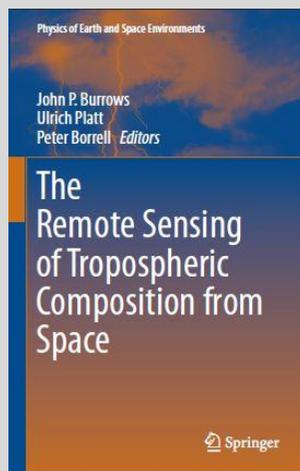


# **The Remote Sensing of Tropospheric Composition from Space**

Editors:

John P. Burrows  
Ulrich Platt  
Peter Borrell

*Pages 451 to 492*



## **Chapter 9**

### **Synergistic Use of Retrieved Trace Constituent Distributions and Numerical Modelling**

Maria Kanakidou, Martin Dameris, Hendrik Elbern,  
Matthias Beekmann, Igor B. Kononov,  
Lars Nieradzic, Achim Strunk and Maarten C. Krol

Publisher: Springer Verlag, Heidelberg

[Springer Book Web Page](#)

[Springer on line Page for the Book](#)

ISBN 978-3-642-14790-6

DOI 10.1007/978-3-642-14791-3

**February 2011**



# Chapter 9

## Synergistic Use of Retrieved Trace Constituent Distributions and Numerical Modelling

**Maria Kanakidou, Martin Dameris, Hendrik Elbern, Matthias Beekmann, Igor B. Konovalov, Lars Nieradzik, Achim Strunk and Maarten C. Krol**

### 9.1 Introduction

Remote sensing of tropospheric constituents from satellite observations of solar irradiance has made significant advances the recent years, opening new horizons in environmental studies and extending observational coverage from individual scarce observations to the global view of short- and long-lived tropospheric constituents. Retrievals of tropospheric trace constituent distributions from satellite observations now provide a concise and global view of the state and of the evolution of the atmosphere. They are valuable for understanding atmospheric responses to natural and human driven emissions, meteorology and climate changes.

For two decades now, the number and type of observations, coupled with improvements in the retrieval algorithms for tropospheric trace constituents and the validation of the retrieved products (trace gases atmospheric columns and profiles, aerosol parameters, fire counts, etc.) has resulted in an increasing confidence in the observations of the Earth's surface and troposphere from space. Thus monitoring air pollution

---

M. Kanakidou  
Environmental Chemical Processes Laboratory, Department of Chemistry, University of Crete, Heraklion, Greece

M. Dameris  
Deutsches Zentrum für Luft- und Raumfahrt, Institut für Physik der Atmosphäre, Oberpfaffenhofen, Germany

H. Elbern, L. Nieradzik and A. Strunk  
Rhenish Institute for Environmental Research at the University of Cologne, Köln, Germany

M. Beekmann  
Laboratoire Interuniversitaire des Systèmes Atmosphériques (LISA) CNRS, Université Paris, Est et Paris 7, Créteil, France

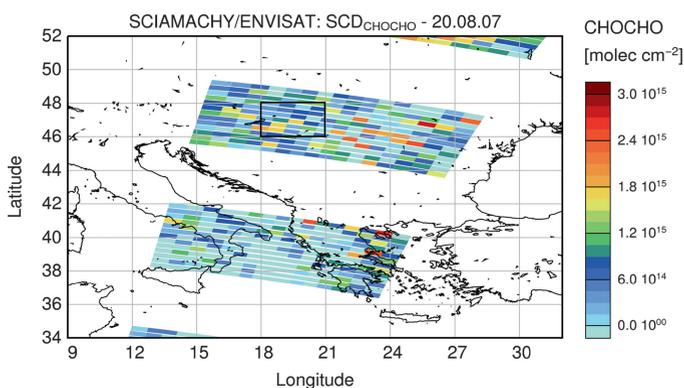
I.B. Konovalov  
Institute of Applied Physics, Russian Academy of Sciences, Nizhniy Novgorod, Russia

M.C. Krol  
Meteorology and Air Quality, Environmental Sciences Group, Wageningen University, Wageningen, The Netherlands

from space is now close to reality and this information can be used for the definition of environmental strategy and control.

Retrieval products of both trace gas and aerosol distributions seen from satellites, are starting to be widely used by the atmospheric modelling community for evaluating models, process studies, improving emissions estimates and estimating environmental and climate impacts occurring in the Earth system due to emissions and chemistry of greenhouse gases and aerosols.

Uncertainties and approximations are associated with both the retrieved data products and the environmental model simulations. These have to be taken carefully into account when using satellite observations jointly with data derived from numerical modelling studies for detecting and quantifying atmospheric changes. In particular, although satellites provide a global view of the atmosphere, this is mostly a composite of several overpasses at specific times of the day and with a specific daily frequency. For instance, SCIAMACHY passes over an area at around 10:30 local time and follows almost the same orbit every 6 days at the equator, whereas GOME-2 monitors an area around 9:30 every 1.5 days and OMI performs more frequent observations with an overpass once a day at 13:30 local time. Similarly the atmospheric models use a variety of horizontal and vertical resolutions ranging for global models from about  $5^\circ \times 5^\circ$  to a few tenths of a degree in latitude by longitude and from nine to several tens of vertical levels that also vary in thickness as a function of altitude and location (Stevenson et al. 2006). The model spatial resolutions mostly do not coincide with the spatial resolutions of the satellite sensor observations (Fig. 9.1). This has to be taken into account when comparing satellite data with model results, particularly for short lived species that have a high spatial and diurnal variability such as  $\text{NO}_2$ , HCHO, and CHOCHO (Vrekoussis et al. 2004; Velasco et al. 2007).



**Fig. 9.1** Tropospheric slant column densities of glyoxal, CHOCHO, retrieved from SCIAMACHY sensor observations over south-eastern Europe on 20th August 2007. To demonstrate the non collocation of satellite pixels and model grid boxes, the satellite pixels, which are of variable size, are shown in colour and a model grid box of  $2^\circ \times 3^\circ$  resolution is marked in black (personal communication from M. Vrekoussis, IUP, University of Bremen).

Using observations of the back scattered solar radiation at the top of the atmosphere and measurements of the extra terrestrial solar irradiance enables the slant column (SC) of a trace gas in a particular wavelength region to be determined. This SC depends on the length of the path of the photons through the atmosphere, the air mass factor (AMF) and the absorptions of the trace gas in a given altitude. So the AMF also depends on the amount of multiple scattering, the fraction of photons reflected at the surface in the direction of the satellite, and therefore the vertical profile of the trace gas (see Chapter 1). The AMF thus provides *a priori* information on the combined effect of all factors that affect the transfer of radiation in the atmosphere and allows conversion of the slant column of atmospheric constituents seen by the satellite sensor to vertical columns or profiles. This information is produced from atmospheric observations and model simulations that take into account the presence of clouds and aerosols, *surface* albedo, the shape of the constituent's vertical profile and temperature in the atmosphere (see in [www.iup.uni-bremen.de/E-Learning/section](http://www.iup.uni-bremen.de/E-Learning/section) on retrieval procedures and column measurements, and Chapter 1).

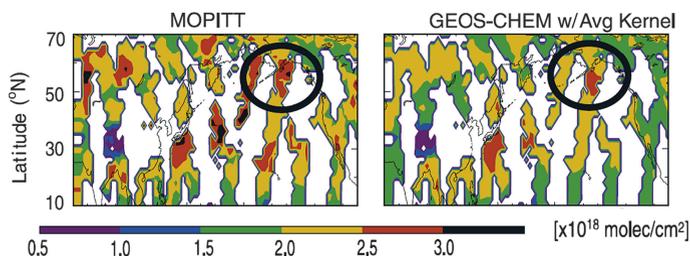
One possibility would be to determine in the model the SC and compare model SC with retrieved SC. This would be consistent. However vertical column amounts are easier to comprehend and thus are more likely to be used. It is evident that the quality of the retrievals of the vertical total or tropospheric columns of trace gases and the outcomes of the synergistic use of models with satellite observations strongly depend on the assumptions and *a priori* knowledge being used to determine the AMF.

A similar issue arises when using measurements of the emerging thermal infrared radiation and optimal estimation retrieval techniques. For the retrievals of atmospheric profiles, such as those of CO from the MOPITT instrument, information on the sensitivity of the retrieval to the real profile of the studied atmospheric constituent is required, as well as the *a priori* constituent's profile. This is described by averaging kernel matrixes (Deeter et al. 2003), unique for each retrieval ([www.eos.ucar.edu/mopitt/data](http://www.eos.ucar.edu/mopitt/data)) that have to be taken into consideration when satellite retrievals are synergistically applied with models.

As an example, Fig. 9.2 shows MOPITT observations of CO columns and compares them with the GEOS-Chem model results sampled along the satellite sensor orbit track and using the MOPITT averaging kernels (Hudman et al. 2004). Asian anthropogenic CO is seen in both distributions. Note that UV/vis/NIR retrievals have no or very limited vertical resolution and so only columns can be retrieved and compared to models.

Provided the sensitivity, for example expressed through the averaging kernels of the retrieved data products, when using DOAS or optimal estimation or other retrieval approaches are appropriately taken into account, satellite observations are and have been synergistically used with modelling for several key objectives:

- To improve our understanding of atmospheric chemistry and its evolution with time under anthropogenic and natural stresses and evaluate models, both on global and smaller scales;



**Fig. 9.2** Transpacific Asian pollution event shown by MOPITT observations of CO columns on 4th May 2002 at 00:00 UTC (*left*). The *right panel* shows the corresponding GEOS-Chem model results sampled along the MOPITT orbit tracks and with MOPITT averaging kernels applied. The *black circles* show the plume location (figure adopted from Hudman et al. (2004)).

- To identify the origin and evaluate the environmental consequences of the atmospheric composition changes as seen from space;
- To initiate chemical transport models by using satellite derived distributions of atmospheric constituents as an input for chemistry transport modelling (for instance fire counts observations as a proxy for biomass burning emission distribution; vegetation and chlorophyll-a distributions used to parameterise biogenic emission from the terrestrial and the marine ecosystems, etc.) and thereby to account for processes that are not explicitly resolved in the models;
- To assimilate retrieved data products from satellites to improve model prognostic results;
- To evaluate and improve emission estimates and atmospheric trends (forward and inverse modelling).

This research field, coupling models and satellite data products, is evolving rapidly. In the following sections a flavour of the results obtained from the above scientific applications and their principles rather than an exhaustive list of studies performed is provided. Selected investigations of tropospheric model evaluation, species origin and sources identification are outlined in Section 9.2. Principles, examples and needs for inverse modelling are presented in Section 9.3 and objectives, methods and examples for retrieved data assimilation are discussed in Section 9.4. Overall conclusions on the state of the art in the field and challenges for future research on synergistic use of satellite retrievals and atmospheric models are given in Section 9.5.

## 9.2 Use of Satellite Data for Process Understanding and Model Evaluation

Consistent global information of the chemical composition and the dynamics of the Earth's atmosphere are provided by space-borne instrument measurements. Satellite data are therefore a major corner stone for better understanding of individual

atmospheric processes and feedback mechanisms. In addition, the synergistic use of observations and respective data derived from studies with numerical atmospheric models helps to improve the knowledge of processes driving atmospheric variability and changes on different time scales. Discrepancies between observations and model results can help to identify gaps in our understanding of dynamical, physical and chemical processes in the atmosphere. A detailed evaluation of atmospheric models is necessary to determine their ability to reproduce adequately atmospheric variability and changes. The exact knowledge of strengths and weaknesses of such models is required to enable solid assessments of the future evolutions of atmospheric chemical composition and climate.

### ***9.2.1 Understanding Atmospheric Chemistry***

In the recent decade work has been intensified to develop numerical model systems to describe the whole Earth system taking into account interactions, variations and feedbacks of the various compartments of the Earth system, including the atmosphere. Previously, atmospheric models required a number of input parameters, relevant to processes that are not explicitly resolved by these models and which are used as boundary conditions. As a result of the increasing availability of relevant satellite observations, processes in atmospheric models can now be driven by taking the initial data from satellite observed parameters. For example, in the absence of coupled dynamic vegetation/emissions/fire models, satellite retrievals of terrestrial vegetation and fire counts are used to parameterise biogenic VOC emissions or deposition (Guenther et al. 2006), and biomass burning emissions (Mota et al. 2006; Giglio et al. 2006). Synergistic use of products from several satellites allows the construction of long data series. For instance, van der Werf et al. (2006) provided computations of global biomass burning emissions that are widely used by the scientific community. They have used measurements from MODIS in conjunction with ATSR and VIRS satellite data covering a 7-year period. Chlorophyll-a distributions seen from space are also introduced as input into the models to parameterise emissions to the atmosphere from the marine ecosystems (O'Dowd et al. 2008; Arnold et al. 2009; Myriokefalitakis et al. 2010). Satellite retrievals of trace constituents such as total O<sub>3</sub> column, HNO<sub>3</sub> or CO levels in the upper troposphere/lower stratosphere and aerosol optical thickness (Lelieveld and Dentener 2000; Barret et al. 2008; Ito and Penner, 2005) have also been used in models when the driving processes, such as stratospheric chemistry or aerosol dynamics, are not explicitly resolved.

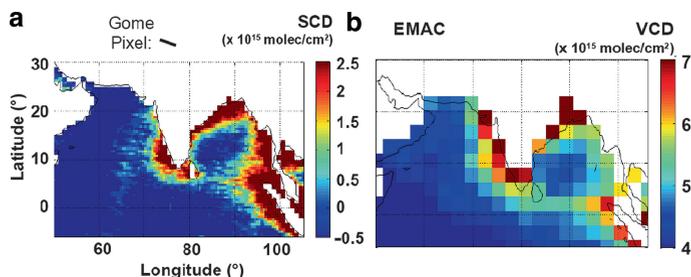
In addition to this one way flow of information from satellite retrievals to models, scientific advances can be achieved through comparisons of model results from targeted simulations with satellite retrievals of trace constituents. Such investigations enable the understanding of atmospheric processes, source identification and quantification, detection and evaluation of the long range transport of pollutants and its impacts as well as model evaluation. Examples of such investigations are discussed in the following sections.

### **a Formaldehyde, HCHO: A Proxy for VOC Emissions**

Formaldehyde, HCHO, is a high-yield oxidation product of numerous VOCs in the atmosphere, including isoprene that is emitted in large amounts from terrestrial vegetation. Due to the short lifetime of HCHO (globally about 4 h) the measured HCHO tropospheric columns are expected to be correlated with the local VOC emissions weighted by the HCHO yield. The signal is smeared out and displaced in the atmosphere due to horizontal transport (Palmer et al., 2003) and diffusion. However, when focusing on areas of the size of a satellite pixel or even larger as those of a global Chemistry-Transport Model (CTM) i.e. with a horizontal resolution of several hundreds of km<sup>2</sup>, this displacement should be negligible. Palmer et al. (2003) have developed a methodology based on the synergistic use of the global CTM GEOS-Chem and satellite retrievals of HCHO columns to constrain isoprene emissions from the terrestrial biosphere, taking into account the lifetimes of HCHO and VOC. They assumed that HCHO column variability was mainly linked to isoprene emissions. Model results have been sampled along the ensemble of GOME orbit tracks and used to derive linear relationships between HCHO columns and isoprene emissions over North America in the model. To minimise biases and maximise consistency between retrievals and model results, HCHO columns have been retrieved from GOME based on AMF derived from the model. Clouds, the primary error source in AMF calculations (Millet et al., 2006), have been filtered out using the cloud fraction data from the same sensor. Using GEOS-Chem, Palmer et al. (2003) calculated a mean HCHO molar yield of 1.2 to 1.96 that is consistent with laboratory experiments and with aircraft HCHO and isoprene profile observations ( $1.6 \pm 0.5$ ; Millet et al. (2006)). This ratio allows one to estimate the isoprene emissions from the HCHO columns retrieved from GOME. The retrieval errors, combined with uncertainties in the HCHO yield from isoprene oxidation, have been estimated to result in a 40% ( $1\sigma$ ) error in inferring isoprene emissions from HCHO satellite observations (Millet et al., 2006).

Guenther et al. (2006) pointed out the potential importance of biomass burning and anthropogenic emission contributions to HCHO signal that introduces errors to the emission estimates. However, Palmer et al. (2003; 2006) have also shown that VOCs other than isoprene generally either have excessive smearing out or insufficient emission relative to the HCHO detection limit, so that HCHO column observations from space are highly specific to isoprene. Millet et al. (2008) used a similar approach to derive isoprene emissions using the GEOS-Chem model and space observations of HCHO columns by the OMI sensor, with  $13 \times 24$  km<sup>2</sup> nadir footprint and daily global coverage. They also concluded that the spatial distribution of HCHO columns from OMI follows that of isoprene emission except in few urban locations, like Houston, where anthropogenic hydrocarbon emissions are detectable from space. Fu et al. (2007) performed a similar analysis of GOME retrievals focusing on east and south Asia and found an underestimate of biogenic VOC emissions over China in the current estimates.

