SMOKE for Europe – adaptation, modification and evaluation of a comprehensive emission model for Europe

J. Bieser¹,², A. Aulinger¹, V. Matthias¹, M. Quante¹,², and P. Builtjes³,⁴

¹GKSS Research Center, Institute for Coastal Research, Max-Planckstr. 1, 21502 Geesacht, Germany
²Leuphana University Lüneburg, Institute of Ecology and Environmental Chemistry, Scharnhorststraße 1, 21335 Lüneburg, Germany
³Institut fur Meteorologie, Freie Universität Berlin, Carl-Heinrich-Becker Weg 6-10, 12165 Berlin, Germany
⁴TNO Built Environment and Geosciences, Air Quality and Climate Team, Princetonlaan 6, Utrecht, The Netherlands

Received: 9 June 2010 – Accepted: 14 June 2010 – Published: 7 July 2010

Correspondence to: J. Bieser (johannes.bieser@gkss.de)

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

The US EPA regional emission model SMOKE was adopted and modified to create temporally and spatially distributed emission for Europe and surrounding countries based on official reports and public domain data only. The aim is to develop a flexible model capable of creating consistent high resolution emission data for long-term runs of Chemical Transport Models (CTM). This modified version of SMOKE, called SMOKE for EUROPE (SMOKE-EU) was successfully used to create hourly gridded emissions for the timespan 1970–2010.

In this paper the SMOKE-EU model and the underlying European datasets are introduced. Emission data created by SMOKE-EU for the year 2000 are evaluated by comparison to data of three different state of the art emission models. Differences of SMOKE-EU to those models were in the same range as the differences amongst them. Further, concentrations of criteria pollutants calculated by the CTM CMAQ using the four different emission datasets were compared against EMEP measurements with hourly and daily resolution. Using SMOKE-EU emissions O₃, NO₂ and SO₄ could be modelled most reliably. The amount of simulated concentrations within a factor of 2 (F2) of the observations for these species are: O₃ (F2=0.79 N=329 197), NO₂ (F2=0.55 N=11 465), and SO₄ (F2=0.62 N=17 536). The lowest values were found for NH₄ (F2=0.34 N=7400) and NO₃ (F2=0.25 N=6184). NH₄ concentrations were generally overestimated, leading to a fractional bias (FB) averaged over 22 measurement stations of (FB=0.83±0.41) while better agreements with observations were found for SO₄ (FB=0.06±0.38, 51 stations) and NO₃ (FB=0.13±0.75, 18 stations).

CMAQ simulations using the three other emission datasets were similar to those modelled using SMOKE-EU emissions. Highest differences were found for NH₄ while O₃ concentrations were almost identical. The results of this comparison confirm that it is adequate to use emissions created by SMOKE-EU as input for CTMs.
1 Introduction

Chemistry transport models (CTMs) are used for a variety of purposes (air quality modelling, source attribution, assessment of abatement strategies, etc.) with modeling domains reaching from global coverage down to local scales. In addition to the meteorological data, lack of knowledge on emissions introduce a major uncertainty to the CTM modeling results (Russell, 2000; Seaman, 2000; Hanna and Davis, 2001; Anderson and Langner, 2005; Sofiev et al., 2009).

In general there are two ways of modelling emissions. The Bottom-Up approach models emissions by combining sources with activities and emission factors. The source is the spatial location of the emitter, the activity is the temporal emission pattern and the emission factor determines the amount of pollutants emitted (Benkovitz, 2004). This approach is reasonable for uniform sources. Thus, Bottom-up is mostly used for biogenic and mobile sources since they can be lumped into a limited number of source types (e.g. coniferous trees, broadleaf trees for biogenic emissions; diesel vehicles, gasoline vehicles for mobile sources). The opposite methodology is the Top-Down approach. It is used for groups of individual sources which can not be easily lumped but for which regional annual total emissions can be estimated from sales, usage or other statistics (e.g. power plants). These estimated annual total emissions are also called emission inventories. They are usually separated into several source sectors combining chemical processes (e.g. combustion, solvents) and/or economic units (e.g. industry, private households). For the use in CTMs these aggregated emissions are spatially and temporally disaggregated using spatial surrogates and temporal profiles. A spatial surrogate is a proxy for the fraction of the total emissions emitted in each grid cell. Because there is only a limited amount of European emission inventories and surrogates, all emission models use similar types input data. The datasets used for SMOKE-EU are introduced in greater detail in Sect. 2.

Besides proprietary emissions models which are not publicly available, there are several public models. Each of these models has its own restrictions: compatibility to
a certain CTM, temporal coverage, spatial resolution for regional modelling or the focus on a single nation or region. The EMEP emission data provided by MSC-W have a high temporal coverage for all of Europe with spatial resolution of 50×50 km². Temporally disaggregated emissions are not published (Webdab, 2010). The Dutch CTM LOTOS-EUROS developed by TNO and RIVM as well as the French CTM CHIMERE have their own emission models producing suitable emission data (Schaap et al., 2005; Vautard et al., 2007). Yu et al. (2008) adapted the SMOKE model to create emissions for the UK. The Dutch TNO and the German IER emission models are two widely used emission models capable of producing high resolution emissions but are not public. (Friedrich and Reis, 2004; Visschedijk et al., 2007). However, the emission datasets calculated by TNO can be obtained free of cost. The EDGAR emission database contains emissions of air pollutants on a 1×1 degree grid for the years 1990, 1995 and 2000 (Olivier, 2001). The before mentioned models are only some examples and do not cover all European emission models. Looking at the variety of emission models available for Europe the question arises what benefit can be gained by an additional model. The reason to develop this emission model is to provide a flexible tool capable of creating consistent high resolution emission datasets for long term CTM runs over Europe based only on open source data. Flexibility means that the model can be easily altered concerning the input data and output format and that new species or different photochemical splits can be implemented with a minimum amount of work. Consistency means that emissions for each year are calculated using similar input data and the same algorithms. This consistency approach is in contrast to many emission models, which use the best available data for each new report year, with report years usually being every five or ten years. This approach leads to a steady improvement of the emission datasets but comes at the cost of inconsistency with older datasets since these older report years are not available with the new methodologies. The model introduced in this paper is specifically designed for long-term CTM runs and thus needs to overcome these problems.
For the evaluation of SMOKE-EU, datasets from three widely used emission models are used. These are the TNO-GEMS dataset created with the TNO model, a purchased dataset from IER further called IER-GKSS and the official EMEP emissions. These emission datasets are introduced in further detail in Sect. 3.1. The emissions are compared concerning the total emissions, the spatial distribution and the temporal distribution. Furthermore all four emissions datasets are being used as input for the CMAQ (Community Multiscale Air Quality) CTM for the year 2000. The calculated air concentrations of the species \( O_3 \), \( NO_2 \), \( NO_3 \), \( SO_2 \), \( SO_4 \) and \( NH_4 \) are compared with measurements from rural measurement sites. These comparisons are thoroughly described in Sect. 4.

2 Methodology

The emission model SMOKE is the official emission model of the United States Environmental Protection Agency (US EPA) and is one of the most used emission models worldwide (Houyoux et al., 2000; MCNC Environmental Modelling Center, 2008; UNC Carolina Environmental Program, 2005). SMOKE was originally created by the MCNC Environmental Modelling Center (EMC) and developed further by the US EPA. It is the official emission model of the Models-3 Community Modelling and Analysis System (CMAS) and creates emission data suitable for CMAQ (Byun and Ching, 1999; Byun and Schere, 2006). Anthropogenic emissions are calculated using the “Top-Down” methodology while biogenic emissions are calculated by the Bottom-Up model BEIS3 (Guenther et al., 2000; Pierce et al., 1998; Schwede, 2005). Although SMOKE is highly specialized for usage with officially reported data in the US, there have been several successful attempts to use it for other regions. In Europe, for example, SMOKE has been adapted to use the national emission inventories of Spain and the UK (Borge et al., 2008; Yu et al., 2008).

The SMOKE emissions model follows a modular setup (Fig. 1). Area, point, mobile and biogenic sources are calculated by different modules and merged into a single
output file. Short descriptions of the major modules for area and point source processing and their function as well as the modules of the biogenic bottom-up model BEIS3 can be found in Appendix C. In order to run SMOKE, four kinds of data are needed for the different species: The bulk emission inventory, spatial surrogates, speciation profiles, and temporal profiles. For Plume in Grid (PiG) calculations and biogenic emissions certain meteorological input data are needed additionally (eg. temperature, radiation, wind, humidity).

The Smkinven module reads the data in the inventory file which contains the aggregated emissions distinguished by a 6 digit regional code FIPS (US Federal Implementation Planning Standards) and a 10 digit source code SCC (Source Classification Code). In the US the emission inventories are usually published on county level leading to a high spatial resolution. Also the 10 digit SCC code allows for detailed partitioning of source types. The subsequent SMOKE modules search for different profiles matching the FIPS and SCC codes of each emission source and use the best fit if no exact match is possible (Baek et al., 2009).

### 2.1 SMOKE for Europe (SMOKE-EU)

The SMOKE model has been under development for over a decade. Therefore it is highly specialized on the usage of official data of the US. Since this model setup is not directly compatible to European data reporting schemes several adjustments need to be made for the use of SMOKE for Europe.

In order to achieve a high spatial resolution SMOKE uses emission aggregates on county basis and distributes them using static surrogates for each region. This is done by the Grdmat module which creates a single, static gridding matrix (GRDMAT) for each year. When used with European emissions aggregated on the national level these static surrogates lead to a static spatial distribution for each country over the whole year. This is a sound assumption for sources that are spatially static like for example mobile emissions which are connected to the road network throughout the year. For emissions that are influenced by local events, such as combustion for heating,
static surrogates in combination with large or inhomogeneous regions can lead to an unrealistic emission distribution. This is due to the fact that the spatial distribution of heating demand is not static throughout the year but changing depending on the temperature. Furthermore the temporal disaggregation in SMOKE is done via monthly, weekly and hourly profiles. This can lead to strong emission changes between the last day of a month and the first day of the next month.

