary of the mean biases found.

$O_3$ is the measure air quality models often have been “tuned” for. COSMO-ART is no different from other models in its ability to represent this quantity very well. A small but consistent underestimation is visible, but seasonal differences are well captured. In winter 2006 largest (negative) biases are observed, while autumn 2008 matches measurements best (Table 2). Overall biases in the median never exceed 10 ppbv and are often below 5 ppbv. Variability within the distributions is comparable with observations. Overall, a correlation of 0.7 ($r$) with hourly station values shows that the performance of our $O_3$ simulations are in the same range as results from simulations with comparable modeling systems like WRF/Chem in Grell et al. (2005).

The $O_3$ precursors NO and NO$_2$ measured within the AIRBASE network show a much larger variability than $O_3$ itself. The differences between rural and rural/remote stations in concentrations of NO and NO$_2$ are well reproduced by the model. Spring 2009, summer 2006 and autumn 2008 concentrations are in a similar range, while values more than twice the median of the other seasons were measured during the high pollution episode of winter 2006. The model reproduces this finding very well. NO$_2$ concentrations vary strongly between station types and season, which the model also represents. However, a comparably strong underestimation is found in summer 2006. Steinbacher et al. (2007) and Dunlea et al. (2007) showed that the often used molybdenum converter based NO$_2$ measurements are biased high due to the additional conversion of other oxidized nitrogen compounds. This will influence the comparison especially during this period, which is characterized by the high oxidative capacity of the atmosphere due to warm, sunny conditions.
Fig. 3. Boxplots of modelled (red) and measured (blue) median concentrations for afternoon hours (12:00–18:00 local time) of several compounds, classified after Henne et al. (2010). The number of stations used is shown at the top of each comparison. Note that values for NO and NO$_2$ are on a logarithmic scale. Simulations are ordered to represent an annual cycle: spring 2009, summer 2006, autumn 2008 and winter 2006.

“Rural/remote” stations show an overestimation throughout all simulation periods (Table 2), a first indication that shipping emissions might be overestimated.

SO$_2$ levels are generally overestimated, again especially at coastal stations. Only during the summer 2006 period, “rural” stations compare well to modelled results. The increase in SO$_2$ concentrations during the polluted winter 2006 episode is reproduced, though exaggerated. We argue that a missing parameterization in COSMO-ART for wet scavenging of gases and the associated aq.-phase oxidation of SO$_2$ to particulate SO$_4^{2-}$ can explain a large part of this SO$_2$ overestimation. A possible overestimation of SO$_2$ emissions in the TNO/MACC inventory can also contribute to the observed mismatch. Uncertainties in emission inventories for SO$_2$ have been shown to be generally large (de Meij et al.,

Fig. 4. Maps of mean afternoon (12–18) NO$_x$, O$_3$ and SO$_2$ concentrations for the summer 2006 period. Comparison with AIRBASE station concentrations is shown as points: Modelled mean in the outer ring and measured mean in the center.
Fig. 5. Modelled and measured (AIRBASE) mean diurnal cycles of several compounds during the summer 2006 period, disaggregated after Henne et al. (2010). Measured values are shown as colored areas, modelled parameters as lines. Light blue areas show 90 % range of station values, the range between dotted lines the same for modelled values. Blue colored areas compare to the area between dashed lines (70 % range). The median of measured values is a solid, dark blue line, the median of modelled values a solid red line. If less than 5 stations report data, measured and modelled medians are reported as blue and red line pair per station, with different line styles for each pair.

Very few measurements were available for NH$_3$ (3 stations in the Netherlands). At those points, NH$_3$ levels are on average well represented, but show large variability throughout the simulation period (not shown).

NMVOCs, the components missing to assess the tropospheric chemistry as a whole, could not be thoroughly evaluated due to a lack of long-term, European-wide measurements. A preliminary comparison with total NMVOC measured at Duebendorf (CH) showed good agreement (not shown), which gave confidence that our NMVOC levels are in the correct range, but we could not assess the spatial distribution.

3.2.2 Spatial distribution

Maps of mean afternoon (hours 12:00–18:00 UTC) concentrations over the whole simulation period were produced, overlaid with point indicators of the same mean concentrations at each measurement station (Fig. 4 for summer 2006 and in the Supplement for the other periods).

The spatial distribution of O$_3$ and NO$_x$ concentrations corresponds with observed values. Only minor differences are found, as for example a large inter-station variability of measured O$_3$ in Eastern Europe which is not seen in the model, and an underestimation of O$_3$ concentrations over the Iberian Peninsula during the spring 2009 period. NO$_x$ values show no region with exceptional biases over land. Striking, however, are the high modelled values of NO$_x$, but also of SO$_2$, over water, along shipping routes in the Mediterranean Sea and the English Channel. The general overestimation of SO$_2$ concentrations found in evaluation of the mean quantities is clearly visible throughout Europe for the autumn 2008 period, but less so in the other periods. Modelled SO$_2$ concentrations at coastal stations in NE Spain are consistently too high, again pointing towards high shipping emission contributions. Apart from that no distinct spatial pattern of overestimation could be found.

3.2.3 Diurnal cycles

The representation of the diurnal cycle of atmospheric constituents was evaluated by means of ensemble plots. The ensemble consisted of all stations which had measurement data for the compound of interest, disaggregated by the classification of Henne et al. (2010). The distribution of concentrations was then calculated for each hour of day, over the whole simulation period. The median and the range covering 70 and 90 % of all stations are shown in Fig. 5 (see Supplement for plots of the other periods).
The simulated daily cycle of \( O_3 \) is accurate throughout most seasons and station types. The slight underestimation of mean \( O_3 \) concentrations found is visible as a shift of the diurnal cycle to lower values. Only in the autumn 2008 period the modelled diurnal amplitude is noticeably smaller than the measured one.

Simulated \( NO_2 \) diurnal cycles also correspond well with observations in most cases. Important aspects like the peaks during morning and evening hours (“rush-hour”) visible in the spring 2009 and autumn 2008 periods are reproduced. \( NO_2 \) levels during nighttime are overestimated in spring 2009 for rural stations, and in autumn 2008 for rural and rural/remote stations. This overestimation at night could be a consequence of the fact that in reality the station is away from emission sources of \( NO_2 \), though in the model \( NO_2 \) is emitted directly into the grid box the station is located in. In spring 2009 (rural stations) and summer 2006 (rural and rural/remote), an exaggerated diurnal amplitude leads to underestimations of \( NO_2 \) concentrations during daytime. Here again, the positive measurement bias will have an influence on our comparison with high levels of oxidized nitrogen compounds such as peroxyacetyl-nitrates (PAN) and HNO\textsubscript{3} in the afternoon, leading to positive biases in the measured \( NO_2 \) concentrations (Steinbacher et al., 2007; Dunlea et al., 2007). Simulated inter-station-type variability is comparable with measurements.