Marbach et al. (2009) reported the first detection of signal from ship emissions in the GOME derived tropospheric HCHO columns over the Indian Ocean (Fig. 9.3),



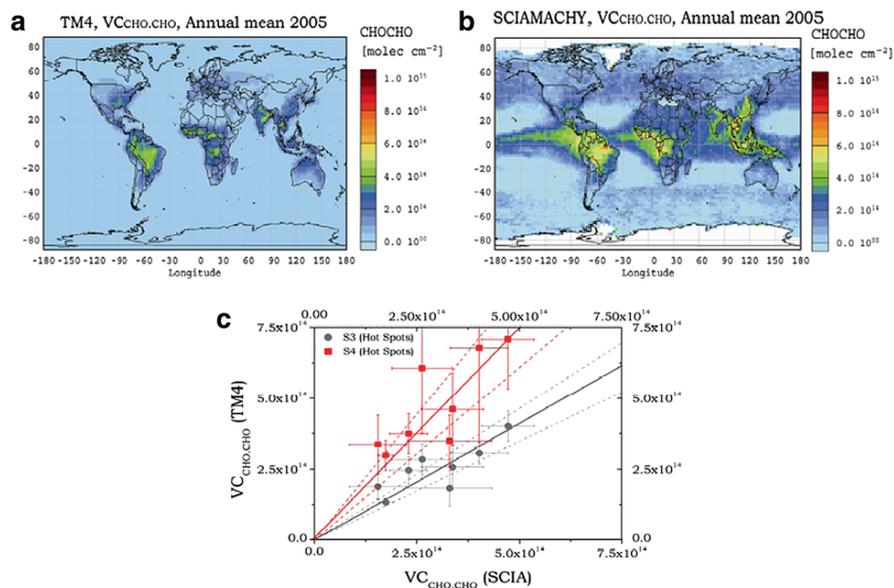
**Fig. 9.3** HCHO distribution over the Indian Ocean (land masses are masked out). Figures have been adapted from Marbach et al. (2009). (a) GOME SCDs during winter (January to March) 1996–2002 with cloud fractions below 20% are averaged. The ship track is visible from Sri Lanka up to about half the distance to Sumatra. For illustration, size and orientation of a single GOME pixel is displayed above panel a. (b) EMAC model results of the mean HCHO VCDs for winter (January to March) 1997–2002, integrated up to 50 hPa (using EMAC local time between 10:00–11:00 a.m.).

where conditions often favour plume detection since ships follow a single narrow track in the same east-west direction as the GOME pixel scanning. From the 7-year composite of cloud free observations, they evaluated a mean HCHO column enhancement over the shipping route of about  $2 \times 10^{15}$  molecules/cm<sup>2</sup>. Although the pattern of this enhancement is reproduced by their Climate-Chemistry model EMAC, the HCHO columns are underestimated by a factor of two, when satellite data and model results are similarly sampled and spatially averaged. The discrepancy is tentatively attributed to an underestimate in the emission inventories and their atmospheric dilution as well as to the rather coarse resolution of the model. This is limiting the proper simulation of the fast high NO<sub>x</sub> ship plume chemistry that enhances oxidation capacity in the marine environment.

## b Glyoxal, CHOCHO: Source Apportionment

Glyoxal, CHOCHO, has recently been observed from space (Wittrock et al. 2006). CHOCHO is known to be mostly a product of biogenic VOC oxidation and has been suggested as indicator of secondary aerosol formation in the troposphere (Volkamer et al. 2005). However, a number of anthropogenic hydrocarbons, such as acetylene and aromatics, have been positively identified as CHOCHO precursors. Myriokefalitakis et al. (2008) investigated the contribution of pollution to the CHOCHO levels using the global 3-dimensional (3-D) CTM TM4-ECPL in conjunction with the respective “Vertical Column Amounts or Densities” of CHOCHO retrieved from the SCIAMACHY sensor observations in 2005.

A series of simulations has been performed accounting for various secondary sources as well as a potential primary source of CHOCHO from biomass burning. The simulations have been evaluated by comparison with the retrieved columns of CHOCHO both on an annual (Fig. 9.4) and on seasonal basis. The observations



**Fig. 9.4** Global annual mean column distribution of glyoxal (CHOCHO) ( $2^\circ \times 3^\circ$  grid) for the year 2005 (in molecules/cm<sup>2</sup>). (a) Simulated by TM4-ECPL, taking into account all known photochemical CHOCHO sources; (b) Retrieved from the measurements made by the satellite based sensor SCIAMACHY; (c) Comparison of annual mean CHOCHO columns from the TM4-ECPL simulations (*black circles*: S3 accounts only for the secondary sources of CHOCHO; *red squares*: S4 accounts for all secondary sources and for a potential primary source of CHOCHO from biomass burning) with the SCIAMACHY data products (in units of molecules/cm<sup>2</sup>). Binned data over continental hot spot areas (figure adapted from Myriokefalitakis et al. (2008) where more details can be found).

have been gridded to  $2^\circ \times 3^\circ$  in order to fit the model's grids and both the retrieved and the calculated columns have been used. When accounting only for the secondary sources of CHOCHO in the model, the model underestimates CHOCHO columns observed by satellites, and the model fails to simulate the high CHOCHO columns retrieved over tropical oceans. This is tentatively attributed to outflow from the continents and local primary oceanic biogenic or secondary sources of HCHO and CHOCHO that are not taken into account in the model or, alternatively, to an overestimate of CHOCHO columns in the retrievals. Elucidation of this discrepancy between satellite observations and model results requires further targeted experiments as well as forward and inverse modelling investigations.

Using primary emissions of about  $7 \text{ Tg yr}^{-1}$  of CHOCHO from biomass burning and anthropogenic combustion sources in the model, leads to an overestimate of CHOCHO columns by the model over areas of intensive emissions (Fig. 9.4). For a global mean lifetime for CHOCHO of about 3 h, their model evaluates the global annual mean CHOCHO burden in the model domain at  $0.02 \text{ Tg}$  equal to the global burden seen by SCIAMACHY over land for the year 2005 (Myriokefalitakis et al., 2008). These results point to the need to understand the presence of CHOCHO over

the tropical oceans. Similar conclusions are drawn by Fu et al. (2008) using the GEOS-Chem global CTM. In contrast to HCHO, the secondary anthropogenic contribution from fossil fuel and industrial VOCs emission oxidation to the CHOCHO columns is found to reach 20–70% in the industrialised areas of the northern hemisphere, suggesting that concurrent observations of HCHO and CHOCHO observations over specific locations could provide proxies for VOC emissions (Myriokefalitakis et al., 2008).

Currently there is much interest in the CHOCHO and IO signals over the oceans. There are indications that in the Pacific the regions of elevated CHOCHO, IO, and also possibly HCHO are over areas which contain phytoplankton and in particular diatoms (personal communication, J. Burrows).

### c Determining Dominant Chemical Pathways: Air Pollution Impact

Correlations between atmospheric trace constituents provide important information about the dominant atmospheric processes, and this novel capability in satellite remote sensing has important implications not only for investigations of air pollution, but also for air pollution control strategy. For instance, satellite derived HCHO and NO<sub>2</sub> columns can be used to investigate tropospheric O<sub>3</sub> photochemistry. Martin et al. (2004) applied the GEOS-Chem model to evaluate the potential of the ratio of HCHO columns to tropospheric NO<sub>2</sub> columns as an indicator of surface ozone – NO<sub>x</sub> (NO<sub>x</sub> = NO + NO<sub>2</sub>) – VOC sensitivity over polluted areas. Relying on these model results, satellite data analysis of the HCHO/NO<sub>2</sub> column ratios have shown the consistency of GOME observations over polluted areas with current understanding of surface O<sub>3</sub> chemistry based on *in situ* observations. The satellite-derived ratios indicate that surface O<sub>3</sub> production is NO<sub>x</sub>-limited throughout most continental regions of the northern hemisphere during summer. Exceptions include major urban and industrial centres such as Los Angeles and industrial areas of Germany that tend to be NO<sub>x</sub>-saturated (and thus VOC limited). The NO<sub>2</sub> derived from GOME also yields a geographical transition to NO<sub>x</sub>-sensitive regime downwind of these centres and a seasonal transition in the autumn when surface O<sub>3</sub> becomes less sensitive to NO<sub>x</sub> and more sensitive to VOCs.

The impact of pollution on the photochemical enhancement of O<sub>3</sub> can be also derived from HCHO and NO<sub>2</sub> columns observed from space used in conjunction with chemical box and Lagrangian (trajectory) models. For instance, Ladstätter-Weißenmayer et al. (2007) using chemical box calculations associated with the GOME-observed NO<sub>2</sub> and HCHO tropospheric columns, found a potential of daily photochemical enhancement in the tropospheric O<sub>3</sub> columns of about 0.8–1 Dobson Units (DU, equivalent to  $2.7 \times 10^{16}$  molecules/cm<sup>2</sup>) and a daily potential of regional photochemical build-up within upwind polluted air masses of about 2–8 DU over Crete in the eastern Mediterranean during spring. At most 10–20 DU of tropospheric O<sub>3</sub> have been attributed to stratosphere-troposphere exchange (STE) whereas the total observed variability in tropospheric O<sub>3</sub> derived from both space and ground based observations was about 25 DU.

CO and aerosol satellite observations also provide information on tropospheric air quality (Chapter 6). Both, CO columns and the aerosol optical depth (AOD) show oxidant-driven seasonal variation since oxidants act both as a source and a sink for CO and a production pathway for secondary aerosols. On global scales, fine mode AOD is driven by sulfate production although carbonaceous particles can be also of importance over several locations (Zhang et al. 2007), particularly during biomass burning events. Edwards et al. (2004) analysed global four year records of concurrent CO and fine mode AOD retrievals from the MOPITT and MODIS observations, both on board the Terra satellite. They concluded that the observed CO and AOD seasonal cycles were several months out of phase, with perturbations occurring during sporadic biomass burning emissions when carbonaceous particles dominate AOD. During such events the retrieved CO columns and AOD are well correlated. Anomalous high pollution observed from space in the northern hemisphere in winter-spring of 2002–2003, has been analysed based on the fire counts from MODIS and on global model simulations with MOZART-2. Artificially releasing pulses of CO over the fire locations in the model and during four other months of the year enabled the evaluation of the persistence of CO in the atmosphere. Edwards et al. (2004) calculated that the build-up of CO in the model for a pulse in October was twice as large as for a pulse in July. This reflects the e-folding time of CO that was calculated to vary over the studied area from about 1.5 months in July to about 3.6 months in October. Thus, the timing of the burning (in late summer-early fall) was favourable for a build-up of CO to anomalously high values in the northern hemisphere in winter compared to other years.

Lelieveld et al. (2009) synergistically used *in situ* aircraft observations of O<sub>3</sub> together with SCIAMACHY and TES satellite sensor observations of NO<sub>2</sub> and O<sub>3</sub>, to perform model simulations with the EMAC model to study the origin of observed high O<sub>3</sub> levels over the Persian Gulf. They concluded that the Persian Gulf region is a hot spot of photochemical smog where air quality standards are violated throughout the year. EMAC simulations allow the identification of long distance transport of air pollution from Europe and the Middle East, natural emissions and stratospheric O<sub>3</sub> to the relatively high background O<sub>3</sub> mixing ratios.

#### **d Understanding Differences Between Retrievals and Model Results**

Comparison of model results with retrievals of atmospheric constituents from satellite observations is not restricted to the analysis of observations but can also point out deficiencies in the retrieval algorithms and thus initiate their improvements, specifically in the assumptions made for the determination of AMF or the averaging kernels.

For example Martin et al. (2002) have compared the distribution of tropical tropospheric O<sub>3</sub> columns retrieved from TOMS with the GEOS-Chem model results together with additional information from *in situ* observations. They found major discrepancies between model results and TOMS retrievals over northern Africa and southern Asia where the TOMS retrieved O<sub>3</sub> columns did not capture

the seasonal enhancements from biomass burning found in the model and in aircraft observations. Martin et al. (2004) attributed this discrepancy to the poor sensitivity of TOMS to Rayleigh scattering that is important for retrieving low troposphere  $O_3$  enhancements by biomass burning. Thus they developed an efficiency correction to the TOMS retrieval algorithm that accounts for the variability of  $O_3$  in the lower troposphere. This correction increased the retrieved  $O_3$  columns over biomass burning regions by 3–5 DU and decreased them by 2–5 DU over oceanic regions, improving the agreement with *in situ* observations. The correction explained about 5 DU of the “tropical Atlantic paradox”, i.e. the enhanced tropical tropospheric column of  $O_3$  over the southern tropical Atlantic retrieved during the northern African biomass burning season in December to February. The remainder of the paradox was reproduced by the model; it was attributed to the combination of upper tropospheric  $O_3$  production from lightning  $NO_x$ , persistent subsidence over the southern tropical Atlantic and cross-equatorial transport of upper tropospheric  $O_3$  from northern mid-latitudes in the African “westerly duct”.

Another recent example is the study by Bergamaschi and Bousquet (2008), who identified a bias in the dry column of  $CH_4$ , retrieved by Frankenberg et al. (2008a) from SCIAMACHY data, based on simultaneous assimilation of surface observations and satellite data. A large latitudinal varying bias-correction of the satellite data was required to make these data compatible with surface observations. The bias is, in fact, absent in the most recent  $CH_4$  retrieval algorithm and the change has been attributed to the previously poor knowledge of spectroscopic absorption lines used in the former retrieval algorithm (Frankenberg et al. 2008a), leading to improved  $CH_4$  emission estimates (Frankenberg et al. 2008b). However, it is worth noting that other retrievals of the dry column do not show the same bias (Schneising et al. 2008). Overall the evolution of the accuracy of the retrieval algorithms through improved instrument calibration and validation exercises using measurements and models enables the maximum information content to be retrieved. This process is essential, both for testing models and improving retrieval approaches and algorithms.

### ***9.2.2 Model Evaluations – Comparison with Observation***

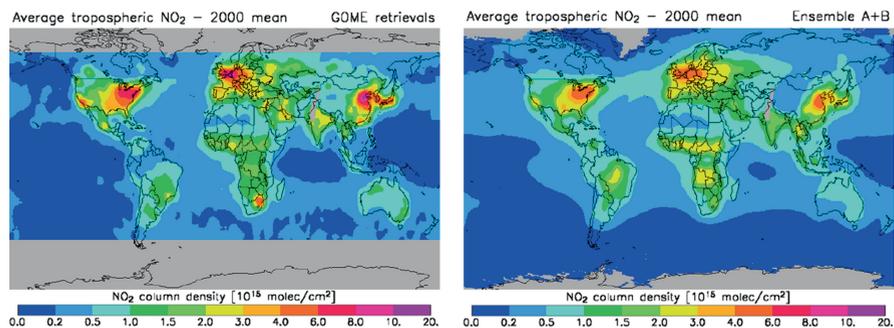
Recently several tropospheric CTM comparison exercises relied on satellite retrievals to supplement the traditional ground-based and aircraft observational data (Velders et al. 2001; van Noije et al. 2006; Shindell et al. 2006; Dentener et al. 2006a; Textor et al. 2006; Kinne et al. 2006). The year 2000 has been used as base year for several global modelling studies. This year benefits from documented emission inventories (Dentener et al. 2006a; 2006b) for both trace gases and aerosols in the framework of the AEROCOM exercise (Aerosol Comparisons between Observations and Models) that is focussed on aerosols (Textor et al. 2006; Kinne et al. 2006; Schulz et al. 2006). These inventories have been also used for the ACCENT intercomparison exercise that focused on tropospheric  $O_3$ ,

NO<sub>2</sub> and CO budgets (van Noije et al. 2006; Dentener et al. 2006b; Stevenson et al. 2006; Shindell et al. 2006). The proper comparison of model results with satellite retrievals allows the evaluation both of the emissions used in the model as well as of the parameterisations of their fate in the atmosphere that reflects our understanding of atmospheric processing. In the following we will illustrate model evaluation procedures by comparing with satellite measurements of a series of relevant species.

## a NO<sub>2</sub>

NO<sub>2</sub> controls tropospheric O<sub>3</sub> production. Its levels in the troposphere show trends that are driven by human activities and these have been observed from space (Richter et al. 2005). The importance of NO<sub>2</sub> for tropospheric O<sub>3</sub> and the consistency of the retrieved tropospheric distributions of NO<sub>2</sub> stimulated their use for model evaluation. Moreover, these satellite observations allow NO<sub>2</sub> global pollution to be evaluated.

Historically the GOME sensor aboard the ERS-2 satellite provided a unique opportunity to compare globally, model calculated NO<sub>2</sub> columns, including, for the troposphere, those from retrievals of remote sensing observations (Lauer et al. 2002). In order to overcome the shortcomings in comparing model results with satellite retrievals outlined in the introduction, van Noije et al. (2006) used 17 different global CTMs that computed daily tropospheric NO<sub>2</sub> column densities for the year 2000. For each model, the computed NO<sub>2</sub> columns were sampled at the satellite overpass time collocated with the measurements to account for sampling biases due to incomplete spatio-temporal coverage by the instrument. The ensemble of the 17 model results has been compared with the mean of the NO<sub>2</sub> columns retrieved from GOME using three different retrieval algorithms (Fig. 9.5).