In order to overcome these restrictions of SMOKE, in SMOKE-EU a new module has been introduced. It’s basic function is to create a distinct gridding matrix (GRDMAT) for each day of the year. This matrix, because it modifies the gridding matrix for each day, is called modification matrix (MODMAT) and the module calculating it Modmat. By definition, unless parts of the surrogate are outside the modelling domain, the sum of each surrogate is always 1 (Eq. 1). This is also true for the average of all modification matrices (Eq. 2) but not for each single daily modification matrix (Eq. 3). The changing sum of each modification matrix for each day represents an annual temporal profile for each grid cell, thus replacing the monthly temporal profiles used by the original SMOKE model.

\[
\sum_{i=1}^{N} \text{GRDMAT}(i) = 1 \quad N = \text{number of grid cells} \quad (1)
\]

\[
\sum_{i=1}^{N} \left[ \sum_{j=1}^{T} \text{MODMAT}(i,j) \right] / T = 1 \quad T = \text{number of time steps (365 days/year)} \quad (2)
\]

\[
\sum_{i=1}^{N} \text{MODMAT}(i,j) \epsilon [0,T] \quad (3)
\]

Equation (4) shows the calculation of gridded emissions by SMOKE. For each species hourly emissions in g/s or mole/s are calculated by multiplying the gridding matrix (GRDAMT), the speciation matrix (SPCMAT), the Plume in Grid matrix (PiG) and the temporal factors (TMPFAC) with the annual total emissions (TOT). Since it is not time
dependent the gridding matrix is calculated only once for each year (Eq. 4).

\[ \mathbf{E}(t,x,y,z) = \mathbf{GRDMAT}(x,y) \cdot \mathbf{SPCMAT}(x,y) \cdot \mathbf{PiG}(z) \cdot \mathbf{TMPFAC}(t) \cdot TOT \]  

(4)

The Modmat module calculates separate gridding matrices for each day as indicated by Eq. (5). For better readability the horizontal dimensions \( x \) and \( y \) have been substituted by the grid cell number \( n \). The change matrix \( \mathbf{CHGMAT}(n,t) \) is calculated from external files. Here, for all emissions from heating, change factors have been calculated using the 2 m temperature as a proxy for heating demand (Aulinger, 2010). For each day, the gridding matrix (GRDMAT) is multiplied with the change matrix (CHGMAT) and normalized. The normalization matrix (NORMAT) is calculated once by multiplying the static gridding matrix with the change matrix (Eq. 6).

\[
\mathbf{MODMAT}(n,t) = \frac{\mathbf{GRDMAT}(n) \cdot \mathbf{CHGMAT}(n,t) \cdot T \cdot \sum_{i=1}^{N} \mathbf{GRDMAT}(i)}{\mathbf{NORMAT}(n)}
\]  

(5)

\[
\mathbf{NORMAT}(n) = \sum_{i=1}^{N} \left[ \mathbf{GRDMAT}(i) \cdot \sum_{j=1}^{T} \mathbf{CHGMAT}(i,j) \right]
\]  

(6)

While the annual total emissions remain unchanged, the spatial as well as the temporal distribution is changed. This leads to a mixture of spatial and temporal disaggregation. Thus, the originally applied monthly profiles are not used anymore, since they are already represented by the 365 daily modification matrices.

Although several changes to the original SMOKE source code have been made SMOKE-EU is not a completely new emission model. It is rather a specific setup of the SMOKE model which is able to create high resolution emissions for Europe. A large part of SMOKE-EU are the numerous input files needed in order to run SMOKE for Europe. These datasets and their usage is described in the following sections.
2.2 Emission Inventories

European emission inventories and datasets are quite inhomogeneous. Most countries use different methodologies to assess the national emissions. This results in different national emission inventories possibly using different emission factors for similar sources and allocation of these to different source categories. Also, some countries do not publish their emission inventories. Furthermore most countries use a national map projection making transformation of the data necessary. For SMOKE-EU it was decided to aim for overall consistency, by using Pan-European datasets when available.

2.2.1 The European Monitoring and Evaluating Program (EMEP)

Initiated by the Convention on Long-range Transboundary Air Pollution (LRTAP), signed in 1979, the European Monitoring and Evaluation Program (EMEP) was implemented. National annual emission estimates are reported by the parties under the LRTAP convention using the standardized methods defined by the CORINAIR (CORe Inventory of AIR emissions) guidebooks (Vestreng, 2007; Webdab, 2010). The officially submitted data are published together with a corrected version that was reviewed by national experts.

EMEP publishes annual national totals for all European countries including Russia and also Turkey and North Africa. The species covered by the EMEP inventory are CO, NO\textsubscript{x}, SO\textsubscript{2}, NH\textsubscript{3}, Non-Methane Volatile Organic Compounds (NMVOC), primary particulate matter (PM) as PM\textsubscript{10} and PM\textsubscript{2.5}, several Heavy Metals (HMs) and some Persistent Organic Pollutants (POPs). The emissions are distributed over 11 SNAP source sectors (Selected Nomenclature for sources of Air Pollution) (Table 1). SNAP is a standard defined by the CORINAIR guidebooks which ensures that emissions reported by different nations are comparable (European Environmental Agency, 2007). EMEP covers the years 1970–2009 with additional projections for 2010, 2015 and 2020. In addition to the national reports also emissions from international shipping are included in the inventory.

957
2.2.2 The European Pollutants Emission Register (EPER)

EPER is the European Pollutant Emission Register, the first Europe-wide register of industrial emissions into air and water, which was established by the European Commission in July 2000 (European Commission, 2000). EPER has been released for two base years. For the EU15 (Austria, Belgium, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Luxembourg, The Netherlands, Portugal, Spain, Sweden and UK) in 2001 and for the EU27 (EU15 + Bulgaria, Cyprus, Czech Republic, Estonia, Hungary, Latvia, Lithuania, Malta, Poland, Romania, Slovakia and Slovenia) in 2004. There are considerable differences between the emission data released in 2001 and 2004. This is mainly due to the fact that the 2004 release is more complete. Here, only the 2004 inventory is used for point source modelling (European Pollution Emission Register, 2010). It covers approximately 12000 industrial point sources with information about annual total emissions, source code and geographical location. The NACE (Nomenclature statistique des activités économiques dans la Communauté européenne) code is a more sophisticated source identifier than the SNAP code. It consists of several hundred different source types, especially distinguishing between different industries. A high percentage of NACE codes are covered by SNAP3 and SNAP4.

2.2.3 Merging EMEP and EPER into a combined emission inventory

Since the EPER inventory includes the exact geographical location of each source no surrogates are needed to estimate the spatial distribution of the emissions. Furthermore the industrial processes of each source are known. This allows for a more precise estimation of the effective emission heights. Because of this, EPER sources are considered more precise than EMEP sources. Since EPER only contains major point sources, the missing emissions are taken from the EMEP inventory which is an estimate of the national total emissions. This is done by subtraction of EPER from EMEP. In very few cases the EPER emissions for a certain species and sector exceed the
EMEP emissions. In those cases EPER emissions are used, leading to slightly higher emissions than reported in the EMEP inventory. The preparation of the SMOKE-ready inventory files is done by a newly written java based preprocessor called InvenCombine (Appendix C). The calculations are done in three steps:

1. Conversion of EPER from NACE to SNAP sectors.
2. Adjustment of the EPER base year 2004 emissions to the modelling year.
3. Merging of the two inventories.

First the EPER data needs to be converted from NACE sectors to SNAP sectors. While most sectors can be converted directly, there still are some incompatibilities between the two systems. NACE has a wide range (more than 100) of industrial sources distinguished by industrial sector. While, concerning industrial sources, SNAP distinguishes between two processes – industrial combustion (SNAP3) and manufacturing and industrial processes (SNAP4).

In order to correctly convert the EPER data, for each region and for each species all NACE classes fitting into SNAP3 and SNAP4 are first combined into a single sector and then redistributed depending on the ratio of SNAP3 to SNAP4. In a second step the 2004 EPER data is attributed to each SNAP sector and each species according to the relative change of EMEP emissions between 2004 and the inventory year. Finally the SNAP converted and adjusted EPER emissions are subtracted from the EMEP emissions.

2.3 Spatial surrogates

Spatial surrogates are the proxies used to disaggregate the national total emissions to the emissions model grid. The sum of each surrogate is, by definition, 1 unless parts of the country for which the emissions have been aggregated are outside the model domain (e.g., Russia). If there are no specific surrogates for a certain region the
population density is used for anthropogenic emissions. Maes et al. showed that disaggregating the combined EPER and EMEP emissions with European datasets leads to spatially distributed emissions comparable to high resolution national emission inventories (2009). A list of datasets used for each SNAP sector is shown in Table 2. All surrogate input datasets are interpolated to the SMOKE-EU modelling domain and converted to the SMOKE format by several preprocessors (Appendix C). In the following the surrogate datasets are briefly described:

*Gridded Population of the World version 3 (GPWv3)* depicts the distribution of human population across the globe. It contains globally consistent and spatially explicit human population information and data. It is released for every fifth year starting in 1990 on a 2.5′ × 2.5′ resolution. Furthermore future projections until 2015 are available (Balk, 2004; Sedac, 2010). The GPWv3 population density dataset is used as default surrogate.

*Corine Air Land Cover (CLC)* dataset was created by the European Environmental Agency (EEA) and is freely available over the official website (Corine Land Cover, 2010). So far the dataset has been released for 1990, 2000 and 2006. CLC distinguishes 45 different land use classes with a spatial resolution of 100×100 m². It covers all member states of the European Union.

*Global Land Cover (GLC2000)* dataset provided by the Land Cover Institute of the United States Geological Survey (USGS) is a global land use database. It was released once, for the year 2000, with 1×1 km² resolution. It distinguishes 24 different land use classes. The GLC2000 data was used as surrogate for all regions without CLC coverage (USGS, 2009).