Nitric oxide compares well to observations during daytime, but is underestimated at night. The relatively high measured concentrations at nighttime could be an indication for local sources affecting the measurement sites since \( NO_x \) is mostly emitted in the form of NO and then rapidly converted to \( NO_2 \) by reaction with ozone. This interpretation is supported by the comparatively high \( NO:NO_2 \) ratios of the measurements. The model, conversely, shows very low NO values as expected for truly remote sites (Carroll et al., 1992; Brown et al., 2004). Overall the diurnal cycle with low values during nighttime, a distinct peak during morning hours and a slow reduction towards evening is captured accurately in all simulated periods.

Only 3 measurement points were available to investigate the simulation quality of \( NH_3 \), and all were located in the (highly \( NH_3 \) loaded) Netherlands, making this comparison relatively uncertain. While \( NH_3 \) mean concentrations were comparable to measurements, the diurnal cycles were not (both not shown). The measured cycles were very variable throughout seasons and stations, and we see a clear deficiency of the modeling system to account for this variability. The main sources of \( NH_3 \) emissions are agricultural activities, especially livestock and manure. \( NH_3 \) concentrations are mostly dominated by local emissions. It is known that the diurnal cycle of \( NH_3 \) emissions strongly depends on the emission source (Reidy et al., 2009). Ellis et al. (2011) showed that bi-directional fluxes between the atmosphere and land surfaces might be needed to accurately simulate \( NH_3 \) (and associated aerosol) levels. All this makes modeling such emissions a major challenge which is currently not accurately addressed in most models (Zhang et al., 2008), as emission inventories based on spatially distributed emission totals and associated, statistically averaged time functions cannot capture such process-based emissions.

### 3.2.4 Satellite observations

For comparison with OMI satellite information, vertical tropospheric columns (VTCs) of \( NO_2 \) were calculated from model output for the hour of the satellite overpass (13:30 local time, approx. 12:30 UTC over Europe). The height of the troposphere was assumed to be fixed over all simulations at 10 km geometric height, the exact choice has little influence on the \( NO_2 \) columns. The comparison was made only where OMI data were available at each overpass.

![Fig. 6. Comparison of modelled \( NO_2 \) VTCs against OMI satellite data for (from top) the spring 2009, summer 2006, autumn 2008 and winter 2006 periods. Modelled values are found in the left column, OMI observations in the right one. Not the whole simulation domain could be compared due to missing coverage of the EOMINO dataset.](image-url)
and the conditions were nearly cloud-free (cloud radiance fraction reported by OMI retrieval <50%, corresponding to approx. <20% cloud coverage). The arithmetic mean over each simulation period was calculated and the results are shown in Fig. 6. The aggregated mean biases for all grid points, land points and sea points can be found in Table 2. We compared the model simulated NO$_2$ columns directly with the respective EOMINO columns without taking into account the averaging kernels which would remove the dependency of the result on the a priori NO$_2$ profiles used in the EOMINO retrieval. Not accounting for the averaging kernels might introduce biases of the order of 30% with EOMINO columns tending to be too high over remote locations and too low over polluted areas (Russell et al., 2011), while differences averaged over Europe are likely to be small (Huijnen et al., 2010).

Spatial distribution and magnitude of NO$_2$ is in good agreement with our modeling results. Highly polluted regions over the Netherlands and southern United Kingdom, as well as the Po Valley (Italy) are accurately captured. Plumes of large urban agglomerations (Paris, Madrid, Berlin, Warszaw) are comparable in extent and magnitude. Also, cleaner regions like for example southern France are reproduced. Notable differences are mostly found in polluted coastal areas, especially in the Mediterranean Sea, where the model tends to overestimate NO$_2$ concentrations over water, particularly in the autumn 2008 and spring 2009 period. This overestimation is also visible in the mean over all grid points over sea in Table 2. Emission estimates for ship traffic are known to have large error margins both in magnitude (Corbett and Koehler, 2003) and spatial allocation (Wang et al., 2008). From the magnitude of the error and the spatial correlation with main shipping routes an overestimation of ship emissions by the inventory used is likely. This would also explain the consistent overestimation of SO$_2$ concentrations at coastal stations in NE Spain. Seasonal differences are captured for spring, summer and autumn, only the model results for the winter 2006 period overestimate NO$_2$ columns notably in Northern and Eastern Europe.

3.3 Aerosol characteristics

All comparisons of measured and modelled particulate matter were made in an as rigorous as possible manner. For PM$_{10}$ and PM$_{2.5}$ bulk mass and NR-PM$_1$ AMS measurements, the modelled log-normal distribution functions were integrated over the respective size ranges, and size cut functions were employed to simulate the size-dependent transmission efficiency that is typically found in the measurement instruments used. See Appendix B for a description of the transmission functions used. For the AMS the modelled quantities were additionally converted to vacuum aerodynamic diameter (DeCarlo et al., 2004). No transmission functions were applied to number size distribution measurements, the modelled values are derived from integration over the exact intervals given: 30 to 50 nm, 50 to 500 nm, 100 to 500 nm and 250 to 500 nm, respectively.

3.3.1 Bulk mass

Continuous bulk aerosol mass measurements are the least available within the measurement dataset, making the ensemble of stations for comparison very small (max. 8 stations). When looking at PM$_{10}$ concentrations (Fig. 3, Table 2), our simulations match observations for rural stations in autumn 2008, and underestimate them in the other periods. Simulated concentrations for rural/remote stations are almost identical to those at rural stations in the model, while in reality large differences are found. In consequence, modelled values are above measurements in spring 2009 and autumn 2008, but below in summer and winter 2006. Rural/coastal station concentrations are underestimated in spring 2009 and summer 2006, match observations in autumn 2008 and are above measurements in winter 2006. All this makes autumn 2008 the period in which PM$_{10}$ is simulated best, and worst in summer 2006 (Table 2).

“Rural”-type stations are deemed the most representative for such a model evaluation, and they show (except in autumn 2008) an underestimation typical for many regional models (see e.g. Stern et al., 2008), probably due to missing sources (e.g. resuspension, secondary organics, local mineral dust sources, missing aq.-phase conversion of SO$_2$ to SO$_4^{2-}$). Stations of type “rural/coastal”, in contrast, have a tendency towards more positive biases, which is reasoned by the high amounts of seasalt aerosols found at these stations in the modeling results. The overestimation could also be an artefact of the limited model resolution: coastal stations may be located in grid cells partly covered by sea where sea salt aerosols are therefore emitted directly. Further investigations, e.g. comparisons with filter samples, are needed to assess if the amount of seasalt from the parameterization in COSMO-ART is realistic. The very high PM$_{10}$ concentrations in winter 2006 are not accurately represented in the model. There is in fact no visible increase in PM$_{10}$ concentrations in the model results compared to the other seasons at all.

