Geosci. Model Dev. Discuss., 3, 949–1007, 2010
www.geosci-model-dev-discuss.net/3/949/2010/
doi:10.5194/gmdd-3-949-2010

Geoscientific Model Development Discussions

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

J. Bieser[1,2], A. Aulinger[1], V. Matthias[1], M. Quante[1,2], and P. Builtjes[3,4]

[1]GKSS Research Center, Institute for Coastal Research, Max-Planckstr. 1, 21502 Geeshacht, Germany
[2]Leuphana University Lüneburg, Institute of Ecology and Environmental Chemistry, Scharnhorststraße 1, 21335 Lüneburg, Germany
[3]Institut fur Meteorologie, Freie Universität Berlin, Carl-Heinrich-Becker Weg 6-10, 12165 Berlin, Germany
[4]TNO 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 O_3, NO_2 and SO_4 could be modelled most reliably. The amount of simulated concentrations within a factor of 2 (F2) of the observations for these species are: O_3 (F2=0.79 N=329 197), NO_2 (F2=0.55 N=11 465), and SO_4 (F2=0.62 N=17 536). The lowest values were found for NH_4 (F2=0.34 N=7400) and NO_3 (F2=0.25 N=6184). NH_4 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 SO_4 (FB=0.06±0.38, 51 stations) and NO_3 (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 where found for NH_4 while O_3 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.

GMDD

3, 949–1007, 2010

SMOKE for Europe – a comprehensive emission model for Europe

J. Bieser et al.

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 knowlegde 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 seperated 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

GMDD
3, 949–1007, 2010

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 $50 \times 50\,km^2$. 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.

GMDD
3, 949–1007, 2010

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 Unites States Environmental Protection Agency (US EPA) and is one of the most used emission models world wide (Houyoux et al., 2000; MCNC Environmental Modelling Center, 2008; UNC Carolina Environmental Program, 2005). SMOKE was originally created by the MCNC Environmental Modeling 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

GMDD
3, 949–1007, 2010

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,

GMDD
3, 949–1007, 2010

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.

$$\sum_{i=1}^{N} \mathbf{GRDMAT}(i) = 1 \qquad N = \text{number of grid cells} \tag{1}$$

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

$$\sum_{i=1}^{N} \mathbf{MODMAT}(i,j)\varepsilon[0,T] \tag{3}$$

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

GMDD
3, 949–1007, 2010

dependent the gridding matrix is calculated only once for each year (Eq. 4).

$$E(t,x,y,z) = \mathbf{GRDMAT}(x,y) \cdot \mathbf{SPCMAT}(x,y) \cdot \mathbf{PiG}(z) \cdot \mathbf{TMPFAC}(t) \cdot TOT \quad (4)$$

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 **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).

$$\mathbf{MODMAT}(n,t) = \frac{\mathbf{GRDMAT}(n) \cdot \mathbf{CHGMAT}(n,t) \cdot T \cdot \sum_{i=1}^{N} [\mathbf{GRDMAT}(i)]}{\mathbf{NORMAT}(n)} \quad (5)$$

$$\mathbf{NORMAT}(n) = \sum_{i=1}^{N} \left[\mathbf{GRDMAT}(i) \cdot \sum_{j=1}^{T} [\mathbf{CHGMAT}(i,j)] \right] \quad (6)$$

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.

Altough 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.

GMDD
3, 949–1007, 2010

2.2 Emission Inventories

European emission inventories and datasets are quite inhomogeneous. Most countries use different methodologies to asses 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_x, SO_2, NH_3, Non-Methane Volatile Organic Compounds (NMVOC), primary particulate matter (PM) as PM_{10} and $PM_{2.5}$, 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.

GMDD
3, 949–1007, 2010

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

GMDD
3, 949–1007, 2010

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

GMDD
3, 949–1007, 2010

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 $2.5' \times 2.5'$ 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 $100 \times 100\,m^2$. 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 $1 \times 1\,km^2$ 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).

GMDD
3, 949–1007, 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-

GMDD
3, 949–1007, 2010

tics (NUTS) geocode standard which is the official European system for referencing subdivitions 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

GMDD
3, 949–1007, 2010

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 *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.

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 *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 *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.

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). Besides 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

GMDD
3, 949–1007, 2010

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_{10}) and particles smaller than 2.5 µm ($PM_{2.5}$). For CMAQ $PM_{2.5}$ 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 it's 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 photochemical 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 disaggregation 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).

GMDD
3, 949–1007, 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 seperatly for the 6 inventory species: NO_x, SO_2, CO, PM_{10}, NH_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_x, SO_2, NH_3, PM_{10}, $PM_{2.5}$ 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 $50 \times 50\,km^2$ domain together with SNAP specific vertical profiles are published by EMEP (Webdab, 2010). For comparison with SMOKE emissions the EMEP dataset was interpolated to $54 \times 54\,km^2$. The LOTOS-EUROS temporal profiles have been used for temporal disaggregation.

IER-GKSS: An emission dataset for the GKSS $54 \times 54\,km^2$ 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

GMDD
3, 949–1007, 2010

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 $54 \times 54\,km^2$.

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

GMDD
3, 949–1007, 2010

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 SO_2, 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: NO_x (12 500 Gg±6.8%), SO_2 (10 600 Gg±9.1%), CO (38 900 Gg±16.7%), PM_{10} (2830 Gg±7.1%), NH_3 (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 NH_3 emissions are 30% lower than the EMEP values. The SMOKE-EU emissions are somewhat

GMDD
3, 949–1007, 2010

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_x, SO_2, PM_{10}, NH_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_2 emissions show the best agreement for all four datasets. This is due to the fact that SO_2 emissions are well known concerning the total amount emitted as well as their spatial and temporal distribution. NH_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_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

GMDD
3, 949–1007, 2010

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 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

$$f(h) = \frac{(z(x+h) - z(x))^2}{2} \qquad x = \text{reference grid cell; } h = \text{distance to origin} \tag{7}$$

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 *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 SO_2 follow a similar pattern (Fig. 8). Still some differences can be seen. Looking at the variograms for 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 SO_2 emissions in the 500–700 km circles (30–40%) for the IER-GKSS datasets. This indicates that the ~8% higher total SO_2 emissions in this dataset are due to higher emissions in a certain area rather than a general overestimation (Fig. 3).

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

GMDD
3, 949–1007, 2010

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 $54 \times 54\,km^2$ grid are quite similar for all four emission datasets, while looking at certain grid cells reveals some obvious differences. Thus, spatial variablity 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.

GMDD
3, 949–1007, 2010

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_x: 400–600 m 30%>600 m 20%) (NO_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^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.

GMDD
3, 949–1007, 2010

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 (NO_2, SO_2, O_3) and three aerosol components (SO_4, NH_4, NO_3). 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 $54 \times 54\,km^2$ 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 NO_x emissions in the SMOKE-EU and EMEP datasets potentially change the ozone regime in certain regions from VOC limited to NO_x limited (Fig. 11b). However, since O_3 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

GMDD
3, 949–1007, 2010

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

GMDD
3, 949–1007, 2010

regional analysis would go beyond the scope of this paper and will be further discussed elsewhere.

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

GMDD
3, 949–1007, 2010

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 $70\times70\,km^2$ down to $10\times10\,km^2$. The underlying datasets allow for a spatial resolution of up to $1\times1\,km^2$ (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 $54\times54\,km^2$ 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

GMDD
3, 949–1007, 2010

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 = Predicted value from Model
O_i = Observed value from Measurement
N = Sample size

1. Mean

$$\bar{O} = \frac{1}{N}\sum_{i=1}^{N} O_i\,, \quad \bar{P} = \frac{1}{N}\sum_{i=1}^{N} P_i \tag{A1}$$

2. Fractional Bias (FB)

$$\mathrm{FB} = \frac{\bar{P} - \bar{O}}{0.5(\bar{P} + \bar{O})} \tag{A2}$$

3. Mean Normalized Bias (NMB)

$$\mathrm{MNB} = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{P_i - O_i}{O_i}\right) \tag{A3}$$

GMDD
3, 949–1007, 2010

4. Mean Normalized Error (MNE)

$$\mathrm{MNE} = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{|P_i - O_i|}{O_i}\right) \tag{A4}$$

5. Normalized Mean Error (NME)

$$\mathrm{NME} = \frac{\left|\bar{P} - \bar{O}\right|}{\bar{O}} \tag{A5}$$

6. Standard Deviation

$$\sigma_0 = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(O_i - \bar{O}\right)^2} \tag{A6}$$

7. Correlation coefficient

$$r = \frac{\frac{1}{N}\sum_{i=1}^{N}(O_i - \bar{O})(P_i - \bar{P})}{\sigma_0 \sigma_p} \tag{A7}$$