*Openstreetmaps (OSM)* is a public domain vector database combining GPS (Global Positioning System) data from thousands of volunteers around the world containing a free global street and land use map. Since the start of the project in 2004 nearly complete coverage of streets and railroads in the EU could be achieved. The 2009/12 version of OSM has been used to create surrogates of motorways, major rural roads and railways (Openstreetmap, 2010).
Digital Chart of the World (DCW) is a public domain vector database developed by the Environmental Systems Research Institute, Inc. (ESRI) for the US Defense Mapping Agency (DMA). Besides others it contains data about roads, railways and waterways. The DCW is freely available for the year 1992 (Digital Chart of the World, 1992). This dataset has been used to disaggregate mobile emissions before 1993. Between 1993 and 2000 an interpolated dataset consisting of OSM and DCW is used.

GSfM Land Use Database is a compilation of different land use datasets. Besides other land use data it contains the for UBA (Federal Environment Agency) created Forest database (JRC/TNO) which distinguishes 136 different tree types and the CLC2000 landuse dataset (Smiatek, 1998). Since the CLC dataset distinguishes only between 5 forest types, the UBA forest database is used to determine the tree coverage for the biogenic emissions model BEIS3. Land use dependent emissions like NO are calculated using the CLC database.

TREMOVE is a policy assessment model, designed to study the effects of different transport and environment policies on the emissions of the transport sector (EC, 2007). The model estimates for policies as road pricing, public transport pricing, emission standards, subsidies for cleaner cars etc., the transport demand, modal shifts, vehicle stock renewal and scrappage decisions as well as the emissions of air pollutants and the welfare level. It models both passenger and freight transport, and covers the period 1995–2030 (TREMOVE, 2010). The v2.7b Basecase dataset of the TREMOVE bottom-up emission model has been used to split the EMEP emissions estimated for sector SNAP7 (Road transport) into motorway, rural and urban subsectors as well as to distinguish between different vehicle and fuel types. The EMEP sector SNAP8 (Other mobile transport) is split into the subsectors transport by rail, inland shipping and airplanes.

EUROSTAT is the statistical service of the European Union. It releases statistics concerning the economy, environment, society, industry, agriculture and regional development (EUROSTAT, 2010). Some EUROSTAT statistics date back as far as 1953. All statistical values are reported using the Nomenclature of Units for Territorial Statis-
tics (NUTS) geocode standard which is the official European system for referencing subdivisions of countries (European Commission, 2003). NUTS regions are defined by the amount of inhabitants (Table 3). The EUROSTAT data are usually available as monthly national or annual regional values, with regional values going down to NUTS3 level. The EUROSTAT regional statistics on NUTS2 level are used to further disaggregate industrial and agricultural emissions depending on the number of employees in certain industries, number of employees in agriculture and animal stocks for NH$_3$ emissions from animals.

2.4 Vertical distribution

For the use in CTMs it is still common to apply static vertical distribution factors to the emissions of each sector or even to put all emissions into the lowest layer. With effective emission heights of industrial sources in the range of 100 m to 600 m Plume in Grid (PiG) calculations can have a strong impact on the calculated air concentrations and depositions. Emissions in higher layers are likely to be transported further away from the source, wet depositions are lower if a higher amount of pollutants is above the cloud layer and particles need longer until they reach the ground by dry deposition giving them more time for interaction with other species. For example, comparisons of different CTM runs showed a change in the SO$_4$ to SO$_2$ ratio depending on the emission height.

All non-VOC emissions from the SNAP sectors 1, 3, 4, 5 and 9 are treated as elevated sources. VOC emissions from dump sites (SNAP9) are interpreted as surface evaporations and thus are not elevated. Data for stack height, stack diameter, exit velocity and exit temperature are applied to all EPER sources depending on NACE sector following Pregger and Friedrich (2009). All emissions not covered by EPER are first horizontally distributed as described in Sect. 2.3 and then supplemented with average stack data depending on SNAP sector. For countries covered by EPER it is assumed that the remaining sources are only minor sources thus having lower average stack heights than their corresponding EPER sources. For those countries not
covered by EPER a sectoral emission-weighted average is built using stack data for major sources. The vertical distribution of emissions by point sources is calculated using the SMOKE module \textit{Laypoint}. It calculates the emission plumes using the Briggs PiG equation (Briggs, 1972; Houyoux, 1998). This leads to different effective emission heights depending on the meteorological fields used as input for the PiG calculations.

\textbf{2.5 Temporal distribution}

SMOKE-EU uses the LOTOS-EUROS monthly, weekly and diurnal profiles which features distinct profiles for each SNAP sector (Builtjes 2003). For SNAP sector 2 (Non-industrial combustion plants) the 2 m temperature is used to create the annual temporal profiles using the \textit{Modmat} module (Aulinger, 2010). This leads to a more realistic, year specific temporal disaggregation. While currently all other SNAP sectors use the static LOTOS-EUROS profiles for temporal disaggregation, there are other possible applications for \textit{Modmat}. For example, it seems promising to use the soil moisture as an additional proxy for NH$_3$ emissions from agricultural areas.

The biogenic emissions which are calculated by the bottom-up model BEIS3 are temporally disaggregated using meteorological fields. VOC emissions of trees are depending on the near surface temperature (2 m–10 m) and the incoming radiation. Biogenic NO emissions are depending on soil moisture and soil temperature.

\textbf{2.6 Chemical speciation}

Some substances in the emission inventories are composites of many different distinct species. For all CTMs volatile organic compounds (VOC) need to be distributed into several organic species, depending on the photochemical mechanism in use. Nitrogen oxides are usually reported as NO$_x$ and need to be split into NO and NO$_2$. SMOKE-EU currently splits all NO$_x$ emissions into 90\% NO and 10\% NO$_2$ (EPA, 2010). Beside this there can be other substances which need to be speciated, such as primary particulate matter for CMAQ. SMOKE is able to split any species from the bulk emission inventory.
into arbitrary subspecies. This makes it easy to adjust the emission model to match
different chemical mechanisms and other user demands.

*Primary Particulate Matter* (PM) in the bulk emission inventory is separated into two
size classes. These are particles smaller than 10 µm (PM<sub>10</sub>) and particles smaller than
2.5 µm (PM<sub>2.5</sub>). For CMAQ PM<sub>2.5</sub> needs to be further speciated into primary elemental
carbon (PEC), primary organic aerosols (POA), primary nitrate aerosols (PNO3),
primary sulfate aerosols (PSO4) and other particles (PMFINE). Each of the 10 SNAP
sectors has its own PM split, while some sectors also have splits on sub-sector level.
Vehicles for example have different PM splits depending on vehicle type (Heavy Duty
Vehicles, Light Duty Vehicles, Buses) and fuel type (Diesel, Gasoline). The PM splits
were adopted from the SMOKE emission model (EPA, 2010). Additionally, split factors
for emissions from international shipping have been implemented (Agrawal, 2008).

*Volatile Organic Compounds* (VOCs) need to be speciated according to the photo-
chemical mechanisms used by the CTM. At this point SMOKE Europe supports VOC
splits for the mechanisms Carbon Bond 4 (CB-IV) and Carbon Bond 5 (CB05) (Gery
et al., 1989). New photochemical mechanisms can be easily implemented by supplying
the split factors for each SNAP sector. The split factors have been calculated using the
chemical VOC analysis of Passant (2002).

### 3 Evaluation of the emission data

First of all the impact of the *Modmat* module on the spatial and temporal disaggrega-
tion of the emissions is assessed. This is done by comparison of two different datasets
created with SMOKE-EU. The first emission dataset, the default case, uses only static
temporal profiles and surrogates. The second dataset is created using the *Modmat*
module for the calculation of emissions from residential heating (SNAP2). In this case
*Modmat* uses the 2 m temperature from meteorological input fields as a proxy for heating
demand (Aulinger, 2010).
In a second step the SMOKE-EU emissions for the year 2000 are statistically compared to three state of the art emission datasets. The comparison is done separately for the 6 inventory species: NO\textsubscript{x}, SO\textsubscript{2}, CO, PM\textsubscript{10}, NH\textsubscript{3}, VOC. First, the total emissions for the EU27 countries are compared, then the horizontal, vertical and temporal distributions of the different emission datasets are compared. Only exemplary figures are shown for each statistical comparison.

3.1 Emission datasets used for comparison

In order to validate the emissions created by SMOKE-EU three emission datasets calculated by widely used models have been used for comparison. These datasets will be referred to as EMEP, IER-GKSS and TNO-GEMS.

**EMEP**: The EMEP emission dataset created by the Meteorological Synthesizing Center – West (MSC-W) is based on the EMEP emission inventory. Species covered are CO, NO\textsubscript{x}, SO\textsubscript{2}, NH\textsubscript{3}, PM\textsubscript{10}, PM\textsubscript{2.5} and NMVOC. Usage of point sources is not implemented. Spatial distribution and speciation is carried out with average profiles for each SNAP sector. EMEP uses the IER temporal profiles for AQ calculations with the EMEP unified model (Benedictow et al., 2009; Simpson et al., 2003). Still only gridded annual totals on a 50×50 km\textsuperscript{2} domain together with SNAP specific vertical profiles are published by EMEP (Webdab, 2010). For comparison with SMOKE emissions the EMEP dataset was interpolated to 54×54 km\textsuperscript{2}. The LOTOS-EUROS temporal profiles have been used for temporal disaggregation.

**IER-GKSS**: An emission dataset for the GKSS 54×54 km\textsuperscript{2} modelling domain over Europe was purchased from the University of Stuttgart Institute for Rational Use of Energy (IER) and is here referred to as IER-GKSS. The IER emissions model is based on the EMEP/CORINAIR emission guidebooks. It features distinct temporal profiles for each country and SNAP sector as well as VOC and PM splits. The dataset purchased by GKSS has no vertical distribution (Friedrich and Reis, 2004).