The diurnal cycles for PM$_{10}$ show that simulated concentrations are often in the same order as the measured values, both in variability and evolution in time, although overall the simulated values are mostly too low. Winter 2006, the period with very high PM levels, has no observable diurnal cycle. In spring 2009 and summer 2006, the diurnal cycles at rural stations show a PM$_{10}$ maximum during night and a minimum at noon, which is – although shifted to lower values – reproduced by the model. The diurnal cycle for rural stations in autumn 2008 is characterized by high but constant PM$_{10}$ levels during nighttime and a drop in concentrations during the day. The model reproduces this finding to a certain degree, although the amplitude of the drop is underestimated.
Only 7 stations, from 3 different categories, had measurements for PM$_{2.5}$ for our simulation periods. From this uncertain data basis we see equally large disagreements as have been found for PM$_{10}$. McKeen et al. (2007) compared PM$_{2.5}$ measurements with several air quality forecast models in North America and concluded that, while most of the models are able to accurately represent daily average PM$_{2.5}$ concentrations, there are substantial inconsistencies in representing the diurnal cycle. Most models show a negative bias and exaggerate the diurnal variability, something we can observe also for the (single) rural PM$_{2.5}$ station in our comparison (Fig. 5).

The errors are in a similar range as found in other model simulations. Vautard et al. (2007) showed similar performance problems in simulating PM$_{10}$ in Europe. Stern et al. (2008) saw better agreement with measurements (i.e. less underestimation) for PM$_{2.5}$ simulations than for PM$_{10}$, which we could not confirm with the dataset mentioned above.

### 3.3.2 Aerosol optical depth

For comparison with MODIS AOD data, a similar procedure was employed as for OMI NO$_2$ vertical tropospheric columns, only using grid points for which satellite data were available and which were cloud-free also in the model. The whole vertical column in the model was used in the calculation of aerosol optical depth with the method described in Vogel et al. (2009). All aerosol categories (internally and externally mixed Aitken and accumulation modes, soot, mineral dust and sea salt modes) contribute to calculated AOD. Then, the median was calculated over the whole simulation period. We chose the median instead of the mean to be more robust against outliers. Figure 7 presents the results. Furthermore, as for the comparison with OMI NO$_2$ VTCs, aggregated biases have been calculated and can be found in Table 2.

For all AERONET stations in the model domain, timelines of AOD at 550 nm were calculated from model output and compared against measured values. AERONET data were interpolated (if no direct measurement at 550 nm was available) linearly in log-log space. In case MODIS data were available also this information was added to the plots. The results for selected stations are shown in Fig. 8, plots for the remaining stations can be found in the Supplement, and Table 2 shows the mean biases for these comparisons.

The comparison against these two independent sets of AOD measurements leaves a mixed picture: compared with MODIS, the model shows consistently lower values than derived from the satellite. We can capture regions with continuously high AOD values like the Po valley (northern Italy) or Saharan dust events like e.g. in the summer 2006 period over the western Mediterranean Sea. The magnitude of the dust event is underestimated, which might be explained by the fact that modelled “dust” is only created within the region of the model domain which covers only a small part of the Sahara.

At the boundaries only very low dust concentrations are prescribed due to the way aerosol boundaries are treated (see Sect. 2.1). Contribution of sea salt to AOD is visible over the Atlantic ocean, but the absolute values are much lower than MODIS derived values, except for winter 2006. Some very polluted regions in south-eastern Europe are captured in location and magnitude (e.g. in Northern Croatia/Southern Hungary), while several other “hot-spots” visible from the satellite (e.g. Eastern UK coast) are missed.
Comparison with AERONET station data reveals additional details. Although the absolute levels are often too low, which is consistent with our comparison with MODIS data, the temporal evolution is often well represented and most high AOD events visible in station data are also observed in our simulations. Differences between MODIS and AERONET derived AOD on the other hand are at several occasions as big as the differences between model and AERONET, and non-negligible on average (up to 10% compared to up to 60% difference between model and measurements, see Table 2). We suggest that the water in the aerosol (both simulated and in reality) will play a major role in the differences found. Both, MODIS and AERONET data, are “cloud-screened”, i.e. data points contaminated by clouds were removed, as they give erroneously high AOD values. Capturing the onset of a cloud is difficult, so some increase in AOD due to aerosol water might be left in the dataset. These effects are visible within the satellite data shown in Fig. 7 (e.g. over Germany in autumn 2008 or west of Ireland in spring 2009) near regions with missing (cloud-screened) pixels. Also in several AERONET stations the sudden steep increase of AOD just before measurements are filtered (for clouds) can be found. While we tried to remove this error by using median values instead of the arithmetic mean to calculate the MODIS-model comparison, we probably could not exclude all of those situations. As the effect is non-linear and acts towards very high AOD values, this will probably bias AOD results. Secondly, differences in simulated and real aerosol chemical composition will also have an effect on AOD. The next section addresses a comparison of aerosol chemical composition.

A clear negative bias in absolute AOD is seen in our model when compared with two independent measurement datasets which appears to be consistent with the too low simulated PM$_{10}$ and PM$_{2.5}$ levels. Fair correlation of the evolution
in time is visible from the AERONET comparison. Performance of our AOD simulations is well in range of results for comparable modeling systems (e.g. Zhang et al., 2010; Aan de Brugh et al., 2011). We argue that both missing aerosol mass at the lateral boundaries and inaccuracies of simulated aerosols within the domain contribute to the underestimated AOD. Especially for aerosol components from natural sources (Saharan dust) the missing lateral contribution could be substantial. Although we tried to remedy this by using averaged profiles from a previous run, we could not – especially for those categories – represent the absolute mass contributions correctly. The impact of the missing pathway to form sulfate in clouds and the known too small yield of SOA in the SORGAM model are additional sources of error that impact the overall accuracy of the comparison.

### 3.3.3 Chemical composition

Aerosol chemical composition was evaluated by comparison with AMS data. In summer 2006, AMS measurements were available at Payerne (CH) and Bush (UK) (Lanz et al., 2010). Several AMS instruments were deployed during the 2008 (autumn) and 2009 (spring) periods at stations throughout Europe. Timelines of the composition of NR−PM$_1$ are presented for both measurement and simulation at these stations. Shown in Figs. 9, 10a, and 10b are the timelines for the autumn 2008 and spring 2009 periods. The comparison for summer 2006 (3 stations) can be found in the Supplement. In the figures, colors typically used in the AMS community are used to represent each species: ammonium (NH$_4^+$) in orange, sulfate (SO$_4^{2-}$) in red, and nitrate (NO$_3^-$) in blue. Organic aerosols (OA) are represented as shades of green. Charges are omitted intentionally for the AMS in the figure legends, as also contributions from organosulfates, organonitrates are included which are not ions (Farmer et al., 2010). In case of modelled values, a distinction can be made between anthropogenic primary organics (aPOA), secondary organics from anthropogenic (aSOA) and biogenic (bSOA) sources. Table 2 presents the mean biases for each species over all stations in each season.