8. Index of Agreement (IOA)

$$\mathrm{IOA} = 1 - \frac{\sum_{i=1}^{N}(P_i - O_i)^2}{\sum_{i=1}^{N}(|P_i - \bar{P}| + |O_i - \bar{O}|)^2} \tag{A8}$$

9. Factor of 2 (F2)

$$\mathrm{FAC2} = \frac{1}{N}\sum_{i=1}^{N} n_i \quad \text{with} \quad n_i = 1 \quad \text{for} \quad 0.5 < \left|\frac{P_i}{O_i}\right| \leq 2 \tag{A9}$$

GMDD
3, 949–1007, 2010

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

GMDD
3, 949–1007, 2010

NACE	Nomenclature statistique des activités économiques dans la Communauté européenne
NUTS	Nomenclature of Units for Territorial Statistics
OMS	OpenStreetMaps
PiG	Plume in Grid
PM	Particulate Matter
$PM_{2.5}$	Particulate Matter smaller than 2.5 µm
PM_{10}	Particulate Matter smaller than 10 µm
POP	Persistent Organic Pollutants
RIVM	National Institute for Public Health and the Environment (NL)
SCC	Source Classification Code
SNAP	Selected Nomenclature for sources of Air Pollution
SMOKE-EU	SMOKE for Europe
SMOKE	Sparse Matrix Operator Kernel Emissions
TNO	Netherlands Organization for Applied Scientific Research (NL)
UCAR	University Cooperation for Atmospheric Research
UBA	Federal Environmental Agency (DE)
UNC	University of North Carolina
USGS	United States Geological Survey
VOC	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

GMDD
3, 949–1007, 2010

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.

GMDD
3, 949–1007, 2010

References

Andersson, C. and Langner, J.: Inter-annual variations of ozone and nitrogen dioxide over Europe during 1958–2003 simulated with a regional CTM, Water Air Soil Poll., 7, 15–23, 2005.

Agrawal, H., Malloy, Q. G. J., Welch, W. A., Miller, J. W., and Cocker, D. R.: In-use gaseous and particulate matter emissions from a modern ocean going container vessel, Technical note, Atmos. Environ., 42, 5504–5510, 2008.

Aulinger, A., Matthias, V., and Quante, M.: An approach to temporally disaggregate benzo(a)pyrene emissions and their application to a 3-D Eulerian atmospheric chemistry transport model, Water Air Soil Poll., submitted, 10 February 2010.

Baek, B. H., Seppanen, C., and Houyoux, M.: SMOKE v2.6 user's manual, http://www.smoke-model.org/version2.6/, last access: 1 May 2009.

Balk, D. and Yetman, G.: The global distribution of population: evaluating the gains in resolution refinement, Center for International Earth Science Information Network (CIESIN) Columbia University, Columbia, USA, 2004.

Benedictow, A., Fagerli, H., Gauss, M., Jonson, J. E., Nyiri, A., Simpson, D., Tsyro, S., Valdebenito, A., Valiyaveetil, S., Wind, P., Aas, W., Hjelbrekke, A., Marechova, K., Wankmueller, R., Harmens, H., Cooper, D., Norris, D., Schroeder, W., Pesch, R., and Holy, M: Transboundary acidification, eutrophication and ground level ozone in Europe in 2007, EMEP Status Report 2009, 16 July 2009, Norwegian Meteorological Institute (NMI), Oslo, Norway, ISSN 1504-6109 (print), ISSN 1504-6192 (on-line), 2009.

Benkovitz, M. C., Akimoto, H., Corbett, J. J., Mobley, J. D., Ohara, T., Olivier, J. G. J., van Ardenne, J. A., and Vestreng, V.: Compilation of Regional to Global Inventories of Anthropogenic Emissions, in: Emissions of Atmoshperic Trace Compounds, edited by: Granier, C., Artaxo, P., and Reeves, C. E., Kluwer Academic Publishers, Dordrecht, 17–69, 2004.

Briggs, G. A.: Discussion on chimney plumes in neutral and stable surroundings, Atmos. Environ. 6, 507–510, 1972.

Borge, R., Lumbreras, J., and Encarnacion, R.: Development of a high-resolution emission inventory for Spain using the SMOKE modelling system: a case study for the years 2000 and 2010, Environ. Modell. Softw., 23, 1026–1044, 2008.

Builtjes, P. J. H., van Loon, M., Schaap, M., Teeuwisse, S., Visschedijk, A. J. H., and Bloos, J. P.: Abschlussbericht zum FE-Vorhaben 298 41 252: "Modellierung und Prüfung von Strate-

GMDD
3, 949–1007, 2010

gien zur Verminderung der Belatung durch Ozon", Contribution of TNO-MEP, TNO-report R2003/166, 2003.
Byun, D. W. and Ching, J. K. S.: Science algorithms of the EPA Models-3 Community Multiscale Air Quality (CMAQ) modeling system, EPA/600/R-99/030, US EPA National Exposure Research Laboratory, Research Triangle Park, NC, 1999.
Byun, D. W. and Schere, K. L.: Review of the governing equations, computational algorithms, and other components of the Models-3 community Multiscale Air Quality (CMAQ) modeling system, Appl. Mech. Rev., 59(2), 51–77, 2006
Digital chart of the world, http://www.maproom.psu.edu/dcw/, last access: 1 January 2010.
Corine land cover, http://www.eea.europa.eu/data-and-maps/data/, last access: 1 March 2010.
EPA: Technology Transfer Network, Clearinghouse for Inventories & Emission Factors, http://www.epa.gov/ttn/chief/, last access: 1 March 2010.
European Commission: Commission decision 2000/479/EC of 17 July 2000 on the implementation of a European pollutant emission register (EPER) according to Article 15 of Council directive 96/61/EC concerning integrated pollution prevention and control (IPPC), Official Journal of the European Communities, EC, 1049 Brussel, Belgium, L. V192, 36–43, 2000.
European Commission: Regulation (EC) No 1059/2003 of the European Parliament and of the Council of 26 May 2003 on the establishment of a common classification of territorial units for statistics (NUTS), Official Journal of the European Communities, EC, 1049 Brussel, Belgium, L. V154, 1–41, 2003.
European Commission: TREMOVE service contract for the further development and application of the transport and environmental TREMOVE model Lot 1 (Improvement of the data set and model structure), EC, 1049 Brussel, Belgium, Final Report, Service Contract 070501/2005/420798/MAR/C1, 2007.
European Environmental Agency: EMEP/CORINAIR emission inventory guidebook, EC, 1049 Brussel, Belgium, Technical report No. 16/2007, 2007.
European Pollution Emission Register, http://www.eper.ec.europa.eu/eper/extract_data.asp, last access: 1 January 2010.
EUROSTAT, http://epp.eurostat.ec.europa.eu/portal/page/portal/statistics/search_databasehttp://epp.eurostat.ec.europa.eu/portal/page/portal/statistics/search_database, last access: 1 March 2010.

GMDD
3, 949–1007, 2010

Friedrich, R. and Reis, S.: Emissions of air pollutants, Springer, Berlin Heidelberg New York, 2004.
Gery, M. W., Whitten, G. Z., Killus, J. P., and Dodge, M. C.: A photochemical kinetics mechanism for urban and regional scale computer modeling, J. Geophys. Res., 94(D10), 12925–12956, 1989.
Guenther, A., Geron, C., Pierce, T., Lamb, B., Harley, P., and Fall, R.: Natural emissions of non-methane volatile organic compounds, carbon monoxide, and oxides of nitrogen from North America, Atmos. Environ., 34, 2205–2230, 2000.
Hanna, S. R. and Davis, J. M.: Uncertainties in predicted ozone concentrations due to input uncertainties for the UAM-V photochemical grid model applied to the July 1995 OTAG domain, Atmos. Environ., 35(5), 891–903, 2001.
Horowitz, L. W., Walters, S., Mauzerall, D. L., Emmons, L. K., Rasch, P. J., Granier, C., Tie, X., Lamarque, J.-F., Schultz, M. G., Tyndall, G. S., Orlando, J. J., and Brasseur, G. P.: A global simulation of tropospheric ozone and related tracers: description and evaluation of MOZART, version 2, J. Geophys. Res, 108, 4784, doi:10.1029/2002JD002853, 2003.
Houyoux, M. R.: Technical report: plume rise algorithm summary for the Sparse Matrix Operator Modeling System (SMOKE), Prepared for North Carolina Department of Environment and Natural Resources, UNC, Chapel Hill, North Carolina, ENV-98TR004eTR0v1.0, 1998.
Houyoux, M. R., Vukovich, J. M., Coats Jr., C. J., et al.: Emission inventory development and processing for the Seasonal Model for Regional Air Quality (SMRAQ) project, J. Geophys. Res., 105(D7), 9079–9090, 2000.
Maes, J., Vliegen, J., van de Vel, K., Janssen, S., Deutsch, F., de Ridder, K., and Mensink, C.: Spatial surrogates for the disaggregation of CORINAIR emission inventories, Atmos. Environ., 43, 1246–1254, 2009.
Matthias, V., Aulinger, A., and Quante, M.: Determination of the optimum MM5 configuration for long term CMAQ simulations of aerosol bound pollutants in Europe, Atmos. Environ., 9(1), 91–108, 2009.
MCNC-Environmental Modeling Center: Sparse Matrix Operational Kernel Emissions Modeling System, http://www.cmascenter.org/download/forms/step_1.cfm?CFID=2077553&CFTOKEN=19843259, last access: 8 May 2008.
Niemeier, U., Granier, C., Kornblueh, L., Walters, S., and Brasseur, G. P.: Global impact of road traffic on atmospheric chemical composition and on ozone climate forcing, J. Geophys. Res., 111, D09301, doi:10.1029/2005JD006407, 2006.