**TNO-GEMS**: The Netherlands Organization for Applied Scientific Research (TNO) GEMS emissions are a 0.125×0.0625 degrees dataset created by the TNO emission
model for the EU FP7 project GEMS (Global and regional Earth-system Monitoring using Satellite and in-situ data). The TNO model uses hourly, daily and monthly emission factors for each species and country. The emissions are vertically distributed using the SNAP dependent EMEP profiles. The TNO-GEMS dataset is scaled to match the EMEP emissions for 2003 (Visschedijk and Denier van der Gon, 2005, Visschedijk et al., 2007). For comparison with SMOKE emissions the dataset has been interpolated to 54×54 km².

3.2 Evaluation of the impact of the Modmat module

SMOKE-EU has been set up to process anthropogenic emissions from sector SNAP2. The default scenario uses the population density as static surrogate for SNAP2 sources and LOTOS-EUROS temporal profiles. SNAP2 emissions are mostly due to residential heating and thus correlated to the near surface temperature. The modified scenario uses the 2 m temperature from meteorological fields as input data for the Modmat module, which in this case calculates daily gridding matrices using the average heating demands which are related to specific emissions (Aulinger, 2010). This changes the spatial as well as the annual temporal distribution.

Comparing the two emission datasets revealed two major impacts of the Modmat module. As expected its impacts correlate with the size of the aggregated region. The biggest differences between the default and the modified scenario could be observed for the spatial disaggregation of big regions or regions with strong temperature gradients. For Switzerland, which is one of the smallest European countries, differences of up to 20% in annual total emissions have been found in certain grid cells. Also the annual temporal disaggregation is not following monthly average profiles any more. This leads to a smoothing of the annual profiles and avoids the sometimes strong emission changes at the end of each month (Fig. 2a). Additionally each year now has a unique temporal profile, making the Modmat module particularly interesting for long term runs. It can be seen that in the year 2000 more heating occurred in January than in December, while the years 1999 and 2001 show the opposite behaviour (Fig. 2b). The inter
annual variability of the temporal profiles is as high as the deviation between the default and modified SMOKE-EU version.

In order to assess the impact of Modmat to the default SMOKE-EU version on air concentrations, the emissions from both the default and the modified scenario were used as input for the CMAQ CTM. For 250 rural grid cells daily average calculated air concentrations of SO₂, NO and CO, the three main emitted substances in SNAP2, have been compared with one another. The statistical indicators used for comparison are the Mean Normalized Error (MNE) and the Mean Normalized Bias (MNB) (Appendix A). The average MNE is 20% (6% to 56%) with a MNB of 9% (−18% to 50%). When comparing the concentrations calculated using the complete emission datasets with all EMEP and EPER emission sources, values are: MNE=3.5% (0.8% to 49%) and MNB=1% (~9% to 38%). The annual total emissions for the whole domain remain unchanged. This shows that the usage of the Modmat module, even for a single SNAP sector, has a significant impact on the calculated air concentrations in certain regions.

3.3 Comparison of annual total emissions

First of all the annual total emissions of the four emission datasets have been compared. The SMOKE-EU, EMEP and IER-GKSS datasets were created for the year 2000 while the TNO-GEMS emissions are for 2003. Figure 3 shows the absolute annual anthropogenic emissions in Gg/a for the EU27. Biogenic emissions as well as emissions from international shipping have been excluded from this comparison since they are not included in all datasets. Due to biogenic emissions the total NMVOC emissions in the SMOKE-EU dataset are higher by 18 000 Gg/a. The annual averages of all datasets and their deviations are: NOₓ (12 500 Gg±6.8%), SO₂ (10 600 Gg±9.1%), CO (38 900 Gg±16.7%), PM₁₀ (2830 Gg±7.1%), NH₃ (4000 Gg±39.8%), VOC (10 500 Gg±10%).

Figure 4 shows that most inventories have annual total emissions similar to those reported by EMEP with differences less than 10%. Only the IER-GKSS NH₃ emissions are 30% lower than the EMEP values. The SMOKE-EU emissions are somewhat
higher than the EMEP reports since in some countries EPER emissions exceed EMEP emissions. Since the total emissions of the four datasets are similar, no further investigation concerning the aggregated emissions have been made.

3.4 Comparison of horizontal disaggregation

All spatial statistics have been calculated using the EU27 emissions only. The values compared are gridded annual total emissions for the species CO, NO\textsubscript{x}, SO\textsubscript{2}, PM\textsubscript{10}, NH\textsubscript{3} and NMVOC. All figures in this section show the best fit (Figs. 5a, 6a, 8a) and the worst fit cases (Figs. 5b, 6b, 8b). Generally SO\textsubscript{2} emissions show the best agreement for all four datasets. This is due to the fact that SO\textsubscript{2} emissions are well known concerning the total amount emitted as well as their spatial and temporal distribution. NH\textsubscript{3} emissions on the other hand have the highest uncertainties and thus generally show the largest differences. Three statistical methods have been chosen in order to compare the spatial disaggregation of the four different emissions datasets:

3.4.1 The frequency distribution of emissions

First the frequency distributions of the emissions have been compared. They give an impression of the overall distribution of the emissions, whether there are more high emission point sources or more low emission areas in a dataset. In general the distribution of all species is very similar with a strong peak for low values. For most species almost no difference in the frequency distribution can be seen (Fig. 5). This leads to correlations between 0.8 and 0.99. Only for NH\textsubscript{3} a shift towards lower emission can be seen for the IER-GKSS emissions.

3.4.2 The frequency distribution of the deviation using EMEP as standard

The deviations of the annual total emissions for all grid cells have been calculated and plotted as frequency distributions. The EMEP dataset has been used as standard. This statistical measure actually compares the spatial surrogates of the different emission
datasets. A shift of all emissions by one grid cell for example would give high deviations for two identical frequency distributions of emissions. Again it could be shown that all four datasets are very similar concerning their spatial distribution. As expected the lower \( \text{NH}_3 \) emissions in the IER-GKSS data leads to slight shift towards negative deviations (Fig. 6).

3.4.3 The spatial variability as indicated by variograms

\[
f(h) = \frac{(z(x + h) - z(x))^2}{2}\quad x = \text{reference grid cell}; \quad h = \text{distance to origin}
\]  

As a third measure for the spatial distribution variograms have been calculated (Eq. 7). The interval size is 100 km. Since it is not possible to show the variograms for every grid cell, a representative origin has been chosen. The variograms shown here have their origin in a central cell of the EU27. As Eq. (7) indicates the values of a variogram are dependent on the emissions in the origin grid cell. To eliminate the influence of the concentration of the origin grid cell and therefore create a more representative comparison also average total emission have been calculated. These \textit{spatial averages} show the annual average concentrations within concentric circles around the origin with 100 km distance (Fig. 7). It can be seen that the spatial distributions as well as the variograms for \( \text{SO}_2 \) follow a similar pattern (Fig. 8). Still some differences can be seen. Looking at the variograms for \( \text{SO}_2 \) it can be seen that the EMEP dataset shows the lowest square differences. This indicates a lower amount of grid cells with much higher emissions than the origin cell, which is most probably due to the lack of point sources in this dataset. The spatial averages show higher \( \text{SO}_2 \) emissions in the 500–700 km circles (30–40%) for the IER-GKSS datasets. This indicates that the \( \sim 8\% \) higher total \( \text{SO}_2 \) emissions in this dataset are due to higher emissions in a certain area rather than a general overestimation (Fig. 3).

\( \text{NH}_3 \) shows the biggest differences with a much higher squared difference in the 600 km and 900 km circles for the SMOKE-EU emissions, while the spatial averages
show only slightly higher NH$_3$ emissions in these areas of the SMOKE-EU dataset (Fig. 9). This could be due to a stronger partitioning of high and low emission grid cells in this area. A possible reason is the spatial disaggregation by EUROSTAT NUTS2 statistics. The IER-GKSS dataset shows lower emissions of NH$_3$ throughout the domain compared to the other datasets.

In general it can be stated that the large scale spatial distribution on the 54 × 54 km$^2$ grid are quite similar for all four emission datasets, while looking at certain grid cells reveals some obvious differences. Thus, spatial variability is a point where the emission datasets deviate from each other.

### 3.5 Temporal distribution

Temporal profiles were available for the SMOKE-EU, IER-GKSS and the TNO-GEMS emissions. These temporal profiles are not directly comparable. The SMOKE profiles are available for each SNAP sector, the original IER-GKSS profiles are not available and the TNO-GEMS profiles are available for each region and species. In order to gain comparable temporal profiles for all three datasets, the average emissions for all grid cells of the EU27 were used to create species-dependent temporal profiles with daily resolution.

For most species these annual time series show deviations of less than 20% for all 365 daily temporal factors. Figure 10 shows an example plot for NO. The biogenic NO emissions which mainly occur during summer lead to a slightly different temporal profiles in the SMOKE-EU dataset (Fig. 10a). Temporal profiles of NO$_x$, PM$_{10}$ and CO are similar. The highest deviations were found for NH$_3$ (Fig. 10b). Here the strong, sudden changes between months of the original SMOKE temporal disaggregation can be seen.
3.6 Vertical distribution

The vertical distributions of the SMOKE-EU emissions were compared to the EMEP vertical distributions. For this purpose annual average vertical profiles for each species were calculated. Also the 5 emission layers of the EMEP profile were interpolated to the 30 layers of the SMOKE-EU dataset. As in Sect. 3.5 this does not necessarily represent the actual profiles used by the emission models. In Fig. 11 it can be seen that the SMOKE-EU plume in grid calculations result in lower emission heights than the official EMEP vertical distribution. EMEP distinguishes 10 static vertical profiles, one for each SNAP sector. The SMOKE-EU effective emission heights are determined using temperature and wind dependent plume in grid calculations, thus leading to different emission heights for each source throughout the year. For some species EMEP uses significant emissions in high layers (SO\textsubscript{x}: 400–600 m 30%>600 m 20%) (NO\textsubscript{x}: 400–800 m 10%). The SMOKE-EU plume in grid calculations show almost no emissions higher than 600 m with less than 10% above 400 m.