At all stations the time evolution of NR−PM$_1$ is represented well by our simulations, sometimes however for the wrong reasons due to a mismatch in chemical composition. Single events with higher aerosol concentrations (e.g. in Vavilhill, 2008, Fig. 9) correspond in time and magnitude with the observations in most cases. Several model deficiencies can also be seen throughout the comparison, namely an overestimation of nitrate components and an underestimation of sulfate and, sometimes, organic mass. In the following we will briefly discuss the result for each station.

In Switzerland, measurements at Payerne were available for three periods. The time evolution of total aerosol mass corresponds best in spring 2009, and worst in the summer 2006 period. The weak correlation in summer 2006 is mostly due to a severe underestimation of OA, especially during daytime hours, and an overestimation of nitrate during nighttime. In spring 2009, several abrupt changes in aerosol mass concentrations were observed. Although the timing is not the same each time, the model reproduces those changes. A tendency to retain too much nitrate in the aerosol phase during daytime is apparent. Sulfate is underestimated. During autumn 2008, an episode of high aerosol concentrations is observed in the middle of the observation period. This

Fig. 10a. Timeline of aerosol chemical composition as in Fig. 9 but for stations measuring during the spring 2009 period (1 of 2).
is also reported by the model. OA are, however, underestimated, and nitrate aerosols overestimated. Here also, modelled aerosol nitrate shows a persistence to remain in the aerosol phase during daytime that is not found in the observed values. Melpitz in Germany differs from Payerne in a generally higher sulfate content. Otherwise those stations report similar aerosol composition. Striking is the stronger overestimation of nitrate aerosols at Melpitz, compared to Payerne, in both periods. Simulated sulfate is in the same range as in Payerne, and therefore even more strongly underestimated. The concentrations of organics are lower in Melpitz, and simulated values are comparable here. The third station with more than one period of measurements is Vavihill (SE). Generally low aerosol concentrations alternate with isolated peaks in aerosol mass with high contents of inorganic secondary components. This burst pattern is captured in our simulations, and also the timing fits mostly well. Especially in spring 2009 the model lacks, though, the OA mass necessary to fit the measurements. While ammonia levels are comparable in autumn 2008, they are above measured levels in spring 2009. The AMS deployed at Hytiälä reports very low NR−PM$_1$ concentrations, with large contributions by sulfate and OA, and, in spring 2009, almost no nitrate. The model can represent the overall level of aerosol concentration. However, the simulations significantly underestimate sulfate and overestimate nitrate.

Fig. 10b. Timeline of aerosol chemical composition as done in Fig. 9 but for stations measuring during the spring 2009 period (2 of 2).

Fig. 11. Modelled and measured aerosol size distributions at EUSAAR stations for the autumn 2008 period. Modelled distributions are shown as area shaded by mass contributions of different species, calculated as median over the simulation period. Measured values are a statistical size distribution from data of the whole season and are shown as lines. Solid lines depict median values, dashed lines the 67 % and dotted lines the 90 % percentile range. Grey background denotes areas with higher measurement uncertainty (Wiedensohler et al., 2010).
Table 3. Comparison of number concentrations in different size ranges after Asmi et al. (2011) for number size distributions during the autumn 2008 simulation. \(N_{30\text{to}50}\): 30 to 50 nm, \(N_{50}\): above 50 nm, \(N_{100}\): above 100 nm, \(N_{250}\): above 250 nm. Note that the \(N_{250}\) parameter has a larger uncertainty than the others due to very low sampling rates.

<table>
<thead>
<tr>
<th>station name</th>
<th>category</th>
<th>(N_{30\text{to}50})</th>
<th>(N_{50})</th>
<th>(N_{100})</th>
<th>(N_{250})</th>
</tr>
</thead>
<tbody>
<tr>
<td>Birkenes (NO)</td>
<td>rural/remote</td>
<td>133 374</td>
<td>250 517</td>
<td>96 171</td>
<td>6 42</td>
</tr>
<tr>
<td>Cabauw (NL)</td>
<td>suburban</td>
<td>2787 2605</td>
<td>4649 4048</td>
<td>1170 1706</td>
<td>63 250</td>
</tr>
<tr>
<td>Harwell (UK)</td>
<td>rural</td>
<td>705 2181</td>
<td>1097 2247</td>
<td>410 713</td>
<td>51 92</td>
</tr>
<tr>
<td>Ispra (IT)</td>
<td>suburban</td>
<td>1732 4915</td>
<td>7172 6293</td>
<td>3829 2632</td>
<td>373 423</td>
</tr>
<tr>
<td>K-Puszta (HU)</td>
<td>rural</td>
<td>597 2698</td>
<td>4626 4247</td>
<td>2609 2013</td>
<td>390 313</td>
</tr>
<tr>
<td>Mace Head (IE)</td>
<td>rural/remote</td>
<td>48 29</td>
<td>158 124</td>
<td>104 95</td>
<td>41 59</td>
</tr>
<tr>
<td>Melpitz (DE)</td>
<td>rural</td>
<td>1031 1377</td>
<td>2386 3447</td>
<td>1306 1650</td>
<td>312 305</td>
</tr>
<tr>
<td>Kosetice (CZ)</td>
<td>rural/remote</td>
<td>508 866</td>
<td>2346 2724</td>
<td>1423 1340</td>
<td>262 265</td>
</tr>
<tr>
<td>Vavilhill (SE)</td>
<td>rural</td>
<td>577 1314</td>
<td>1187 2761</td>
<td>365 1010</td>
<td>51 146</td>
</tr>
</tbody>
</table>

Table 4. Comparison of number concentration comparisons, like in Table 3, but for the spring 2009 simulation.

<table>
<thead>
<tr>
<th>station name</th>
<th>category</th>
<th>(N_{30\text{to}50})</th>
<th>(N_{50})</th>
<th>(N_{100})</th>
<th>(N_{250})</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aspvreten (SE)</td>
<td>rural/coastal</td>
<td>201 1084</td>
<td>645 1932</td>
<td>302 808</td>
<td>63 152</td>
</tr>
<tr>
<td>Cabauw (NL)</td>
<td>suburban</td>
<td>1466 2780</td>
<td>1894 3399</td>
<td>433 1248</td>
<td>20 187</td>
</tr>
<tr>
<td>Harwell (UK)</td>
<td>rural</td>
<td>746 2517</td>
<td>1453 2285</td>
<td>634 762</td>
<td>103 125</td>
</tr>
<tr>
<td>Ispra (IT)</td>
<td>suburban</td>
<td>904 2341</td>
<td>2921 2136</td>
<td>1451 776</td>
<td>157 116</td>
</tr>
<tr>
<td>K-Puszta (HU)</td>
<td>rural</td>
<td>855 2133</td>
<td>3104 3890</td>
<td>1673 1670</td>
<td>203 267</td>
</tr>
<tr>
<td>Mace Head (IE)</td>
<td>rural/remote</td>
<td>324 542</td>
<td>779 1199</td>
<td>418 569</td>
<td>108 127</td>
</tr>
<tr>
<td>Melpitz (DE)</td>
<td>rural</td>
<td>508 1380</td>
<td>1343 2868</td>
<td>762 1279</td>
<td>219 242</td>
</tr>
<tr>
<td>Kosetice (CZ)</td>
<td>rural/remote</td>
<td>467 1342</td>
<td>2032 3748</td>
<td>1282 1751</td>
<td>210 319</td>
</tr>
<tr>
<td>Vavilhill (SE)</td>
<td>rural</td>
<td>402 1272</td>
<td>1496 2850</td>
<td>607 1266</td>
<td>171 236</td>
</tr>
<tr>
<td>Waldhof (DE)</td>
<td>rural/remote</td>
<td>652 1209</td>
<td>1744 2678</td>
<td>935 1241</td>
<td>227 248</td>
</tr>
</tbody>
</table>