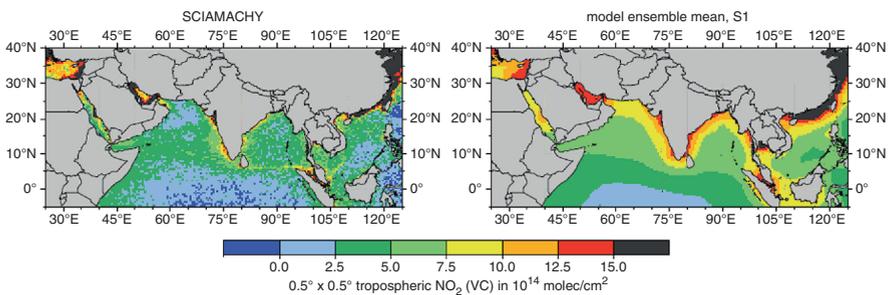
**Fig. 9.5** Ensemble average annual mean tropospheric NO<sub>2</sub> column density for three different GOME retrievals (*left panel*) and the full model ensemble (A + B; *right panel*). These quantities have been calculated after smoothing the data to a horizontal resolution of 5° × 5° (figure from van Noije et al. (2006)).

Three major continental regions of high  $\text{NO}_2$  tropospheric column densities are indicated in the mean columns derived both from the 17 models and from the three GOME retrievals: North America, western Europe, and China (Fig. 9.5). These regions are subject to high pollution emissions. The averaged model maxima of  $6\text{--}8 \times 10^{15}$  molecules/ $\text{cm}^2$  are smaller than the GOME observed values, which exceed  $10 \times 10^{15}$  molecules/ $\text{cm}^2$ . These discrepancies between models and retrievals could neither be explained by *a priori* profile assumptions made in the retrievals, nor by diurnal variations in anthropogenic emissions (van Noije et al. 2006). They have been attributed (Dentener et al. 2006b) to the assumed  $\text{NO}_x$  emissions that may be unrealistically low in these regions, in particular over the rapidly developing parts of eastern China and South Africa. Similar conclusions for the emissions over Asia were drawn by an earlier and less extensive model intercomparison with GOME observations (Velders et al. 2001). In regions dominated by biomass burning, such as in Africa and South America, the models overestimate the retrievals during the dry season. The comparison improves when using biomass burning emissions specific to the year 2000 instead of a 5-year average inventory used for the base simulations (van Noije et al. 2006), pointing to the importance of the inter-annual variability in the emissions. Another significant finding is that the differences in the GOME retrievals are in many instances as large as the spread in model results (10–50% in the annual mean over polluted regions). This means that in only a few cases, such as China, can robust statements on the underestimation of  $\text{NO}_x$  emissions be made (Dentener et al. 2006b; van Noije et al. 2006). The findings imply that top-down estimations of  $\text{NO}_x$  emissions from satellite retrievals of tropospheric  $\text{NO}_2$  are strongly dependent on the choice of model and retrieval.

Recently, Boersma et al. (2009) detected diurnal variations of  $\text{NO}_2$  over Israel and Egypt synergistically using two different satellite sensor retrievals and analysing them with CTM results: those from SCIAMACHY that observed the atmosphere at 10:00 and those from OMI that overpasses at 13:45. They demonstrated that  $\text{NO}_2$  temporal variability over source regions can be followed from space. They derived  $\text{NO}_2$  columns about twice as high in winter as in summer and a strong weekly cycle with  $\text{NO}_2$  almost twice as low on Saturdays than on weekdays. The diurnal difference between SCIAMACHY (10:00) and OMI (13:45)  $\text{NO}_2$  is seen to maximise in summer when SCIAMACHY is up to 40% higher than OMI, and minimise in winter when OMI slightly exceeds SCIAMACHY. The model simulations indicated that a much stronger photochemical loss of  $\text{NO}_2$  in summer than in winter is needed to explain these observations.

Another source of  $\text{NO}_x$  that can be seen from space under certain conditions is that from ships over the remote oceans. Eyring et al. (2007) studied the impact of ship emissions on atmospheric chemistry and climate, using multi-model simulations of Chemistry-Climate Models (CCMs) that have been evaluated for their response to ship emissions. Part of this evaluation used satellite data, particularly the recently observed enhanced tropospheric  $\text{NO}_2$  columns over the Red Sea and along the main shipping lane to the southern tip of India, to Indonesia and northwards towards China and Japan (Beirle et al. 2004; Richter et al. 2004).

The tropospheric  $\text{NO}_2$  columns derived from SCIAMACHY nadir measurements from August 2002 to April 2004 (Richter et al. 2004) have been compared to ensemble means of the models representing the year 2000. The ensemble mean was derived from eight models that provided tropospheric  $\text{NO}_2$  columns at 10:30 a.m. local time, which is close to the overpass time of the ERS-2 satellite. For this comparison, individual model results and SCIAMACHY data were interpolated to a common grid ( $0.5^\circ \times 0.5^\circ$ ). Fig. 9.6 shows that the ensemble of the models reproduces the magnitude and the general pattern of the tropospheric  $\text{NO}_2$  columns over the remote ocean observed by SCIAMACHY. However, the shipping signal that is clearly visible in the satellite data with a high horizontal resolution ( $30 \times 60 \text{ km}^2$ ), does not appear in the model results with a much lower typical resolution of  $5^\circ \times 5^\circ$ . In addition, shipping routes in that area are rather close to land and thus the models grid boxes close to the coast are dominated by  $\text{NO}_x$  emissions from land sources which are much higher (see also Franke et al. (2009)).



**Fig. 9.6**  $\text{NO}_2$  tropospheric columns retrieved from SCIAMACHY observations (*left*) and ensemble of model simulations (*right*) interpolated to  $0.5^\circ \times 0.5^\circ$  grid (figure from Eyring et al. (2007)).

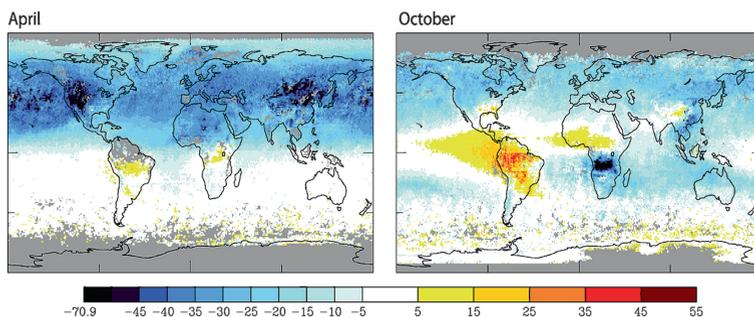
There is clearly a need for model data comparisons with satellite observations over remote oceans. However, such data are inhibited by the distributed nature of the ship emissions over the remote oceans leading to dilution of the emissions, which makes it difficult to distinguish the shipping signal from the effect of long-range transport of polluted air, such as that from the United States towards Europe. The emissions between India and Indonesia present a unique emission pattern. Increasing spatial resolution both of models and of satellite observations might allow us to resolve ship emissions over other oceanic locations.

## b CO

In the frame of the ACCENT AT2 global model intercomparison exercise, Shindell et al. (2006) compared near-global satellite observations from the MOPITT instrument and local surface measurements with present-day CO simulations by 26 state-of-the-art atmospheric CTMs and CCMs. For this purpose, they used monthly mean

daytime values derived from version 3 retrievals from MOPITT gridded at a resolution of  $1^\circ \times 1^\circ$ ; the models' coarser grids were sub-sampled at this same resolution for comparison. In addition the models were sampled like MOPITT observations, using space and time varying averaging kernels from the MOPITT retrievals. This procedure minimises potential biases due to the *a priori* information used in the retrievals. Biases may also be induced by the differences between the models' full temporal and spatial averaging and the satellite's limited sampling time and exclusion of cloudy areas. Such biases are expected to be significant for short-lived atmospheric constituents like aerosols but quite small for CO given its relatively long lifetime.

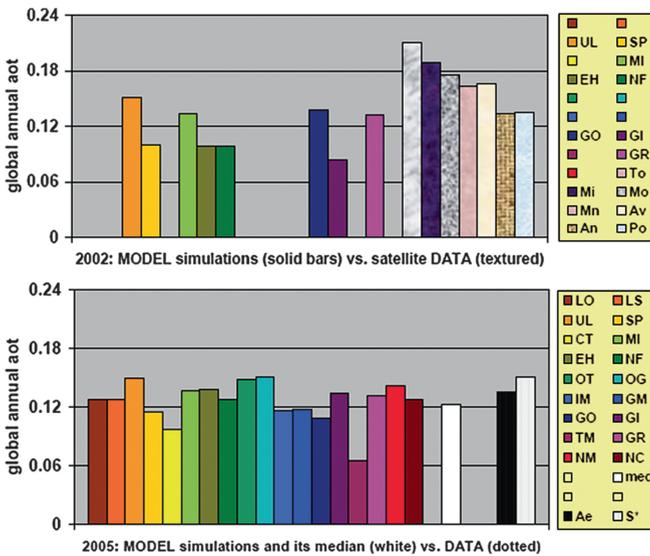
In general, Shindell et al. (2006) pointed out that the models do not adequately capture CO accumulation during the OH-poor winter. The models underestimate both magnitude and seasonality of the CO retrievals throughout the entire extra tropical troposphere in the northern hemisphere, indicating that the biases do not merely reflect an erroneous vertical structure of modelled CO (Fig. 9.7). However, they typically perform reasonably well elsewhere. These results suggest that yearly emissions, probably from fossil fuel burning in eastern Asia and seasonal biomass burning emissions in south-central Africa are greatly underestimated in current inventories such as IIASA and EDGAR3.2. Arellano et al. (2006) performed inverse modelling of CO emissions from various geographical regions and sources from fossil fuel/bio fuel use in Asia based on MOPITT CO data and found that these emissions are almost twice as high as recent bottom-up estimates. The underestimate of the Asian CO emissions in the national estimate-based inventory is also consistent with the under-reporting of  $\text{NO}_x$  emissions (van Noije et al. 2006). In a more recent study, Arellano and Hess (2006) conducted a sensitivity analysis on the differences in the model treatment of transport on top-down estimates of CO sources. They showed that differences between CO model values are due to atmospheric transport and are of the order of 10–30%, with the highest discrepancies for Indonesia, South America, Europe and Russia.



**Fig. 9.7** Differences between multi-model average and MOPITT 2000–2004 average CO (ppbv). Values are shown for April (*left*) and October (*right*) for the 500 hPa pressure level (figure from Shindell et al. (2006)).

**c Aerosol**

Evaluation of aerosol simulations on a global scale is now customarily made through comparison of simulated annual global aerosol optical thickness (AOT) values with those obtained from remote sensing. Fig. 9.8 demonstrates how model simulations for the annual and globally averaged mid-visible AOT (at 550 nm) have changed from the work by Kinne et al. (2003) to the work by Kinne et al. (2006) in the frame of the AEROCOM exercise, and how they compare with data from remote sensing. In the lower panel the number of remote sensing references is reduced to two selections, though of higher quality; a satellite composite, which combines individual satellite retrievals (S\*) and an estimate based on statistics at AERONET ground sites (Ae) (Fig. 9.8).



**Fig. 9.8** The upper panel shows diversity in 2002 among models and satellite data (Kinne et al. 2003). The lower panel compares global annual aerosol AOT median value from the 18 models (med) with the satellite data composite (S\* – see text) and the Aeronet sunphotometer network (Ae) observations. Other symbols correspond to individual models – for more explanations see text and Kinne et al. (2006).

In the earlier work, fewer models were available and the simulated AOTs exhibited a larger variability between models than in the AEROCOM exercise. The upper panel of Fig. 9.8 presents adjusted global annual averages from TOMS, MISR, MODIS, AVHRR and POLDER retrievals. The composite value (S\*) is based on  $3^\circ \times 3^\circ$  longitude/latitude monthly averages, where preference is given to year 2000 data. At 0.11–0.14, simulated AOT values are at the lower end of global averages suggested by remote sensing from ground (AERONET about 0.135) and space (satellite composite about 0.15). More detailed comparisons, however, reveal

that larger differences in regional distribution and significant differences in compositional mixture remain.

Critical for this exercise has been the production of a composite remote sensing dataset of AOT for comparison. Since all ground based remote sensing data are spatially incomplete, adjustments were needed to make global average AOT values comparable. The adjustments involved the spatially and temporally complete median field from modelling. Details are given in Kinne et al. (2006) who summarise in a table the contributing time-periods, retrieval references and known biases.

### 9.3 Inverse Modelling

The examples given in the previous section involved a mostly qualitative analysis of the mismatch between models and satellite observations. To exploit satellite data in a quantitative sense, formal techniques of data assimilation and inverse modelling are required. Solving the *inverse problem* is a common task in many branches of science, where the values of model variables need to be obtained from observations (see 9.6, Appendix). Inverse modelling techniques are in widespread use today in atmospheric science for three major applications: (1) retrieval of atmospheric concentrations from observed radiances (Chapter 4), (2) optimal estimation of atmospheric model parameters and in particular of emissions, and (3) chemical data assimilation. The formal framework for these three problems is similar (Rodgers 2000). Applications presented here will focus on inverse modelling of emissions using satellite observations and also improvements to atmospheric model performances.

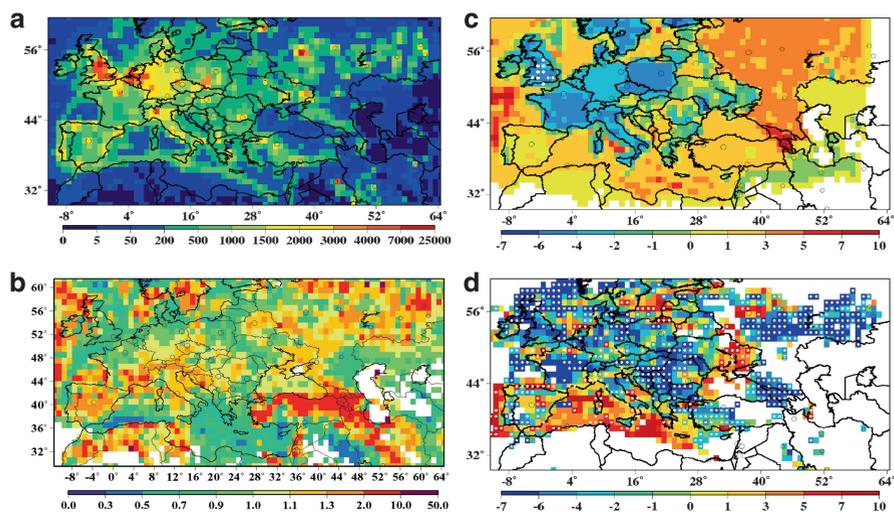
#### 9.3.1 *Inversions for Short-Lived Species*

The use of satellite data to improve emission estimates is illustrated using NO<sub>2</sub> observations from space, which comprise a large part of inverse modelling applications. NO<sub>x</sub> finds favour for several reasons. First, NO<sub>2</sub> satellite data products are abundant and, since the NO<sub>2</sub> lifetime is short, the gradients are large and easily observed from space. Second, the relationship between NO<sub>2</sub> columns derived from satellite measurements and NO<sub>x</sub> emissions is direct and easy to interpret because NO<sub>x</sub> emissions are the major driver of variability on NO<sub>x</sub> columns. Finally, the interest for inverse modelling of NO<sub>x</sub> emissions is fostered by the fact that these emissions are one of the key factors responsible for air pollution problems.

Many early applications of satellite measurement for estimating NO<sub>x</sub> emissions did not explicitly involve any CTM, but used a simple mass balance method which assumed a constant lifetime of the emitted NO<sub>x</sub>. For example, Leue et al. (2001) provided estimates of continental and global NO<sub>x</sub> emissions, Beirle et al. (2003) investigated weekly variations of anthropogenic NO<sub>x</sub> emissions, Beirle et al. (2004)

estimated  $\text{NO}_x$  emissions from shipping in a specific region of the Indian Ocean, Bertram et al. (2005) investigated daily variations in soil  $\text{NO}_x$  emissions. Very recently, Hayn et al. (2009) investigated in detail the spatio-temporal patterns of the global  $\text{NO}_2$  distribution retrieved from GOME satellite observations using a generalised additive model.

Martin et al. (2003) were first to apply the Bayesian inverse modelling approach to providing inventories for  $\text{NO}_x$  emissions using satellite observations. Specifically, they performed a probabilistic combination of “top-down” and “bottom-up” emission estimates and provided a global  $\text{NO}_x$  emission inventory constrained by satellite measurements, which was claimed to be more accurate than the “bottom-up” inventory. They used the global GEOS-Chem CTM in order to define a local linear relationship between  $\text{NO}_x$  emissions. The transport of  $\text{NO}_x$  between different grid cells was disregarded. Typically they found correction factors (for *a posteriori* with respect to *a priori* emissions) of 10–20% for most of the regions, with maximum values up to a factor of two. While the use of a global CTM in the inverse modelling scheme may help to improve global emission inventories, models with much higher spatial resolution are needed in order to elaborate the constrained emission inventories for use in air quality studies. Thus, Konovalov et al. (2006a) used a regional CTM with the resolution of  $0.5^\circ \times 0.5^\circ$  in combination with GOME and SCIAMACHY measurements to improve  $\text{NO}_x$  emission estimates on a regular model grid for western Europe (Fig. 9.9). Other novel features of their study were:



**Fig. 9.9** (a) *A priori* estimates of summertime anthropogenic  $\text{NO}_x$  emission rates used in CHIMERE (in  $10^8$  molecules/ $\text{cm}^2/\text{s}$ ); (b) *a posteriori* correction factors to them. Decadal trends (percent per year) in summertime anthropogenic  $\text{NO}_x$  emissions; (c) estimated with the EMEP data and (d) derived from satellite measurements. *Blank dots* in “d” mark grid cells for which the difference between the emission trends is significant in terms of  $1\sigma$  (from Konovalov et al. (2006b; 2008)).