4 Comparison of CTM calculated concentrations to observations

The CTM CMAQ4.6 of the US EPA was used to simulate atmospheric concentrations of air pollutants for the year 2000 (US EPA, 2009). Figure 7 shows the modelling domain containing Europe and surrounding countries. The spatial resolution is 54×54 km\textsuperscript{2} with 30 vertical layers, the photochemical mechanism used is CB-IV. Meteorological fields are taken from the COSMO-CLM model (Rockel and Geyer, 2008; Rockel et al., 2008). Monthly average boundary conditions were derived from the MOZART global model (Horowitz et al., 2003; Niemeier et al., 2006). With this setup, four CMAQ runs using different emission datasets were calculated. The three emission datasets for comparison with SMOKE-EU have been used as described in Sect. 3.1. Additionally, VOC and PM emissions were split using the same distribution as SMOKE-EU. SMOKE-EU is the only one among these models considering biogenic emissions.
The calculated atmospheric concentrations in the lowest model layer were compared to observations from EMEP measurement stations. From 242 available rural measurement stations those with more than 90% data coverage for the year 2000 were used for comparison (Fig. 7). Mountain stations which are not representative for a model grid cell have been skipped (e.g. CH01 Jungfraujoch at 3573 m). Six different compounds are used for comparison, three gaseous species (NO₂, SO₂, O₃) and three aerosol components (SO₄, NH₄, NO₃). Ozone concentrations are given as hourly values while all other values are reported as daily averages. Table 4 shows all used EMEP measurement sites and provides information on their location and species observed. Some sites consistently disagree with modelled values for all species and emission models (e.g. IT04 Ispra). This can be caused by strong topographic gradients not resolved by the CTM as well as by the meteorological model, local sources influencing the station or instrumental reasons. It should be kept in mind that a single observation site is not necessarily representative for the average concentrations in a 54 × 54 km² grid cell with a height of the lowest layer of 36 m.

The statistical measures used for comparison of simulated and observed values were chosen following Schlünzen and Sokhi (2009) and are described in further detail in Appendix A. Table 5 provides all used statistical values averaged over all relevant measurement stations as well as their standard deviation. The general picture when comparing the CMAQ results with measurements is that the four emission datasets produce comparable concentrations for all species.

The SMOKE-EU and EMEP based CTM runs predict slightly higher ozone values than the other models (Fig. 12a). One reason for this is the implementation of biogenic emissions in SMOKE-EU, leading to higher VOC and NO emissions during summer. Also the vertical distribution of NOₓ emissions in the SMOKE-EU and EMEP datasets potentially change the ozone regime in certain regions from VOC limited to NOₓ limited (Fig. 11b). However, since O₃ is strongly influenced by the meteorology (Andersson and Langner, 2005), the correlations and factor of 2 (F2) percentages for all four emission datasets are almost identical (Fig. 12c, Table 5). Only the Index of
Agreement (IOA) for the SMOKE-EU scenario is slightly higher (Fig. 12b). The Taylor diagram in Fig. 14a presents a similar picture. Although some regional differences can be seen, most measurement stations form a tight cluster between correlations of 0.5 and 0.8. In this setup the ozone concentrations calculated by CMAQ are generally 10% higher than the observations. Tests with meteorological fields created with a different meteorological model (MM5) (Matthias et al., 2009) produced 20% lower O$_3$ concentrations.

Looking at the Sulfuroxide species the highest mean SO$_4$ concentrations are predicted in the SMOKE-EU case (Mean=0.66 µg S/m$^3$) followed by the EMEP case with 0.61 µg S/m$^3$ while the other two cases underestimate SO$_4$ (Mean=0.57 and 0.54 µg S/m$^3$) (Fig. 13). Similar results can be seen for SO$_2$ were higher values are simulated in the SMOKE-EU case compared to the CTM runs using the other three emission datasets (Table 5). Since the total emissions as well as the spatial and temporal distribution of the SO$_2$ emissions are very similar in all four datasets, these differences can be explained by the different vertical distributions. In the EMEP and the TNO-GEMS datasets SO$_2$ is emitted in higher altitudes and partially above the boundary layer. This leads to less SO$_2$ in the surface layer because the emissions are distributed over a larger area and thus gives them more time to form particles before they reach the surface. Additionally, meteorological aspects are different in higher altitudes influencing chemical reactions. In the IER-GKSS dataset on the other hand all SO$_2$ is emitted in the surface layer, leading to a faster deposition and therefore to lower atmospheric SO$_2$ and SO$_4$ concentrations. CTM calculations using a version of the EMEP and TNO-GEMS datasets without vertical distribution agree with this assumption (Table 6). In most cases the emissions with vertical distribution show higher values for correlation, F2 and IOA. Looking at the Taylor diagram (Fig. 14b) some strong regional differences can be seen. Generally Scandinavian (green) measurement sites, with exception of NO42 (Spitzbergen), have the highest correlations. Central European (blue) sites on the other hand have the lowest biases, while the concentrations over the Spanish peninsula (orange) are systematically underestimated. However, a detailed
For all four emission datasets, modelled NH$_4$ concentrations are overestimated (Fig. 15a) and show the least agreement with observations of all species compared (Table 5). This is in agreement with the fact that the NH$_3$ emissions have the highest uncertainties of all species in the emission datasets. The lowest concentrations and best agreements with observations were simulated using the IER-GKSS emissions. This can be explained by the ~30% lower NH$_3$ emissions of this dataset (Figs. 3 and 4). On the other hand, the low NH$_3$ emissions also lead to an underestimation of NO$_3$ concentrations in the IER-GKSS case. Accordingly the higher NH$_4$ values in the SMOKE-EU case lead to an overestimation of NO$_3$ (Fig. 15b). Other than expected, the smoother temporal profiles of the IER-GKSS NH$_3$ emissions do not lead to better correlations on the annual scale.

For NO$_2$, CTM results show much higher Fractional Biases (FB) for the SMOKE-EU case (Fig. 15c). Since NO$_2$ is generally underestimated this leads to a higher amount of values within a factor of 2 (Table 5). The mean NO$_2$ concentration over all measurement stations given in Table 5 is dominated by high values at two stations IT04 (Ispra) and NL10 (Vredepeel). The comparison of simulated and observed NO$_2$ concentrations show strong spatial differences. Especially over the Spanish peninsula, where 5 of 33 measurement stations are located NO$_2$ concentrations are generally underestimated by a factor of 5.

5 Conclusions

The US-EPA SMOKE emission model has been successfully adopted and modified to use publicly available pan-European datasets in order to create high resolution emission data for Europe. Several preprocessors were developed which transform these datasets to create the input data necessary to run the SMOKE for Europe (SMOKE-EU) model. SMOKE-EU is capable of creating CMAQ ready emissions for the whole
of Europe, including Western Russia, Turkey and North Africa (Fig. 7a). Currently it is used to create emission datasets with spatial resolution in the range of $70 \times 70 \text{ km}^2$ down to $10 \times 10 \text{ km}^2$. The underlying datasets allow for a spatial resolution of up to $1 \times 1 \text{ km}^2$ (Table 2). Effective emission heights are determined via plume in grid algorithms. The species calculated by the model are CO, SO$_2$, NO$_x$, NH$_4$, PM, and NMVOC split according to the CB-IV or CB05 chemical reaction schemes.

The SMOKE-EU emissions were compared to datasets from three widely used emission models. These are the TNO-GEMS dataset created by TNO, a dataset from IER purchased by GKSS and the official gridded EMEP emissions provided by the MSC-W. Comparisons with SMOKE-EU emissions on a $54 \times 54 \text{ km}^2$ grid for the year 2000 showed similar total emissions, spatial and temporal distributions. The most significant differences were identified to be the NH$_3$ emissions (Fig. 4) as well as the vertical distributions (Fig. 11). Biogenic emissions lead to significantly higher NMVOC emissions as well as slightly higher NO emission during summers (Fig. 10). For the other species (CO, SO$_2$, NO$_x$, PM) total emissions differed less than 10% and temporal distributions differed less than 20%.

CMAQ has been used to calculate atmospheric concentrations of air pollutants using the four different emission datasets. Comparison of simulated values with observations from EMEP measurement stations showed that each of the four CTM runs produced sound results (Table 5). It could be shown that the vertical distribution has a strong influence on the simulated SO$_4$ and SO$_2$ concentrations (Table 6). Generally, SO$_2$ emissions in higher altitudes have lead to higher SO$_4$ concentrations near the surface and a better agreement with observations (Fig. 13). The largest differences were found for NH$_4$ and NO$_3$ concentrations (Fig. 15a,b). NH$_4$ was systematically overestimated while NO$_2$ was strongly underestimated over the Spanish peninsula (Fig. 15b,c). Ozone concentrations which are strongly influenced by the meteorology were almost identical for all datasets (Fig. 12).

By comparison with other emission datasets for the years 2000 and 2003 it could be shown that the project to create high resolution European emission data with the use
of open source data only was successful. Emission data created by SMOKE-EU will now be used for European long-term CTM runs for the timespan 1970–2010. Being a very flexible tool, SMOKE-EU will be further enhanced in the future. Improvements planned are different temporal profiles for each country, implementation of further photochemical mechanisms and the implementation of additional species (benzo[a]pyrene, mercury).

Appendix A

Statistical measures used for comparisons

\[ P_i = \text{Predicted value from Model} \]
\[ O_i = \text{Observed value from Measurement} \]
\[ N = \text{Sample size} \]

1. Mean

\[
\bar{O} = \frac{1}{N} \sum_{i=1}^{N} O_i, \quad \bar{P} = \frac{1}{N} \sum_{i=1}^{N} P_i
\]  
(A1)

2. Fractional Bias (FB)

\[
\text{FB} = \frac{\bar{P} - \bar{O}}{0.5(\bar{P} + \bar{O})}
\]  
(A2)

3. Mean Normalized Bias (NMB)

\[
\text{MNB} = \frac{1}{N} \sum_{i=1}^{N} \left( \frac{P_i - O_i}{O_i} \right)
\]  
(A3)

976
4. Mean Normalized Error (MNE)

\[
\text{MNE} = \frac{1}{N} \sum_{i=1}^{N} \left( \frac{|P_i - O_i|}{O_i} \right)
\]  

(A4)

5. Normalized Mean Error (NME)

\[
\text{NME} = \frac{|\bar{P} - \bar{O}|}{\bar{O}}
\]  

(A5)

6. Standard Deviation

\[
\sigma_0 = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (O_i - \bar{O})^2}
\]  

(A6)

7. Correlation coefficient

\[
r = \frac{1}{N} \sum_{i=1}^{N} (O_i - \bar{O})(P_i - \bar{P})
\]

\[
\sigma_0 \sigma_P
\]  

(A7)

8. Index of Agreement (IOA)

\[
\text{IOA} = 1 - \frac{\sum_{i=1}^{N} (P_i - O_i)^2}{\sum_{i=1}^{N} (|P_i - \bar{P}| + |O_i - \bar{O}|)^2}
\]  

(A8)

9. Factor of 2 (F2)

\[
\text{FAC2} = \frac{1}{N} \sum_{i=1}^{N} n_i \quad \text{with} \quad n_i = 1 \quad \text{for} \quad 0.5 < \left| \frac{P_i}{O_i} \right| \leq 2
\]  

(A9)
Appendix B  Abbreviations

CLC  Corine Air Land Cover database
CMAQ  Community Modelling Air Quality
CMAS  Community Modelling Air quality System
CORINAIR  Core Inventory of Air emissions
CLM  Climate version of the Lokal Model
CTM  Chemical Transport Model
DCW  Digital Chart of the World
DMA  Defense Mapping Agency
EMC  Environmental Modelling Center (USA)
EMEP  European Monitoring and Evaluation Program
EEA  European Environmental Agency
EPA  Environmental Protection Agency (USA)
EPER  European Pollutants Emission Register
ESRI  Environmental Systems Research Institute
EU15  European Union 15 Member states
EU27  European Union 27 Member states
EUROSTAT  European Statistical Service
FIPS  US Federal Implementation Planning Standards
GEMS  Global and regional Earth-system Monitoring using Satellite and in-situ data
GLC  Global Land Cover database
GPS  Global Positioning System
GPW  Gridded Population of the World
HM  Heavy Metals
IER  Institute for Rational use of Energy
LRTAP  Convention on Long-Range Transport of Air Pollutants
MM5  The Fifth-Generation NCAR/Penn State Mesoscale Meteorological Model
MSC-W  Meteorological Synthesizing Center – West
NMVOC  Non-Methane Volatile Organic Compounds
Appendix C  Short description of SMOKE and BEIS3 core modules

SMKINVEN: Reads in the raw input data, sorts the records, and creates the SMOKE inventory files that are required by most of the SMOKE programs.

GRDMAT: Reads the surrogate files and produces the matrix that contains the factors for spatially allocating the emission sources to the modeling domain.

SPCMAT: Calculates the matrices containing split factors for the specie speciation

CNTLMAT: The Cntlmat program uses control packets to create a growth matrix, and/or a multiplicative control matrix, and/or a reactivity control matrix
TEMPORAL: Reads the temporal profiles and produces a file with hourly inventory pollutant emissions. Unlike the SMOKE matrices produced by Cntlmat, Grdmat, and Spcmat, the output file from Temporal contains the actual emissions data.

ELEVPOINT: Select elevated point sources and to prepare certain input files for special elevated source or PinG processing.

LAYPOINT: Uses the SMOKE point-source inventory file with gridded and hourly meteorology data to compute hourly plume rise for all point sources. The plume rise is expressed in terms of layer fractions for each source.

SMKMERGE: Combines the matrices produced by the other SMOKE programs to produce the emissions files for input to the AQM. The Smkmerge program may be run on any combination of source types and may incorporate temporal, speciation, projection, and spatial processing.

NORMBEIS: Reads gridded land use data and emissions factors and produces gridded normalized biogenic emissions.

METSCAN: Determines winter and summer seasons depending on surface temperature.

TMPBEIS3: Uses temperature, surface pressure and radiation data from meteorological files to calculate hourly biogenic emissions.

Supplementary material related to this article is available online at: http://www.geosci-model-dev-discuss.net/3/949/2010/gmdd-3-949-2010-supplement.pdf.

Acknowledgements

US EPA is gratefully acknowledged for the use of SMOKE and CMAQ. We are thankful to Beate Geyer for providing the CLM meteorological fields and to Hugo Denier van der Gon for the GEMS emission data. We want to thank EMEP and IER Stuttgart for providing emission data. Finally, the authors want to thank the numerous open source online databases used for this publication.
References


Digital chart of the world, http://www.maproom.psu.edu/dcw/, last access: 1 January 2010.


Table 1. SNAP: Selected Nomenclature for sources of Air Pollution.

<table>
<thead>
<tr>
<th>Sector</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>SNAP 1</td>
<td>Combustion in energy and transformation industries</td>
</tr>
<tr>
<td>SNAP 2</td>
<td>Non-industrial combustion plants</td>
</tr>
<tr>
<td>SNAP 3</td>
<td>Combustion in manufacturing industry</td>
</tr>
<tr>
<td>SNAP 4</td>
<td>Production processes</td>
</tr>
<tr>
<td>SNAP 5</td>
<td>Extraction and distribution of fossil fuels and geothermal energy</td>
</tr>
<tr>
<td>SNAP 6</td>
<td>Solvent use and other product use</td>
</tr>
<tr>
<td>SNAP 7</td>
<td>Road transport</td>
</tr>
<tr>
<td>SNAP 8</td>
<td>Other mobile sources and machinery</td>
</tr>
<tr>
<td>SNAP 9</td>
<td>Waste treatment and disposal</td>
</tr>
<tr>
<td>SNAP 10</td>
<td>Agriculture</td>
</tr>
<tr>
<td>SNAP 11</td>
<td>Other sources and sinks</td>
</tr>
</tbody>
</table>

986
**Table 2.** Spatial surrogates used for different SNAP sectors and biogenic emissions. A list of abbreviations can be found in Appendix B.

<table>
<thead>
<tr>
<th>Sector</th>
<th>Datasets used for spatial disaggregation</th>
</tr>
</thead>
<tbody>
<tr>
<td>SNAP 1</td>
<td>EPER, CLC (commercial and industrial units), GLC (urban area), GPWv3</td>
</tr>
<tr>
<td>SNAP 2</td>
<td>GPWv3, 2 m temperature</td>
</tr>
<tr>
<td>SNAP 3</td>
<td>EPER, CLC (commercial and industrial units), GLC (urban area), EUROSTAT (employees in industry), GPWv3</td>
</tr>
<tr>
<td>SNAP 4</td>
<td>EPER, CLC (commercial and industrial units), GLC (urban area), EUROSTAT (employees in industry), GPWv3</td>
</tr>
<tr>
<td>SNAP 5</td>
<td>EPER, CLC (ports), GPWv3</td>
</tr>
<tr>
<td>SNAP 6</td>
<td>GPWv3</td>
</tr>
<tr>
<td>SNAP 7</td>
<td>TREMOVE, OSM and DCW (motorways, roads), CLC (urban area), GLC (urban area)</td>
</tr>
<tr>
<td>SNAP 8</td>
<td>TREMOVE, CLC and GLC (airports, agricultural areas), OSM and DCW (railways, waterways, roads)</td>
</tr>
<tr>
<td>SNAP 9</td>
<td>CLC (dump sites), GPWv3</td>
</tr>
<tr>
<td>SNAP 10</td>
<td>CLC (agricultural areas, pastures), GLC (agricultural areas), EUROSTAT (employees in agriculture, animal stocks)</td>
</tr>
<tr>
<td>Biogenic</td>
<td>GsfM (Tree distribution), CLC (land use), GLC (land use)</td>
</tr>
</tbody>
</table>
**Table 3.** NUTS level definition.

<table>
<thead>
<tr>
<th>NUTS 1</th>
<th>3 million–7 million inhabitants</th>
</tr>
</thead>
<tbody>
<tr>
<td>NUTS 2</td>
<td>800 000–3 million inhabitants</td>
</tr>
<tr>
<td>NUTS 3</td>
<td>150 000–800 000 inhabitants</td>
</tr>
</tbody>
</table>
Table 4. EMEP measurement stations used for comparison with modelled air concentrations.

<table>
<thead>
<tr>
<th>ID</th>
<th>Name</th>
<th>Longitude</th>
<th>Latitude</th>
<th>Altitude</th>
<th>O&lt;sub&gt;3&lt;/sub&gt;</th>
<th>NO&lt;sub&gt;2&lt;/sub&gt;</th>
<th>SO&lt;sub&gt;2&lt;/sub&gt;</th>
<th>SO&lt;sub&gt;4&lt;/sub&gt;</th>
<th>NO&lt;sub&gt;x&lt;/sub&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>AT02R</td>
<td>Illmitz</td>
<td>47 46 0 N</td>
<td>16 46 0 E</td>
<td>117</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>CH02R</td>
<td>Payrnes</td>
<td>46 48 47 N</td>
<td>6 56 41 E</td>
<td>489</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>CH03R</td>
<td>Tänikon</td>
<td>47 28 47 N</td>
<td>8 54 17 E</td>
<td>539</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>CZ03R</td>
<td>Kosetice</td>
<td>49 35 0 N</td>
<td>15 50 0 E</td>
<td>534</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>DE01R</td>
<td>Westerland</td>
<td>54 55 32 N</td>
<td>8 18 35 E</td>
<td>12</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>DE02R</td>
<td>Langenbrügge</td>
<td>52 48 8 N</td>
<td>10 45 34 E</td>
<td>74</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>DE04R</td>
<td>Deuselbach</td>
<td>49 45 53 N</td>
<td>7 37 7 E</td>
<td>480</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>DE07R</td>
<td>Neuglobsow</td>
<td>53 10 0 N</td>
<td>13 2 0 E</td>
<td>62</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>DE09R</td>
<td>Zingst</td>
<td>54 26 0 N</td>
<td>12 44 0 E</td>
<td>1</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>DK03R</td>
<td>Tange</td>
<td>56 21 0 N</td>
<td>9 36 0 E</td>
<td>13</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>DK05R</td>
<td>Keldsnor</td>
<td>54 44 0 N</td>
<td>10 44 0 E</td>
<td>10</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>DK08R</td>
<td>Anholt</td>
<td>56 43 0 N</td>
<td>11 31 0 E</td>
<td>40</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>EE09R</td>
<td>Lahemaa</td>
<td>59 30 0 N</td>
<td>25 54 0 E</td>
<td>32</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>ES03R</td>
<td>Roquetas</td>
<td>40 49 14 N</td>
<td>0 29 29 E</td>
<td>44</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>ES04R</td>
<td>Logroño</td>
<td>42 27 28 N</td>
<td>2 30 11 W</td>
<td>445</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>ES08R</td>
<td>Niembro</td>
<td>43 26 32 N</td>
<td>4 51 1 W</td>
<td>134</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>ES10R</td>
<td>Cabo de Creus</td>
<td>42 19 10 N</td>
<td>3 19 1 E</td>
<td>23</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>ES11R</td>
<td>Barcarrola</td>
<td>38 28 33 N</td>
<td>6 55 22 W</td>
<td>393</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>FI09R</td>
<td>Utö</td>
<td>59 46 45 N</td>
<td>21 22 38 E</td>
<td>7</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>FI17R</td>
<td>Virolahti II</td>
<td>60 31 36 N</td>
<td>27 41 0 E</td>
<td>4</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>FI22R</td>
<td>Oulanka</td>
<td>66 19 13 N</td>
<td>29 24 6 E</td>
<td>310</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>FI37R</td>
<td>Ahtari II</td>
<td>62 35 0 N</td>
<td>24 11 0 E</td>
<td>180</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>FR03R</td>
<td>La Crouzille</td>
<td>45 50 N</td>
<td>1 16 0 E</td>
<td>497</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>FR05R</td>
<td>La Hague</td>
<td>49 37 0 N</td>
<td>1 49 59 W</td>
<td>133</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>FR09R</td>
<td>Revin</td>
<td>49 54 0 N</td>
<td>4 38 0 E</td>
<td>390</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>FR13R</td>
<td>Peyrusse Vieille</td>
<td>43 37 0 N</td>
<td>0 11 0 E</td>
<td>2</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>GB02R</td>
<td>Eskdalemuir</td>
<td>55 18 47 N</td>
<td>3 12 15 W</td>
<td>243</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>GB04R</td>
<td>Stoke Ferry</td>
<td>52 34 0 N</td>
<td>0 30 0 E</td>
<td>15</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>GB06R</td>
<td>Lough Navar</td>
<td>54 26 35 N</td>
<td>7 52 12 W</td>
<td>126</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>GB07R</td>
<td>Barcombe Mills</td>
<td>50 52 0 N</td>
<td>0 15 9 W</td>
<td>8</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>GB13R</td>
<td>Yarner Wood</td>
<td>50 35 47 N</td>
<td>3 42 47 W</td>
<td>119</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>GB14R</td>
<td>High Muffles</td>
<td>54 20 4 N</td>
<td>0 48 27 W</td>
<td>267</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>
### Table 4. Continued.

<table>
<thead>
<tr>
<th>ID</th>
<th>Name</th>
<th>Longitude</th>
<th>Latitude</th>
<th>Altitude</th>
<th>O₃</th>
<th>NO₂</th>
<th>SO₂</th>
<th>SO₄</th>
<th>NO₃</th>
</tr>
</thead>
<tbody>
<tr>
<td>GB15R</td>
<td>Strath Vaich Dam</td>
<td>57 44 4 N</td>
<td>4 46 28 W</td>
<td>270</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GB16R</td>
<td>Glen Dye</td>
<td>56 58 0 N</td>
<td>2 25 0 W</td>
<td>85</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GB39R</td>
<td>Sibton</td>
<td>52 17 38 N</td>
<td>1 27 47 E</td>
<td>46</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GR01R</td>
<td>Aliartos</td>
<td>38 22 0 N</td>
<td>23 5 0 E</td>
<td>110</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HU02R</td>
<td>K-pusztasz</td>
<td>46 58 0 N</td>
<td>19 35 0 E</td>
<td>125</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>IE02R</td>
<td>Turlough Hill</td>
<td>53 2 12 N</td>
<td>6 24 0 W</td>
<td>420</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IE31R</td>
<td>Mace Head</td>
<td>53 10 0 N</td>
<td>9 30 0 W</td>
<td>15</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IS02R</td>
<td>Irafo</td>
<td>64 5 0 N</td>
<td>21 1 0 W</td>
<td>66</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IT01R</td>
<td>Montelibretti</td>
<td>42 6 0 N</td>
<td>12 38 0 E</td>
<td>48</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>IT04R</td>
<td>Ispra</td>
<td>45 48 0 N</td>
<td>8 38 0 E</td>
<td>209</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>LT15R</td>
<td>Preila</td>
<td>55 21 0 N</td>
<td>21 4 0 E</td>
<td>5</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>LV10R</td>
<td>Rucava</td>
<td>56 13 0 N</td>
<td>21 13 0 E</td>
<td>5</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>LV16R</td>
<td>Zoseni</td>
<td>57 8 0 N</td>
<td>25 55 0 E</td>
<td>183</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NL09R</td>
<td>Kollumerwaard</td>
<td>53 20 2 N</td>
<td>6 16 38 E</td>
<td>1</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NL10R</td>
<td>Vredepeel</td>
<td>51 32 28 N</td>
<td>5 51 13 E</td>
<td>28</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NO01R</td>
<td>Birkenes</td>
<td>58 23 0 N</td>
<td>8 15 0 E</td>
<td>190</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>NO08R</td>
<td>Skreádal</td>
<td>58 49 0 N</td>
<td>6 43 0 E</td>
<td>475</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NO15R</td>
<td>Tustervatn</td>
<td>65 50 0 N</td>
<td>13 55 0 E</td>
<td>439</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NO39R</td>
<td>Kárvatn</td>
<td>62 47 0 N</td>
<td>8 53 0 E</td>
<td>210</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NO41R</td>
<td>Osen</td>
<td>61 15 0 N</td>
<td>11 47 0 E</td>
<td>440</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NO42G</td>
<td>Spitsbergen, Zeppelinfjell</td>
<td>78 54 0 N</td>
<td>11 53 0 E</td>
<td>474</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NO55R</td>
<td>Karasjok</td>
<td>69 28 0 N</td>
<td>25 13 0 E</td>
<td>333</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>PL02R</td>
<td>Jarczew</td>
<td>51 49 0 N</td>
<td>21 59 0 E</td>
<td>180</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>PL04R</td>
<td>Leba</td>
<td>54 45 0 N</td>
<td>17 32 0 E</td>
<td>2</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PL05R</td>
<td>Diabla Gora</td>
<td>54 9 0 N</td>
<td>22 4 0 E</td>
<td>157</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PT04R</td>
<td>Monte Velho</td>
<td>38 5 0 N</td>
<td>8 48 0 W</td>
<td>43</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RU01R</td>
<td>Janiskoski</td>
<td>68 56 0 N</td>
<td>28 51 0 E</td>
<td>118</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RU18R</td>
<td>Danki</td>
<td>54 54 0 N</td>
<td>37 48 0 E</td>
<td>150</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SE02R</td>
<td>Rörvik</td>
<td>57 25 0 N</td>
<td>11 56 0 E</td>
<td>10</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SE11R</td>
<td>Vavihill</td>
<td>56 1 0 N</td>
<td>13 9 0 E</td>
<td>175</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SE12R</td>
<td>Aspvretten</td>
<td>58 48 0 N</td>
<td>17 23 0 E</td>
<td>20</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SE13R</td>
<td>Esrange</td>
<td>67 53 0 N</td>
<td>21 4 0 E</td>
<td>475</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SE32R</td>
<td>Norra-Kvill</td>
<td>57 49 0 N</td>
<td>15 34 0 E</td>
<td>261</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 5. Statistical comparison of CMAQ results using four different emission datasets. Values are averages over all measurement stations and their standard deviations. For more detailed results see Figs. 12–15.

<table>
<thead>
<tr>
<th></th>
<th>EMEP</th>
<th>TNO-GEMS</th>
<th>IER-GKSS</th>
<th>SMOKE-EU</th>
<th>OBSERVATION</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(O_3) – 40 stations ((N=329,197))</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MEAN</td>
<td>77.43±6.05</td>
<td>74.58±6.26</td>
<td>75.86±6.08</td>
<td>78.25±6.44</td>
<td>57.79±6.76</td>
</tr>
<tr>
<td>FB</td>
<td>0.29±0.13</td>
<td>0.26±0.13</td>
<td>0.27±0.13</td>
<td>0.3±0.14</td>
<td>–</td>
</tr>
<tr>
<td>NME</td>
<td>0.36±0.19</td>
<td>0.31±0.18</td>
<td>0.33±0.18</td>
<td>0.37±0.19</td>
<td>–</td>
</tr>
<tr>
<td>FAC2</td>
<td>0.79±0.79</td>
<td>0.79±0.79</td>
<td></td>
<td></td>
<td>–</td>
</tr>
<tr>
<td>CORR</td>
<td>0.62±0.08</td>
<td>0.61±0.06</td>
<td>0.62±0.07</td>
<td>0.63±0.08</td>
<td>–</td>
</tr>
<tr>
<td>IOA</td>
<td>0.45±0.28</td>
<td>0.47±0.25</td>
<td>0.46±0.27</td>
<td>0.47±0.26</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>(NO_2) – 33 stations ((N=11,465))</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MEAN</td>
<td>1.31±1.01</td>
<td>1.37±1.48</td>
<td>1.33±1.23</td>
<td>1.57±1.32</td>
<td>2.31±1.74</td>
</tr>
<tr>
<td>FB</td>
<td>−0.47±0.46</td>
<td>−0.51±0.53</td>
<td>−0.49±0.46</td>
<td>−0.28±0.48</td>
<td>–</td>
</tr>
<tr>
<td>NME</td>
<td>0.37±0.25</td>
<td>0.44±0.27</td>
<td>0.38±0.25</td>
<td>0.33±0.23</td>
<td>–</td>
</tr>
<tr>
<td>FAC2</td>
<td>0.46</td>
<td>0.46</td>
<td>0.45</td>
<td>0.55</td>
<td>–</td>
</tr>
<tr>
<td>CORR</td>
<td>0.44±0.31</td>
<td>0.42±0.3</td>
<td>0.45±0.3</td>
<td>0.45±0.3</td>
<td>–</td>
</tr>
<tr>
<td>IOA</td>
<td>0.41±0.41</td>
<td>0.35±0.4</td>
<td>0.41±0.37</td>
<td>0.48±0.33</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>(SO_2) – 36 stations ((N=12,430))</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MEAN</td>
<td>0.98±0.83</td>
<td>0.98±1.03</td>
<td>1.09±1.3</td>
<td>1.27±1.2</td>
<td>0.78±0.63</td>
</tr>
<tr>
<td>FB</td>
<td>0.21±0.71</td>
<td>0.09±0.65</td>
<td>0.1±0.72</td>
<td>0.34±0.73</td>
<td>–</td>
</tr>
<tr>
<td>NME</td>
<td>0.8±0.65</td>
<td>0.63±0.58</td>
<td>0.7±0.56</td>
<td>1.03±0.82</td>
<td>–</td>
</tr>
<tr>
<td>FAC2</td>
<td>0.46</td>
<td>0.44</td>
<td>0.44</td>
<td>0.44</td>
<td>–</td>
</tr>
<tr>
<td>CORR</td>
<td>0.4±0.23</td>
<td>0.38±0.23</td>
<td>0.38±0.25</td>
<td>0.4±0.23</td>
<td>–</td>
</tr>
<tr>
<td>IOA</td>
<td>0.42±0.26</td>
<td>0.43±0.25</td>
<td>0.42±0.27</td>
<td>0.37±0.25</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>(SO_2) – 51 stations ((N=17,536))</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MEAN</td>
<td>0.61±0.18</td>
<td>0.57±0.18</td>
<td>0.54±0.17</td>
<td>0.66±0.21</td>
<td>0.71±0.42</td>
</tr>
<tr>
<td>FB</td>
<td>−0.02±0.4</td>
<td>−0.08±0.41</td>
<td>−0.13±0.4</td>
<td>0.06±0.38</td>
<td>–</td>
</tr>
<tr>
<td>NME</td>
<td>0.35±0.33</td>
<td>0.34±0.32</td>
<td>0.33±0.27</td>
<td>0.36±0.38</td>
<td>–</td>
</tr>
<tr>
<td>FAC2</td>
<td>0.61</td>
<td>0.59</td>
<td>0.59</td>
<td>0.62</td>
<td>–</td>
</tr>
<tr>
<td>CORR</td>
<td>0.44±0.16</td>
<td>0.39±0.15</td>
<td>0.42±0.15</td>
<td>0.45±0.16</td>
<td>–</td>
</tr>
<tr>
<td>IOA</td>
<td>0.49±0.26</td>
<td>0.43±0.27</td>
<td>0.44±0.24</td>
<td>0.51±0.26</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>(NH_3) – 22 stations ((N=7,400))</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MEAN</td>
<td>1.28±0.77</td>
<td>1.05±0.6</td>
<td>1.03±0.64</td>
<td>1.44±0.9</td>
<td>0.75±0.78</td>
</tr>
<tr>
<td>FB</td>
<td>0.74±0.45</td>
<td>0.59±0.47</td>
<td>0.57±0.47</td>
<td>0.83±0.41</td>
<td>–</td>
</tr>
<tr>
<td>NME</td>
<td>1.62±1.31</td>
<td>1.24±1.1</td>
<td>1.2±1.15</td>
<td>1.84±1.38</td>
<td>–</td>
</tr>
<tr>
<td>FAC2</td>
<td>0.37</td>
<td>0.4</td>
<td>0.4</td>
<td>0.4</td>
<td>–</td>
</tr>
<tr>
<td>CORR</td>
<td>0.46±0.17</td>
<td>0.38±0.21</td>
<td>0.45±0.18</td>
<td>0.46±0.18</td>
<td>–</td>
</tr>
<tr>
<td>IOA</td>
<td>0.14±0.7</td>
<td>0.25±0.59</td>
<td>0.25±0.58</td>
<td>0.09±0.63</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>(NO_3) – 18 stations ((N=6,184))</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MEAN</td>
<td>0.47±0.41</td>
<td>0.3±0.24</td>
<td>0.32±0.31</td>
<td>0.51±0.46</td>
<td>0.41±0.54</td>
</tr>
<tr>
<td>FB</td>
<td>0.05±0.79</td>
<td>−0.2±0.79</td>
<td>−0.18±0.67</td>
<td>0.13±0.75</td>
<td>–</td>
</tr>
<tr>
<td>NME</td>
<td>0.78±0.76</td>
<td>0.58±0.42</td>
<td>0.37±0.32</td>
<td>0.81±1.02</td>
<td>–</td>
</tr>
<tr>
<td>FAC2</td>
<td>0.25</td>
<td>0.18</td>
<td>0.22</td>
<td>0.25</td>
<td>–</td>
</tr>
<tr>
<td>CORR</td>
<td>0.32±0.27</td>
<td>0.26±0.21</td>
<td>0.32±0.26</td>
<td>0.32±0.27</td>
<td>–</td>
</tr>
<tr>
<td>IOA</td>
<td>0.29±0.34</td>
<td>0.27±0.32</td>
<td>0.34±0.34</td>
<td>0.28±0.25</td>
<td>–</td>
</tr>
</tbody>
</table>
Table 6. Comparison of mean concentrations of SO$_4$ and SO$_2$ with and without vertical distribution. Values are averages over all measurement stations (51 stations for SO$_4$, 33 stations for SO$_2$) and their standard deviations.

<table>
<thead>
<tr>
<th></th>
<th>EMEP 3-D</th>
<th>EMEP 2-D</th>
<th>TNO-GEMS 3-D</th>
<th>TNO-GEMS 2-D</th>
</tr>
</thead>
<tbody>
<tr>
<td>SO$_4$</td>
<td>0.61±0.18</td>
<td>0.58±0.16</td>
<td>0.55±0.19</td>
<td>0.54±0.16</td>
</tr>
<tr>
<td>SO$_2$</td>
<td>0.98±0.83</td>
<td>1.2±1.18</td>
<td>0.99±1.03</td>
<td>1.06±1.2</td>
</tr>
</tbody>
</table>
Fig. 1. SMOKE and BEIS3 (green) core programs including modifications for SMOKE EUROPE (blue). Short descriptions of the most important modules can be found in Appendix C (Baek et al., 2009).
Fig. 2. (a) Comparison of temperature dependent temporal profiles SMOKE default vs. modified version. (b) Inter annual comparison of temperature dependent CO temporal profiles (all values averaged over the whole domain).
Fig. 3. Annual total anthropogenic emissions of the EU27 in 2000 (The SMOKE-EU dataset also includes 18 000 Gg/a biogenic NMVOC emissions).
Fig. 4. Relative annual total emissions of the EU27 (biogenic emissions are not included).
Fig. 5. Frequency distribution of (a) SO$_2$ and (b) NH$_3$ emissions (annual total emissions of EU27).
Fig. 6. Frequency distribution of (a) \( \text{SO}_2 \) and (b) \( \text{NH}_3 \) emissions (annual total emissions of EU27).
Fig. 7. Modelling domain used for CTM calculations with 54×54 km² grid resolution.
Fig. 8. (a) Spatial average annual emissions of SO$_2$. (b) Variograms for SO$_2$. All values for concentric circles with 100 km distance.
Fig. 9. (a) Spatial average annual emissions of NH$_3$. (b) Variograms for NH$_3$. All values for concentric circles with 100 km distance.
Fig. 10. Averaged annual temporal profiles with daily resolution of (a) NO and (b) NH$_3$. The biogenic NO emissions included in the SMOKE-EU dataset lead to higher average emission in summer and lower average emissions in winter.
Fig. 11. Average vertical distribution of (a) SO₂ and (b) NO emissions. For comparison with the SMOKE-EU dataset, the official EMEP vertical profiles were interpolated from 6 to 30 layers. The TNO-GEMS dataset uses the EMEP vertical distributions.
Fig. 12. Comparison of modelled $O_3$ concentrations with hourly observations from 40 rural EMEP measurement sites ($N=329,197$). See also Tables 6 and 7. (a) Fractional bias, (b) index of agreement, (c) percentage of values within a factor of 2.
Fig. 13. Comparison of modelled SO$_4$ concentrations with hourly observations from 51 rural EMEP measurement sites ($N=17\,536$) See also Tables 6 and 7. (a) Fractional bias, (b) index of agreement, (c) percentage of values within a factor of 2.
Fig. 14. Taylor diagrams showing correlation and fractional bias of modelled atmospheric concentrations of O₃ (a) and SO₄ (b) compared to observations. Different shapes indicate the 4 emission datasets used, while colors indicate geographical regions.
Fig. 15. Fractional bias and observed concentrations for \( \text{NH}_4 \), \( \text{NO}_3 \) and \( \text{NO}_2 \).