All other stations only report data for one period. In autumn 2008, measurements of aerosol chemical composition were also available for K-Puszta, Hungary. The station reported high aerosol concentrations with levels up to \(30 \mu g \text{ m}^{-3}\) total mass. While the model represents the buildup of aerosols towards the middle of the observation period, the overall mass is underestimated. Too high nitrate levels are simulated. Organics and ammonium match observations better, but sulfate tends to be underestimated also at this location. Four more stations reported data during spring 2009: Cabauw (NL), Helsinki (FI), Barcelona (ES) and Montseny (ES). Cabauw (NL) has lower concentrations than e.g. Payerne or Melpitz, and a big gap in measurements during the first half of our simulations. There is some resemblance in the peaks of aerosol mass during the second half of the simulation between model and station values. Nitrates are overestimated while ammonium and sulfate are too low. Organics are well captured. Helsinki (FI), an urban background station is, like Hyytiälä (FI) characterized by a strong contribution from sulfate. The simulated total aerosol loadings are comparable to the observed concentrations but do not match in composition. We can tentatively explain this difference by looking beyond the border of the model domain: both stations are in the vicinity of large sources of \(\text{SO}_2\) on the Kola peninsula in Russia (Tuovinen et al., 1993) which are still found to be underestimated in current emission inventories (Prank et al., 2010). Additionally, due to the setup of aerosol boundary conditions in our modeling system, we very likely underestimate direct sulfate inflow in this region. In Barcelona (ES), also an urban background location, a very variable time series is reported, with the highest absolute concentrations of all stations used in this analysis. Several peaks of aerosol concentration each day are common, containing relatively high sulfate levels compared to other stations. The model produces a similar variability, although it overestimates nitrate. Sulfate levels are comparable at this site with a large influence from shipping. OA concentrations are, in contrast to most other stations, overestimated.
at Barcelona and Helsinki. The largest contributor to simulated total organic mass at these stations is primary emitted organics. Statistical analysis of the organic fraction (positive matrix factorization (PMF), Paatero and Tapper, 1994) indicates that organics in urban stations are comprised of similar amounts of SOA and POA, while in the model it is almost exclusively POA. This points towards a strong underestimation of secondary organics in polluted regions as it has been found already by Fast et al. (2009). Finally, the AMS in Montseny (ES) measured a time-series with several periods with increased aerosol loadings, and during the first third a period where almost no aerosols were found due to an episode of strong Atlantic advection. The model captures this period well. Total organics are comparable throughout the simulation period, although a PMF analysis gives about 5% mass contribution from urban primary organics (Minguillon et al., 2011), instead of about 30% as given in the model. Nitrates are too high, and lacking the diurnal cycle visible in the measurement. Simulated sulfate is below measurements.

3.3.4 Number concentrations and size distributions

The dataset compiled by Asmi et al. (2011) provides a homogenized overview of the statistical characteristics of aerosol size distributions in Europe during the years 2008 and 2009. We evaluate different particle dry size separated subsets of the number concentrations, following Asmi et al. (2011). The number of particles from 50 ($N_{50}$) and 100 ($N_{100}$) nm up to 500 nm have been chosen as proxies to study climate effects. Health concerns are related to very small particles, which are assessed by comparing number concentrations of particles between 30 and 50 nm ($N_{30}$). This concentration can also serve as an indicator of new particle formation and emissions from combustion processes. Finally, the number of particles with diameters between 250 and 500 nm ($N_{250}$) are given to show the contribution of larger particles to total aerosol number concentrations. We have calculated the corresponding model values by integrating the aerosol modes over the respective intervals. Data were available in up to hourly resolution, so a direct comparison could be made between modelled and simulated values. Table 3 shows the resulting comparison for the autumn 2008 period, Table 4 for spring 2009. Table 2 gives a summary overview of the mean biases over all stations.

We also studied the histograms (occurrence distribution) of logarithms of the number concentrations in the particle size ranges (not shown). The analysis was done in logarithmic concentration space as most of the aerosol number concentrations are log-normally distributed (Asmi et al., 2011). It shows the model’s ability to produce similar distributions of number concentrations as measured and provides a more detailed way to analyze the differences. We also performed a Mann-Whitney U-test (Higgins, 2004) on the modelled and measured concentration distributions to see with what p-value they could be considered to be from the same distribution with similar mean and distribution shape.

The histograms of number concentrations show that the agreement is better in greater diameter size ranges ($N_{100}$ and $N_{250}$) in comparison to concentrations in $N_{30}$ size range. The model seems to overestimate the number concentrations in the smaller size ranges by a factor of two to five, especially in Harwell (UK), Ispra (IT) and the two Swedish stations (Aspvreten and Vavihill). This overestimation could be explained by a relatively low fraction of new particle formation in the modelled environment. COSMO-ART uses the nucleation parametrization from Kerminen and Wexler (1994), which does not generally produce the observed amounts of nucleated particles in the European boundary layer. Thus the overestimation could be due to a disproportional amount of emitted sulphur to be considered as primary Aitken particles, which have a much higher lifetime in the atmosphere compared to newly nucleated particles in these regions. For the larger particle sizes ($N_{100}$ and $N_{250}$), the model-measurement comparison is more successful. At Central European stations the modelled and measured concentration distributions are generally of similar shape and median, which is well demonstrated by p-values ranging from 0.31 to 0.66 in the U-test test parameter for Kosetice and Melpitz. The overall shapes of the concentration histograms are generally similar in all the stations, although some discrepancies in lower-concentration regions are visible. The agreement is generally poorer in lower-concentration regions of Northern Europe, but also in Cabauw (NL) and Harwell (UK) $N_{250}$ concentrations, where the model overestimated the concentrations by a factor of 2 in 2008.