GMDD
3, 949–1007, 2010

Olivier, J. G. J. and Berdowski, J. J. M.: Global emissions sources and sinks, in: The Climate System, edited by: Berdowski, J., Guichert, R., and Heij, B. J., A. A. Balkema Publishers/Swets & Zeitlinger Publishers, Lisse, The Netherlands, 33–78, 2001.

Openstreetmap, http://planet.openstreetmap.org/planet-100106.osm.bz2, last access: 7 January 2010.

Passant, N.: Speciation of UK emissions of non-methane volatile organic compounds, AEA Technology, Culham Abingdon Oxon, GB, AEAT/R/ENV/0545, February 2002.

Pregger, T. and Friedrich, R.: Effective pollutant emission height for atmospheric transport modelling based on real-world information, Environ. Pollut., 157, 552–560, 2009.

Rockel, B., Will, A., and Hense, A.: The Regional Climate Model COSMO-CLM (CCLM), Meteorol. Z., 17, 347–248, 2008.

Rockel, B. and Geyer, B.: The performance of the regional climate model CLM in different climate regions, based on the example of precipitation, Meteorol. Z., 17(4), 487–498, 2008.

Russell, D.: NARSTO critical review of photochemical models and modelling, Atmos. Environ., 34(12–14), 2261–2282, 2000.

Schlünzen, K. H. and Sokhi, R. S.: Overview of tools and methods for meteorological and air pollution mesoscale model evaluation and user training, WMO, Geneva, Switzerland, Joint Report of COST Action 728 and GURME, GAW Report No. 181, 2008.

Schaap, M., Roemer, M., Sauter, F., Boersen, G., Timmermans, R., Builtjes, P. J. H., and Vermeulen, A. T.: LOTOS-EUROS documentation, http://www.lotos-euros.nl/doc/index.html (last access: 1 January 2010), 2005.

Schwede, D., Pouliot, G., and Pierce, T.: Changes to the Biogenic Emissions Inventory System Version 3 (BEIS3), 4th CMAS Models-3 users' conference, Chapel Hill, NC, 26–28 September 2005.

Sedac: Gridded population of the world v3, http://sedac.ciesin.columbia.edu/gpw/global.jsp, last access: 1 January 2010.

Seaman, N. L.: Meteorological modeling for air-quality assessments, Atmos. Environ., 34(12–14), 2231–2259, 2000.

Simpson, D., Fagerli, H., Jonson, J. E., Tsyro, S., Wind, P., and Tuovinen, J.: Transboundary acidification, eutrophication and ground level ozone in Europe – Part I: Unified EMEP model description, EMEP Status Report 2003, Norwegian Meteorological Institute, NMI, Oslo, Norway, 2007.

GMDD

3, 949–1007, 2010

Smiatek, G.: Mapping land use for modelling biogenic and anthropogenic emissions,, in: The Proceedings of Symposium '98, Boston Southampton 2000, edited by: Borrell, P. M. and Borrell P., WIT press, UK, ISBN 1-85312-743-4, 1765 pp., 1998.

Sofiev, M., Miranda, A. I., and Sokhi, R.: Review of the capabilities of meteorological and chemistry-transport models for describing and predicting air pollution episodes, (WMO/TD-No. 1502), Technical report, COST 728 and GURME, Geneva, 2009.

TREMOVE, http://www.tremove.org/documentation/index.htm, last access: 1 January 2010.

UNC Carolina Environmental Program: Sparse Matrix Operator Kernel Emissions (SMOKE) Modeling System, UNC Chapel Hill, North Carolina, USA, 2005.

US Environmental Protection Agency: Community multiscale air quality modeling system, http://www.epa.gov/asmdnerl/models3/, last access: 12 August 2009.

USGS Global Land Cover Characterization, http://edc2.usgs.gov/glcc/glcc.php, last access: 1 January 2009.

Vautard, R., Beekmann, M., Bessagnet, B., and Menut, L.: Chimere Un simulateur numerique de la qualite de l'air, IPSL, Paris, France, 2007 (in French).

Vestreng, V., Mareckova, K., Kakareka, S., Malchykhina, A., and Kukharchyk, T.: Inventory review 2007, Emission data reported to LRTAP convention and NEC directive, MSC-West, Oslo, Norway, MSC-W Technical Report 1/07, ISSN 1504-6079 (print), ISSN 1504-6206 (online), 2007.

Visschedijk, A. J. H. and Denier van der Gon, H. A. C.: Gridded European anthropogenic emission data for NO_x, SO_2, NMVOC, NH_3, CO, PM_{10}, $PM_{2.5}$ and CH_4 for the year 2000, TNO, Apeldoorn, Netherlands, TNO-report B&O-A R2005/106 version 2, 2005.

Visschedijk, A. J. H., Zandveld, P., and Denier van der Gon, H. A. C.: A high resolution gridded European emission database for the EU integrated project GEMS, TNO, Apeldoorn, Netherlands, TNO-report 2007-A-R0233/B, 2007.

Webdab, http://www.ceip.at/emission-data-webdab/emissions-used-in-emep-models/, last access: 1 January 2010.

Yu, Y., Sokhi, R. S., Kitwiroon, N., Middleton, D. R., and Fisher, B.: Performance characteristics of MM5-SMOKE-CMAQ for a summer photochemical episode in Southeast England, UK, Atmos. Environ., 42, 4870–4883, 2008.

GMDD

3, 949–1007, 2010

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

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

GMDD
3, 949–1007, 2010

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

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

GMDD
3, 949–1007, 2010

Table 3. NUTS level definition.

NUTS 1	3 million–7 million inhabitants
NUTS 2	800 000–3 million inhabitants
NUTS 3	150 000–800 000 inhabitants

GMDD
3, 949–1007, 2010

Table 4 . EMEP measurement stations used for comparison with modelled air concentrations.

ID	Name	Longitude	Latitude	Altitude	O_3	NO_2	SO_2	SO_4	NO_3
AT02R	Illmitz	47 46 0 N	16 46 0 E	117	X	X	X	X	
CH02R	Payerne	46 48 47 N	6 56 41 E	489	X	X	X	X	
CH03R	Tänikon	47 28 47 N	8 54 17 E	539	X	X			
CZ03R	Kosetice	49 35 0 N	15 5 0 E	534	X				
DE01R	Westerland	54 55 32 N	8 18 35 E	12	X	X			
DE02R	Langenbrügge	52 48 8 N	10 45 34 E	74	X	X			
DE04R	Deuselbach	49 45 53 N	7 3 7 E	480	X				
DE07R	Neuglobsow	53 10 0 N	13 2 0 E	62	X	X	X	X	
DE09R	Zingst	54 26 0 N	12 44 0 E	1	X	X	X	X	
DK03R	Tange	56 21 0 N	9 36 0 E	13			X	X	
DK05R	Keldsnor	54 44 0 N	10 44 0 E	10			X	X	
DK08R	Anholt	56 43 0 N	11 31 0 E	40		X	X	X	
EE09R	Lahemaa	59 30 0 N	25 54 0 E	32		X		X	
ES03R	Roquetas	40 49 14 N	0 29 29 E	44		X	X	X	X
ES04R	Logroño	42 27 28 N	2 30 11 W	445		X	X	X	X
ES08R	Niembro	43 26 32 N	4 51 1 W	134		X	X		X
ES10R	Cabo de Creus	42 19 10 N	3 19 1 E	23			X	X	X
ES11R	Barcarrola	38 28 33 N	6 55 22 W	393	X		X	X	X
FI09R	Utö	59 46 45 N	21 22 38 E	7	X		X	X	
FI17R	Virolahti II	60 31 36 N	27 41 10 E	4	X		X	X	
FI22R	Oulanka	66 19 13 N	29 24 6 E	310	X		X	X	
FI37R	Ahtari II	62 35 0 N	24 11 0 E	180			X	X	
FR03R	La Crouzille	45 50 N	1 16 0 E	497				X	
FR05R	La Hague	49 37 0 N	1 49 59 W	133				X	
FR09R	Revin	49 54 0 N	4 38 0 E	390	X		X	X	
FR13R	Peyrusse Vieille	43 37 0 N	0 11 0 E	2			X	X	
GB02R	Eskdalemuir	55 18 47 N	3 12 15 W	243	X				
GB04R	Stoke Ferry	52 34 0 N	0 30 0 E	15				X	
GB06R	Lough Navar	54 26 35 N	7 52 12 W	126				X	
GB07R	Barcombe Mills	50 52 0 N	0 1 59 W	8				X	
GB13R	Yarner Wood	50 35 47 N	3 42 47 W	119				X	
GB14R	High Muffles	54 20 4 N	0 48 27 W	267	X		X	X	

GMDD
3, 949–1007, 2010

SMOKE for Europe – a comprehensive emission model for Europe

J. Bieser et al.

Table 4 . Continued.

ID	Name	Longitude	Latitude	Altitude	O_3	NO_2	SO_2	SO_4	NO_3
GB15R	Strath Vaich Dam	57 44 4 N	4 46 28 W	270	X			X	
GB16R	Glen Dye	56 58 0 N	2 25 0 W	85				X	
GB39R	Sibton	52 17 38 N	1 27 47 E	46	X				
GR01R	Aliartos	38 22 0 N	23 5 0 E	110		X			
HU02R	K-puszta	46 58 0 N	19 35 0 E	125	X	X	X	X	X
IE02R	Turlough Hill	53 2 12 N	6 24 0 W	420				X	
IE31R	Mace Head	53 10 0 N	9 30 0 W	15	X				
IS02R	Irafoss	64 5 0 N	21 1 0 W	66				X	
IT01R	Montelibretti	42 6 0 N	12 38 0 E	48	X	X		X	X
IT04R	Ispra	45 48 0 N	8 38 0 E	209	X	X	X	X	X
LT15R	Preila	55 21 0 N	21 4 0 E	5	X	X	X	X	
LV10R	Rucava	56 13 0 N	21 13 0 E	5	X	X	X	X	X
LV16R	Zoseni	57 8 0 N	25 55 0 E	183		X	X	X	X
NL09R	Kollumerwaard	53 20 2 N	6 16 38 E	1	X	X		X	X
NL10R	Vredepeel	51 32 28 N	5 51 13 E	28		X		X	X
NO01R	Birkenes	58 23 0 N	8 15 0 E	190	X	X	X	X	X
NO08R	Skreådalen	58 49 0 N	6 43 0 E	475		X	X	X	X
NO15R	Tustervatn	65 50 0 N	13 55 0 E	439		X	X	X	X
NO39R	Kårvatn	62 47 0 N	8 53 0 E	210		X	X	X	X
NO41R	Osen	61 15 0 N	11 47 0 E	440		X	X	X	X
NO42G	Spitsbergen, Zeppelinfjell	78 54 0 N	11 53 0 E	474			X	X	X
NO55R	Karasjok	69 28 0 N	25 13 0 E	333	X	X	X	X	X
PL02R	Jarczew	51 49 0 N	21 59 0 E	180	X	X	X	X	X
PL04R	Leba	54 45 0 N	17 32 0 E	2	X	X	X	X	X
PL05R	Diabla Gora	54 9 0 N	22 4 0 E	157		X	X	X	
PT04R	Monte Velho	38 5 0 N	8 48 0 W	43	X				
RU01R	Janiskoski	68 56 0 N	28 51 0 E	118			X	X	X
RU18R	Danki	54 54 0 N	37 48 0 E	150	X			X	X
SE02R	Rörvik	57 25 0 N	11 56 0 E	10	X	X	X	X	
SE11R	Vavihill	56 1 0 N	13 9 0 E	175	X				
SE12R	Aspvreten	58 48 0 N	17 23 0 E	20	X				
SE13R	Esrange	67 53 0 N	21 4 0 E	475	X				
SE32R	Norra-Kvill	57 49 0 N	15 34 0 E	261	X				

GMDD
3, 949–1007, 2010

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.

	EMEP	TNO-GEMS	IER-GKSS	SMOKE-EU	OBSERVATION
			O_3 – 40 stations (N=329 197)		
MEAN	77.43±6.05	74.58±6.26	75.86±6.08	78.25±6.44	57.79±6.76
FB	0.29±0.13	0.26±0.13	0.27±0.13	0.3±0.14	–
NME	0.36±0.19	0.31±0.18	0.33±0.18	0.37±0.19	–
FAC2	0.79	0.8	0.79	0.79	–
CORR	0.62±0.08	0.61±0.06	0.62±0.07	0.63±0.08	–
IOA	0.45±0.28	0.47±0.25	0.46±0.27	0.47±0.26	–
			NO_2 – 33 stations (N=11 465)		
MEAN	1.31±1.01	1.37±1.48	1.33±1.23	1.57±1.32	2.31±1.74
FB	−0.47±0.46	−0.51±0.53	−0.49±0.46	−0.28±0.48	–
NME	0.37±0.25	0.44±0.27	0.38±0.25	0.33±0.23	–
FAC2	0.49	0.46	0.45	0.55	–
CORR	0.44±0.31	0.42±0.3	0.45±0.3	0.45±0.3	–
IOA	0.41±0.41	0.35±0.4	0.41±0.37	0.48±0.33	–
			SO_2 – 36 stations (N=12 430)		
MEAN	0.98±0.83	0.98±1.03	1.09±1.3	1.27±1.2	0.78±0.63
FB	0.21±0.71	0.09±0.65	0.1±0.72	0.34±0.73	–
NME	0.8±0.65	0.63±0.58	0.7±0.56	1.03±0.82	–
FAC2	0.46	0.44	0.44	0.44	–
CORR	0.4±0.23	0.38±0.23	0.38±0.25	0.4±0.23	–
IOA	0.42±0.26	0.43±0.25	0.42±0.27	0.37±0.25	–
			SO_4 – 51 stations (N=17 536)		
MEAN	0.61±0.18	0.57±0.18	0.54±0.17	0.66±0.21	0.71±0.42
FB	−0.02±0.4	−0.08±0.41	−0.13±0.4	0.06±0.38	–
NME	0.35±0.33	0.34±0.32	0.33±0.27	0.36±0.38	–
FAC2	0.61	0.59	0.59	0.62	–
CORR	0.44±0.16	0.39±0.15	0.42±0.15	0.45±0.16	–
IOA	0.49±0.26	0.43±0.27	0.44±0.24	0.51±0.26	–
			NH_4 – 22 stations (N=7400)		
MEAN	1.28±0.77	1.05±0.6	1.03±0.64	1.44±0.9	0.75±0.78
FB	0.74±0.45	0.59±0.47	0.57±0.47	0.83±0.41	–
NME	1.62±1.31	1.24±1.1	1.2±1.14	1.84±1.38	–
FAC2	0.37	0.4	0.41	0.34	–
CORR	0.46±0.17	0.38±0.21	0.45±0.18	0.46±0.18	–
IOA	0.14±0.7	0.25±0.59	0.25±0.58	0.09±0.63	–
			NO_3 – 18 stations (N=6184)		
MEAN	0.47±0.41	0.3±0.24	0.32±0.31	0.51±0.46	0.41±0.54
FB	0.05±0.79	−0.2±0.79	−0.18±0.67	0.13±0.75	–
NME	0.78±0.76	0.58±0.42	0.37±0.32	0.81±1.02	–
FAC2	0.25	0.18	0.22	0.25	–
CORR	0.32±0.27	0.26±0.21	0.32±0.26	0.32±0.27	–
IOA	0.29±0.34	0.27±0.32	0.34±0.34	0.28±0.25	–

GMDD
3, 949–1007, 2010

SMOKE for Europe – a comprehensive emission model for Europe

J. Bieser et al.

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.

	EMEP 3-D	EMEP 2-D	TNO-GEMS 3-D	TNO-GEMS 2-D
SO_4	0.61±0.18	0.58±0.16	0.55±0.19	0.54±0.16
SO_2	0.98±0.83	1.2±1.18	0.99±1.03	1.06±1.2

GMDD
3, 949–1007, 2010

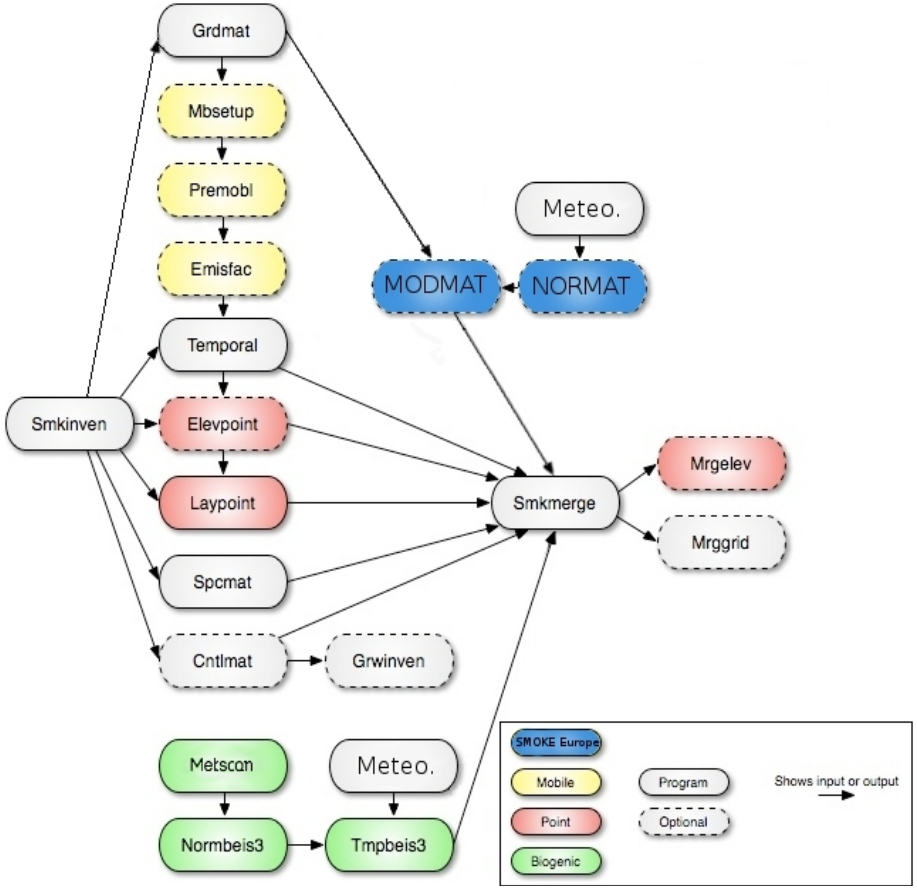
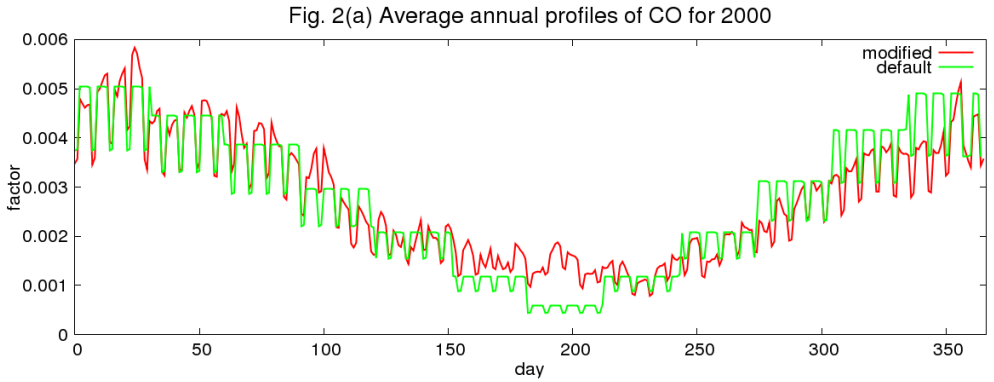


Fig. 1. SMOKE and BEIS3 (green) core programs including modifications for SMOKE EUROPE (blue). Short descriptions of the most important modules can be found in Appendix C (Baek et al., 2009).

GMDD
3, 949–1007, 2010

SMOKE for Europe – a comprehensive emission model for Europe

J. Bieser et al.



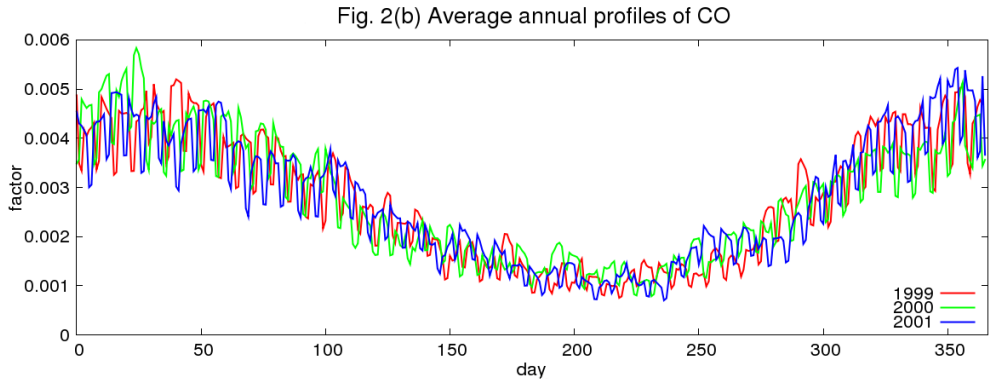


Fig. 2. **(a)** Comparison of temperature dependent temporal profiles SMOKE default vs. modified version. **(b)** Inter annual comparison of temperature dependent CO temporal profiles (all values averaged over the whole domain).

GMDD
3, 949–1007, 2010

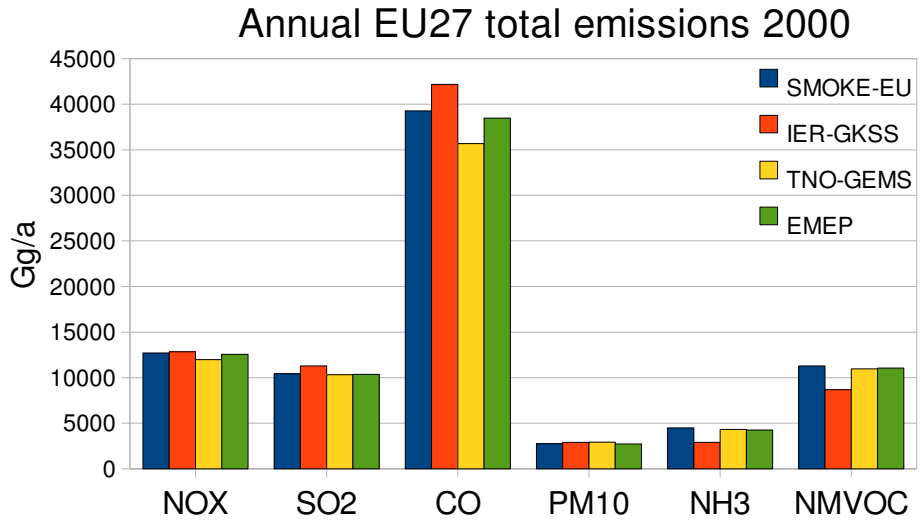


Fig. 3. Annual total anthropogenic emissions of the EU27 in 2000 (The SMOKE-EU dataset also includes 18 000 Gg/a biogenic NMVOC emissions).