(1) an original inversion method enabling a partial accounting for the horizontal transport of  $\text{NO}_x$ , (2) the measurement based estimation of the error variances involved in the Bayesian cost function, and (3) the demonstration that the  $\text{NO}_x$  emissions constrained by satellite measurements improve the simulated near-surface concentrations of  $\text{NO}_2$ . In a later study Konovalov et al. (2006b) (Fig. 9.9) found that, on average, the uncertainties in total  $\text{NO}_x$  emissions are estimated to be about 1.7 in terms of the geometric standard deviation in Europe and about 2.1 outside Europe. The corrected emission estimates provided better agreement of the modelled results with observations for both  $\text{NO}_2$  columns and near surface concentrations of  $\text{O}_3$ .

Although in general satellite data products do not themselves distinguish between anthropogenic and biogenic emissions, it is sometimes possible to use additional sources of information, by taking into account temporal evolution of the measured  $\text{NO}_2$  columns and by selecting regions with dominating types of  $\text{NO}_x$  sources. For example, Jaeglé et al. (2004) estimated NO emissions from soils in Africa by combining inversions of GOME  $\text{NO}_2$  columns with space-based observations of fires and bottom-up estimates of fossil fuel and bio fuel emissions. A similar approach was used by Wang et al. (2007) to quantify NO emissions from soils in eastern China. Martin et al. (2007) estimated  $\text{NO}_x$  emissions from lightning in tropical regions using tropospheric  $\text{NO}_2$  columns from SCIAMACHY together with tropospheric  $\text{O}_3$  columns from OMI and MLS, and upper tropospheric  $\text{HNO}_3$  from ACE-FTS. In their study, the CTM GEOS-Chem was used to identify locations and time periods in which lightning would be expected to dominate the trace gas observations. Multi-annual satellite measurements already available can be used, not only to constrain different sources of emissions and improve their spatial allocation, but also to study their long-term changes. Such studies provide valuable opportunities to verify air pollution control strategies and to monitor changes in air pollution sources in the regions where the ground based monitoring networks are either sparse or absent.

Richter et al. (2005) found a highly significant increase of  $\text{NO}_2$  columns over the industrial areas of China in a decadal period from 1996 to 2005 using a combined time series of GOME and SCIAMACHY measurements. They used the MOZART CTM to justify that these changes are caused by similar increases in  $\text{NO}_x$  emissions. Note that this study is not an inverse modelling study in a classical sense, since it does not involve any inversion of a mathematical relationship between observations and emissions.

Similarly, Kim et al. (2006) and van der A et al. (2006) provided some useful insights into  $\text{NO}_x$  emission changes in the United States and China, respectively, without employing any inverse modelling technique. Kim et al. (2009) extended their work to the western USA where they found evidence that changing legislation and changing populations have a strong impact on the  $\text{NO}_x$  emissions in cities in this region. Another true inverse modelling study of multi-annual changes in  $\text{NO}_x$  emissions was performed by Konovalov et al. (2008). They used the time series of  $\text{NO}_2$  columns derived from the GOME and SCIAMACHY measurements in combination with the CHIMERE CTM in a Bayesian inverse modelling scheme,

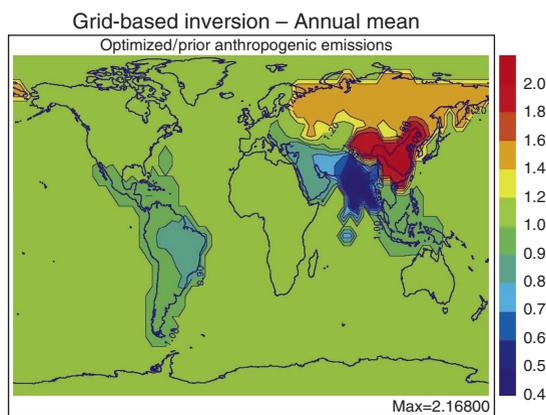
to estimate decadal  $\text{NO}_x$  emission trends in Europe and the Mediterranean (Fig. 9.9). Instead of looking for deviations from “expert” estimates of emissions (as is common in atmospheric inversion studies), Kononov et al. (2008) constrained *a priori* only the minimum and maximum values of emission trends in each grid cell. Accordingly, the top-down estimates obtained can be regarded as a measurement-based alternative to trends derived from bottom-up emission cadastres. An even more direct approach involving a simple combination of satellite measurements with CTM simulations was proposed later by Kononov et al. (2010) to estimate multi-annual  $\text{NO}_x$  emission trends in megacity regions.  $\text{NO}_x$  emission reductions in the last decade over western and central Europe are confirmed by the top-down approach, as well as increases in Spain and emissions related to shipping. Differences between the bottom-up and top-down approaches are notable, especially over south-eastern and eastern European countries. Importantly, the estimates of emission trends obtained were found to be consistent with the available surface measurements of  $\text{NO}_x$  and  $\text{O}_3$  (available mainly over western Europe).

Satellite measurements have also been used to estimate emissions of other important short-lived species such as isoprene. Such estimates can be performed by inversion of the modelled relationship between isoprene emissions and HCHO column measurements. For example, Millet et al. (2008) performed the inversion of OMI measurements and found that the derived isoprene emissions in North America are spatially consistent with a normal bottom-up isoprene emission inventory (MEGAN) ( $R^2 = 0.48\text{--}0.68$ ) but, on average, lower by 4–25%. Corresponding work was performed for Europe (Dufour et al. 2009) using SCIAMACHY derived HCHO columns, facing the difficulty of much smaller isoprene emissions in Europe compared with the US.

Recently Stavrakou et al. (2009), motivated by the large underestimate from the global CTMs of the observed CHOCHO columns from SCIAMACHY, investigated the possible existence of an additional CHOCHO source of biogenic origin over the land. They performed two inverse modelling scenarios for CHOCHO sources. The first included an additional primary source of CHOCHO over land and the second assumed secondary formation through the oxidation of an unspecified CHOCHO precursor with a lifetime of 5 days. As well as the extra source, the inversion scheme optimised the primary CHOCHO and HCHO emissions as well as their secondary production from other identified NMVOC precursors of anthropogenic, pyrogenic and biogenic origin. The best performance is achieved in the second scenario with the inferred total global continental CHOCHO source estimated to be  $108 \text{ Tg yr}^{-1}$ , almost twice as high as the global *a priori* source. The extra secondary source is the largest contribution to the global CHOCHO budget (50%), followed by the production from isoprene (26%) and from anthropogenic NMVOC precursors (14%). The updated emissions allowed for a satisfactory agreement of the model with both satellite and *in situ* CHOCHO observations. The large CHOCHO column amounts observed over the tropical oceans are still not explained.

### 9.3.2 Inversions for CO and CH<sub>4</sub>

The dense, high-quality CO observations performed by the MOPITT satellite instrument gave rise to several remarkable studies applying the inverse modelling technique to estimate CO emissions. For example, Pétron et al. (2004) used the MOZART CTM and the MOPITT CO retrievals to perform a sequential Bayesian estimation of CO sources in 15 large regions of the world. The largest correction factor (about 50 %) for anthropogenic emissions was needed for eastern Asia. *A posteriori* emissions significantly improved the agreement between the simulated CO distributions and independent ground based measurements. A similar study was presented by Arellano et al. (2006). Their *a posteriori* estimate is also much higher (about a factor of 2) than the corresponding earlier estimate for eastern Asia, and significant differences were found between the earlier and post source emissions in many other regions. Stavrou and Müller (2006) performed an inversion of MOPITT data using the adjoint of the IMAGE model (Fig. 9.10). The main goal of their study was a comparison of the large region and grid-based Bayesian inversion methods. Both methods gave similar average estimates, but the grid-based approach brings the model columns much closer to the observations because of its better spatial representativity. An interesting novel feature of the method used by Stavrou and Müller (2006) is optimisation of the sources of the main biogenic VOC compounds simultaneously with the CO sources.



**Fig. 9.10** Ratio of optimised to earlier anthropogenic emissions estimated by the inversion of MOPITT observations (adopted from Stavrou and Müller (2006)).

The CH<sub>4</sub> and CO<sub>2</sub> retrievals are clearly improving but they still differ between different groups (Schneising et al. 2009). As CH<sub>4</sub> has a reasonably uniform distribution over the troposphere, due to its large lifetime (8–9 years), accurate satellite observations, with perhaps less than a few percent of error, are needed to obtain information on CH<sub>4</sub> fluxes, which are primarily emissions. Meirink et al. (2006) applied four-dimensional variational (4D-var) data assimilation

method to synthetic measurements of atmospheric CH<sub>4</sub> to investigate the utility of SCIAMACHY observations for CH<sub>4</sub> source estimation. They concluded that SCIAMACHY observations with a precision of 1–2% can contribute considerably to uncertainty reduction in CH<sub>4</sub> source strengths, but that systematic observation errors well below 1% would have a dramatic impact on the quality of the derived emission; thus identification and removal of these biases is crucial. Bergamaschi et al. (2007) presented initial results of a synthesis inversion of coupled surface and satellite CH<sub>4</sub> observations. The use of surface measurements allowed the inverse system to compensate for potential systematic biases in satellite retrievals. They found, in particular, that a coupled inversion yields significantly larger tropical emissions compared to the *a priori* estimates or the inversion estimates based on the surface measurements only. These discrepancies have been reduced with corrected spectroscopic data that lead to more accurate CH<sub>4</sub> retrievals as discussed in Section 9.2.1. Meirink et al. (2008a) demonstrated the advantage of 4D-var in reducing aggregation errors by optimising emissions at the grid scale of the transport model using the 1 year surface observations of CH<sub>4</sub>, whereas Meirink et al. (2008b) applied the 4D-var system to analyse SCIAMACHY observations with a focus over South America.

Recently, Bloom et al. (2010) have developed a simple model to combine satellite observations of CH<sub>4</sub> from SCIAMACHY and of gravity anomalies from the Gravity recovery and Climate Experiment (GRACE) satellite, used as proxy for water table depth, together with surface temperature field. Using *a priori* information about rice paddy distribution to isolate wetland regions from their emission estimates they found that tropical wetlands contribute 52–58% of global emissions, with the remainder coming from the outside the tropics. They also estimated a 7% rise in wetland CH<sub>4</sub> emissions over the period 2003–2007, due to warming of mid-latitude and Arctic wetland regions, a figure consistent with recent changes in atmospheric CH<sub>4</sub>.

### 9.3.3 *Need for Future Developments*

Atmospheric inverse modelling based on satellite data is a new field of research, and it is obvious that the potential for satellite measurements in the given context is yet far from fully exploited. Probable future developments could include an applicable extension of a list of the atmospheric constituents, both gaseous and aerosol, for which sources can be estimated from observations (for example, VOC emissions from CHOCHO and/or HCHO observations, CO<sub>2</sub>, SO<sub>2</sub> emissions, etc.). Inverse methods (in particular, Bayesian probabilistic methods) are already widely used in retrievals of satellite data. Future developments should aim at combining satellite retrieval and emission inversion into a coherent framework, because both rely on atmospheric models. Inversions of satellite measurements by means of atmospheric models can also help in estimating parameters of atmospheric processes other than emissions, particularly loss processes in regions without noticeable emissions or when sources have known (say, weekly) patterns of variation. An interesting demonstration of the

potential of satellite observations in this sense was provided by Beirle et al. (2003) who estimated the atmospheric lifetime of nitrogen oxides by using GOME NO<sub>2</sub> measurements and a simple box model approach. In addition to chemical sinks, wet and dry deposition rates could also be inverted in future studies. Certainly, observations from a geostationary platform would allow a much finer analysis: in particular, the diurnal variability of emissions or other parameters could be inverted.

## 9.4 Data Assimilation

### 9.4.1 *Objectives and State of the Art Approaches*

Coupling models with data in a mathematically sound fashion inevitably requires data assimilation (DA) (Chapter 7). There is a dual requirement in confronting models with data: either, improved forecasting, or control of consistency between observations and model results. Improved forecasting is expected from the assimilation of meteorological and atmospheric chemistry observations, thereby creating a dynamically consistent and complete “movie” of optimal quality in some objective sense of estimation theory. Alternatively, from a scientific viewpoint, evidence is provided about whether model results and measurements are mutually consistent within predefined margins, corroborating or rejecting our system knowledge as sufficient (Bennett 2002).

Advanced DA algorithms incorporate the following subtasks: (1) Filtering the signal from noisy observations, (2) interpolation in space and time, and (3) completion of state variables that are not sampled by the observation network (Cohn, 1997). By doing this, DA serves the classical objective to estimate complete parameter fields from sparse data by chemical and physical laws, estimates forcing-fields acting on the system under investigation, tests scientific hypotheses, helps design optimal observation system configurations, and solves mathematically ill-posed modelling problems (Bennett 1992). DA can further be applied for data validation, field experiments, and climate signal detection.

DA algorithms must be designed to satisfy objective quality criteria. Spatio-temporal DA or inversion techniques are candidates for advanced methods, which are able to combine model information with data in a consistent way, while, at the same time, are able to provide a Best Linear Unbiased Estimate (BLUE), i.e. a linear unbiased estimator of the data to be assimilated by the model having the smallest dispersion matrix. Past attempts to analyse tracer fields were based on monovariate kriging techniques in the troposphere (a basic version of BLUE, Fedorov (1989)), and other purely spatial methods in the stratosphere (Stajner et al. 2001; Struthers et al. 2002).

In many cases, these approaches are equivalent to Optimal Interpolation (OI) (Daley 1991) and satisfy the BLUE property, on the spatial scale but not the temporal scale. The latter fact implies that repeated applications of OI do not

force models to maintain a BLUE compatible evolution, which means that chemical imbalances, much larger than observed, cannot be avoided.

Generally, chemical DA methods produce chemical state estimates, frequently referred to as *analyses*, typically on the model grid, after assimilation of observations in model-simulated fields as background information. In contrast to meteorological conditions, chemical DA deals with a vast manifold of chemical species (in the order of 100 species in CTMs) and the tiny number of different observed compounds (less than 5 in most cases: O<sub>3</sub>, NO<sub>2</sub>, particulate matter, CO, SO<sub>2</sub>). Therefore a chemical model to be used as constraint is a key for chemical DA. For this spatio-temporal DA, the BLUE property is provided by two main families of techniques, the 4D-var DA algorithm and the Kalman filter.

For 4D-var, a first successful demonstration was provided by Fisher and Lary (1995) applying a stratospheric chemical box model with a small number of constituents. The authors assessed the applicability of a variational DA method for atmospheric chemistry applications. Eskes et al. (1999) applied the variational method to a two-dimensional model for the assimilation of total satellite columns. For the troposphere, the usefulness of the variational method has been shown by Elbern et al. (1997), applying the box model version of the chemistry mechanism RADM (Regional Acid Deposition Model) (Stockwell et al. 1990). Further, the successful extension to a full chemical 4D-var DA system was demonstrated in the context of identical twin experiments and for an O<sub>3</sub> case study (Elbern and Schmidt 2001), using the University of Cologne EURAD regional CTM. Additional chemistry applications of the 4D-var technique were provided for both the troposphere (Chai et al. 2006) and the stratosphere (Errera and Fonteyn 2001).

Comprehensive DA setups for the troposphere have to account for the fact that, in contrast to stratospheric constituent DA and general meteorological DA, after some time of integration, the evolution of the tropospheric model state is not primarily controlled by the initial state. Rather, emission rates act as strong controlling factors, and exert a direct influence even over short timescales (ranging from seconds close to sources to days in remote areas). Furthermore, emission rates are still not sufficiently well known. Thus, emission rates must be considered as one parameter to be optimised in the DA process. More generally, the parameter to be optimised by DA must be tailored to the simulation objectives. To this end, those parameters must be chosen for optimisation by DA, which exert a strong influence on the simulation or forecast skill and, at the same time, are not sufficiently well known. Given a model parameter, the degree of priority for optimisation is indicated by the product of impact on forecast skill, as quantified by suitable sensitivity tests, and the paucity of knowledge, as quantified by error margins.

For tropospheric chemistry DA, a generalisation with respect to emissions needs to be implemented. This can be done in the incremental formulation of 4D-var by augmenting the state vector by inclusion of deviations from the underlying emission inventory, as well as deviations from a background chemical state, as shown by the work of Elbern and Schmidt (2001) and Elbern et al. (2007) with EURAD-IM. Details on this approach are given in the Appendix, Section 9.6.

An example of a non-chemical implementation of DA, targeted at optimising sources and sinks of CO<sub>2</sub>, is provided by the CarbonTracker system (Peters et al., 2007).

#### **9.4.2 Example Results for Tropospheric O<sub>3</sub> assimilation**

Since most of the O<sub>3</sub> is in the stratosphere, obtaining accurate tropospheric O<sub>3</sub> measurements from space is challenging. The combination of observations with CTM results is a suitable procedure to obtain global estimates of tropospheric O<sub>3</sub>. Two examples are presented here. de Laat et al. (2009) estimated the tropospheric O<sub>3</sub> columns by the subtraction of assimilated O<sub>3</sub> profile observations from total column observations, the so-called Tropospheric O<sub>3</sub> Re-Analysis or TORA method. They evaluated the tropospheric O<sub>3</sub> columns so derived, with space-borne O<sub>3</sub> observations. Six years (1996–2001) of ERS-2 GOME/TOMS total O<sub>3</sub> and GOME O<sub>3</sub> profile observations have been used in the TM5 model with a linearised chemistry parameterisation for the stratosphere.

GOME O<sub>3</sub> profile observations improve the comparisons between model results and ozone-sondes in the tropical UTLS region but slightly degrade the comparisons in the extra-tropical UTLS for both day-to-day variability and monthly means. The large ground pixel size of the GOME O<sub>3</sub> measurements (960 × 100 km) in combination with retrieval and calibration errors have been suggested to be the main causes of this degradation. Results are expected to improve with higher resolution observations from space.

Stajner et al. (2008) included retrievals from the MLS and the OMI on EOS-Aura in the GEOS-4 O<sub>3</sub> data assimilation system to derive tropospheric O<sub>3</sub>. Independent ozone-sondes and MOZAIC data were used for evaluation of the tropospheric O<sub>3</sub> columns. In the troposphere, OMI and MLS provide constraints on the O<sub>3</sub> column, but the O<sub>3</sub> profile shape results from the parameterised O<sub>3</sub> chemistry and the resolved and parameterised transport. Assimilation of OMI and MLS data improves tropospheric column estimates in the Atlantic region but leads to an overestimation in the tropical Pacific, as well as an underestimation in the northern high and middle latitudes in winter and spring. Comparisons of assimilated tropospheric O<sub>3</sub> columns with ozone-sonde data reveal differences of 2.9–7.2 Dobson Units (DU), which are smaller than the model-sonde differences of 3.2–8.7 DU.

Geer et al. (2006) analysed eleven sets of O<sub>3</sub> from seven different DA systems. In most analyses, MIPAS O<sub>3</sub> data are assimilated; two studies assimilate SCIAMACHY observations instead. Analyses are compared to independent O<sub>3</sub> observations (e.g. ozone-sondes) covering the troposphere, stratosphere and lower mesosphere during the period July to November 2003. Where the model results diverge, the main explanation was the way O<sub>3</sub> was modelled. Two analyses used numerical weather prediction (NWP) systems based on general circulation models (GCMs); the other five used CTMs. The systems examined contain either linearised

or detailed  $O_3$  chemistry, or no chemistry at all. The result points to the need of further physically based model improvements.

$O_3$  assimilation also offers the possibility of providing more accurate initial guesses for  $O_3$  retrieval algorithms than are currently available. Other major applications of  $O_3$  assimilation with regard to the troposphere are:

- The provision of vertically resolved global maps of  $O_3$ ;
- An improvement in the radiative transfer calculations needed to retrieve information from many satellite instruments that require accurate representation of  $O_3$ ;
- Improvement in the predictions of UV radiation fluctuations at the surface of the Earth, since UV is absorbed by  $O_3$  in the atmosphere; and
- Provision of constraints on other observed constituents that are affected by  $O_3$  chemistry (Rood 2007).

### 9.4.3 Example Results for $NO_2$ Tropospheric Column Assimilation

Retrieval results from tropospheric  $NO_2$  columns are ingested into the model by means of averaging kernels, where the observation operator, commonly denoted as  $\mathbf{H}$  (see Appendix, Section 9.6), is constructed by the scalar product of the averaging kernel with the  $NO_2$  molecular density of the model profile (Eskes et al. 2005). The average fraction of the averaging kernel at the surface is roughly 10% of the maximum amplitude in most cases (Fig. 9.11). Recalling the significance of the averaging kernel shape as a sensitivity profile, corrections by the DA procedure are implemented in the same proportion as the sensitivity. The practical meaning is that the modification of

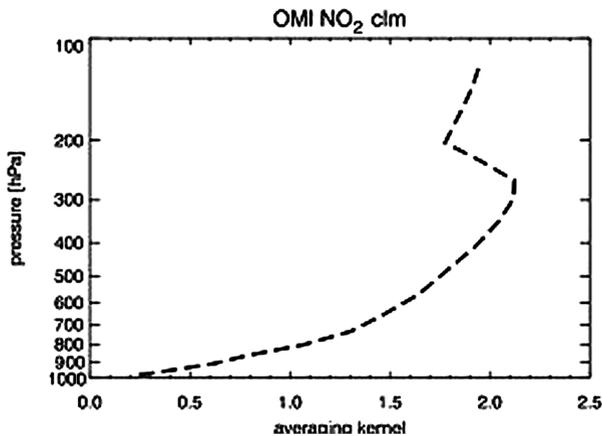
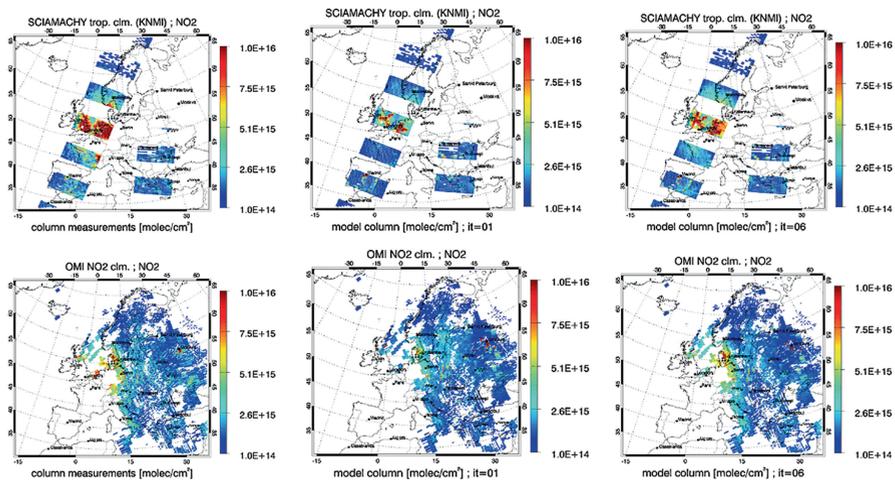


Fig. 9.11 Mean averaging kernel over the European continental scale model domain for a 2 week case study.

the surface layer is hardly affected by this correction in comparison to the free troposphere. This must be remembered when considering the following results.

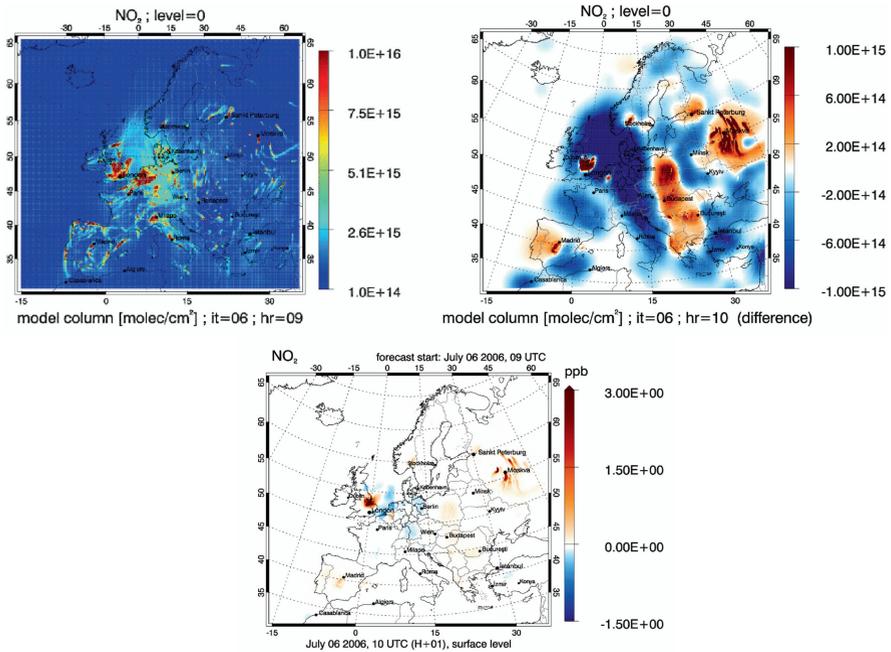
An example of combined assimilation of SCIAMACHY and OMI NO<sub>2</sub> tropospheric columns with the EURAD-IM CTM for July 2006 is given for illustration. Both SCIAMACHY and OMI satellite retrievals from KNMI were assimilated with averaging kernels, using error information from the appropriate data provider. In an attempt to provide a horizontal model resolution comparable to the minimum OMI 24 × 13 km<sup>2</sup> footprints, the horizontal model resolution was refined to 15 × 15 km<sup>2</sup>. In this case study, the DA configuration has a time window of 3 h (09:00–12:00 UTC) to include SCIAMACHY satellite sensor data with a late morning overpass over Europe and OMI overpass in the early afternoon over eastern Europe. A longer assimilation window turned out to be unaffordable at 15 × 15 km<sup>2</sup> horizontal resolution. After assimilation, a 24-h forecast is made, starting at 09:00 UTC. The analysis produced by the assimilation is the initial field, and an emission rate correction factor is applied. Numerical experiments suggest that about 10 iterations are sufficient to ensure convergence to the observations. After assimilation, an *a posteriori* analysis was performed (Talagrand 2003).

The affordable short assimilation interval from 9–12 UTC enforced a fairly disjointed footprint pattern for SCIAMACHY and OMI. While the former covers the western model domain due to its late morning orbit, the latter covers eastern parts due to its early afternoon overpass. Fig. 9.12 illustrates these conditions, along with retrievals (**y**), forecasted retrievals of NO<sub>2</sub> columns (**Hx<sub>b</sub>**), and analysed tropospheric NO<sub>2</sub> columns (**Hx<sub>a</sub>**). The analysis result can be clearly identified as a weighted combination of all information sources, retrievals and forecast acting as background information.



**Fig. 9.12** Comparison of NO<sub>2</sub> tropospheric columns in molecules/cm<sup>2</sup> for 6th July 2006. *Left panel* column: KNMI retrieved and assimilated values (**y**); *middle panel* column: EURAD forecasted values (**Hx<sub>b</sub>**); *right panel* column analyses (**Hx<sub>a</sub>**).

The effect of NO<sub>2</sub> column DA for 6th July 2006 is presented in Fig. 9.13. The field obtained is supplemented by difference fields for the tropospheric columns and the concentrations. Clearly major increments can be observed in western England and in the area of north-western Russia. Both these signals are visible for surface concentrations.



**Fig. 9.13** Data assimilation result in terms of tropospheric columns for 6th July 2006. NO<sub>2</sub> model columns based on OMI and SCIAMACHY assimilation within the assimilation interval, 9–12 UTC. Units in molecules/cm<sup>2</sup> (*left panel*). Difference field giving implied changes for tropospheric columns by assimilation (*right panel*), and induced surface concentration changes by NO<sub>2</sub> in ppb (*bottom panel*).

#### 9.4.4 Aerosol Satellite Data Assimilation

A 4D-var approach for aerosol modelling is hampered by the construction of the adjoint model, as the algorithm contains numerous cases which can not be differentiated. Hence, no adjoint of a full-fledged aerosol model with a 3D model is available at present. Sandu et al. (2004) gives a discussion on numerical aspects; however, a 3D-var example is presented here. In addition to difficulties of modelling the variety of processes affecting aerosols, aerosol DA is further hampered by the fact, that the only measurements of lumped particulate matter abundances or aerosol optical depth are possible. Neither quantity is modelled as a space state variable as such.

Enforced by the sparseness of data sources of a single type, the combination of information from remote sensing and *in situ* measurement, that is the use of heterogeneous data sources, is therefore crucial for success. Unlike conditions with satellite retrieved NO<sub>2</sub> tropospheric columns, in satellite retrievals for aerosol data the strongest signals originate from lower atmospheric layers. However, as optical information must be processed, the central problem for DA is bridging information from the optical to the chemical domain. An example of a recent discussion on MODIS data inverse modelling is given in Dubovik et al. (2008). Strictly, related observation operators have to be devised for ingestion of the information.

Aerosol data used in this study are obtained from both satellite retrievals and routine *in situ* observations.

(a) Aerosol data products retrieved from satellite:

SYnergetic Aerosol Retrieval (SYNAER) satellite retrievals: Satellite data are from retrievals based on the SYNAER method (Holzer-Popp et al. 2002a; 2003b). The retrieval procedure has been developed to make a synergistic use of simultaneous GOME and ATSR-2 measurements. It has then been modified to use SCIAMACHY and AATSR data. The retrieval principle is to utilise two complementary properties of (A)ATSR (high spatial resolution) and GOME/SCIAMACHY (high spectral resolution). It is able to deliver various PM<sub>x</sub> as integrated values.

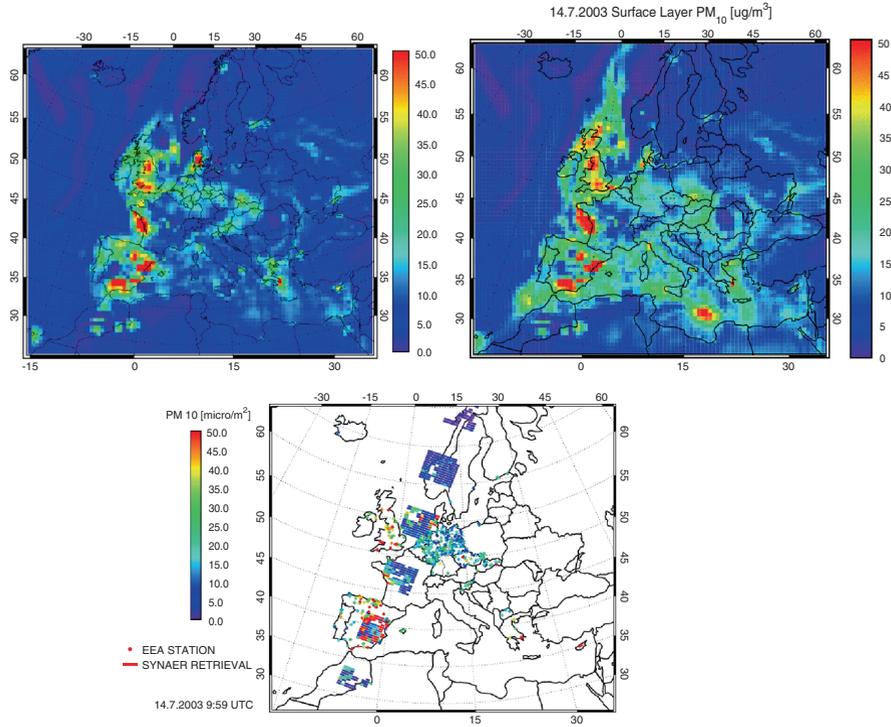
(b) *In situ* observations:

The European Environmental Agency collected aerosol data that originate from about 445 routine measurement sites, operated by national and regional environmental protection agencies. From these data, available from the EEA database AirBase, hourly PM<sub>10</sub> mass concentrations have been used.

The algorithm is highly adaptable to non-isotropic and inhomogeneous aerosol radii of influence, while at the same time, it allows for efficient processing of ensemble model runs for background error covariance matrix estimation. For the study presented here, aerosol dynamics are perturbed to generate ensembles with nine members, modifying gaseous precursors and direct aerosol emission independently.

The demonstration case features 14th July 2003 (Fig. 9.14). The SCIAMACHY footprint track covers parts of the Iberian Peninsula, where the retrievals show elevated PM10 levels, which proved to be biomass burning products from large wildfire events at that time. This case has been selected to demonstrate the capabilities and limits for assimilation under conditions of unpredictable emission events. The special challenge is that extraordinary events like wildfires engender higher level aerosol maxima, which usually differ substantially from modelled profiles, with profile maxima often higher than usual.

Table 9.1 quantifies the beneficial impact of satellite DA in addition to *in situ* DA, with independent *in situ* data, (i.e. direct surface measurements) withheld from the assimilation procedure. A substantial improvement can be claimed for typical conditions (without measured notable wildfires) signified by reduction of root mean



**Fig. 9.14** Signal duration of SYNAER data assimilation results. Comparison of simulated PM10 contents: assimilation based analysis after assimilation 14th July 2003, at 10:00 UTC without any previous assimilation (*top left panel*), and with data assimilation processing continuously from 1st to 14th July 2003 (*top right panel*). *In situ* and satellite/SYNAER observation base available for 14th July 2003 (*bottom panel*). The accumulation of information during the 2 weeks is clearly visible from the discrepancies.

**Table 9.1** Root mean square errors for assimilation fields for two consecutive dates validated by unassimilated *in situ* observations within satellite footprints. Improvements are given with respect to no assimilation

Date	Number of ground stations withheld	Special characteristics	RMS error [ $\mu\text{g m}^{-3}$ ] (% improvement)		
			No assimilation	<i>In situ</i> data	Satellite + <i>in situ</i>
13th July 2003	20	None	6.3	2.5 (61%)	1.9 (70%)
14th July 2003	49	Wildfires	7.3	5.5 (24%)	5.3 (28%)

square errors, especially with SYNAER satellite data. In the case of wildfires, the improvement is markedly reduced, but still clearly visible. In both cases, satellite data from SYNAER demonstrate efficient support for DA improvements.

## 9.5 Summary: Perspectives

Earth system models, describing dynamical, physical, chemical and biological processes which determine the conditions at the Earth's surface, have been evolving over the past decades. The increasing computer power is facilitating more accurate simulation of current and past conditions and improving our confidence in the accuracy of prediction. The latter is required for policymakers in their quest to achieve legislation facilitating sustainable environmental conditions. Global data sets are required to test the ability of our Earth system models.

In this context the synergistic use of satellite observations with chemistry-transport modelling has opened new horizons in air pollution control and climate change evaluation. However, models and observations from satellite platforms having *higher spatial and temporal resolution* are required in order to attribute better the air pollution to its different sources. The signals from these sources will not be diluted or mixed with other sources in large ground or satellite observations or larger grid boxes of the CTMs. *Improvements* are also needed in the *temporal sampling* of the atmosphere by the satellite based sensors, which has to be taken into account in the models when satellite retrievals are used.

A *continuous dialogue* between satellite retrieved observations of atmospheric trace constituents and model results enables the improvement of: (1) process understanding, (2) the models, and (3) retrieval algorithms, in order to construct a precise picture of our changing atmosphere.

*Inverse modelling* can contribute to the development of a general observation-based methodology for estimating parameters of the atmosphere that cannot be observed directly. Inverse modelling combining both traditional, i.e. ground-based and airborne (aircraft, balloon) observations and satellite observations can contribute to improving emission inventories of trace constituents, relevant to air pollution and climate change. It also allows verification, by space-based observations, of control strategies for atmospheric emissions.

*Data assimilation* has a unique role to play with respect to forecasting conditions. It has also been shown that it is important for *filtering* the signal from noisy observations, *interpolation* in space and time, and *completion* of state variables that are not sampled by the observation network. Data assimilation can further be applied for data validation, field experiments, and climate signal detection. The assimilation of satellite observations is a critical step forward in impact modelling for scientific and political decisions with regard to environmental changes. It contributes critically to the improvement of prognostic models, such as those used for weather forecasts and more recently for chemical weather forecasts. Such results are of relevance for society (public services) since they *increase the accuracy of the predictions* and our confidence to them.

## 9.6 Appendix

### Inverse Modelling: Principles

The following equation provides a relationship between  $\mathbf{y}$ , a vector grouping a set of observations, and  $\mathbf{x}$ , the vector of variables to be determined (called *state variables*, e.g. emissions), through applying a *forward* model  $\mathbf{F}$  with fixed parameters  $\mathbf{b}$ :

$$\mathbf{y} = \mathbf{F}(\mathbf{x}, \mathbf{b}). \quad (9.1)$$

By inverting this model, analytically or numerically, we can obtain  $\mathbf{x}$  for a given  $\mathbf{y}$ . As errors are associated with both  $\mathbf{x}$  and  $\mathbf{y}$ , a probabilistic treatment of the inversion problem is necessary. Its starting point is the Bayesian theorem.

$$P(\mathbf{x}|\mathbf{y}) = \frac{P(\mathbf{y}|\mathbf{x})P(\mathbf{x})}{P(\mathbf{y})}. \quad (9.2)$$

In the context of an inverse modelling problem,  $P(\mathbf{x})$  and  $P(\mathbf{y})$  represent the probability distribution functions (pdf) for the state vector  $\mathbf{x}$  and the observation vector  $\mathbf{y}$ . In Bayesian framework,  $P(\mathbf{y}|\mathbf{x})$  represents the conditional probability to observe a particular vector  $\mathbf{y}$  for a given value of the set of state variables  $\mathbf{x}$ . While  $P(\mathbf{x})$  represents the *a priori* information on the state vector (for example an emission inventory with a given uncertainty), then  $P(\mathbf{x}|\mathbf{y})$  represents the *a posteriori* probability that the state vector was  $\mathbf{x}$  when  $\mathbf{y}$  was measured reflecting the additional information from measurements. This latter pdf can be used to find a maximum likelihood for an *a posteriori* estimate (MLAP) of  $\mathbf{x}$  and its uncertainty.

For simplicity, we illustrate the main steps to derive a MLAP estimate for a scalar linear case first. Let  $x_a$  be the *a priori* estimate of the state variable,  $\sigma_a$  its uncertainty (assumed to be normal),  $y$  an observation with uncertainty  $\sigma_e$  including both the error in observations and in the forward model (due to the model formulation and uncertainty in model parameters  $b$ ), and  $k$  a the forward model linking  $x$  and  $y$ .

If uncertainties in  $x_a$  and  $y$  are distributed in accordance to a normal distribution,  $P(\mathbf{x}|\mathbf{y})$  are proportional to an exponential expression combining two terms (Eq. 9.3): the departure of an updated state variable from the *a priori*, and the difference between observations  $y$  and simulated values  $kx$  for a given  $x$ . Combining these two terms, weighed by the respective inverse error variances, is exactly the principle of optimal estimation.

$$P(\mathbf{x}|\mathbf{y}) \sim \exp \left[ -\frac{(x - x_a)^2}{2\sigma_a^2} - \frac{(y - kx)^2}{2\sigma_e^2} \right] \quad (9.3)$$

In order to determine the maximum value for  $P(x/y)$ , we define a *cost function*  $J(x)$  given by the terms in brackets in Eq. 9.3. The solution to  $\partial J/\partial x = 0$  gives the optimal estimate  $\hat{x}$ :

$$\hat{x} = x_a + g(y - kx_a). \quad (9.4)$$

where  $g$  is the so called gain factor, which weights the contribution of observations  $y$  to  $\hat{x}$ . It is given by:

$$g = \frac{k\sigma_a^2}{k^2\sigma_a^2 + \sigma_\varepsilon^2} \quad (9.5)$$

For  $\sigma_\varepsilon/k \gg \sigma_a$ ,  $g$  tends to 0, for  $\sigma_\varepsilon/k \ll \sigma_a$ ,  $g$  tends to 1. This latter case is obviously advantageous for inverse modelling because then observations bring much new information significantly reducing the *a posteriori* uncertainty,  $\hat{\sigma}^2$ , given by Eq. 9.6:

$$\frac{1}{\hat{\sigma}^2} = \frac{1}{\sigma_a^2} + \frac{1}{(\sigma_\varepsilon/k)^2}. \quad (9.6)$$

Note however, that  $\hat{\sigma}^2 < \sigma_a^2$  always holds, so even *uncertain* observations add *some* information.

In the more general case of a set of state variables to be optimised (for example emissions at different locations or times) from a set of observations, we need to reformulate equations 9.3–9.6 in matrix form. The cost function  $J(x)$  becomes:

$$J(x) = (x - x_a)^T S_a^{-1} (x - x_a) + (y - Kx)^T S_\varepsilon^{-1} (y - Kx) \quad (9.7)$$

Again, an optimal solution for  $x$  can be found by differentiating  $J$  with respect to  $x$ . However, several practical problems occur. The *a priori* and observational error terms  $S_{ij}$  constituting the matrices  $S_a$  and  $S_\varepsilon$  are in general not well known. Very often, non-diagonal terms representing error correlations for different measurements or state variables are simply neglected. Subjective estimates for the diagonal error terms are made, which can affect the weighting procedure in the optimal estimates.

In Eq. 9.7,  $K$  represents the so called Jacobian matrix with elements  $k_{ij} = \delta y_i / \delta x_j$  indicating the sensitivity of a measurement at  $i$  to a state variable at  $j$ . It can be solved by running an atmospheric model and varying  $x_j$  which needs  $n + 1$  runs,  $n$  being the dimension of  $x$ . For the general case of a non-linear model, this procedure has to be repeated several times starting with  $x_a$ , and iterating for each new update of  $x$ . Two types of solutions are applied to avoid a large computational burden for large  $n$ : (1) the problem is decomposed for sub-domains where each small  $n$ , or state variables are aggregated for large regions reducing their number; and (2) use of the adjoint model (Giering and Kaminski 1998) allowing a calculation in one backward

run all derivatives  $\delta J/\delta x_1 \dots \delta J/\delta x_n$ , and thus to avoiding explicit evaluation of  $\mathbf{K}$ . This procedure has to be applied iteratively (using for example steepest descent algorithms) until an absolute minimum of the cost function  $J$  is found. The adjoint method is also known as 4D-var assimilation, because observations can be distributed over a time window, and is in widespread use in meteorology when a task is to optimise initial conditions. For the implementation of the 4D-var approach, a distance function or objective function, which penalises both discrepancies with observations and *a priori* knowledge of emission rates and initial values, also called cost function, may be defined as follows:

$$J(\mathbf{x}, \mathbf{e}(t)) = \frac{1}{2}(\mathbf{x}_b - \mathbf{x}(0))^T \mathbf{B}^{-1}(\mathbf{x}_b - \mathbf{x}(0)) + \frac{1}{2} \int_0^N (\mathbf{e}_b(t) - \mathbf{e}(t))^T \mathbf{K}^{-1}(\mathbf{e}_b(t) - \mathbf{e}(t)) dt + \frac{1}{2} \int_0^N (\mathbf{y}(t) - \mathbf{H}\mathbf{x}(t))^T \mathbf{R}^{-1}(\mathbf{y}(t) - \mathbf{H}\mathbf{x}(t)) dt \quad (9.8)$$

where  $J$  is a scalar functional defined on the time interval  $0 \leq t \leq N$  dependent on the vector valued state variable  $\mathbf{x}(t)$ , and a parameter  $\mathbf{e}(t)$  to be optimised. Here, observations  $\mathbf{y}$  are compared with their model equivalent  $\mathbf{H}\mathbf{x}$  at time  $t$ , with the operator  $\mathbf{H}$  being the forward observation operator. The error covariance matrices of the forecast  $\mathbf{x}_b$ , the first guess or background emission rates  $\mathbf{e}_b(t)$  and observations  $\mathbf{y}(t)$  are denoted  $\mathbf{B}$ ,  $\mathbf{K}$  and  $\mathbf{R}$ , respectively. The CTM with inclusion of emissions is given by  $d\mathbf{x}/dt = \mathbf{M}(\mathbf{x}) + \mathbf{e}$ , where  $\mathbf{M}$  acts as a generally non-linear model operator and  $\mathbf{e}$  is in our case the vector of emission rates. Both terms uniquely define the state variable  $\mathbf{x}(t)$  at time  $t$ , after an ever fixed initial state  $\mathbf{x}(0)$  is provided.

The variational chemistry DA algorithm is composed by four components: (1) the forward model, (2) the adjoint of its tangent linear version, (3) the background error covariance matrix, making use of the diffusion paradigm (Weaver and Courtier 2001) for anisotropic and inhomogeneous radii of influence, and (4) the minimisation routine, where the quasi-Newton (L-BFGS) method is selected. Further numerical and implementation details are given in Elbern et al. (2007).

An alternative to the 4D-var approach is Kalman filtering (KF) with the theoretical ability to update the forecast error covariance matrix  $\mathbf{B}$  and analysis error covariance matrix  $\mathbf{A}$ . Given the model and analysis space state dimension of order  $10^{6-7}$  in today's CTMs, direct application of KF is far beyond feasibility (and will always be so), and sophisticated numerical complexity reduction measures must be devised. In sloppy parlance often sequential applications of optimal interpolation are termed "reduced KF", despite the fact that neither  $\mathbf{B}$  is updated nor  $\mathbf{A}$  computed. In contrast, Kalman filter implementations with sophisticated complexity reduction techniques are still very rare. Advanced implementations are presented by van Loon et al. (2000), where a reduced rank square-root approach (RRSQR-KF) was selected to factorise covariance matrices by a few principal components (Verlaan

and Heemink, 1995). Further elaboration on this technique by combination with an ensemble Kalman filter method (En-KF) resulted in additional skill (Hanea et al., 2004). Optimisation parameters include emission rates, photolysis rates, and deposition rates, the correction quantities of which are formally introduced as “noise” parameters in the Kalman filter formulation.

## References

- Arnold, S. R., D. V. Spracklen, J. Williams, N. Yassaa, J. Sciare, B. Bonsang, V. Gros, I. Peeken, A. C. Lewis, S. Alvaïn, and C. Moulin, 2009. Evaluation of the global oceanic isoprene source and its impacts on marine organic carbon aerosol, *Atmos. Chem. Phys.*, **9**, 1253–1262.
- Arellano, A. F., Jr., and P. G. Hess, 2006. Sensitivity of top-down estimates of CO sources to GCTM transport, *Geophys. Res. Lett.*, **33**, L21807, doi: 10.1029/2006GL027371.
- Arellano, A. F., P. S. Kasibhatla, L. Giglio, G. R. van der Werf, J. T. Randerson, and G. J. Collatz, 2006. Time-dependent inversion estimates of global biomass-burning CO emissions using Measurement of Pollution in the Troposphere (MOPITT) measurements, *J. Geophys. Res.*, **111**, D09303, doi: 10.1029/2005JD006613.
- Barret B., P. Ricaud, C. Mari, J.-L. Attié, N. Boussez, B. Josse, E. Le Flochmoën, N. J. Livesey, S. Massart, V.-H. Peuch, A. Piacentini, B. Sauvage, V. Thouret, and J.-P. Cammas, 2008. Transport pathways of CO in the African upper troposphere during the monsoon season: a study based upon the assimilation of spaceborne observations, *Atmos. Chem. Phys.*, **8**, 3231–3246.
- Bennett, A. F., 1992. *Inverse methods in physical oceanography*, Cambridge University Press, 347 pp.
- Bennett, A. F., 2002. *Inverse modelling of the ocean and the atmosphere*, Cambridge University Press, 234 pp.
- Beirle, S., U. Platt, M. Wenig, and T. Wagner, 2003. Weekly cycle of NO<sub>2</sub> by GOME measurements: A signature of anthropogenic sources, *Atmos. Chem. Phys.*, **3**, 2225–2232.
- Beirle, S., U. Platt, R. von Glasow, M. Wenig, and T. Wagner, 2004. Estimate of nitrogen oxide emissions from shipping by satellite remote sensing, *Geophys. Res. Lett.*, **31**, L18102, doi: 10.1029/2004GL020312.
- Bergamaschi, P., and P. Bousquet, 2008. Estimating Sources and Sinks of Methane: An Atmospheric View, in *The Continental-Scale Greenhouse Gas Balance of Europe*, edited by A. J. Dolman, et al., pp. 113–133, Springer, New York.
- Bergamaschi, P., C. Frankenberg, J. F. Meirink, M. Krol, F. Dentener, T. Wagner, U. Platt, J. O. Kaplan, S. Körner, M. Heimann, E. J. Dlugokencky, A. Goede, 2007. Satellite cartography of atmospheric methane from SCIAMACHY on board ENVISAT 2. Evaluation based on inverse model simulations, *J. Geophys. Res.*, **112**, D02304, doi: 10.1029/2006JD007268.
- Bertram, T. H., A. Heckel, A. Richter, J. P. Burrows, R. C. Cohen, 2005. Satellite measurements of daily variations in soil NO<sub>x</sub> emissions, *Geophys. Res. Lett.*, **32**(24), L24812.
- Bloom, A.A., P. I. Palmer, A. Fraser, D. S. Reay, Ch. Frankenberg, 2010. Large-Scale Controls of Methanogenesis Inferred from Methane and Gravity Spaceborne Data, *Science* **327**, 322–325, DOI: 10.1126/science.1175176
- Boersma, K. F., Jacob, D. J., Trainic, M., Rudich, Y., DeSmedt, I., Dirksen, R., and Eskes, H. J., 2009. Validation of urban NO<sub>2</sub> concentrations and their diurnal and seasonal variations observed from the SCIAMACHY and OMI sensors using *in situ* surface measurements in Israeli cities, *Atmos. Chem. Phys.*, **9**, 3867–3879.
- Chai, T. F., G. R. Carmichael, A. Sandu, Y. H. Tang, and D. N. Daescu, 2006. Chemical data assimilation of Transport and Chemical Evolution over the Pacific (TRACE-P) aircraft measurements, *J. Geophys. Res.*, **111**, D02301, doi:10.1029/2005JD005883.
- Cohn, S., 1997. An introduction to estimation theory, *J. Met. Soc. Japan*, **75**(1B), 257–288.

- Daley, R., 1991. *Atmospheric Data Analysis*, Cambridge University Press, Cambridge, ISBN: 9780521382151.
- Deeter, M. N., L. K. Emmons, G. L. Francis, D. P. Edwards, J. C. Gille, J. X. Warner, B. Khattatov, D. Ziskin, J.-F. Lamarque, S.-P. Ho, V. Yudin, J.-L. Attié, D. Packman, J. Chen, D. Mao, J. R. Drummond, 2003, Operational carbon monoxide retrieval algorithm and selected results for the MOPITT instrument, *J. Geophys. Res.*, **108** (D14), 4399, doi: 10.1029/2002JD003186.
- de Laat, A. T. J., R. J. van der A, and M. van Weele, 2009. Evaluation of tropospheric ozone columns derived from assimilated GOME ozone profile observations, *Atmos. Chem. Phys.*
- Dentener, F. J., D. Stevenson, K. Ellingsen, T. van Noije, M. Schultz, M. Amann, C. Atherton, N. Bell, D. Bergmann, I. Bey, L. Bouwman, T. Butler, J. Cofala, B. Collins, J. Drevet, R. Doherty, B. Eickhout, H. Eskes, A. Fiore, M. Gauss, D. Hauglustaine, L. Horowitz, I.S.A. Isaksen, B. Josse, M. Lawrence, M. Krol, J.F. Lamarque, V. Montanaro, J.F. Müller, V. H. Peuch, G. Pitari, J. Pyle, S. Rast, J. Rodriguez, M. Sanderson, N.H. Savage, D. Shindell, S. Strahan, S. Szopa, K. Sudo, R. Van Dingenen, O. Wild, G. Zeng, 2006a. The global atmospheric environment for the next generation, *Environ. Sci. Technol.*, **40**, 3586–3594.
- Dentener, F., S. Kinne, T. Bond, O. Boucher, J. Cofala, S. Generoso, P. Ginoux, S. Gong, J. J. Hoelzemann, A. Ito, L. Marelli, J. Penner, J.-P. Putaud, C. Textor, M. Schulz, G. R. van der Werf, and J. Wilson, 2006b. Emissions of primary aerosol and precursor gases for the years 2000 and 1750 prescribed data-sets for AeroCom, *Atmos. Chem. Phys.*, **6**, 4321–4344.
- Dubovik, O., T. Lapyonok, Y. J. Kaufman, M. Chin, P. Ginoux, R. A. Kahn, and A. Sinyuk, 2008. Retrieving global aerosol sources from satellites using inverse modeling, *Atmos. Chem. Phys.*, **8**, 209–250.
- Dufour, G., Szopa, S., Barkley, M. P., Boone, C. D., Perrin, A., Palmer, P. I., and Bernath, P. F.: Global upper-tropospheric formaldehyde: seasonal cycles observed by the ACE-FTS satellite instrument, *Atmos. Chem. Phys.*, **9**, 3893–3910, 2009.
- Edwards, D. P., L. K. Emmons, D. A. Hauglustaine, D. A. Chu, J. C. Gille, Y. J. Kaufman, G. Pétron, L. N. Yurganov, L. Giglio, M. N. Deeter, V. Yudin, D. C. Ziskin, J. Warner, J.-F. Lamarque, G. L. Francis, S. P. Ho, D. Mao, J. Chen, E. I. Grechko, and J. R. Drummond, 2004. Observations of carbon monoxide and aerosols from the Terra satellite: Northern Hemisphere variability, *J. Geophys. Res.*, **109**, 24202, doi: 10.1029/2004JD004727.
- Elbern, H. and H. Schmidt, 2001. Ozone episode analysis by four-dimensional variational chemistry data assimilation, *J. Geophys. Res.*, **106**, 3569–3590.
- Elbern, H., H. Schmidt, and A. Ebel, 1997. Variational data assimilation for tropospheric chemistry modeling, *J. Geophys. Res.*, **102**, 967–985.
- Elbern, H., A. Strunk, H. Schmidt, and O. Talagrand, 2007. Emission rate and chemical state estimation by 4-dimensional variational inversion, *Atmos. Chem. Phys.*, **7**, 3749–3769.
- Errera, Q. and D. Fonteyn, 2001. Four-dimensional variational chemical assimilation of CRISTA stratospheric measurements, *J. Geophys. Res.*, **106**, 12 253–12 265.
- Eskes, H. J., A. J. M. Piters, P. F. Levelt, M. A. F. Allart, and H. M. Kelder, 1999. Variational assimilation of total-column ozone satellite data in a 2D lat-lon tracer-transport model, *J. Atmos. Sci.*, **56**, 3560–3572.
- Eskes, H. J., R. J. van der A, E. J. Brinkma, J. P. Veefkind, J. F. de Haan, and P. J. M. Valks, 2005. Retrieval and validation of ozone columns derived from measurements of SCIAMACHY on Envisat, *Atmos. Chem. Phys. Discuss.*, **5**, 4429–4475.
- Eyring, V., Stevenson, D. S., Lauer, A., Dentener, F. J., Butler, T., Collins, W. J., Ellingsen, K., Gauss, M., Hauglustaine, D. A., Isaksen, I. S. A., Lawrence, M. G., Richter, A., Rodriguez, J. M., Sanderson, M., Strahan, S. E., Sudo, K., Szopa, S., van Noije, T. P. C., and Wild, O., 2007. Multi-model simulations of the impact of international shipping on Atmospheric Chemistry and Climate in 2000 and 2030, *Atmos. Chem. Phys.*, **7**, 757–780.
- Fedorov, V., 1989. Kriging and other estimators of spatial field characteristics (with special reference to environmental studies), *Atmos. Environ.*, **23**, 174–184.

- Fisher, M. and D. Lary, 1995. Lagrangian four-dimensional variational data assimilation of chemical species, *Q. J. Roy. Meteor. Soc.*, **121**, 1681–1704.
- Franke, K., A. Richter, H. Bovensmann, V. Eyring, P. Jockel, P. Hoor, and J.P. Burrows, 2009. Ship emitted NO<sub>2</sub> in the Indian Ocean: comparison of model results with satellite data, *Atmos. Chem. Phys.*, **9**, 7289–7301.
- Frankenberg, C., T. Warneke, A. Butz, I. Aben, F. Hase, P. Spietz, and L. R. Brown, 2008a. Pressure broadening in the 2ν<sub>3</sub> band of methane and its implication on atmospheric retrievals, *Atmos. Chem. Phys.*, **8**, 5061–5075.
- Frankenberg, C., P. Bergamaschi, A. Butz, S. Houweling, J. F. Meirink, J. Notholt, A. K. Petersen, H. Schrijver, T. Warneke, and I. Aben, 2008b. Tropical methane emissions: A revised view from SCIAMACHY onboard ENVISAT, *Geophys. Res. Lett.*, **35**, L15811, doi:10.1029/2008GL034300.
- Fu, T.-M., D. J. Jacob, P. I. Palmer, K. Chance, Y. X. Wang, B. Barletta, D. R. Blake, J. C. Stanton, and M. J. Pilling, 2007. Space-based formaldehyde measurements as constraints on volatile organic compound emissions in east and south Asia and implications for ozone, *J. Geophys. Res.*, **112**, D06312, doi: 10.1029/2006JD007853.
- Fu, T.-M., D. J. Jacob, F. Wittrock, J. P. Burrows, M. Vrekoussis, and D. K. Henze, 2008. Global budgets of atmospheric glyoxal and methylglyoxal, and implications for formation of secondary organic aerosols, *J. Geophys. Res.*, **113**, D15303, doi: 10.1029/2007JD009505.
- Geer, A. J., W. A. Lahoz, S. Bekki, N. Bormann, Q. Errera, H. J. Eskes, D. Fonteyn, D. R. Jackson, M. N. Juckes, S. Massart, V.-H. Peuch, S. Rharmili, and A. Segers, 2006. The ASSET intercomparison of ozone analyses: method and first results, *Atmos. Chem. Phys.*, **6**, 5445–5474.
- Giering, R. and T. Kaminski, 1998. Recipes for adjoint code construction, *ACM Trans. on Math. Software*, **24**, 437–474.
- Giglio, L., G. R. van der Werf, J. T. Randerson, G. J. Collatz, and P. Kasibhatla, Global estimation of burned area using MODIS active fire observations, *Atmos. Chem. Phys.*, **6**, 957–974, 2006.
- Guenther, A., T. Karl, P. Harley, C. Wiedinmyer, P. I. Palmer, and C. Geron, 2006. Estimates of global terrestrial isoprene emissions using MEGAN (Model of Emissions of Gases and Aerosols from Nature), *Atmos. Chem. Phys.*, **6**, 3181–3210.
- Hanea, R. G., G. J. M. Velders, and A. Heemink, 2004. Data assimilation of ground-level ozone in Europe with a Kalman filter and chemistry transport, *J. Geophys. Res.*, **109**, D10302, doi: 10.1029/2003JD004283.
- Hayn, M., Beirle, S., Hamprecht, F. A., Platt, U., Menze, B. H., and Wagner, T.: Analysing spatio-temporal patterns of the global NO<sub>2</sub>-distribution retrieved from GOME satellite observations using a generalized additive model, *Atmos. Chem. Phys.*, **9**, 6459–6477, 2009
- Holzer-Popp, T., M. Schroedter, and G. Gesell, 2002a. Retrieving aerosol optical depth and type in the boundary layer over land and ocean from simultaneous GOME spectrometer and ATSR-2 radiometer measurements, 2. Case study application and validation, *J. Geophys. Res.*, **107**(D24), 4770, doi: 10.1029/2001JD002777.
- Holzer-Popp, T., M. Schroedter, and G. Gesell, 2002b. Retrieving aerosol optical depth and type in the boundary layer over land and ocean from simultaneous GOME spectrometer and ATSR-2 radiometer measurements, 1. Method description, *J. Geophys. Res.*, **107**(D21), 4578, doi: 10.1029/2001JD002013,
- Hudman, R. C., D. J. Jacob, O. R. Cooper, M. J. Evans, C. L. Heald, R. J. Park, F. Fehsenfeld, F. Flocke, J. Holloway, G. Hubler, K. Kita, M. Koike, Y. Kondo, A. Neuman, J. Nowak, S. Oltmans, D. Parrish, J. M. Roberts, and T. Ryerson, 2004. Ozone production in transpacific Asian pollution plumes and implications for ozone air quality in California, *J. Geophys. Res.*, **109**(D23), 14, doi: 10.1029/2004JD004974.
- Ito, A. and J. E. Penner, 2005. Historical emissions of carbonaceous aerosols from biomass and fossil fuel burning for the period 1870–2000, *Global Biogeochem. Cycles*, **19**, GB2028, doi: 10.1029/2004GB002374.

- Jaeglé, L., R. V. Martin, K. Chance, L. Steinberger, T. P. Kurosu, D. J. Jacob, A. I. Modi, V. Yoboué, L. Sigha-Nkamdjou, C. Galy-Lacaux, 2004. Satellite mapping of rain-induced nitric oxide emissions from soils, *J. Geophys. Res.*, **109**, D21310, doi: 10.1029/2004JD004787.
- Kim, S.-W., A. Heckel, S. A. McKeen, G. J. Frost, E.-Y. Hsie, M. K. Trainer, A. Richter, J. P. Burrows, S. E. Peckham, G. A. Grell, 2006. Satellite-observed U.S. power plant NO<sub>x</sub> emission reductions and their impact on air quality, *Geophys. Res. Lett.*, **33**, L22812, doi: 10.1029/2006GL027749.
- Kim, S.-W., A. Heckel, G. J. Frost, A. Richter, J. Gleason, J. P. Burrows, S. McKeen, E.-Y. Hsie, C. Granier, and M. Trainer, 2009. NO<sub>2</sub> columns in the western United States observed from space and simulated by a regional chemistry model and their implications for NO<sub>x</sub> emissions, *J. Geophys. Res.*, **114**, 11301, doi:10.1029/2008JD011343.
- Kinne, S., U. Lohmann, J. Feichter, C. Timmreck, M. Schulz, S. Ghan, R. Easter, M. Chin, P. Ginoux, T. Takemura, I. Tegen, D. Koch, M. Herzog, J. Penner, G. Pitari, B. Holben, T. Eck, A. Smirnov, O. Dubovik, I. Slutsker, D. Tanré, O. Torres, M. Mishchenko, I. Geogdzhayev, D.A. Chu, and Y. Kaufman, 2003. Monthly Averages of Aerosol Properties: A Global comparison among models, satellite data and AERONET ground data, *J. Geophys. Res.*, **108**, 4634.
- Kinne, S., M. Schulz, C. Textor, S. Guibert, Y. Balkanski, S.E. Bauer, T. Bernsten, T.F. Berglen, O. Boucher, M. Chin, W. Collins, F. Dentener, T. Diehl, R. Easter, J. Feichter, D. Illmore, S. Ghan, P. Ginoux, S. Gong, A. Grini, J. Hendricks, M. Herzog, L. Horowitz, I. Isaksen, T. Iversen, A. Kirkevåg, S. Kloster, D. Koch, J. E. Kristjansson, M. Krol, A. Lauer, J. F. Lamarque, G. Lesins, X. Liu, U. Lohmann, V. Montanaro, G. Myhre, J. Penner, G. Pitari, S. Reddy, O. Seland, P. Stier, T. Takemura, and X. Tie, 2006. An AeroCom initial assessment – optical properties in aerosol component modules of global models, *Atmos. Chem. Phys.*, **6**, 1815–1834.
- Konovalov, I. B., M. Beekmann, A. Richter, J. P. Burrows, 2006a, Inverse modelling of the spatial distribution of NO<sub>x</sub> emissions on a continental scale using satellite data, *Atmos. Chem. Phys.*, **6**, 1747–1770.
- Konovalov, I. B., M. Beekmann, A. Richter, J. P. Burrows, 2006b, The use of satellite and ground based measurements for estimating and reducing uncertainties in the spatial distribution of emissions of nitrogen oxides, arXiv: physics/0612144 (www.arxiv.org).
- Konovalov, I. B., M. Beekmann, J. P. Burrows, and A. Richter, 2008. Satellite measurement based estimates of decadal changes in European nitrogen oxides emissions, *Atmos. Chem. Phys.*, **8**, 2623–2641.
- Konovalov, I. B., M. Beekmann, A. Richter, A. Hilboll, and J. P. Burrows, 2010. Multi-annual changes of NO<sub>x</sub> emissions in megacity regions: nonlinear trend analysis of satellite measurement based estimates, *Atmos. Chem. Phys.*, **10**, 8481–8498, doi:10.5194/acp-10-8481-2010, 2010.
- Ladstätter-Weissenmayer A., M. Kanakidou, J. Meyer-Arnek, E. V. Dermizaki, A. Richter, M. Vrekoussis, F. Wittrock, and J. P. Burrows, 2007. Pollution events over the East Mediterranean: Synergistic use of GOME, ground based and sonde observations and models, *Atmos. Environ.*, **41**, 7262–7273, doi:10.1016/j.atmosenv.2007.05.031.
- Lauer, A., M. Dameris, A. Richter, J. P. Burrows, 2002. Tropospheric NO<sub>2</sub> columns: A comparison between model and retrieved data from GOME measurements, *Atmos. Chem. Phys.*, **2**, 67–78.
- Lelieveld, J. and F. J. Dentener, 2000. What controls tropospheric ozone?, *J. Geophys. Res.*, **105**, 3531–3551.
- Lelieveld, J., P. Hoor, P. Jöckel, A. Pozzer, P. Hadjinicolaou, J.-P. Cammas, and S. Beirle, 2009. Severe ozone air pollution in the Persian Gulf region, *Atmos. Chem. Phys.*, **9**, 1393–1406.
- Leue, C., M. Wenig, T. Wagner, O. Klimm, U. Platt, and B. Jahne, 2001. Quantitative analysis of NO<sub>x</sub> emissions from GOME satellite image sequences, *J. Geophys. Res.*, **106**, 5493–5505.
- Marbach, T., S. Beirle, U. Platt, P. Hoor, F. Wittrock, A. Richter, M. Vrekoussis, M. Grzegorski, J. P. Burrows, and T. Wagner, 2009. Satellite measurements of formaldehyde from shipping emissions, *Atmos. Chem. Phys.*, **9**, 8223–8234.

- Martin, R. V., D. J. Jacob, J. A. Logan, I. Bey, R. M. Yantosca, A. C. Staudt, Q. Li, A. M. Fiore, B. N. Duncan, H. Liu, P. Ginoux, V. Thouret, 2002. Interpretation of TOMS observations of tropical tropospheric ozone with a global model and *in situ* observations, *J. Geophys. Res.*, **107(D18)**, 4351, doi: 10.1029/2001JD001480.
- Martin, R. V., D. J. Jacob, K. Chance, T. Kurosu, P. I. Palmer, M. J. Evans, 2003. Global inventory of nitrogen oxide emissions constrained by space-based observations of NO<sub>2</sub> columns, *J. Geophys. Res.*, **108**, 4537, doi: 10.1029/2003JD003453.
- Martin, R. V., A. M. Fiore, and A. Van Donkelaar, 2004. Space-based diagnosis of surface ozone sensitivity to anthropogenic emissions, *Geophys. Res. Lett.*, **31**, L06120, doi: 10.1029/2004GL019416.
- Martin, R. V., B. Sauvage, I. Folkens, C. E. Sioris, C. Boone, P. Bernath, J. Ziemke, 2007. Space-based constraints on the production of nitric oxide by lightning, *J. Geophys. Res.*, **112**, D09309, doi: 10.1029/2006JD007831.
- Meirink, J. F., H. J. Eskes, and A. P. H. Goede, 2006. Sensitivity analysis of methane emissions derived from SCIAMACHY observations through inverse modeling, *Atmos. Chem. Phys.*, **6**, 1275–1292.
- Meirink J. F., P. Bergamaschi, and M. C. Krol, 2008a. Four-dimensional variational data assimilation for inverse modelling of atmospheric methane emissions: method and comparison with synthesis inversion, *Atmos. Chem. Phys.*, **8(21)**, 6341–635.
- Meirink, J. F., P. Bergamaschi, C. Frankenberg, M. T. S. d'Amelio, E. J. Dlugokencky, L. V. Gatti, S. Houweling, J. B. Miller, T. Rockmann, M. G. Villani, and M. Krol, 2008b. Four-dimensional variational data assimilation for inverse modeling of atmospheric methane emissions: Analysis of SCIAMACHY observations, *J. Geophys. Res.*, **113**, D17301, doi:10.1029/2007JD009740.
- Millet, D. B., D. J. Jacob, S. Turquety, R. C. Hudman, S. Wu, A. Fried, J. Walega, B. G. Heikes, D. R. Blake, H. B. Singh, B. E. Anderson, and A. D. Clarke, 2006. Formaldehyde distribution over North America: Implications for satellite retrievals of formaldehyde columns and isoprene emission, *J. Geophys. Res.*, **111(24)**, 17, doi: 10.1029/2005JD006853.
- Millet, D. B., D. J. Jacob, K. F. Boersma, T. M. Fu, T. P. Kurosu, K. Chance, C. L. Heald, A. Guenther, 2008. Spatial distribution of isoprene emissions from North America derived from formaldehyde column measurements by the OMI satellite sensor, *J. Geophys. Res.*, **113**, D02307, doi: 10.1029/2007JD008950.
- Mota, B. W., J. M. C. Pereira, D. Oom, M. J. P. Vasconcelos, and M. Schultz, 2006. Screening the ESA ATSR-2 World Fire Atlas (1997–2002), *Atmos. Chem. Phys.*, **6**, 1409–1424.
- Myriokefalitakis, S., M. Vrekoussis, K. Tsigaridis, F. Wittrock, A. Richter, C. Brühl, R. Volkamer, J. P. Burrows, M. Kanakidou, 2008. The influence of natural and anthropogenic secondary sources on the glyoxal global distribution, *Atmos. Chem. Phys.*, **8**, 4965–4981.
- Myriokefalitakis, S., E., Vignati, K., Tsigaridis, C., Papadimas, J., Sciare, N., Mihalopoulos, M. C., Facchini, M., Rinaldi, F. J., Dentener, D., Ceburnis, N., Hatzianastasiou, C. D., O'Dowd, M., van Weele, M., Kanakidou, 2010. Global modelling of the oceanic source of organic aerosols, *Adv. Meteo.*, <http://www.hindawi.com/journals/amet/aip.939171.html>.
- O'Dowd, C. D., B. Langmann, S. Varghese, C. Scannell, D. Ceburnis, and M. C. Facchini, 2008. A combined organic-inorganic sea-spray source function, *Geophys. Res. Lett.*, **35**, L01801, doi: 10.1029/2007GL030331.
- Palmer, P. I., D. J. Jacob, A. Fiore, R. V. Martin, K. Chance, and T. P. Kurosu, 2003. Mapping isoprene emissions over North America using formaldehyde column observations from space, *J. Geophys. Res.*, **108(D6)**, 4180, doi: 10.1029/2002JD002153.
- Palmer P. I., D. S. Abbot, T.-M. Fu, D. J. Jacob, K. Chance, T. P. Kurosu, A. Guenther, Ch. Wiedinmyer, J. C. Stanton, M. J. Pilling, S. N. Pressley, B. Lamb, A.-L. Sumner, 2006. Quantifying the seasonal and interannual variability of North American isoprene emissions using satellite observations of the formaldehyde column, *J. Geophys. Res.*, **111**, D12315, doi:10.1029/2005JD006689.
- Peters W., A. R. Jacobson, C. Sweeney, A. E. Andrews, Th. J. Conway, K. Masarie, J. B. Miller, L. M. P. Bruhwiler, G. Pétron, A. I. Hirsch, D. E. J. Worthy, G. R. van der Werf, J. T.

- Randerson, P. O. Wennberg, M. C. Krol, and P. P. Tans, 2007. An atmospheric perspective on North American CO<sub>2</sub> exchange: Carbon Tracker, *Proc Natl Acad Sci U S A*, **2007** **104**(48): 18925–18930. doi: 10.1073/pnas.0708986104. <http://www.esrl.noaa.gov/gmd/ccgg/carbontracker/>
- Pétron, G., C. Granier, B. Khattatov, V. Yudin, J.-F. Lamarque, L. Emmons, J. Gille, D. P. Edwards, 2004. Monthly CO surface sources inventory based on the 2000–2001 MOPITT satellite data, *Geophys. Res. Lett.*, **31**, L21107, doi: 10.1029/2004GL020560.
- Richter, A., H. Nüß, J. P. Burrows, C. Granier, and U. Niemeier, 2004. Long term measurements of NO<sub>2</sub> and other tropospheric species from space, *Proceedings of the XX Quadrennial Ozone Symposium*, 1–8 June 2004, Kos, Greece, 213–214.
- Richter, A., J. P. Burrows, H. Nüß, C. Granier, U. Niemeier, 2005. Increase in tropospheric nitrogen dioxide over China observed from Space, *Nature*, **437**, 129–132, doi: 10.1038/nature04092.
- Rodgers, C. D., 2000. *Inverse Methods for Atmospheric Sounding*, World Scientific.
- Rood R. B., 2007. *Fundamentals of Modeling, Data Assimilation, and High-Performance Computing*, in *Observing Systems for Atmospheric Composition*, Springer, New York, 10.1007/978-0-387-35848-2.
- Sandu, A., W. Liao, G. R. Carmichael, D. Henze, J. H. Seinfeld, T. Chai, D. Daescu, 2004. Computational aspects of data assimilation for aerosol dynamics, *Lecture notes in computer science*, V **3038**, 709–716.
- Schneising, O., Buchwitz, M., Burrows, J. P., Bovensmann, H., Bergamaschi, P., and Peters, W., 2009. Three years of greenhouse gas column-averaged dry air mole fractions retrieved from satellite – Part 2: Methane, *Atmos. Chem. Phys.*, **9**, 443–465.
- Schulz, M., C. Textor, S. Kinne, Y. Balkanski, S. Bauer, T. Bernsten, T. Berglen, O. Boucher, F. Dentener, S. Guibert, I. S. A. Isaksen, T. Iversen, D. Koch, A. Kirkevåg, X. Liu, V. Montanaro, G. Myhre, J. E. Penner, G. Pitari, S. Reddy, Ø. Seland, P. Stier, and T. Takemura, 2006. Radiative forcing by aerosols as derived from the AeroCom present-day and pre-industrial simulations, *Atmos. Chem. Phys.*, **6**, 5225–5246.
- Shindell, D. T., G. Faluvegi, D. S. Stevenson, M. C. Krol, L. K. Emmons, J.-F. Lamarque, G. Petron, F. J. Dentener, K. Ellingsen, M. G. Schultz, O. Wild, M. Amann, C. S. Atherton, D. J. Bergmann, I. Bey, T. Butler, J. Cofala, W. J. Collins, R. G. Derwent, R. M. Doherty, J. Drevet, H. J. Eskes, A. M. Fiore, M. Gauss, D. A. Hauglustaine, L. W. Horowitz, I. S. A. Isaksen, M. G. Lawrence, V. Montanaro, J.-F. Müller, G. Pitari, M. J. Prather, J. A. Pyle, S. Rast, J. M. Rodriguez, M. G. Sanderson, N. H. Savage, S. E. Strahan, K. Sudo, S. Szopa, N. Unger, T. P. C. van Noije, and G. Zeng, 2006. Multimodel simulations of carbon monoxide: Comparison with observations and projected near-future changes, *J. Geophys. Res.*, **111**, D19306, doi: 10.1029/2006JD007100.
- Stajner, I., L. P. Riishøjgaard, and R. B. Rood, 2001. The GEOS ozone data assimilation system: Specification of error statistics, *Q. J. Roy. Meteor. Soc.*, **127**, 1069–1094.
- Stajner, I., K. Wargan, S. Pawson, H. Hayashi, L.-P. Change, R. C. Hudman, L. Froidevaux, N. Livesey, P. F. Levelt, A. M. Thompson, D. W. Tarasick, R. Stubi, S. B. Andersen, M. Yela, G. König-Langlo, F. J. Schmidlin, and J. C. Witte, 2008. Assimilated ozone from EOS Aura: Evaluation of the tropopause region and tropospheric columns, *J. Geophys. Res.*, **113**, D16S32, doi:10.1039/2007JD008863.
- Stavrakou, T., J.-F. Müller, 2006. Grid-based versus big region approach for inverting CO emissions using Measurement of Pollution in the Troposphere (MOPITT) data, *J. Geophys. Res.*, **111**, D15304, doi: 10.1029/2005JD006896.
- Stavrakou, T., J.-F. Müller, I. De Smedt, M. Van Roozendael, M. Kanakidou, M. Vrekoussis, F. Wittrock, A. Richter, and J. P. Burrows, 2009. The continental source of glyoxal estimated by the synergistic use of spaceborne measurements and inverse modelling, *Atmos. Chem. Phys.*, **9**, 8431–8446.
- Stevenson, D. S., F. J. Dentener, M. G. Schultz, K. Ellingsen, T. P. C. van Noije, O. Wild, G. Zeng, H. J. Eskes, A. M. Fiore, M. Gauss, D. A. Hauglustaine, L. W. Horowitz, I. S. A. Isaksen, M. C. Krol, J.-F. Lamarque, M. G. Lawrence, V. Montanaro, J.-F. Müller, G. Pitari, M. J. Prather,

- J. A. Pyle, S. Rast, J. M. Rodriguez, M. G. Sanderson, N. H. Savage, D. T. Shindell, S. E. Strahan, K. Sudo, S. Szopa, 2006. Multi-model ensemble simulations of present-day and near-future tropospheric ozone, *J. Geophys. Res.*, **111**, D08301, doi: 10.1029/2005JD006338.
- Stockwell, W. R., P. Middleton, and J. S. Chang, 1990. The second generation regional acid deposition model chemical mechanism for regional air quality modeling, *J. Geophys. Res.*, **95**, 16343–16367.
- Struthers, H., R. Brugge, W. A. Lahoz, A. O'Neill, and R. Swinbank, 2002. Assimilation of ozone profiles and total column measurements into a General Circulation Model, *J. Geophys. Res.*, **107**, 4438, doi: 10.1029/2001JD000957.
- Talagrand, O., 2003. *A posteriori* validation of assimilation algorithms, In: Swinbank, R., V. Shutyaev, and W. A. Lahoz, (eds) Data Assimilation for the Earth System, NATO ASI Series, Kluwer, Dordrecht, The Netherlands.
- Textor, C., Schulz, M., Kinne, S., Guibert, S., Balkanski, Y., Bauer, S. E., Bernsten, T., Berglen, T., Boucher, O., Chin, M., Dentener, F., Diehl, T., Easter, R., Feichter, H., Fillmore, D., Ghan, S., Ginoux, P., Gong, S., Grini, A., Hendricks, J., Horowitz, L., Huang, P., Isaksen, I., Iversen, T., Kirkevag, A., Kloster, S., Koch, D., Kristjansson, E., Krol, M., Lauer, A., Lamarque, J. F., Liu, X., Montanaro, V., Myhre, G., Penner, J., Pitari, G., Reddy, S., Seland, Ø., Stier, P., Takemura, T., and Tie, X.: 2006. Analysis and quantification of the diversities of aerosol life cycles within AeroCom, *Atmos. Chem. Phys.*, **6**, 1777–1813.
- van der A, R. J. M., D. H. M. U. Peters, H. Eskes, K. F. Boersma, M. van Roozendael, I. De Smedt, H. M. Kelder, 2006. Detection of the trend and seasonal variation in tropospheric NO<sub>2</sub> over China, *J. Geophys. Res.*, **111**, 12317, doi: 10.1029/2005JD006594.
- van der Werf, G. R., Randerson, J. T., Giglio, L., Collatz, G. J., Kasibhatla, P. S., and Arellano Jr., A. F., 2006. Interannual variability in global biomass burning emissions from 1997 to 2004, *Atmos. Chem. Phys.*, **6**, 3423–3441.
- van Noije, T. P. C., Eskes, H. J., Dentener, F. J., Stevenson, D. S., Ellingsen, K., Schultz, M. G., Wild, O., Amann, M., Atherton, C. S., Bergmann, D. J., Bey, I., Boersma, K. F., Butler, T., Cofala, J., Drevet, J., Fiore, A. M., Gauss, M., Hauglustaine, D. A., Horowitz, L. W., Isaksen, I. S. A., Krol, M. C., Lamarque, J.-F., Lawrence, M. G., Martin, R. V., Montanaro, V., Müller, J.-F., Pitari, G., Prather, M. J., Pyle, J. A., Richter, A., Rodriguez, J. M., Savage, N. H., Strahan, S. E., Sudo, K., Szopa, S., and van Roozendael, M., 2006. Multi-model ensemble simulations of tropospheric NO<sub>2</sub> compared with GOME retrievals for the year 2000, *Atmos. Chem. Phys.*, **6**, 2943–2979.
- van Loon, M., P. J. H. Bultjes, and A. J. Segers, 2000. Data assimilation of ozone in the atmospheric transport chemistry model LOTOS, *Environ. Model. Software*, **15**, 603–609.
- Velasco, E., Lamb, B., Westberg, H., Allwine, E., Sosa, G., Arriaga-Colina, J. L., Jobson, B. T., Alexander, M. L., Prazeller, P., Knighton, W. B., Rogers, T. M., Grutter, M., Herndon, S. C., Kolb, C. E., Zavala, M., de Foy, B., Volkamer, R., Molina, L. T., and Molina, M. J., 2007. Distribution, magnitudes, reactivities, ratios and diurnal patterns of volatile organic compounds in the Valley of Mexico during the MCMA 2002 & 2003 field campaigns, *Atmos. Chem. Phys.*, **7**, 329–353.
- Velders, G. J. M., C. Granier, R. W. Portmann, K. Pfeilsticker, M. Wenig, T. Wagner, U. Platt, A. Richter, and J. P. Burrows, 2001. Global tropospheric NO<sub>2</sub> column distributions: Comparing three-dimensional model calculations with GOME measurements, *J. Geophys. Res.*, **106** (12), 12,643–12,660.
- Verlaan, M. and A. W. Heemink, 1995. Reduced rank square root filters for large scale data assimilation problems, in: Second International Symposium on Assimilation of Observations in Meteorology and Oceanography.
- Volkamer, R., L. T. Molina, M. J. Molina, T. Shirley, and W. H. Brune, 2005. DOAS measurement of glyoxal as an indicator for fast VOC chemistry in urban air, *Geophys. Res. Lett.*, **32**, L08806, doi:10.1029/2005GL022616.

- Vrekoussis, M., M. Kanakidou, N. Mihalopoulos, P. J. Crutzen, J. Lelieveld, D. Perner, H. Berresheim, E. Baboukas, 2004. Role of the NO<sub>3</sub> radicals in oxidation processes in the eastern Mediterranean troposphere during the MINOS campaign, *Atmos. Chem. Phys.*, **4**, 169–182.
- Wang, Y., M. B. McElroy, R. V. Martin, D. G. Streets, Q. Zhang, and T.-M. Fu, 2007. Seasonal variability of NO<sub>x</sub> emissions over east China constrained by satellite observations: Implications for combustion and microbial sources, *J. Geophys. Res.*, **112**, D06301, doi: 10.1029/2006JD007538.
- Weaver, A. and P. Courtier, 2001. Correlation modelling on the sphere using a generalized diffusion equation, *Q. J. Roy. Meteor. Soc.*, **127**, 1815–1846.
- Wittrock, F., A. Richter, H. Oetjen, J. P. Burrows, M. Kanakidou, S. Myriokefalitakis, R. Volkamer, S. Beirle, U. Platt, and T. Wagner, 2006. Simultaneous global observations of glyoxal and formaldehyde from space, *Geophys. Res. Lett.*, **33**, L16804, doi: 10.1029/2006GL026310.
- Zhang, Q., J. L. Jimenez, M. R. Canagaratna, J. D. Allan, H. Coe, I. Ulbrich, M. R. Alfarra, A. Takami, A. M. Middlebrook, Y. L. Sun, K. Dzepina, E. Dunlea, K. Docherty, P. F. DeCarlo, D. Salcedo, T. Onasch, J. T. Jayne, T. Miyoshi, A. Shimono, S. Hatakeyama, N. Takegawa, Y. Kondo, J. Schneider, F. Drewnick, S. Borrmann, S. Weimer, K. Demerjian, P. Williams, K. Bower, R. Bahreini, L. Cottrell, R. J. Griffin, J. Rautiainen, J. Y. Sun, Y. M. Zhang, and D. R. Worsnop, 2007. Ubiquity and dominance of oxygenated species in organic aerosols in anthropogenically-influenced Northern Hemisphere midlatitudes, *Geophys. Res. Lett.*, **34**, L13801, doi: 10.1029/2007GL029979.