A second dataset available from Asmi et al. (2011) is seasonal statistics of aerosol number size distributions. We have calculated a distribution function as mean over all modelled values in each simulation period, and compared it against the measured distribution statistics of the corresponding season (Fig. 11 for autumn 2008, plots for spring 2009 can be found in the Supplement). Note that there is no exact match between the time periods covered by the measurements and the simulations (3 weeks out of the 3 months). Overall, the modelled size distributions are at several stations close to the observed ones. At most stations, simulated size distributions were within the central 67% percentiles of the values reported by Asmi et al. (2011) when comparing the 20 to 200 nm size range, for which the instruments were reported to compare the best (Wiedensohler et al., 2010). Concerning the shape of the size distributions, stations with the best match between model and measurements were Melpitz (DE), Waldhof (DE) and Kosetice (CZ), with only very small deviations in both years. Aerosol number size distributions at the rather polluted sites Ispra (IT) and K-Pustza (HU) show distribution functions with comparable peak values but opposite skewness. While model values lean towards smaller diameters, measurements have their peak in number concentration at much larger aerosol diameters. For Ispra (IT) this is probably
due to the influence of the Milan urban agglomeration. Due to the coarse horizontal resolution, fresh emissions (with smaller diameter) contribute much more to aerosol composition at Ispra in the model than in reality, where the aerosol had more time to age. This ageing would shift the size distribution towards larger diameters via coagulation as observed in the measured distributions. A similar explanation might hold for K-Puszta (HU), which is located near Budapest, the capital of Hungary. Cloud processing of aerosols is missing in COSMO-ART and might be responsible in general for a bias towards small peak diameters. Cabauw (NL) and Vavilhill (SE) show comparable shape but model and measurements disagree in number concentration. Both, Birkenes (NO) and Harwell (UK) show a tendency towards a bimodal size distribution, which is captured by the model in 2008, but missed in the 2009 case. Finally, Mace Head (IE), with its large variability in number concentrations reasoned by the stations setting at the coast in western Ireland, representing mostly clean maritime air masses, occasionally interrupted by continental influences, shows acceptable agreement in terms of total number concentrations, but no clear agreement in size distribution.

In general the model has an acceptable representation of the variability of number concentrations between stations (Table 3). In most cases the model overestimates number concentrations throughout the size range covered. The station with the best agreement was Aspvreten (SE) in autumn 2008 and Birkenes (NO) in spring 2009, while Ispra (IT) compared worst in 2008, K-Puszta in 2009. Several stations showed acceptable agreement for number concentrations, for example Melpitz (DE) and Birkenes (NO) in 2008 or Waldhof (DE) and Vavilhill (SE) (except for the N_{30,050} range) in 2009. The agreement found was generally better during the autumn 2008 than during the spring 2009 period.

4 Discussion of aerosol characteristics

4.1 Sulfate

This aerosol species is virtually always underestimated. Several factors contribute to this error: Besides some minor direct emissions of sulfate particles, most of the aerosol sulfate is secondary, created from oxidation of SO_2 in the gas-phase and within the aqueous-phase in cloud droplets. Studies have shown that the amount of sulfate produced in clouds is substantial and even dominating (Walcek and Taylor, 1986; Rasch et al., 2000). COSMO-ART currently lacks a parameterization for this pathway. Therefore, especially during periods with cloudy conditions, the underestimation of SO_4^{2-} is likely explained by this missing process. The missing conversion of SO_2 to SO_4^{2-} is also consistent with too high levels of SO_2 in our model. Aksoyoglu et al. (2011) simulated the summer 2006 period with another modeling system including in-cloud oxidation of SO_2 and found better agreement. Whether this can be attributed to in-cloud sulfur oxidation is unclear. In addition to the oxidation issue it was shown that the regional (e.g. Wagstrom and Pandis, 2011) and even intercontinental (e.g. Liu and Mauzerall, 2007) contributions to sulfate aerosol mass are higher than for other aerosol categories like nitrate. Inflow of aerosol concentrations at the lateral boundaries is realized by a smooth transition to values from a given profile or a coarser grid model, this is called relaxation. While we do relax our model at the lateral boundaries against data from a global chemistry transport model (CTM) for gas-phase species, we could not provide similar boundary conditions for aerosol species. Instead we relax against a mean profile from a previous run (which is also low in sulfate). Therefore, only very little long-range transport of sulfate is simulated (approx. 0.2 to 0.4 µg m^{-3} surface concentration), contributing to this underestimation. A sensitivity study with strongly increased lateral sulfate showed a noticeable but insufficient increase of sulfate at the grid boxes of the AMS measurement stations. Finally, oceanic emissions of dimethyl sulfate (DMS) have also been shown to contribute to aerosol SO_4^{2-} levels (Gondwe et al., 2003). A parameterization has recently been included (Lundgren, 2010) in COSMO-ART but was not yet used in our studies. Sensitivity studies showed, though, that sulfate originating from maritime DMS emissions has no substantial influence over continental regions, which again indicates the importance of cloud processing of SO_2. Oxidation of sulfates in clouds will be included via a comprehensive wet scavenging and aqueous-phase chemistry scheme, currently under development at Empa.

4.2 Organics

Often also organic aerosol contributions are underestimated. This is a well-known problem of current CTMs (Volkamer et al., 2006; Hodzic et al., 2009; Hallquist et al., 2009), in our case reasoned by the use of an older parameterization of the conversion of condensable organic vapours to secondary organic aerosols (SOA) (Schell et al., 2001), based on the two-product method by Odum et al. (1996).

Our total OA underestimations are substantial and reach factors of 2. Underestimations for SOA alone by a factor of 10 or more were summarized by Volkamer et al. (2006) and Hodzic et al. (2010) for multiple polluted regions in 3 continents using SOA modules similar to ours. Compared to the current state of knowledge our SOA parameterization has too low yields, and is lacking the description of semi-volatile and intermediate volatility species as implemented in e.g. the volatility basis set approach (Donahue et al., 2006; Murphy et al., 2011). The particular SOA module used in this work (MADE/SORGAM) has been shown to underpredict SOA formation by about a factor of 10 in the Mexico City region (Fast et al., 2009). Thus it is very likely that a strong underprediction of pollution-related SOA is compensated by an overprediction of anthropogenic POA.
(De Gouw and Jimenez, 2009), to result in a lower underprediction of total OA. The comparisons with the AMS deployed in Barcelona/Helsinki further support this hypothesis: these two stations were located within an urban area. There, the model overestimates total organics, and attributes the large majority of the mass to primary organics, while the measurements show a major fraction of secondary organics, as is typical of most urban areas (Zhang et al., 2007; Jimenez et al., 2009). Furthermore, emissions from forest fires were not included in our simulations, although it is known that they can be a major OA contributor (e.g. Aiken et al., 2010). Emissions of biogenic SOA precursors, and the effectiveness of the conversion pathways, are still in discussion and will also contribute to the discrepancies found. Finally, domestic wood burning has been shown to release substantial amounts of OA in wintertime (Alfarra et al., 2007), but also these emissions were not included. Work is currently underway to integrate all these recent developments in SOA and emissions modeling in COSMO-ART.