GMDD
3, 949–1007, 2010

SMOKE for Europe – a comprehensive emission model for Europe

J. Bieser et al.

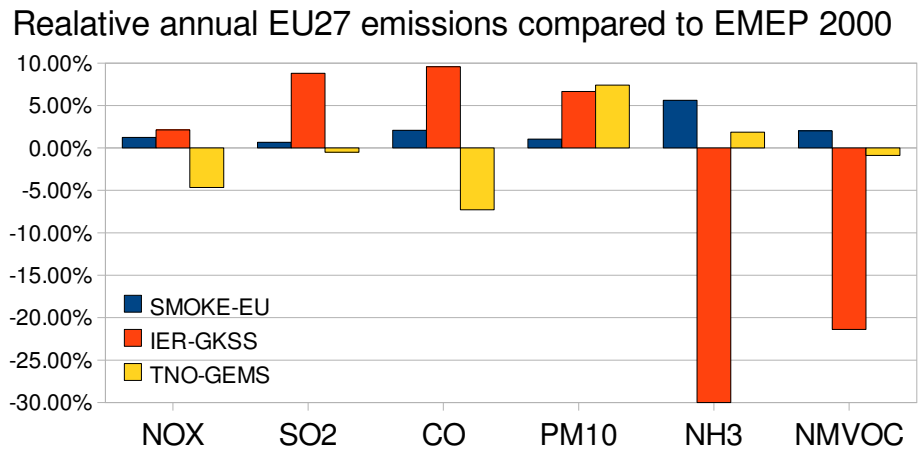
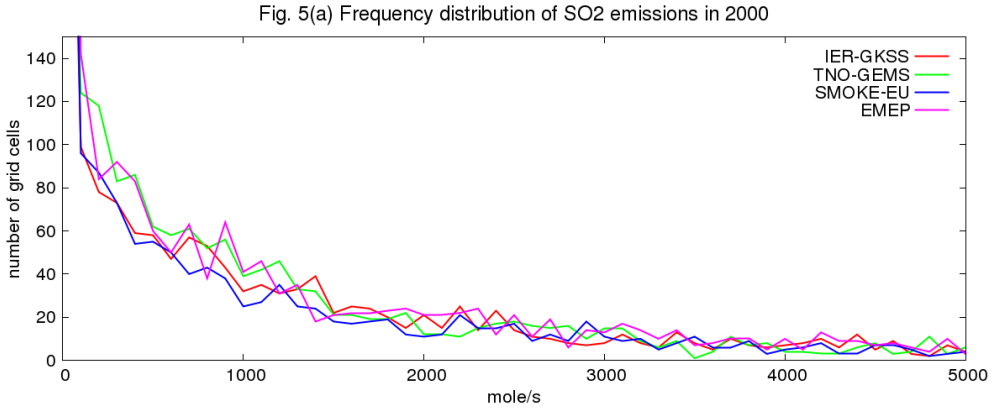


Fig. 4. Relative annual total emissions of the EU27 (biogenic emissions are not included).

GMDD
3, 949–1007, 2010



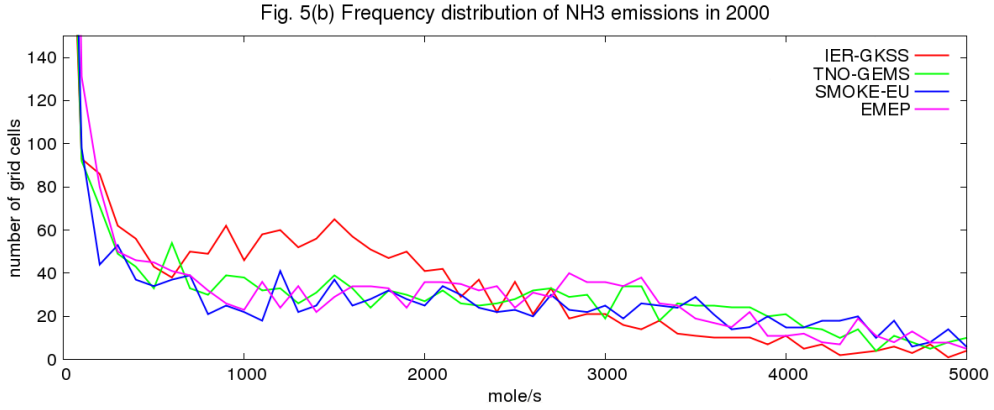


Fig. 5. Frequency distribution of **(a)** SO_2 and **(b)** NH_3 emissions (annual total emissions of EU27).

GMDD
3, 949–1007, 2010

SMOKE for Europe – a comprehensive emission model for Europe

J. Bieser et al.

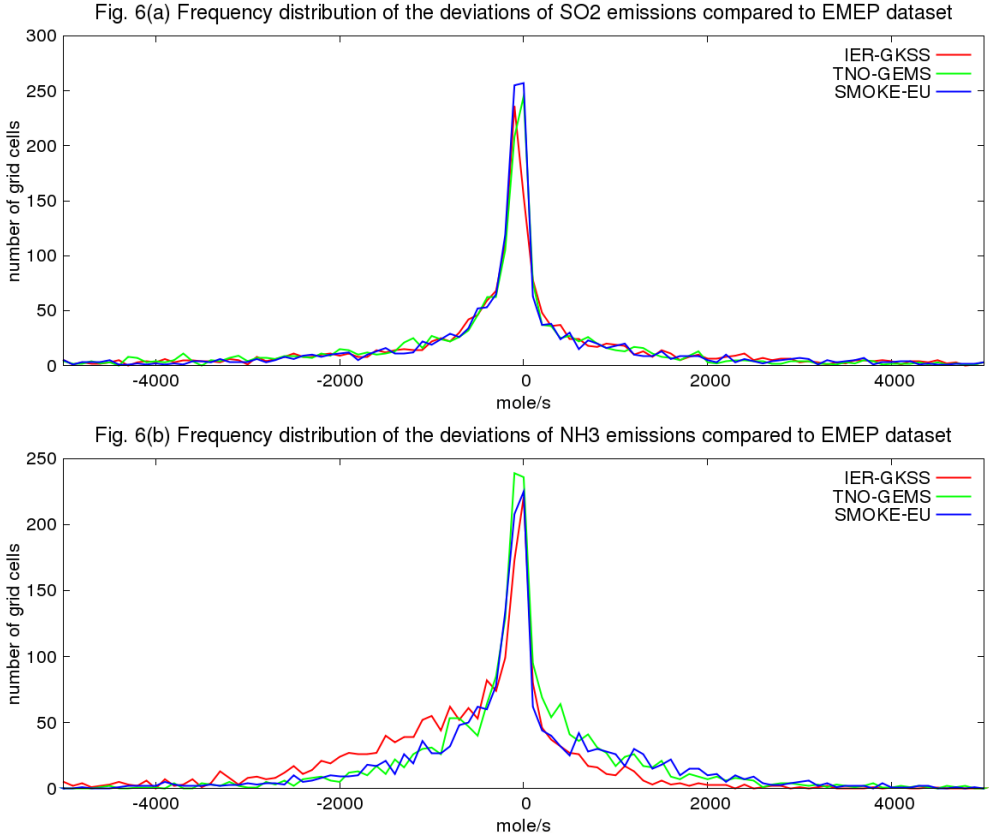


Fig. 6. Frequency distribution of **(a)** SO_2 and **(b)** NH_3 emissions (annual total emissions of EU27).

GMDD
3, 949–1007, 2010

SMOKE for Europe – a comprehensive emission model for Europe

J. Bieser et al.

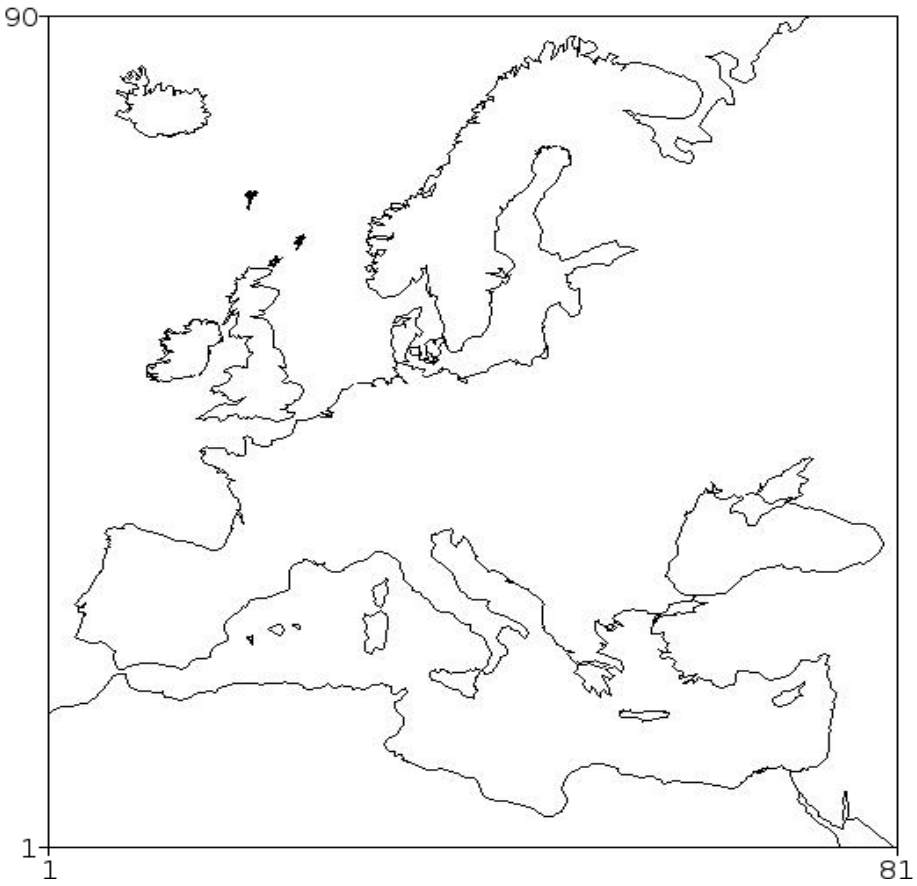


Fig. 7. Modelling domain used for CTM calculations with $54 \times 54\,km^2$ grid resolution.

GMDD
3, 949–1007, 2010

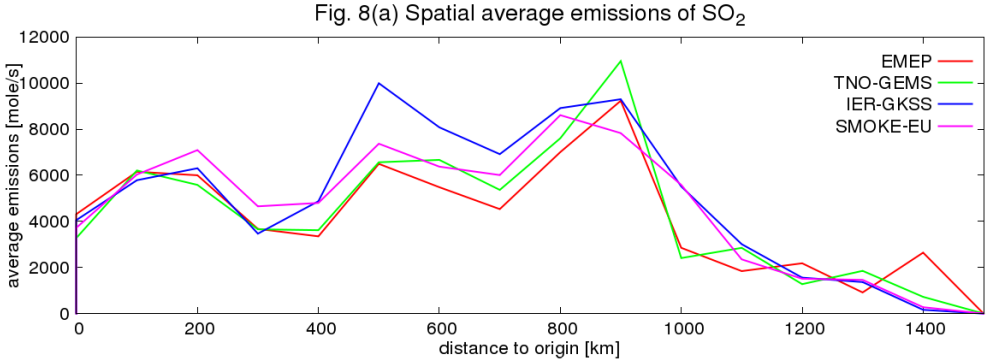


Fig. 8. **(a)** Spatial average annual emissions of SO_2. **(b)** Variograms for SO_2. All values for concentric circles with 100 km distance.

GMDD
3, 949–1007, 2010

SMOKE for Europe – a comprehensive emission model for Europe

J. Bieser et al.

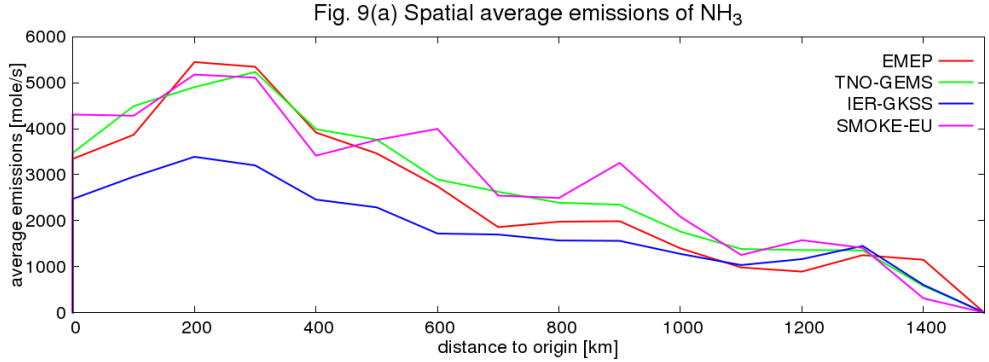


Fig. 9. **(a)** Spatial average annual emissions of NH_3. **(b)** Variograms for NH_3. All values for concentric circles with 100 km distance.

GMDD
3, 949–1007, 2010

SMOKE for Europe – a comprehensive emission model for Europe

J. Bieser et al.

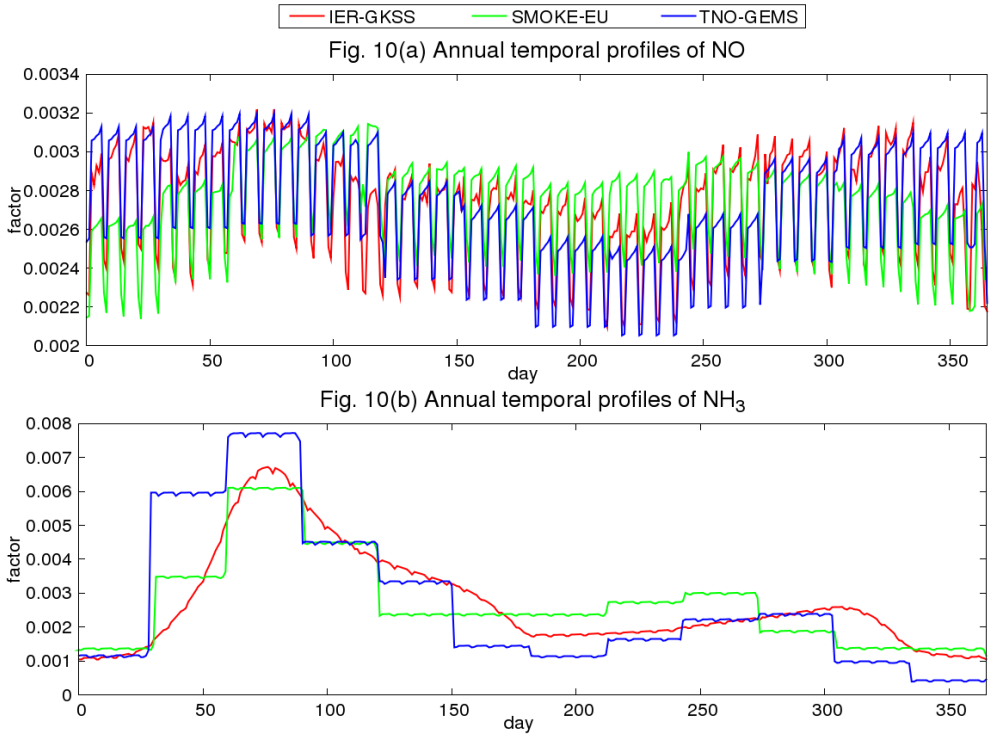


Fig. 10. Averaged annual temporal profiles with daily resolution of **(a)** NO and **(b)** NH_3. The biogenic NO emissions included in the SMOKE-EU dataset lead to higher average emission in summer and lower average emissions in winter.

GMDD
3, 949–1007, 2010

SMOKE for Europe – a comprehensive emission model for Europe

J. Bieser et al.

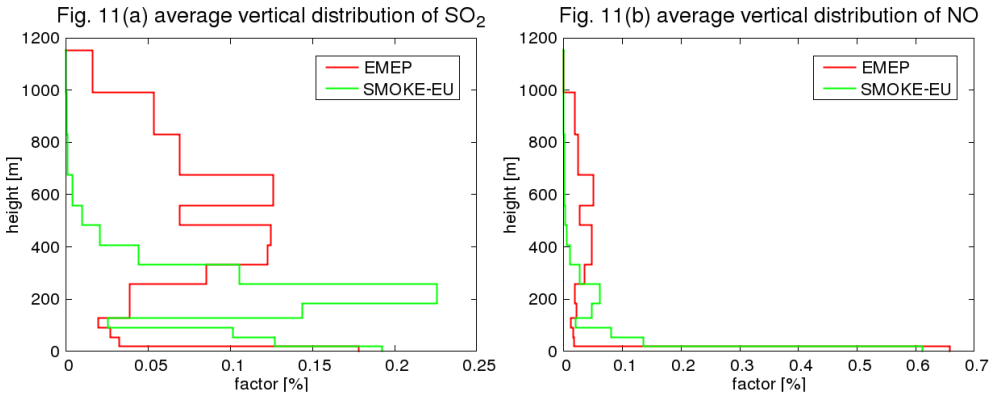


Fig. 11. Average vertical distribution of **(a)** SO_2 and **(b)** NO emissions. For comparison with the SMOKE-EU dataset, the official EMEP vertical profiles were interpolated from 6 to 30 layers. The TNO-GEMS dataset uses the EMEP vertical distributions.

GMDD
3, 949–1007, 2010

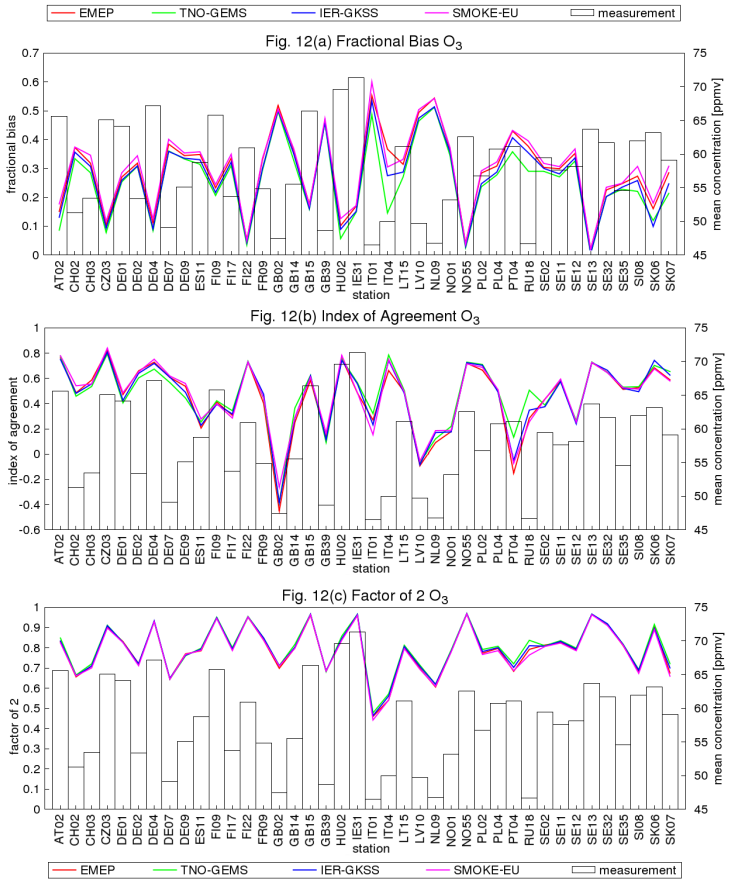


Fig. 12. Comparison of modelled O_3 concentrations with hourly observations from 40 rural EMEP measurement sites (N=329 197) See also Tables 6 and 7. **(a)** Fractional bias, **(b)** index of agreement, **(c)** percentage of values within a factor of 2.

GMDD
3, 949–1007, 2010

SMOKE for Europe – a comprehensive emission model for Europe

J. Bieser et al.

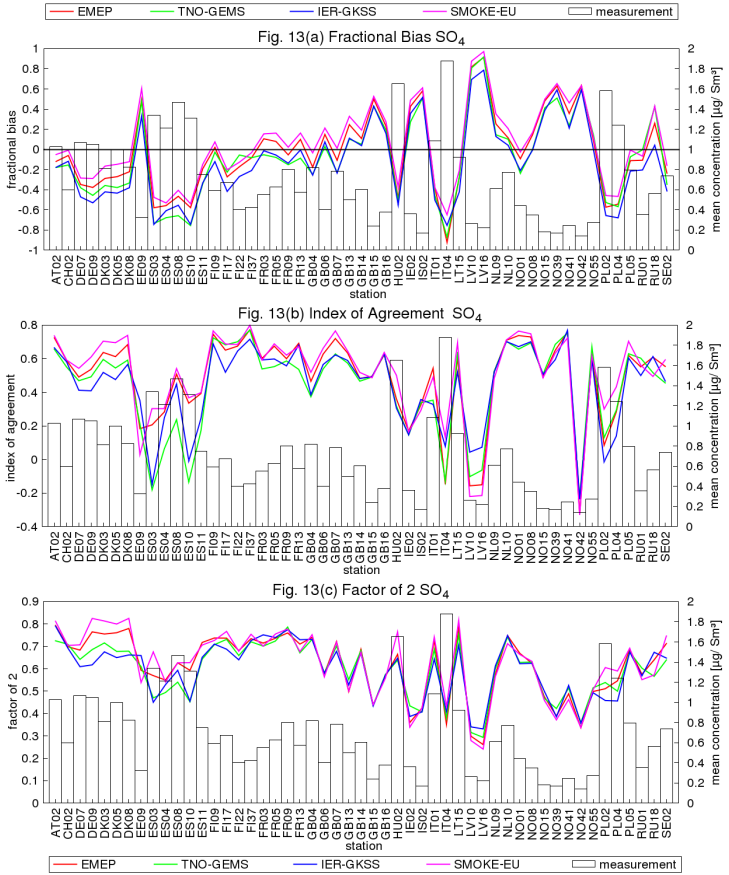


Fig. 13. Comparison of modelled SO_4 concentrations with hourly observations from 51 rural EMEP measurement sites (N=17 536) See also Tables 6 and 7. **(a)** Fractional bias, **(b)** index of agreement, **(c)** percentage of values within a factor of 2.

GMDD
3, 949–1007, 2010

SMOKE for Europe – a comprehensive emission model for Europe

J. Bieser et al.

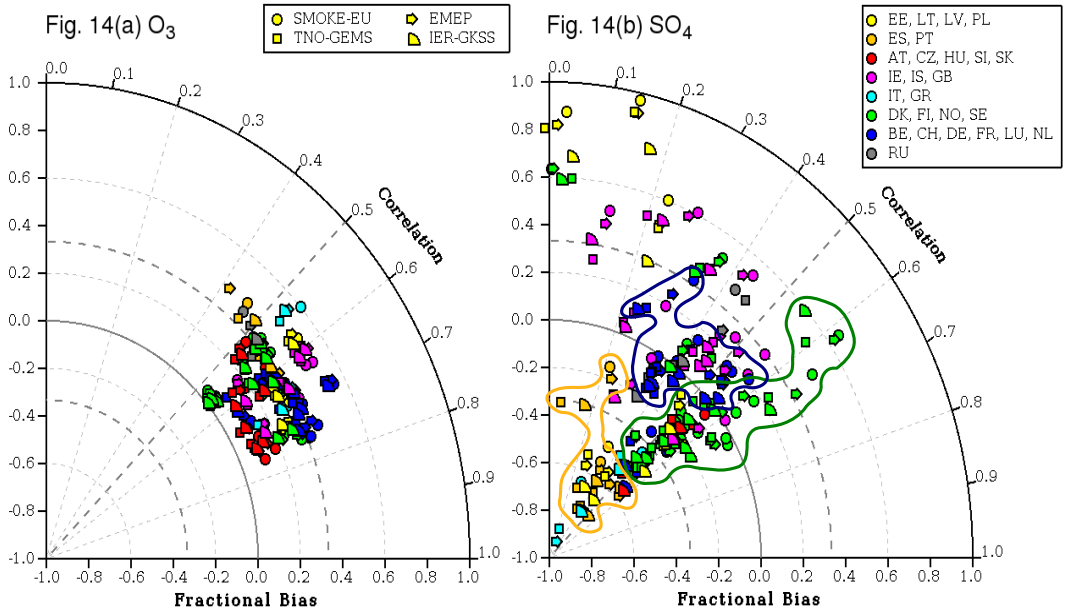


Fig. 14. Taylor diagrams showing correlation and fractional bias of modelled atmospheric concentrations of O_3 **(a)** and SO_4 **(b)** compared to observations. Different shapes indicate the 4 emission datasets used, while colors indicate geographical regions.

GMDD
3, 949–1007, 2010

SMOKE for Europe – a comprehensive emission model for Europe

J. Bieser et al.

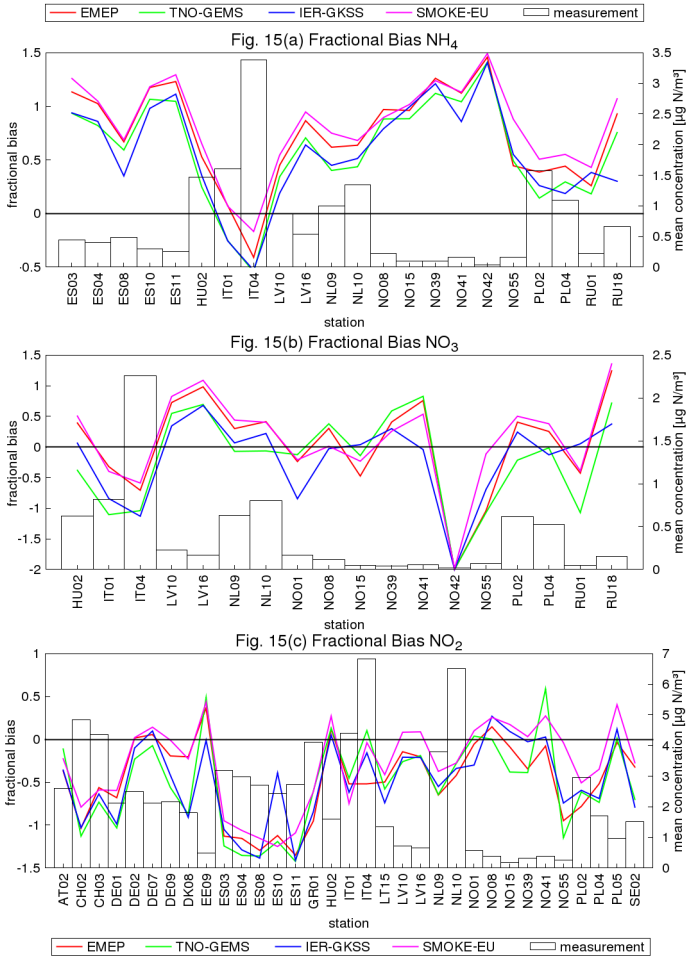


Fig. 15. Fractional bias and observed concentrations for NH_4, NO_3 and NO_2.

GMDD
3, 949–1007, 2010

SMOKE for Europe – a comprehensive emission model for Europe

J. Bieser et al.