4.3 Nitrate

The most substantial bias found in our simulations is an overestimation of nitrate aerosol components. This is not a new phenomenon and seen also in other model evaluations (e.g. Stern et al., 2008). Accurately modeling this species is challenging (Dentener and Crutzen, 1994), as it represents the result of a dynamic, coupled system between gas- and aerosol-phase, depending on the amount of gas-phase precursors, temperature, relative humidity and aerosol composition (cf. Chapter 9 in Seinfeld and Pandis, 2006). We have tested several hypotheses to understand this deficiency in our model. We could exclude an erroneous nighttime or daytime chemistry (e.g. providing too much HNO₃) and emission sources (too high levels of NOx). Evaluation against nitrate totals (daily averages of gas-phase HNO₃ + particulate NOx from impregnated filter packs at the station Payene (CH)) showed some high bias, but the overestimation is much smaller than for nitrate alone. Three hypotheses seem likely: the lack of sulfate, missing wet deposition of HNO₃ and inaccuracies in the model’s ability to reproduce relative humidity and temperature well enough.

In experiments, available ammonia is first neutralized by sulfuric acid and only if no more sulfuric acid is available, nitric acid serves as a replacement to form NH₄NO₃ aerosols (Seinfeld and Pandis, 2006). Suppose now that ammonia is overestimated, to result in a lower underestimation of available NH₃, to which sulfuric acid is not removed from the atmosphere as effectively in reality. As HNO₃ is a very hydrophilic substance it will easily transfer to the aqueous-phase and is therefore efficiently scavenged. This could lead to the observed too high levels of total nitrate, which are then subject to gas-aerosol partitioning.

Secondly, due to the missing wet scavenging of gases, HNO₃ is not removed from the atmosphere as effectively as in reality. As HNO₃ is a very hydrophilic substance it will easily transfer to the aqueous phase and is therefore efficiently scavenged. This could lead to the observed too high levels of total nitrate, which are then subject to gas-aerosol partitioning.

Thirdly, the gas-aerosol partitioning for nitrates has a strong temperature dependence. If our model cannot represent daytime temperature maxima and minima to a high degree of accuracy, this will lead to errors in the partitioning. It is also known that the phase-state solid/liquid is a strongly non-linear function of aerosol chemical composition and relative humidity (see Figs. 2, 5 and 7 in Nenes et al., 1998), consequently changing the partitioning behaviour between gas and aerosol phase. As the lifetime of gas-phase HNO₃ is much shorter than for particulate NH₄NO₃, too strong partitioning to the aerosol phase results in too strong survival of total HNO₃ + nitrate in the atmosphere, which contributes to overpredictions of nitrate at later times.

In summary: firstly, if the model underestimates sulfate more ammonia is available to form nitrate aerosols, which leads to an overestimation. Secondly, if not enough HNO₃ is removed more total nitrate is available for partitioning. Thirdly, even small differences in modelled temperature and relative humidity compared to the situation at the instrument could change the nitrate gas/aerosol partitioning. Finally, if NH₃ concentrations are overestimated themselves, even more NH₃ is available to neutralize NO₃⁻, and even more ammonium-nitrate is formed. We think these processes together explain a large part of our overestimation of nitrate aerosols. Other models are better able to simulate mean nitrate concentrations, for example the Comprehensive Air Quality Model with extensions (CAMx)/particulate matter CAMx (PMCAMx) (e.g. Aksoyoglu et al., 2011; Andreani-Aksoyoglu et al., 2007) or WRF/Chem (e.g. Li et al., 2010), but once the diurnal cycle of gas-/aerosol partitioning of nitrate is looked at, also these modeling systems exhibit problems.

We have seen that there are substantial model deficiencies to accurately describe nitrate aerosols and that this is a general problem also for other, comparable model systems. One straightforward way to improve the situation will be to increase the horizontal resolution of the simulations to better represent the variability in temperature and relative humidity. A better knowledge on NH₃ emissions and concentrations is needed. The impact of the implementation of a comprehensive wet phase chemistry scheme will be investigated in a future study.
4.4 Number size distributions and concentrations

The evaluation showed a small but consistent high bias of modelled number concentrations in both periods. As it can be seen from comparison with the number size distributions this is often caused by overestimated particle numbers in smaller diameter regions, pointing towards either a too high number of particles emitted in the Aitken mode, overestimated nucleation rates, or underestimated coagulation. We consider the nucleation scheme of Kerminen and Wexler (1994) used to contribute to these differences. No explicit nucleation mode exists in COSMO-ART, hence secondary generated sulfate particles are transferred directly into the Aitken mode. A fixed factor is applied here to reasonably reduce number concentrations that get lost through coagulation during the growth from freshly nucleated clusters to Aitken-mode-size particles. In case the number of existing particles (e.g. in strongly polluted regions) does not match the assumptions made for this conversion factor, formation rates of Aitken-mode-sized particles through nucleation of SO$_4^{2-}$ are underestimated.

The distribution of emitted particle mass on accumulation and Aitken mode follows a recent publication of Elleman and Covert (2010), which used a similar aerosol module. They did not find high variability in used emission diameters for different categories in North American simulations, and therefore allowed a very simple description of the size distribution of emitted particles (a time and emission source category invariant split based on total emitted mass). However, their study did not consider modern new particle formation parametrizations. As a result they could have considered at least some of the particle precursors as primary emissions. A study by Spracklen et al. (2010) over Europe came to a different conclusion, indicating substantial variability in emitted number size distributions and high importance of adequate representation of new particle formation.

Even though COSMO-ART does not consider some of the more recent findings on new particle formation or nucleation (e.g. Kerminen et al., 2010), the overall ability of the model to reproduce the measured size distributions was adequate. We consider this to be an evidence that using properly derived emission factors, the overall transformation from emissions to CCN sized particles can be somewhat captured using mostly primary-emission based methodologies. However, the overall agreement between modelled and measured values were in general poorer in $N_{300050}$ range in comparison to larger size ranges, especially in more remote areas, suggesting a need for a better mechanism to account for the differences between primary and secondary formation. We did not assess, however, the relative contributions of nucleation (and condensation) versus primary emissions (and condensation) to number concentrations in smaller diameters, so these findings will need further study. Finally, the missing description of cloud processing will influence number size distributions and likely shift the distribution to larger diameters. This could explain part of the overestimation for small particles found.

We conclude for the number size comparison that the approach used is comparable with methods currently taken by other modeling groups, but that there is considerable uncertainty that needs to be better understood for future simulations. The overall acceptable agreement between modelled and measured $N_{100}$ and $N_{250}$ concentrations suggests that the pool of CCN-sized particles simulated in these periods are generally well captured in comparison with measurements.

5 Conclusions

Our goal has been to thoroughly evaluate the online-coupled, regional-scale chemistry-transport-model COSMO-ART for its ability to simulate trace gas concentrations and aerosol characteristics. The evaluation dataset we have collected allows for a comprehensive assessment of model performance at the surface throughout Europe. Comparison with only recently available measurements of aerosol chemical composition (AMS) and aerosol size distribution data was particularly valuable. Not included in our work has been an evaluation of vertical profiles and upper-air variables of chemical and meteorological parameters with aircraft or radiosonde measurements.

Surface meteorological conditions are very well simulated in all periods investigated without any need for tuning. However, there is room left for improvements through data assimilation and nudging. Results for gas-phase tracer and bulk aerosol mass concentrations are encouraging, also for rather difficult periods like winter 2006. Both, temporal and spatial distributions of O$_3$ and NO$_x$ are in good agreement with observations. The lack of a coordinated, european-wide measurement network for NH$_3$ and NMVOC impairs our ability to wholly evaluate gas-phase chemistry, and a missing homogenized elemental carbon dataset hinders evaluation of this aerosol component. From the more advanced datasets, aerosol chemical composition and size distributions, we can conclude that the modeling system is able to represent those quantities with an acceptable degree of accuracy, although nitrate aerosols tend to be overestimated and sulfate underestimated. Not only is the temporal evolution of aerosol mass correctly reproduced, including distinct peaks seen on several occasions and places, but also the chemical composition is quite comparable to reality though some deficiencies have been found. In addition, we could show that the modeling system is able to represent these quantities in an acceptably size-resolved manner – a quantity that is indispensable for correct quantification of climate and health effects. Some deficiencies have been identified in the model system. Most of them will be addressed in the near future by already ongoing developments.
These are:

- wet scavenging for gases and wet-phase chemistry/parameterization of in-cloud oxidation of SO$_2$ to SO$_4^{2-}$;
- update of the representation of secondary organic aerosol components;
- realistic lateral boundary conditions for aerosol species;
- representation of number size distribution and concentrations in primary emissions of aerosol particles;
- inclusion of forest fire emissions.

Some discrepancies found are more likely related to the simulation setup rather than the model system itself. An increase in horizontal resolution will be key to address those issues. Continuous assimilation of meteorological measurement data is another method which will likely result in improvements.

The coupling to a meteorological core that is actively used and developed for both short-term weather forecasting as well as climate simulations is regarded as a key benefit. We conclude that the model is suitable for air-quality assessments and the framework is set to evaluate the accuracy of aerosol-climate interactions. Only after our evaluation results are known, more complex studies of e.g. climate impacts, can be conducted reliably.

Appendix A

INT2COSMO-ART

INT2COSMO (formally known as INT2LM, see Schättler, 2009) is the official preprocessor for COSMO which, among other tasks, interpolates initial and boundary conditions (IC/BC) for meteorology to a given COSMO grid. We have extended this preprocessor by the ability to interpolate and combine several emission and boundary data sets for gas-phase and aerosol species. Additional boundary conditions for parameterized emissions can now also be interpolated. The interpolation procedures are mass-conserving on the total domain.

COSMO-ART thereby gained the ability to be used easily for different domain and grid setups, with different kinds of chemistry IC/BC and emission datasets. Additionally, operational usage is now made feasible. This addition is called INT2COSMO-ART and is available for other users of COSMO-ART.

Appendix B

Transmission functions

A correction for the size cut/ transmission characteristics of the inlet system for the AIRBASE/EMEP bulk mass and the AMS measurements was performed. For measurements within the AIRBASE/EMEP network (PM$_{10}$, PM$_{2.5}$ bulk mass) we could not determine the inlets used at each measurement station separately. Therefore we settled on applying the transmission characteristics of typically used inlets for such instruments. In case of PM$_{2.5}$ we used the transmission characteristics of the United States Environmental Protection Agency (US EPA) PM$_{2.5}$ Well Impactor Ninety-Six (WINS) as described in Peters et al. (2001). We could not find studies on transmission functions for any PM$_{10}$ inlet used. Instead we applied a function that resembles the maximum allowed tolerances for a PM$_{10}$ measurement method that would be accepted as US EPA reference method (Table D-3 in 53.43, Part 53, Title 40, Electronic Code of Federal Regulations, http://www.gpoaccess.gov/ecfr/). For the AMS, additionally to the correction for vacuum aerodynamic diameter, we applied a transmission function which consists of:

- 0 % transmission below 40 nm $d_{va}$;
- linear increase in transmission vs log($d_{va}$), from 0 % at $d_{va} = 40$ nm to 100 % at $d_{va} = 100$ nm;
- 100 % transmission from $d_{va} = 100$ nm up to $d_{va} = 550$ nm;
- linear decrease in transmission vs log($d_{va}$), from 100% at 550 nm to 0 % at $d_{va} = 2 \mu$m.

This transmission function is an average of the transmission curves used in several AMS studies (DeCarlo et al., 2004; Cross et al., 2007; Vaden et al., 2011; Park et al., 2004).

Supplementary material related to this article is available online at: http://www.geosci-model-dev.net/4/1077/2011/gmd-4-1077-2011-supplement.pdf.

Acknowledgements. We acknowledge the EBAS team at NILU, Norway, for providing measurement data through their web interface. Pirmin Kaufmann, MeteoSuisse, made SYNOP measurement data available. The Swiss National Supercomputing Center (CSCS) and the Ipazia-team at Empa provided computing resources for our simulations. The Swiss National Science Foundation is acknowledged for partly financing the IPAZIA computational cluster (project 206021_128754). We thank the AERONET PI investigators and their staff for establishing and maintaining the 36 sites used in this investigation. Louisa Emmons provided MOZART-NCEP output which we have used as initial and boundary conditions in our simulations, and we are also thankful for
advice and help she gave us via email. The work of the AMS measurement group from Manchester has been funded by ACCENT and the UK Natural Environment Research Council (NERC). The Colorado AMS group was supported by NSF ATM-0919189 and NOAA NA08OAR4310565. The AMS measurements at Melpitz were supported by the Umweltbundesamt (UBA) grants no. 351 01 031 and no. 351 01 038, and UFOPLAN contract 3703 43 200.

Edited by: A. Stenke

References


Ellis, R. A., Murphy, J. G., Markovic, M. Z., and Van den Boer, T. C., Elleman, R., and Covert, D.: Aerosol size distribution modeling

docking, D., Cunningham, J., Damokosh, A., Neas, L., Spengler, J., Laepple, T., Orlando, J., Tie, X., Tyndall, G., Wiedinmyer, C.,


Lundgren, K.: Direct Radiative Effects of Sea Salt on the Regional


Murphy, B. N., Donahue, N. M., Fountoukis, C., and Pandis, S. N.: Simulating the oxygen content of ambient organic aerosol with the 2D volatility basis set.


Stanelle, T., Vogel, B., Vogel, H., Bäumer, D., and Kottmeier, W.:


