ULB

UNIVERSITE LIBRE DE BRUXELLES

Faculté des Sciences Appliquées

Laboratoire de Chimie Quantique et Photophysique

Long-Term Global Observations of Tropospheric Formaldehyde Retrieved from Spaceborne Nadir UV Sensors

Directeurs de thèse:
Michel Herman et Michel Van Roozendael

Dissertation présentée par
Isabelle De Smedt en vue de l'obtention du
grade de Docteur en Sciences de l'Ingénieur

Année académique 2010 - 2011

UNIVERSITE LIBRE DE BRUXELLES

Faculté des Sciences Appliquées

Laboratoire de Chimie Quantique et Photophysique

Long-Term Global Observations of Tropospheric Formaldehyde Retrieved from Spaceborne Nadir UV Sensors

Directeurs de thèse:
Michel Herman et Michel Van Roozendael

Dissertation présentée par
Isabelle De Smedt en vue de l'obtention du grade de Docteur en Sciences de l'Ingénieur

Année académique 2010 - 2011

Remerciements

Les travaux de recherche présentés dans cette thèse ont été réalisés à l'Institut d'Aéronomie Spatiale de Belgique (BIRA-IASB). Je tiens à remercier tout particulièrement Michel Van Roozendael, chef du groupe de recherche UV-Visible de l'IASB, qui m'a accueillie au sein de son équipe. Il m'a offert sa confiance et la possibilité de faire une thèse dans un cadre privilégié. Il n'a jamais économisé son temps, ses conseils, ses critiques et m'a fait progresser tout au long de ces années. Je mesure tous les jours la chance que j'ai de travailler à ses côtés. Je remercie aussi très chaleureusement mon promoteur Michel Herman, directeur du laboratoire de chimie quantique et photophysique de l'ULB, ainsi que Frank Dubois, Michel Godefroid, Jean Vander Auwera, Emile Biémont, Gaëlle Dufour et Florence Goutail qui ont très gentiment accepté de faire partie de mon jury.

Tout au long de ce travail, j'ai bénéficié de l'apport scientifique et de l'aide précieuse de Jenny Stavrakou et de Jean-François Müller, en particulier au moment de la publication des résultats. Je les en remercie grandement. L'enthousiasme et l'énergie de Jenny m'ont plus d'une fois remonté le moral. Ces derniers mois, j'ai pu compter sur le soutien de Nicolas Theys qui a corrigé patiemment et avec beaucoup de soin ce travail dès les premières ébauches. Merci Nico! Un tout grand merci aussi à Christophe Lerot et à François Hendrick qui ont également pris sur leur temps pour lire et critiquer mon travail. De nombreuses discussions et échanges avec Nicolas et Christophe, ainsi qu'avec Corinne Vigouroux, Katrijn Clémer et Gaia Pinardi, ont beaucoup contribué à améliorer les résultats. Merci aussi à Caroline Fayt qui m'aide depuis le début pour l'utilisation de QDOAS et Matlab, ainsi qu'aux membres du service informatique, en particulier Vincent Letocart et Olivier Rasson, pour l'indispensable support technique.

Je partage avec bonheur mon bureau avec Gaia, Nicolas et Christophe, avec qui, tous les jours, je peux boire mon café, rire, discuter de tout et de rien, soupirer et me plaindre à souhait. Je les remercie de rendre mes journées plus légères. Merci aussi à Corinne, Caro, Christian, Alexis, Manu, Fabien, Arnaud, Jeroen, Katrijn, et tous les autres collègues de l'Institut.

Enfin, j'ai la chance de pouvoir m'appuyer sur l'affection de mes amis et de ma famille. Merci à Céline et Bruno, Alex et Nathalie, Christine, Djamila, Antal, Denis, Pierre, Délia, Thomas, Stéphanie et les autres, pour vos encouragements et votre compréhension. Un tendre merci à mes parents, qui me soutiennent depuis toujours, et en particulier à ma maman, qui depuis presque 7 ans, joue les supers nounous pour mes deux petits garçons.

Merci à David, Maximilien et Noé, pour votre amour, vos sourires, votre patience et votre confiance.

Résumé

Le formaldéhyde (H_2CO) joue un rôle central dans la chimie de la troposphère en tant que produit intermédiaire commun à la dégradation de la plupart des composés organiques volatils dans l'atmosphère. L'oxydation du méthane est responsable de plus de la moitié de la concentration moyenne globale du formaldéhyde. Sur les continents en revanche, les hydrocarbures non-méthaniques (NMVOCs) émis par la végétation, les feux de biomasse et les activités humaines, augmentent de façon significative et localisée la concentration de H_2CO. Les récents senseurs satellitaires à visée nadir offrent la possibilité de quantifier à l'échelle globale l'abondance du formaldéhyde dans la troposphère et de ce fait, d'améliorer notre connaissance des émissions de NMVOCs. Ceci est essentiel à la compréhension des mécanismes contrôlant la production et l'évolution de l'ozone troposphérique, élément clé pour la qualité de l'air et les changements climatiques, mais aussi du composé hydroxyle OH, le principal agent nettoyant de notre troposphère. C'est pourquoi, une méthode de plus en plus répandue pour améliorer les inventaires d'émissions des NMVOCs consiste en l'utilisation d'observations satellitaires de H_2CO en combinaison avec un modèle de chimie et de transport troposphérique, dans une approche appelée modélisation inverse. Ce genre d'application demande des produits satellitaires bien caractérisés et dérivés de façon cohérente sur de longues périodes de temps.

Le travail présenté dans ce manuscrit porte sur l'inversion des colonnes de formaldéhyde à partir de spectres de la radiation solaire diffusée par l'atmosphère terrestre, mesurés par les senseurs nadir GOME, SCIAMACHY et GOME-2, lancés successivement en 1995, 2002 et 2006. La méthode d'inversion est basée sur la spectroscopie d'absorption optique différentielle (DOAS). Les concentrations de formaldéhyde intégrées le long du chemin optique moyen dans l'atmosphère sont dérivées à partir des spectres mesurés, et ensuite transformées en colonnes verticales par le biais de facteurs de conversion appelés facteurs de masse d'air. Ces derniers sont calculés à l'aide d'un modèle de transfert radiatif, en tenant compte de la présence de nuages, des propriétés de la surface terrestre et la distribution verticale supposée du formaldéhyde, fournie par le modèle IMAGES. Un des objectifs principaux de la thèse a été d'optimiser les paramètres d'inversion pour H_2CO, et ceci pour les trois senseurs, tout en tenant compte des spécificités de chaque instrument. Ces efforts ont conduit à la création d'un jeu de données homogène, couvrant la période de 1996 à 2010. Les colonnes sont fournies avec un bilan d'erreur complet, incluant les erreurs liées à l'inversion des concentrations dans les spectres, ainsi que celles provenant de l'évaluation des facteurs de masse d'air. La série temporelle des observations de GOME, SCIAMACHY et GOME-2 présente une bonne cohérence et stabilité sur toute la période. Nous montrons aussi que la meilleure couverture terrestre de GOME-2 entraîne une réduction significative du bruit sur les observations moyennées, permettant une meilleure identification des sources d'émission. Notre jeu de données est exploité pour étudier la distribution régionale du formaldéhyde, ainsi que ses variations saisonnières et interannuelles, principalement liées aux variations de température et aux feux de végétation, mais aussi aux activités anthropiques. De plus, en s'appuyant sur la qualité de la série temporelle de 15 ans, nous présentons la première analyse des variations à long terme des concentrations de H_2CO. Des tendances positives, de l'ordre de 1.5 à 4% par an, sont observées en Asie, en particulier dans l'est de la Chine et en Inde, liées à l'augmentation des émissions anthropiques d'hydrocarbures dans ces régions. Finalement, nos données ont été largement exploitées par le groupe de modélisation de l'IASB pour faire des études de modélisation inverse des émissions de NMVOCs. Les résultats démontrent le haut potentiel des données satellitaires pour contraindre les inventaires d'émissions dues à la végétation et aux feux de biomasse, particulièrement dans les écosystèmes tropicaux, en Asie du sud-est, et dans le sud-est des Etats-Unis.

iv

Abstract

Atmospheric formaldehyde (H_2CO) is an intermediate product common to the degradation of many volatile organic compounds and therefore it is a central component of the tropospheric chemistry. While the global formaldehyde background is due to methane oxidation, emissions of non-methane volatile organic compounds (NMVOCs) from biogenic, biomass burning and anthropogenic continental sources result in important and localised enhancements of the H_2CO concentration. Recent spaceborne nadir sensors provide an opportunity to quantify the abundance of tropospheric formaldehyde at the global scale, and thereby to improve our knowledge of NMVOC emissions. This is essential for a better understanding of the processes that control the production and the evolution of tropospheric ozone, a key actor in air quality and climate change, but also of the hydroxyl radical OH, the main cleansing agent of our troposphere. For this reason, H_2CO satellite observations are increasingly used in combination with tropospheric chemistry transport models to constrain NMVOC emission inventories in so-called top-down inversion approaches. Such inverse modelling applications require well characterised satellite data products consistently retrieved over long time periods.

This work reports on global observations of formaldehyde columns retrieved from the successive solar backscatter nadir sensors GOME, SCIAMACHY and GOME-2, respectively launched in 1995, 2002 and 2006. The retrieval procedure is based on the differential optical absorption spectroscopy technique (DOAS). Formaldehyde concentrations integrated along the mean atmospheric optical path are derived from the recorded spectra in the UV region, and further converted to vertical columns by means of calculated air mass factors. These are obtained from radiative transfer simulations, accounting for cloud coverage, surface properties and best-guess H_2CO profiles, the latter being derived from the IMAGES chemistry transport model. A key task of the thesis has consisted in the optimisation of the H_2CO retrieval settings from multiple sensors, taking into account the instrumental specificities of each sounder. As a result of these efforts, a homogeneous dataset of formaldehyde columns covering the period from 1996 to 2010 has been created. This comes with a comprehensive error budget that treats errors related to the spectral fit of the columns as well as those associated to the air mass factor evaluation. The time series of the GOME, SCIAMACHY and GOME-2 H_2CO observations is shown to be consistent and stable over time. In addition, GOME-2 brings a significant reduction of the noise on spatiotemporally averaged observations, leading to a better identification of the emission sources. Our dataset is used to study the regional formaldehyde distribution, as well as its seasonal and interannual variations, principally related to temperature changes and fire events, but also to anthropogenic activities. Moreover, building on the quality of our 15-year time series, we present the first analysis of long-term changes in the H_2CO columns. Positive trends, in the range of 1.5 to 4% yr^{-1}, are found in Asia, more particularly in Eastern China and India, and are related to the known increase of anthropogenic NMVOC emissions in these regions. Finally, our dataset has been extensively used in several studies, in particular by the BIRA-IASB modelling team to constrain NMVOC emission fluxes. The results demonstrate the high potential of satellite data as top-down constraint for biogenic and biomass burning NMVOC emission inventories, especially in Tropical ecosystems, in Southeastern Asia, and in Southeastern US.

Table of Contents

Chapter 1: Introduction

Chapter 2: Introduction to Atmospheric Chemistry

Chapter 3: Volatile Organic Compounds

Chapter 4: Satellite Remote Sensing of Tropospheric Composition

Chapter 5: Differential Absorption Spectroscopy

Chapter 6: Global Observation of Formaldehyde in the Troposphere Using GOME and SCIAMACHY Sensors

Chapter 7: Formaldehyde Columns Retrieved from GOME-2

Chapter 8: Trend Detection in Satellite Observations of Formaldehyde Tropospheric Columns

Chapter 9: Constraining Non-Methane Volatile Organic Compound Emissions with Satellite Formaldehyde Observations

Chapter 10: Conclusions and Perspectives

Publications

References

List of Acronyms

x

1. Introduction

Despite their low concentrations, atmospheric trace gases are of critical importance in Earth environmental issues. They are responsible for phenomena ranging from urban photochemical smog and acid deposition to stratospheric ozone depletion and climate change. Over the past century, the composition of the atmosphere has changed drastically due to human activities and more changes are awaited for the future. The global concentrations of greenhouse gases like carbon dioxide (CO_2), methane (CH_4) and nitrous oxide (N_2O) have increased far beyond their natural values, with fundamental implications for the radiative balance of the Earth's atmosphere. Furthermore, the greenhouse effect has been amplified by the increase of man-made halocarbons emissions, and of tropospheric ozone (O_3) levels, produced by photochemical oxidation of carbon monoxide (CO), methane (CH_4) and non-methane volatile organic compounds (NMVOCs) in the presence of nitrogen oxides (NO_x =$NO+NO_2$). Concurrently, the long-lived species CH_4, N_2O and halocarbons (*e.g.* chlorofluorocarbons or CFCs) are able to reach the stratosphere where their degradation products contribute to ozone chemical depletion. The various atmospheric species interact through a wealth of physical and chemical processes and are interconnected in a very complex and non-linear way, often causing feedbacks that can amplify or mitigate the original perturbation. For example, there is scientific evidence that climate change and air quality are strongly coupled. Changes in climate can affect the chemical composition of the atmosphere, for instance by altering emissions or atmospheric chemical processes. Changes in the distribution of ozone and its precursors can also have an impact on the radiative forcing and on the lifetime of greenhouse gases [*Forster et al.*, *IPCC*, 2007].

In many respects, the most important trace gas in the atmosphere is ozone. In the stratosphere, ozone screens living organism from biologically harmful solar ultraviolet radiation. In the troposphere, it is one of the four most important contributors to greenhouse radiative forcing, while it is also a major source of the hydroxyl radical (OH), the primary oxidizing agent and scavenger of pollutants in our troposphere. At the surface, atmospheric ozone produces harmful effects on human health and vegetation when present at too elevated levels (photochemical smog). At the global scale, understanding both the natural ozone chemistry in the troposphere and the causes of continually increasing background tropospheric ozone levels is a major issue. At the urban and regional scales, important policy questions deal with how to decrease ozone levels by controlling the ozone precursors, *i.e.* NO_x, CH_4, CO, and NMVOCs. Understanding the photochemical production of tropospheric ozone and its complicated relationship to the precursor sources can only be achieved using chemistry transport models (CTM) that describes the complex non-linear tropospheric chemistry and its coupling to atmospheric transport [*Seinfeld and Pandis*, 2006].

Non-methane volatile organic compounds in particular have a strong influence on the tropospheric composition because of their contribution to ozone production in polluted areas and because of their impact on the oxidizing capacity of the global troposphere [*Houweling et al.*, 1998]. NMVOCs also influence the climate due to their role as precursors of secondary organic aerosols (SOA) [*Kanakidou et al.*, 2005]. In contrast to CH_4, CO and NO_x emissions, NMVOCs are still largely dominated by natural emission from vegetation (~1150 TgC yr^{-1}). The remainder is due to anthropogenic activities (~160 TgC yr^{-1}) [*Olivier et al.*, 2001] and vegetation fires (~50 TgC yr^{-1}) [*Andreae and Merlet*, 2001]. Biogenic isoprene is the most important NMVOC as its emissions represent almost half of the total biogenic source, and almost 40% of total NMVOC emissions on the global scale [*Guenther et al.*, 1995]. Isoprene emissions depend on vegetation types, light

intensity and temperature. It has been shown that climate change can potentially induce large long-term changes in global biogenic emissions [*Guenther et al.*, 2006] and that meteorological variability, and in particular El Niño events, induce a significant interannual variability of these emissions [*Müller et al.*, 2008].

Despite their importance, the quantification of the NMVOC emissions from biogenic, pyrogenic and anthropogenic sources, as well as their repartition in species (speciation), is far from being complete and accurate. Emission inventories are generally based on bottom-up methodologies, in which ground-based and laboratory measurements are extrapolated to provide regional and global emission estimates, using geographical and statistical data together with modelled dependences of emissions to environmental variables. However, the diversity of the NMVOCs, the high spatiotemporal variability of theirs sources and sinks, and the lack of a dense monitoring network, result in large uncertainties in the emission inventories [*Stavrakou et al.*, 2009a]. For example, global emissions of isoprene are generally estimated by extrapolating limited laboratory and field measurements to the prescribed global ecosystems and to meteorological conditions. Despite recent efforts, large differences exist between the various isoprene emission estimates, particularly in tropical ecosystems which collectively account for more than half of global isoprene emissions [*Guenther et al.*, 2006; *Barkley et al.*, 2008; *Müller et al.*, 2008].

In this general context, the motivation of this PhD thesis is to contribute to a better knowledge of the NMVOC emissions through long-term global observations of formaldehyde (H_2CO), one of the most important intermediate compounds in the degradation of many NMVOCs. We use nadir spaceborne sounders to quantify the formaldehyde abundance at the global scale, and to provide new constraints on the emissions of NMVOCs.

Although the main source of formaldehyde in the background troposphere is the oxidation of methane, the oxidation of NMVOCs from biogenic, biomass burning and anthropogenic continental sources results in important and localised enhancements of the H_2CO concentration. For example, strong H_2CO signals are observed over densely forested areas due to the combination of biomass burning and high isoprene emission fluxes, and to a lesser extent over regions with sustained human activities. The lifetime of formaldehyde being relatively short (a few hours), H_2CO is not significantly affected by atmospheric transport, so that its concentrations can be directly related to its NMVOC precursor emission sources. In emission regions, the bulk of formaldehyde lies in the boundary layer, which can only be sounded from space using nadir-viewing instruments. Satellite observations of formaldehyde started with the launch of GOME on the ERS-2 satellite in April 1995 and are now possible with several nadir sun-synchronous spaceborne spectrometers that measure the solar radiation backscattered by the Earth atmosphere. The H_2CO total columns (*i.e.* the vertically integrated concentrations) are derived from these radiance spectra through the exploitation of the H_2CO absorption bands in the near ultraviolet (UV), using the technique of differential optical absorption spectroscopy (DOAS) [*Platt and Stutz*, 2008]. Infrared limb and occultation satellite instruments are also used to retrieve H_2CO concentrations in the upper troposphere and in the stratosphere.

Making available long-term global satellite observations of formaldehyde helps addressing a number of key scientific issues in relation to the current and future changes in air quality and in the feedbacks between air quality and climate:

- How accurate are the current biogenic, biomass burning and anthropogenic NMVOC emission inventories?
- How are the NMVOC emissions changing with time, on a global and on a regional scale?

- What is the relative contribution of biogenic, anthropogenic, and biomass burning emissions to these changes?
- How will climate change impact biogenic emissions, atmospheric processes, and global concentrations of NMVOCs?
- What local and global effects have land use changes, *e.g.* deforestation in the Tropical regions or afforestation in China, on biogenic emissions?
- How do levels of NMVOC compounds change in cities/megacities and what is the possible impact on air quality?
- What are the factors controlling air quality in the United States, Europe or Asia? Are the emission controls in these industrialized areas sufficiently efficient to reduce tropospheric ozone levels?
- What are the intercontinental effects of regional changes in NMVOC emissions?
- What is the impact of NMVOCs on aerosols?

A powerful technique to address such questions is the so-called top-down or inverse modelling approach that consists in adjusting the surface emissions used as input in state-of-the-art chemistry transport models, in order to minimize the discrepancy between a set of atmospheric observations and the model predictions. In this method, it is implicitly assumed that the relationship between surface fluxes and atmospheric abundances is reasonably well predicted by the model, so that the biases between the model and the data are mostly due to errors in the emission inventories [*e.g. Müller and Stavrakou*, 2005]. The usefulness of satellite observations of CO, NO_2 and H_2CO columns as a mean to constrain the emissions of reactive ozone precursors (CO, NO_x and NMVOCs) through inverse modelling has been demonstrated in several studies [*i.a. Kopacz et al.*, 2010; *Stavrakou et al.*, 2008; *Palmer et al.*, 2003].

Inverse modelling studies rely on accurate, stable and well characterized satellite measurements. The main objective of our work is to provide state-of-the-art, high quality H_2CO column measurements from the European spaceborne UV-Visible spectrometers, GOME, SCIAMACHY and GOME-2, including proper characterization of the uncertainty and information content of the retrieved datasets. Using the synergy with model developments conducted at BIRA-IASB on the subject of NMVOC chemistry and inverse modelling of sources, we aim to demonstrate the usefulness of our satellite retrieval for tropospheric chemistry investigations.

The manuscript is structured as follows: An overview of the basics of atmospheric chemistry and related environmental issues is given in **chapter 2**, with focus on tropospheric ozone chemistry and its precursors. In **chapter 3**, the current knowledge on the global budgets of carbon-containing compounds in the atmosphere and their sources and sinks is presented. The large uncertainties of NMVOC emission inventories are discussed, as well as the potential of satellite formaldehyde observations to provide quantitative information on NMVOC emissions. In **chapter 4**, the fundamental principles of satellite remote sensing are presented, and the capability of nadir instruments to monitor air quality is illustrated with examples of satellite observations of CO, NO_2 and NMVOCs in the lower troposphere. The GOME, SCIAMACHY and GOME-2 instruments are described in this chapter. **Chapter 5** deals with the observation principles in the UV region. The differential optical absorption spectroscopy (DOAS) technique is described in details, with an emphasis on the advantages, the assumptions and the limitations of the DOAS method.

The development of our algorithms to retrieve tropospheric formaldehyde columns from GOME (1996-2002) and SCIAMACHY (2003-2010) measurements is presented in **chapter 6**. A special effort has been made to ensure the coherence and quality of the H_2CO dataset for the entire period 1996-2010. Optimised DOAS settings are developed in order to reduce important error sources in

the derivation of H_2CO columns and a comprehensive error analysis is presented. The features of the retrieved formaldehyde column distribution are discussed and compared with previous H_2CO datasets over the major emission regions.

In **chapter 7**, we present formaldehyde retrievals performed using GOME-2 observations between 2007 and 2010, based on an improved version of the algorithms developed for the GOME and SCIAMACHY instruments. The retrieval performances obtained with GOME, SCIAMACHY and GOME-2 are compared by taking into account the differences in the spatial and temporal resolutions of these sensors. The better Earth sampling of GOME-2 is shown to allow for a better identification of the emission sources. Furthermore, we present first attempts to validate the satellite data using ground-based measurements in Cabauw (The Netherlands) and Beijing (China).

On the basis of this newly developed GOME-SCIAMACHY-GOME-2 dataset, an analysis of the long-term trends in H_2CO columns is presented in **chapter 8**. A linear model with a seasonal component is used to fit the time series of monthly averaged columns in Asian countries and around major urban agglomerations worldwide. The error and statistical significance of the inferred trends are estimated, taking into account uncertainties on satellite observations.

The results of the studies conducted by the BIRA-IASB tropospheric modelling team based on our formaldehyde observations are summarized in **chapter 9**. These have been performed with an updated version of the IMAGES model, including an optimized chemical scheme with respect to H_2CO production. The first objective is to test the ability of the model to reproduce the observed features of the H_2CO distribution and to provide an assessment on whether the more recent biomass burning and isoprene inventories lead to a better agreement with the observations. The second objective is to derive updated NMVOC emission fluxes using the IMAGESv2 model in an inverse mode and the satellite formaldehyde columns as top-down constraints.

Conclusions and perspectives are discussed in the last chapter.

2   Introduction to Atmospheric Chemistry

This chapter presents a general introduction to the Earth atmosphere chemistry, pointing to the molecules and processes involved in the main current environmental issues, in particular for tropospheric chemistry and air quality. The principal factors controlling the chemical composition of the atmosphere and the distribution of its components are first presented in sections 2.1 and 2.2. Then the global budget of the solar and terrestrial radiation is estimated in section 2.3. The main production and destruction processes of ozone in the stratosphere are briefly introduced in section 2.4, while tropospheric chemistry is presented with more details in section 2.5, with a focus on the production of tropospheric ozone. The chapter is largely based on the following books: *D. Jacob*, 1999 (Introduction to atmospheric chemistry); *R. Delmas, G. Megie and V.-H. Peuch*, 2005 (Physique et chimie de l'atmosphère); *G. Brasseur, R. Prinn and A. Pszenny,* 2003 (Atmospheric chemistry in a changing world) and *J. Seinfeld and S. Pandis*, 2006 (Atmospheric chemistry and physics: from air pollution to climate change).

2.1   Chemical composition of the Earth atmosphere

The Earth atmosphere consists in a thin layer of gases surrounding the planet. The density of this layer decreases with altitude, and although there is no definite boundary between the atmosphere and the outer space, its thickness is of the order of 100 km. As illustrated on Figure 2-1, compared to the 6400 km of the Earth radius, the Earth atmosphere is very thin and is often compared to an apple skin.
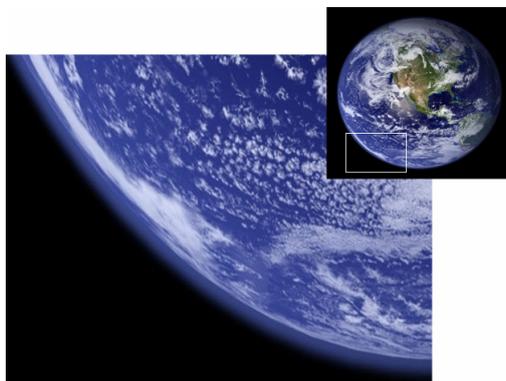
Figure 2-1: The Earth atmosphere [http://www.meteo.psu.edu/~nese/1_1_apple_skin.png].

Vertically, the atmosphere is partitioned into several layers. This partition is based on variation of the temperature with altitude[1]. The first layer is called the troposphere and extends from the surface to 8-18 km altitude depending on the latitude and the season. The second layer is the stratosphere, which extends up to 50 km. These two layers contain 99% of the total atmospheric mass, and are the domains of main interest from an environmental perspective.

The chemical composition of the atmosphere results from the Earth biogeochemical cycles, *i.e.* from the continuous exchanges of matter between the different terrestrial spheres (the atmosphere, the hydrosphere, the biosphere, the pedosphere, the cryosphere and the lithosphere). Many of the atmospheric chemical compounds are constantly taken up and regenerated by biological processes

---

[1] A more detailed definition of the atmospheric layers based on the temperature profile is given in section 2.2.1

(microbial activity in soils, photosynthesis and respiration, foliage and ocean emissions, etc…). Typical mixing ratios[2] of some important atmospheric gases are given in Table 2-1 for dry air. The most abundant are nitrogen (78%) and molecular oxygen (21%), while water vapour is the third component (0.33% in average). Carbon dioxide (CO_2) plays a special role because it contributes to photosynthesis, which is the driving force of the biosphere (see chapter 3). The abundances of those gases are regulated by long timescale biogeochemical processes. Noble gases (argon, neon, helium) are chemically inert and do not take part in biogeochemical cycles. The other gases are minor compounds, present in the atmosphere at extremely low concentrations (from a few ppm to several tens of ppt) and are called trace gases. Despite their low concentrations, trace gases are of critical importance for Earth environment and climate. The natural greenhouse effect and the stratospheric ozone layer are fundamental elements of the planet equilibrium which allow life on Earth. Today however, some trace gases undergo important concentration increases because of human activities. They are responsible for phenomena ranging from urban photochemical smog and acid deposition to stratospheric ozone depletion and potential climate change.

Table 2-1: Averaged global concentrations and lifetimes of some important atmospheric gases [*Delmas et al.*, 2005], * updated with 2005 value [*Forster et al.*, *IPCC*, 2007].

Constituent	Molecular Weight	Mixing ratio[2]	Mean atmospheric lifetime
Nitrogen (N_2)	28.03	0.78	15×10^{16} years
Oxygen (O_2)	32.00	0.21	8×10^3 years
Argon (Ar)	39.95	0.0093	infinite
Carbon dioxide (CO_2)	44.01	380 ppm*	15 years
Neon (Ne)	20.18	18 ppm	infinite
Helium (He)	04.00	5.2 ppm	infinite
Methane (CH_4)	16.04	~1.77 ppm*	9 years
Ozone (O_3)	48.00	~0.02-2 ppm	1 to 2 months
Hydrogen (H_2)	02.02	~500 ppb	10 years
Nitrous oxide (N_2O)	44.01	~319 ppb*	120 years
Carbon monoxide (CO)	28.01	~40-200 ppb	2 months
Ammonia (NH_3)	17.03	~1-10 ppb	20 days
Nitrogen dioxide (NO_2)	46.00	~1-10 ppb	1 day
VOC ($C_xH_yO_z$)	variable	~1-50 ppb	hours-days
Sulfur dioxide (SO_2)	64.06	~1-20 ppb	1 day
Not included in above dry atmosphere:			
Water vapour (H_2O)	18.01	Variable	6 to 15 days

The cycle "emission - chemical transformation – transport - deposition", schematically described in Figure 2-2, is the atmospheric phase of the biogeochemical cycles and describes how chemical compounds are emitted from the surfaces of the oceans and the continents, how they evolve in the atmosphere and how they are eliminated from the atmosphere to be recycled in other Earth environmental reservoirs. The sources of the minor atmospheric compounds are essentially located at the Earth surface. After being emitted in the atmosphere, species are chemically transformed and transported horizontally and vertically by the winds. Depending on their lifetime in the atmosphere, minor compounds have an impact on very different geographic and temporal scales. The concentrations of chemical compounds and their effects need to be known across a wide range of temporal and spatial scales. Short-lived compounds have primarily a local impact, causing for

---

[2] The mixing ratio of a gas x (C_x) is the number of moles of x per mole of air [mol mol^{-1}]. C_x is often expressed in ppm: parts per million [10^{-6}], ppb: parts per billion [10^{-9}] or ppt: parts per trillion [10^{-12}].

The number density of a gas x (n_x) is the number of molecules of x per unit volume of air [molec.cm^{-3}] and $n_x = C_x n_{air}$, where n_{air} is the air number density.

example atmospheric pollution problems in urban and industrialised areas, while long-lived species are transported along large distances or to higher altitudes and have therefore a more global impact, typically on climate or on stratospheric ozone.

In many respects, ozone is one of the most important minor gases in the atmosphere. Stratospheric ozone screens living organism from biologically harmful solar ultraviolet radiation while tropospheric ozone is a greenhouse gas and produces adverse effects on human health and plants when present at too elevated levels. Ozone is also a major source of the OH radical, the main agent of chemical transformation in the troposphere.
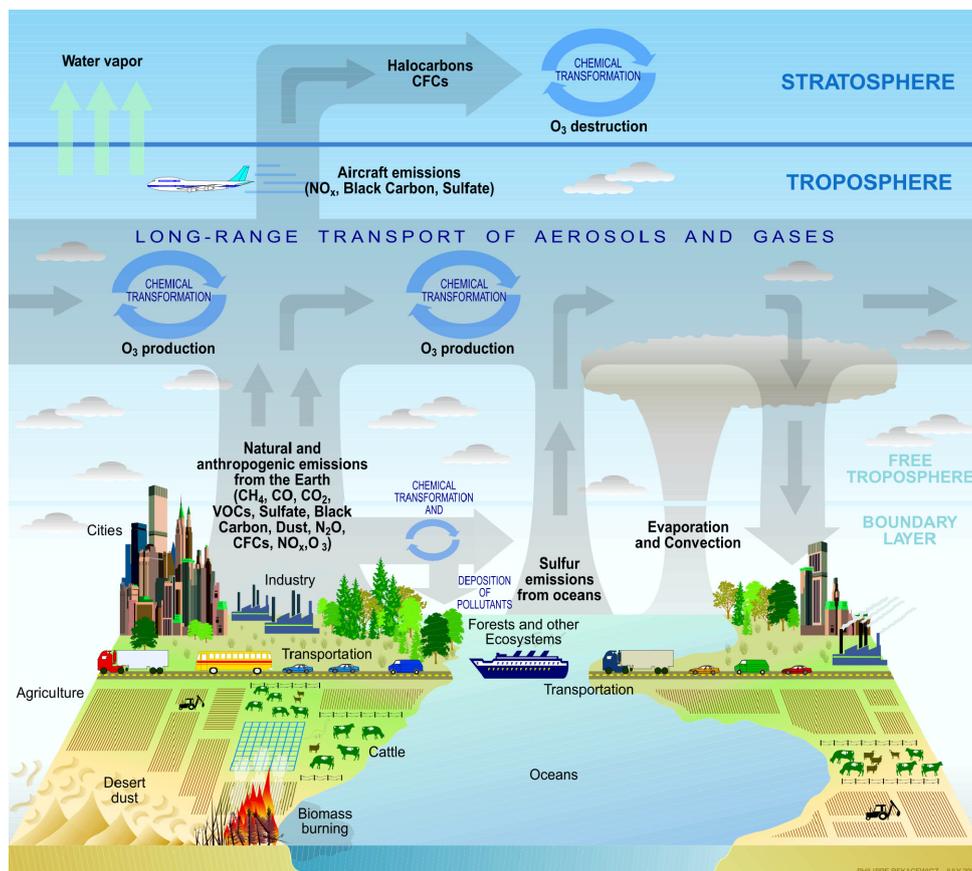
Figure 2-2: Emission, chemical transformation, transport and deposition processes related to atmospheric composition. **These processes link the atmosphere with other components of the Earth system, including the oceans, lands, and terrestrial and marine plants and animals [© P. Rekacewicz, US CCSP].**

2.2 Distribution of atmospheric trace gases

The distribution of gases in the atmosphere, *i.e.* the spatial variability of their concentrations on a vertical and horizontal plane, is directly linked to their lifetime and to the typical timescales of vertical and horizontal transport in the atmosphere. The description of air circulation is the subject of meteorology and dynamic of the atmosphere. We will only outline these processes in order to explain the typical transport timescales in the vertical and horizontal (longitudinal and meridional) planes. These movements have three origins:

- the vertical thermal structure of the atmosphere,
- the Earth rotation on its axis,

- the inhomogeneity of the solar energy distribution, with a maximum in the equatorial belt and a minimum at the poles.

These factors cause dynamic and thermodynamic instabilities that generate horizontal and vertical air movements, which define the general atmospheric circulation in the longitudinal and meridional directions.

2.2.1 Pressure and temperature vertical profiles

A typical temperature profile is represented with the yellow curve on Figure 2-3. Reversals of the temperature gradient are commonly used to define the atmospheric layers. The troposphere is characterized by a decrease of temperature with altitude, caused by the bottom-up heating of the atmosphere. Above troposphere, the influence of the Earth surface is strongly reduced. The top of the troposphere, the tropopause, is the point where air ceases to cool with height, and becomes almost completely dry. The ozone layer is located in the stratosphere, with a maximum concentration around 20 km. The lower stratosphere is almost isotherm, but the absorption of solar radiation by ozone is a new heating source, causing an increase of the temperature with altitude (between 30 and 50 km). The stratosphere extends from the tropopause to the stratopause, where a local maximum in temperature occurs. In the mesosphere, above the ozone layer, the temperature decreases again with altitude. The mesosphere extends up to 85 km (mesopause), above which lies the thermosphere where the temperature increases again with altitude due to absorption of strong ultraviolet solar radiation by N_2 and O_2.
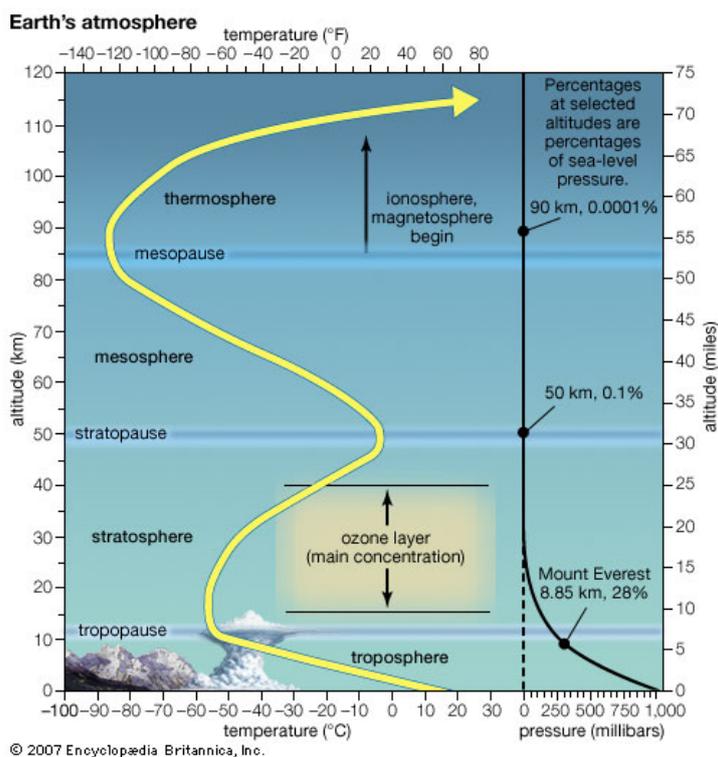
Figure 2-3: Layers of the Earth atmosphere. The yellow line shows a typical vertical profile of temperature, while the black curve represents the pressure profile [© 2007 Encyclopaedia Britannica, Inc.].

Pressure varies with altitude following the barometric law:

$$P(z) = P(0)e^{-z/H} \tag{2-1}$$

where **H** is defined as the scale height for the atmosphere[3] (about 8 km) and the temperature is considered to be constant, which is true in first approximation as **T** varies only by 30% below 80 km, while the pressure changes by six orders of magnitude. This exponential behaviour of the atmospheric pressure is represented with the black curve in Figure 2-3. The fraction of total atmospheric mass[4] located above altitude **z** is **P(z)/P(0)**. The global mean pressure surface being 984 hPa, the troposphere and stratosphere account together for 99.9% of the total atmospheric mass. The troposphere alone accounts for 85% of total atmospheric mass (pressure at tropopause level is about 150 hPa) and contains almost the totality of water vapour, clouds and precipitations. It is the part of the atmosphere where the influence of the Earth surface is predominant. The troposphere is further divided into two sub layers, the planetary boundary layer which is directly influenced by the surface of the Earth (the thickness of this layer varies between a few hundred meters and two kilometres), and the free troposphere above.

2.2.2 Time scales of vertical and horizontal exchanges
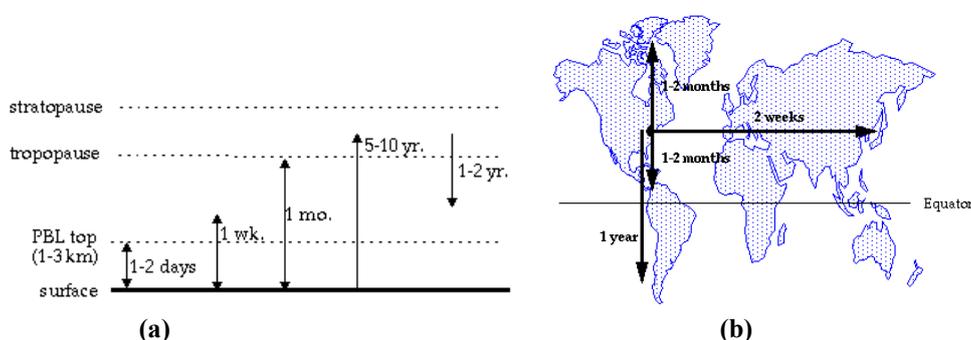
Figure 2-4: Typical time scales for (a) vertical transport and (b) horizontal transport of chemical compounds in the atmosphere [© *Jacob et al.*, 1999].

Figure 2-4 represents the typical timescales of vertical and horizontal transport in the atmosphere. In the troposphere, convection is an efficient process of vertical transport because of the temperature decrease with altitude. The reversion of the temperature gradient between the troposphere and the stratosphere makes the tropopause act as a dynamical barrier and constitutes a transition zone for most trace gases. As illustrated on Figure 2-4, it takes one month in average for the transport of air between the surface and the tropopause. In the planetary boundary layer, the air circulation is strongly influenced by interactions with the ground. Because of the turbulent fluctuations occurring in this layer, vertical and horizontal mixing is intense and takes only 1 to 2 days. Transport of the air into the free troposphere takes one week. Vertical exchanges between the troposphere and the stratosphere are limited, yet they exist at all latitudes. The typical timescale of exchanges between the troposphere and the stratosphere is several years. The stratosphere is characterized by much stronger stratification than the troposphere because the temperature increases with altitude. This stratification inhibits vertical movements, while horizontal movements can be fast. Mixing in the stratosphere takes months to year.

The largest scales of motion are in the horizontal direction and form the basis for the general circulation of the atmosphere. Figure 2-5 illustrates the general patterns of the tropospheric

---

[3] $H = \dfrac{RT}{M_{air}\,g}$, where M_{air} is the molecular weight of air (28.96 g·mol^{-1}), **g** is the acceleration of gravity (9.8 m·s^{-2}), **R** is the gas constant (8.31 J·mol^{-1}·K^{-1}) and **T** is the mean atmospheric temperature.

[4] The total mass of the atmosphere is 5.14x10^{18} kg.

circulation. Longitudinal transport results from West-East air movements, modified by the Coriolis force to a variable degree, depending on the latitude and the altitude. Meridional movement is made of cells, among which the best known is the Hadley cell. Vertical exchanges between the troposphere and the stratosphere are strongest at the level of permanent ascending branches of tropical cells, in the inter-tropical convergence zone. In the stratosphere, the mean meridional circulation is known as the Brewer-Dobson circulation (not shown on Figure 2-5). Air that enters the stratosphere in the tropics is transported poleward and downward and finally exits the stratosphere at middle and high latitudes.

As illustrated on Figure 2-4 (b), transport of air around the Earth in the longitudinal direction takes a few weeks. Meridional transport is slower, it takes typically 1-2 months for air at mid-latitudes to reach the tropics or Polar Regions. Inter-hemispheric transport is even slower; it takes about one year for air to exchange between the Northern and Southern hemispheres. The typical length of these movements impacts directly the concentration distribution of trace gases in the atmosphere.
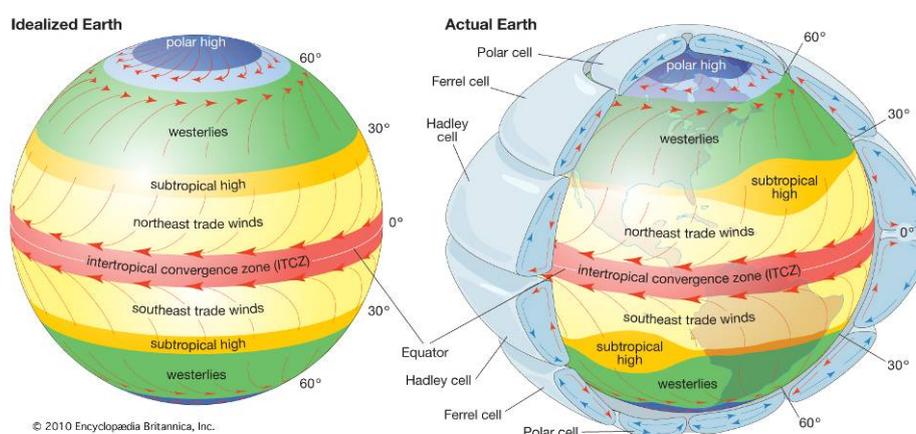
Figure 2-5: Horizontal and vertical patterns of tropospheric circulation over an idealized Earth with a uniform surface (left) and the actual Earth (right) [© 2010 Encyclopaedia Britannica, Inc.].

2.2.3 Lifetime of species in the atmosphere

Permanent exchanges of matter occur between the different environmental spheres, which are assimilated to reservoirs. Concentrations of chemical species in the atmospheric reservoir are controlled by four types of processes:

- Emissions: chemical species are emitted by a variety of sources.
- Chemistry: reactions in the atmosphere can lead to the formation and removal of species.
- Transport: winds transport atmospheric species away from their point of origin.
- Deposition: all materials in the atmosphere is eventually deposited back to the Earth's surface.

A simple box model allows describing the concentration of one species in a selected atmospheric domain (it can be an urban area, the troposphere or the global atmosphere). The temporal variation of the mass *m* of the species in the box is obtained from the mass balance equation:

$$\frac{\partial m}{\partial t} = \sum sources - \sum sinks = F_{in} + E + P - F_{out} - L - D \qquad (2\text{-}2)$$

F_{in} and F_{out} are the transport terms. They represent the variation induced by the incoming and outgoing fluxes of the species at the edges of the considered domain. If the domain is the global atmosphere then $F_{in}=F_{out}=0$. The production and loss rates of the species inside the domain may include contributions from emissions (E), chemical production (P), chemical loss (L) and deposition (D). F_{in}, E and P are the source terms while F_{out}, L and D are the sinks of the species. The species lifetime τ can be defined as the average stay time of the species in the considered domain. From the mass balance equation, the lifetime is defined as the ratio between the mass m and the sinks terms:

$$\tau = \frac{m}{-(\frac{\partial m}{\partial t})_{sinks}} = \frac{m}{F_{out} + L + D} \qquad\qquad (2\text{-}3)$$

This definition allows calculating the mean lifetimes of chemical species in the atmospheric reservoir. For a given species, the lifetime varies with the localisation in the atmosphere (boundary layer/free troposphere, continental area/ocean area …) and depends on the sinks efficiency. It is interesting to determine the relative importance of the different sinks contributing to the overall removal of a species. Sink-specific lifetimes can be defined against export (τ_{out}), chemical loss (τ_c) and deposition (τ_d):

$$\tau_{out} = \frac{m}{F_{out}}, \; \tau_c = \frac{m}{L}, \; \tau_d = \frac{m}{D}. \text{ The sinks apply in parallel so that: } \frac{1}{\tau} = \frac{1}{\tau_{out}} + \frac{1}{\tau_c} + \frac{1}{\tau_d}.$$

The term *lifetime* is mainly used when the loss term is dominated by a chemical process while the term *residence time* is often used when the dominant sinks are physical processes such transport and deposition. The typical mean lifetimes of the major atmospheric compounds are given in Table 2-1. For the species with a very short lifetime (less than a few days), τ is deduced from the kinetics of their main oxidation reactions.

2.2.4 Spatial and temporal scales of variability of atmospheric constituents

The spatial and temporal variability of atmospheric species is determined by the spatial scale covered by an air parcel during the lifetime of a considered species. Figure 2-6 presents schematically the lifetimes of several compounds with the associated spatial scales. It is only a simplified representation of the interactions, often subtle, between the transport time scales and the lifetime of the compounds but it allows appreciating the spatial and temporal variability of atmospheric constituents. Figure 2-7 shows typical vertical profiles of long-lived gases.

Gases with very long lifetimes (hundreds of years), *i.e.* much larger that the typical timescale of horizontal and vertical transports, present spatially homogenous distributions (vertically and horizontally). Their mixing ratio profile is constant while their number density profile follows, like dry air, the perfect gas law and varies in a logarithmic way. Gases presenting long lifetimes (tens of years), like nitrous oxide (N_2O), chlorofluorocarbons (CFCs) and methane (CH_4), present quasi-homogeneous distributions, yet with an inter-hemispheric gradient of several percents, associated with the larger density of sources in the Northern hemisphere. Gases presenting lifetime of the order of the month, i.e. smaller than the typical timescale of horizontal transports, present less homogeneous distributions. It is for example the case of carbon monoxide (CO) or tropospheric ozone. Most of the compounds released at the Earth's surface do not reach the stratosphere, instead they are either decomposed by the main tropospheric oxidants, broken down by sunlight, or

deposited back to the surface with rain or as particles. Short lifetime compounds (less than 1 week) present generally large concentrations in the boundary layer near the sources and low concentrations in the free troposphere. Finally, gases with very short lifetimes, like for example nitrogen oxides and NMVOCs (hours) or the OH radical (seconds), present very high spatial and temporal variability (with often a pronounced diurnal cycle), and their concentrations depend directly on the proximity of the sources. Ozone is a particular case because its main source is located in the stratosphere and drives the chemistry of the stratosphere. Most of the other stratospheric gases are either really long lived compounds emitted originally into the troposphere, such as CFCs, or are brought in by severe volcanic eruptions (generally sulphur containing compounds and aerosols).
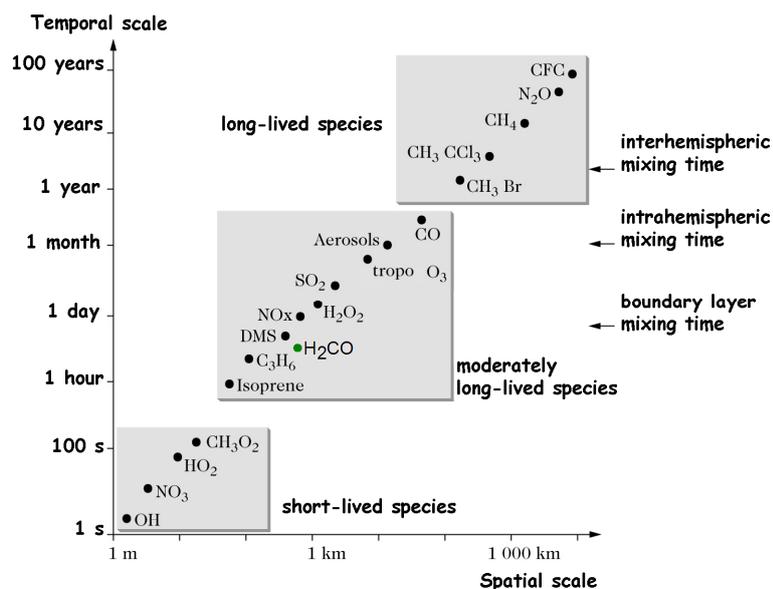
Figure 2-6: Lifetimes and spatial scales for some atmospheric species [© *Seinfeld and Pandis*, 2006].
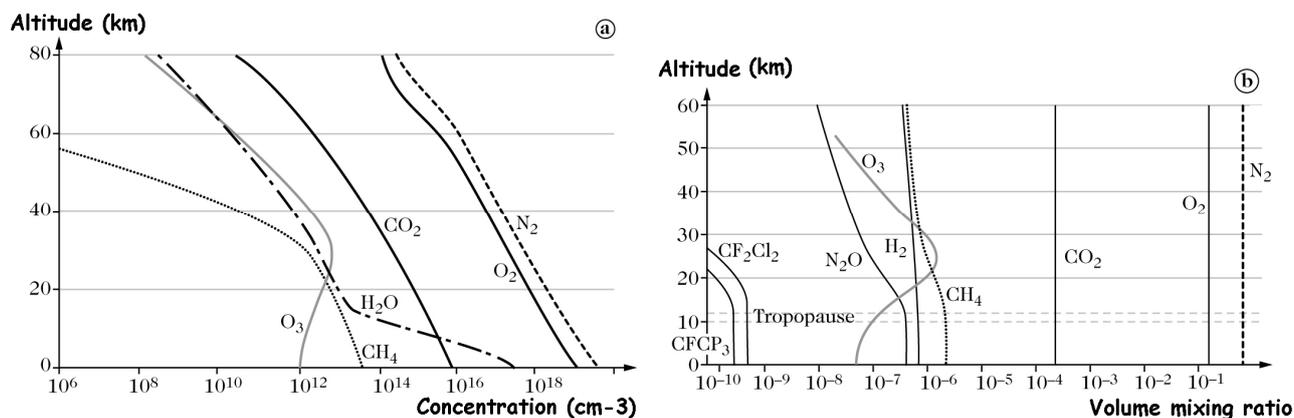
Figure 2-7: Vertical distribution in (a) number density [molec.cm^{-3}] and (b) mixing ratio of the main atmospheric species [© *Delmas et al.*, 2005].

2.3 *Solar and terrestrial radiation*

The main source of heat to the Earth is solar energy, which is transmitted from the Sun to the Earth by radiation and is converted to heat at the Earth's surface. The spectrum of the solar radiation measured outside the Earth's atmosphere (Figure 2-8) matches closely that of a black body at 5800

K[5], except for the solar Fraunhofer lines which are due to absorption by the Sun photosphere. Solar radiation peaks in the visible wavelength range (Vis: λ=0.4-0.7 μm). About half of the solar radiation flux is at infrared wavelengths (IR: λ>0.7 μm) and a small fraction is in the ultraviolet region (UV: λ<0.4 μm). The solar radiation flux at sea level is weaker than at the top of atmosphere, in part because of diffusion by clouds, particles and molecules, and because of major atmospheric absorption features by O_2 and O_3 in the UV and by H_2O and CO_2 in the IR.
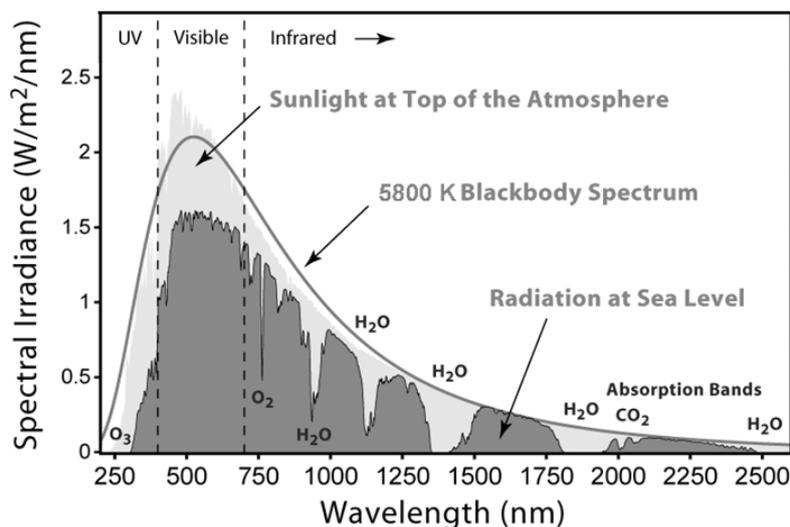
Figure 2-8: Solar radiation spectrum for direct light at both the top of the Earth's atmosphere and at sea level [© R. Rohde / Global Warming Art].
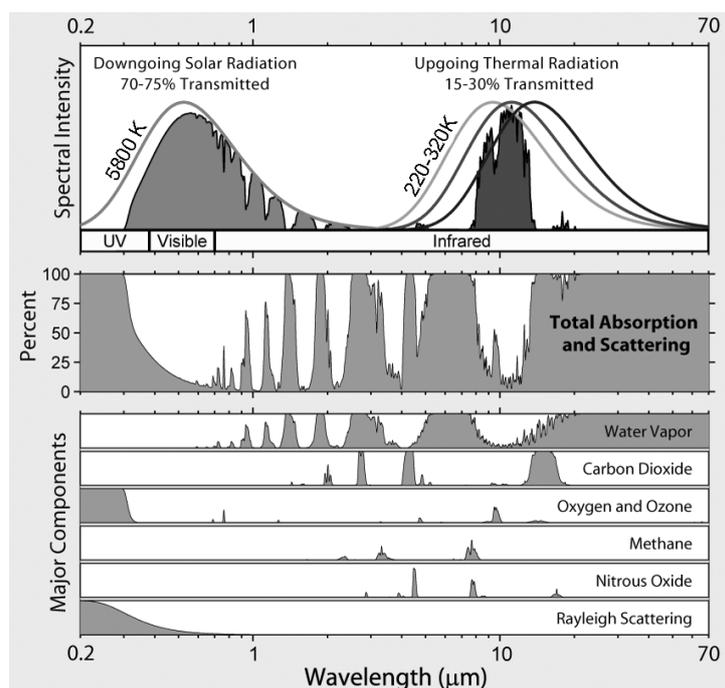
Figure 2-9: Top panel: normalised solar and upgoing thermal radiation. Middle panel: Atmospheric absorption bands. Lower panel: individual absorption spectra for major greenhouse gases plus Rayleigh scattering [© R. Rohde / Global Warming Art].

---

[5] The total emission flux emitted by a black body is $\phi_T = \sigma T^4$ where σ is the Stefan-Boltzmann constant (5.67x10^{-8} W·m^{-2}·K^{-4}) and T is the temperature of the body.

In response to the solar radiation, the Earth itself emits radiation to space. On Figure 2-9, the normalised solar radiance is displayed together with the normalised terrestrial radiation spectrum, which is a combination of blackbody spectra for different temperatures, ranging from 220 to 320 K. The wavelength range of maximum terrestrial emission is 5-20 μm. The efficiency of radiation absorption by the atmosphere is plotted in the second panel of Figure 2-9 as a function of wavelength. Absorption is ~100% efficient in the UV due to O_2 and O_3 absorption. The atmosphere is largely transparent at visible wavelengths. At IR wavelengths the absorption is again almost 100 % efficient because of greenhouse gases, mainly water vapour and CO_2. There is however a window between 8 and 13 μm, near the peak of terrestrial emission, where the atmosphere is only a weak absorber except for a strong O_3 feature at 9.6 μm. This atmospheric window allows direct escape of radiation from the surface of the Earth to space and is of great importance for defining the temperature of the Earth surface.

In order to maintain a stable climate, the Earth must be in energetic equilibrium between the radiation it receives from the Sun and the radiation it emits out of space. Figure 2-10 illustrates the radiative budget of the Earth. Because of the planetary reflectivity[6], 30% of the incoming solar radiation ($F_S/4$ = 342 W·m^{-2})[7] is reflected either by the atmosphere or by the Earth surface. The remaining of the solar radiation is absorbed by the atmosphere and by oceanic and continental surface. Outgoing infrared radiation compensates exactly the absorption of solar radiation. To a large part, it comes from the radiation of the atmosphere itself. Only a small part of the radiation emitted by the surface goes trough the atmosphere without being absorbed, because the efficiency of absorption of the radiation is very high in the IR region (see Figure 2-9). The Earth radiative budget is closed by taking into account the thermal and evapo-transpiration fluxes at the surface. Note that the temperature of radiative equilibrium of the Earth would be 255K without atmosphere[8] but the atmospheric greenhouse gases rise the mean surface temperature up to 288K (+15°C).
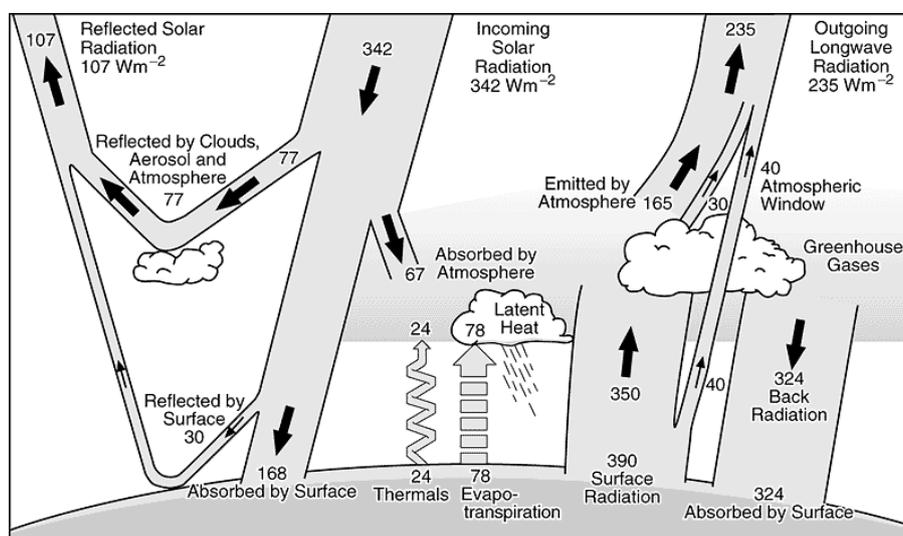
Figure 2-10: Estimate of the Earth's annual and global mean energy balance. The amount of incoming solar radiation absorbed by the Earth and atmosphere is balanced by the Earth and atmosphere releasing the same amount of outgoing longwave radiation. Radiation fluxes are expressed in W·m^{-2} [*Le Treut et al., IPCC*, 2007].

[6] The planetary reflectivity or albedo *A* is the ratio of up and down-welling radiation.

[7] At the distance between the Earth and the Sun, the solar radiation flux F_S is 1370 W·m^{-2}. F_S is intercepted by the Earth over a disk of cross-sectional area πR_E^2 and the radiation received by the Earth per unit area is therefore $F_S/4$.

[8] If the emission flux from the Earth is approximated as that of a blackbody of temperature T_E: $\sigma T_E^4 = \dfrac{F_S(1-A)}{4}$.

Important greenhouse gases are those present at concentrations sufficiently high to absorb a significant fraction of the radiation emitted by the Earth. By far, the most important greenhouse gas is water vapour because of its abundance and its extensive IR absorption features. There is presently much concern that anthropogenic increases in greenhouse gases could be inducing rapid surface warming of the Earth. The naturally occurring greenhouse gases CO_2, CH_4 and N_2O showed large increases over the past century due to human activity (Figure 2-11). Additional greenhouse gases produced by chemical industry have accumulated in the atmosphere over the past decades and added to the greenhouse effect. Tropospheric O_3 is also an important anthropogenic greenhouse gas because of its determining absorption in the atmospheric window. The Kyoto protocol was adopted in December 1997 with the aim to reduce greenhouse gas emissions. By the end of 2009, 187 states had signed and ratified the protocol (with the exception of the United States). Under the protocol, 39 industrialized countries and the European Union committed themselves to a reduction of greenhouse gases, such as CO_2, CH_4, N_2O or HCFCs.
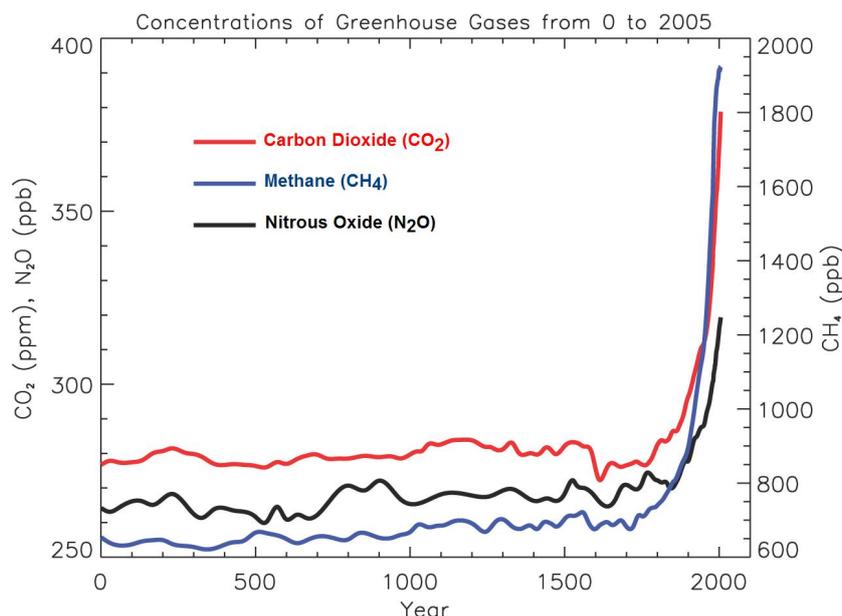
Figure 2-11: Atmospheric concentrations of important long-lived greenhouse gases over the last 2000 years. Increases since about 1750 are attributed to human activities in the industrial era [*Forster et al.*, *IPCC*, 2007].

2.4 Stratospheric ozone

Figure 2-12 presents a typical ozone vertical profile. The ozone amount is small in the troposphere (~ tens of ppb), increases abruptly crossing the tropopause and reaches its maximum in the lower stratosphere (a few ppm). Above 30 km, the ozone amount decreases exponentially. The stratospheric ozone layer accounts almost 90% of the total ozone in the atmosphere. It is formed from the direct photolysis of oxygen. Ozone in the stratosphere plays a vital role as a shield against low wavelengths ultraviolet light, which is damaging to many life forms. The troposphere contains only about 10% of atmospheric ozone. However, tropospheric ozone is harmful for live beings and vegetation and is a greenhouse gas. The mechanisms of tropospheric ozone production will be the main subject of the next section about tropospheric chemistry.
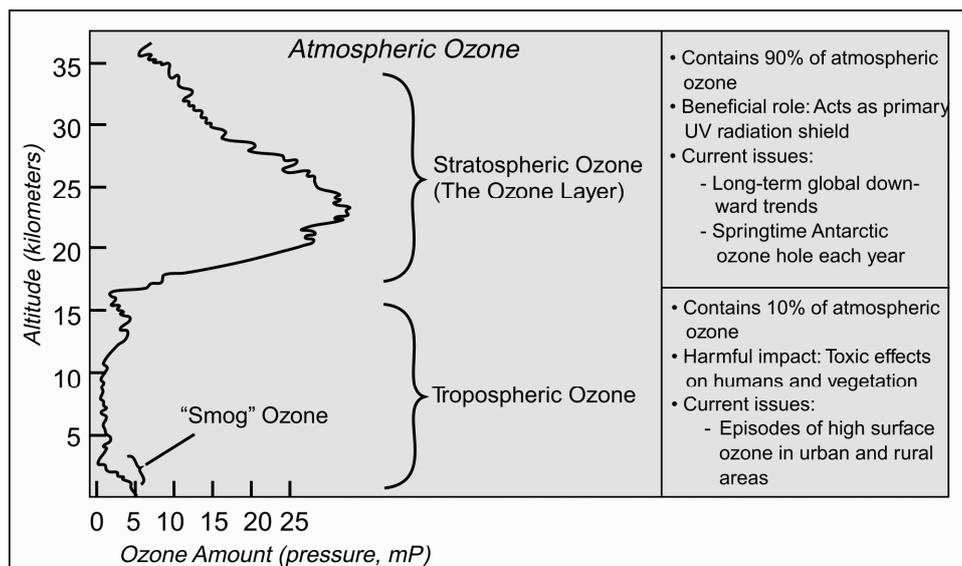
Figure 2-12: The distribution of atmospheric ozone in partial pressure[9] as a function of altitude [*WMO*, 1998].

The process of stratospheric ozone production is known as the Chapman mechanism[10] and can be described with the following equations:

$$O_2 + h\upsilon \rightarrow O + O \qquad\qquad (\lambda<240nm) \qquad\qquad \text{R 1}$$

$$O + O_2 + M \rightarrow O_3 + M \qquad\qquad (+\Delta Q) \qquad\qquad \text{R 2}$$

$$O_3 + h\upsilon \rightarrow O(^1D) + O_2 \qquad\qquad (\lambda<320nm) \qquad\qquad \text{R 3-1}$$

$$O_3 + h\upsilon \rightarrow O(^3P) + O_2 \qquad\qquad (\lambda<1180nm) \qquad\qquad \text{R 3-2}$$

$$O(^1D) + M \rightarrow O(^3P) + M \qquad\qquad\qquad \text{R 4}$$

$$O + O + M \rightarrow O_2 + M \qquad\qquad\qquad \text{R 5}$$

$$O(^3P) + O_3 \rightarrow 2O_2 \qquad\qquad\qquad \text{R 6}$$
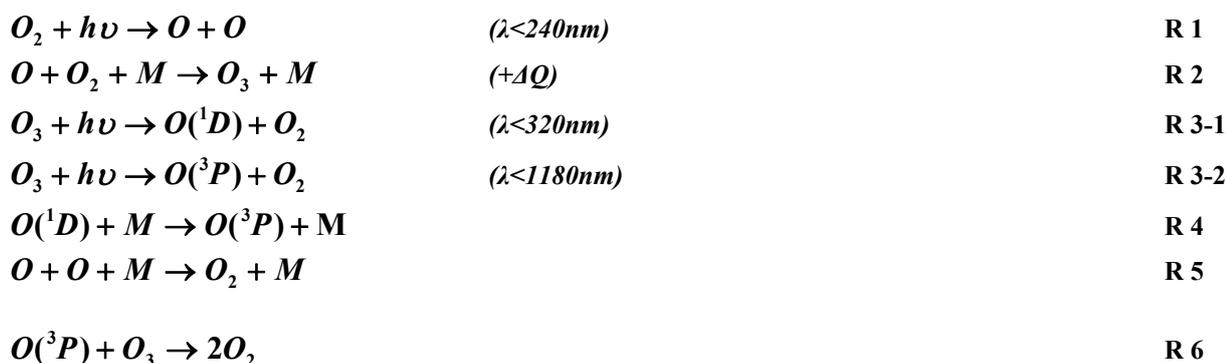
Ozone is formed in the stratosphere by photodissociation of molecular oxygen O_2 yielding two oxygen atoms, which combine rapidly with O_2 to form ozone (R1-R2). This reaction is very fast (less than a fraction of second) and releases heat, which explains the positive gradient of temperature in the stratosphere. The O_3 molecules are then photodissociated, yielding $O(^1D)$, the oxygen atom in an excited singlet state (R3-1), or $O(^3P)$, the oxygen atom in the fundamental state (R3-2). This Chapman cycle leads to a chemical steady state for O_3 (R1-R5). For O_3 to actually be lost, the oxygen atom must undergo another reaction where O_3 recombines with O to form two molecules of oxygen (R6). There is a rapid cycle between O and O_3 and a slower cycle between O_2 and (O+O_3), it is therefore convenient to refer to the sum as the chemical family of odd oxygen ($O_x \equiv O_3 + O$), which is produced by R1 and consumed by R6. While the photolytic lifetime of stratospheric O_3 is less than half an hour, the O_x lifetime varies from about several months to a year in the lower stratosphere (at the altitude of the maximum ozone concentration). Ozone is mostly produced in the equatorial region, where the solar insulation is maximal, and the odd-oxygen

---

[9] The total pressure *P*, the partial pressure P_x and the mixing ratio C_x of a gas *X* are related by: $P_X = C_X P$.

[10] In 1930, Sydney Chapman proposed a theory for the origin of the ozone layer, which was first detected in the 1920s from observations of the solar UV spectrum.

species are transported from the equatorial region toward the poles, as a result of the Brewer-Dobson circulation.

Although the Chapman mechanism is successful in reproducing the general shape of the O_3 layer, it overestimates the observed O_3 concentrations by a factor of 2 or more. In reality, stratospheric ozone is not only chemically destroyed by photolysis or reaction by atomic oxygen but also by catalytic ozone depletion cycles. The hydrogen cycle was the first identified catalytic cycle [*Bates and Nicolet*, 1950]:

$$O(^1D) + H_2O \rightarrow 2OH \qquad\qquad\qquad\qquad\qquad\qquad \textbf{R 7}$$

$$OH + O_3 \rightarrow HO_2 + O_2 \qquad\qquad\qquad\qquad\qquad\qquad \textbf{R 8}$$

$$HO_2 + O_3 \rightarrow OH + 2O_2 \qquad\qquad\qquad\qquad\qquad\qquad \textbf{R 9}$$

$$OH + HO_2 \rightarrow H_2O + O_2 \qquad\qquad\qquad\qquad\qquad\qquad \textbf{R 10}$$
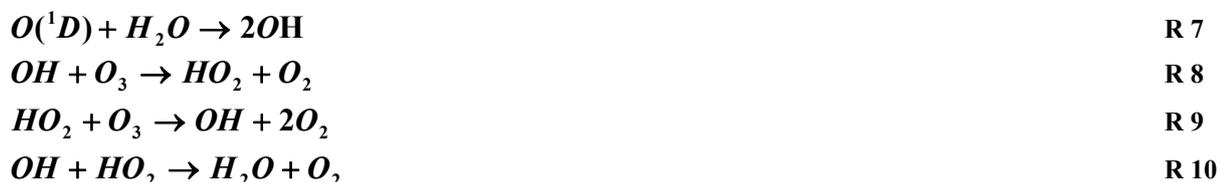
Water vapour is oxidized by $O(^1D)$ (produced from R3-1) to form the hydroxyl radical OH (R7). The ensemble of OH and HO_2 is referred to as the HO_x chemical family. The reactions R8 and R9 consume O_3 while conserving HO_x. Termination of the cycle requires loss of HO_x by a reaction such as R10. HO_x catalysis is a significant O_3 sink but not sufficient to reconcile the chemical budget of O_3. Other catalytic reactions involving nitrogen, chlorine and bromine species can trigger chain reactions destroying O_3 [*WMO*, 2002]. The concentrations of OH_x, NO_x, ClO_x and BrO_x in the atmosphere are determined by the level of emission of their main precursors at the Earth surface (water vapour and methane for HO_x, nitrous oxide for NO_x, CFCs for ClO_x, methyl bromide (CH_3Br) and halons for BrO_x). These are long-lived gases able to reach the stratosphere and therefore to perturb the stratospheric ozone cycle. Since industrialisation, the emissions of CFCs and halons had considerably increased and caused the depletion of the ozone layer, discovered in the early eighties (with the ozone hole in Antarctica) [*Farman et al.*, 1985]. In response, the Montreal Protocol agreement was negotiated in 1987. This agreement regulated the production of CFCs and halons. Many current issues concerning stratospheric ozone are described in the WMO reports including long-term global trends and Artic and Antarctic ozone loss, involving polar stratospheric clouds and heterogeneous chemistry [*WMO*, 2002]. These processes are behind the scope of this thesis and will not be discussed further here. Only the hydrogen catalytic cycle has been described because reaction R7 is the main source of the OH radical, which provides the principal sink for a large number of species emitted in the atmosphere and is therefore a central component of tropospheric chemistry.

2.5 Tropospheric chemistry

The main subject of this section is to describe the processes that control ozone concentration and distribution in the troposphere. Tropospheric ozone is the precursor of OH and plays therefore a key role in maintaining the oxidizing power of the troposphere. It is also of environmental importance as greenhouse gas and as toxic pollutant. An obvious source is the large reservoir of stratospheric ozone that enters the troposphere when air is transferred across the tropopause. However, production of O_3 in the troposphere involves much more complex processes in which chemistry plays a major role, resulting *e.g.* in photochemical smog near the surface as shown in Figure 2-12. It can occur anywhere in the troposphere, depending however on the available near-UV radiation, and on the concentration of various ozone precursors such as volatile organic compounds, nitrogen oxides and water vapour.

The main processes of tropospheric ozone chemistry are summarised in Figure 2-13 and will be detailed in the following sections. In brief, like in the stratosphere, the process of ozone destruction

by sunlight produces excited atomic oxygen $O(^1D)$ (at $\lambda<320$ nm), a small fraction of which reacts with atmospheric water vapour to produce OH radicals. Photons at wavelength shorter than 290 nm are not present in the troposphere so that O_2 cannot be photolysed and O_3 cannot be produced via the Chapman mechanism. Instead, the oxygen atom needed to form O_3 is provided by the photolysis of NO_2. However, NO will in turn react with O_3 to reform NO_2, leading to a photochemical stationary state between NO, NO_2 and O_3. It is the oxidation of VOCs by OH, yielding HO_2, which leads to a net production of ozone by shifting the equilibrium of the NO_x cycle.
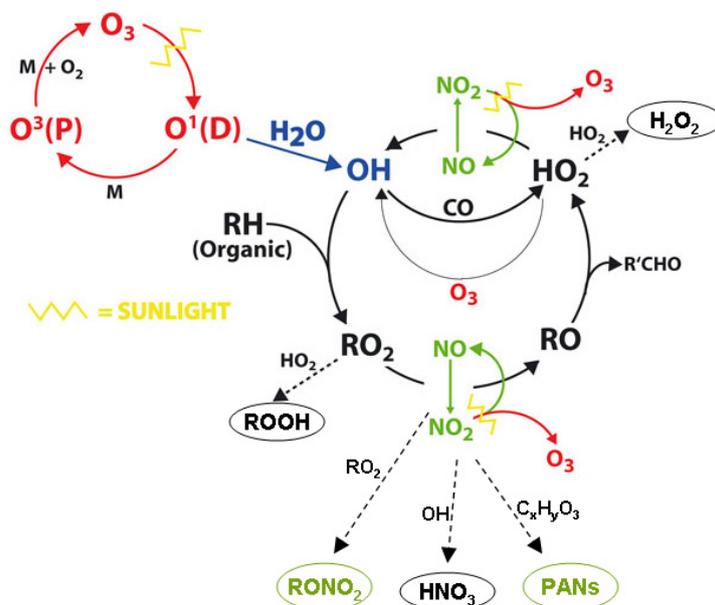
Figure 2-13: Chemical processes affecting tropospheric ozone.

In the remote atmosphere, ozone formation is sustained by the oxidation of CO and CH_4, while in the continental and urban atmosphere it is driven by much shorter-lived non-methane volatile organic compounds (NMVOCs), emitted from natural and anthropogenic sources. Since two classes of precursor compounds are involved in tropospheric ozone formation, a key question is how variations of VOC and NO_x levels affect the amount of ozone formed. The OH formation and the NO_x photochemical cycle will be described in sections 2.5.1 and 2.5.2 respectively. The CO and CH_4 oxidation mechanisms will then be described (2.5.3 and 2.5.4), illustrating the central role of formaldehyde. The global tropospheric ozone budget and the relative roles of NO_x and NMVOCs for continental air pollution will be evaluated in sections 2.5.5 and 2.5.6. A more complete description of tropospheric chemistry should also include sulphur compounds chemistry and heterogeneous chemistry, taking into accounts the presence of condensed water and solid particles. However, these processes will not be presented here, because we are mostly interested in the role of NMVOCs, particularly formaldehyde, in tropospheric chemistry.

2.5.1 Oxidizing capacity of the troposphere

Oxidation in the troposphere is of key importance because many environmentally important gases are principally removed from the atmosphere by oxidation, like greenhouse gases such as CH_4, combustion gases such as CO, agents for stratospheric O_3 depletion such as HCFCs, and others. Reactive molecules and free radicals[11], which provide the oxidising power of the atmosphere, are

---

[11] Radicals are defined as chemical species with an unpaired electron in the outer (valence) shell.

formed almost entirely by photochemistry and often involve coupling between ozone and water vapour. Main oxidants are hydroxyl radicals (OH), peroxy radicals (both inorganic HO_2 and organics RO_2) and peroxides (H_2O_2 and ROOH).

The OH radical is the primary oxidizing species in the troposphere. It reacts rapidly with almost all trace gases (with the notable exception of CFCs and N_2O which are mostly removed by photolysis in the stratosphere). Production of OH is achieved through the reaction of water vapour with $O(^1D)$ (R7). The production of $O(^1D)$ atoms by photolysis of O_3 (R3-1) is therefore critical to the generation of OH. Despite the near-total absorption of UV radiation by stratospheric ozone (see Figure 2-8), the production of $O(^1D)$ in the troposphere is not negligible. It is possible in a narrow wavelength band between 290 and 320 nm. Although considerably slower than in the stratosphere, the production of $O(^1D)$ in the troposphere is compensated in terms of OH production by the larger H_2O mixing ratios (10^2-10^3 times higher than in the stratosphere). Basic model calculations accounting for the penetration of UV radiation at 290-320 nm result in tropospheric OH concentrations of the order of 10^6 molec.cm^{-3} during daylight hours. The accurate measurement of OH is an extremely difficult problem because of its low concentrations. Furthermore, due to its short lifetime (of the order of one second), OH atmospheric concentrations are highly variable and respond very rapidly to changes in sources and sinks.

CO is the dominant sink of OH in most of the troposphere and CH_4 is next in importance. These two gases therefore play a critical role in controlling OH concentrations and more generally in driving radical chemistry in the troposphere. One concern is that sufficient OH is present in the troposphere to allow for oxidation of species such as CO, CH_4 and other pollutants. The chemical lifetime of O_3 in the lower stratosphere is of the order of the year, sufficiently long to allow transport of O_3 to the troposphere (at the rate of 1-2x10^{13} moles yr^{-1}). Each O_3 molecule crossing the tropopause can yield at most two OH molecules (R3-1, R7). The resulting maximum source of OH is therefore 2-4x10^{13} moles yr^{-1}. In comparison, the global estimated annual emissions of CO and CH_4 are respectively 1550 and 598 Tg yr^{-1} [*Ehhalt et al., IPCC*, 2001, see also chapter 3], corresponding to about 5.5 and 3.7x10^{13} moles yr^{-1}. There are therefore more molecules of CO and CH_4 emitted to the atmosphere each year than can be oxidized by OH molecules originating solely from O_3 transported across the tropopause [*Jacob et al.*, 1999]. There is therefore a major *in situ* source of O_3 and OH in the troposphere. A key factor is the presence in the troposphere of trace levels of NO_x that induces catalytic cycles regenerating OH and concurrently provides a major source of O_3.

2.5.2 Basic photochemical cycle of NO, NO_2 and O_3

In the troposphere, important nitrogen oxides are NO and NO_2 because they are much more reactive than N_2O. The estimated global source of NO_x amounts to 44±5 TgN yr^{-1} [*Denman et al., IPCC,* 2007]. Fossil fuel combustion (including aircraft emissions) accounts for more than half of the global source. Biomass burning, mostly from tropical agriculture and deforestation, accounts for another 13% while the lightning source is particularly uncertain. Other sources of NO_x include microbial nitrification and denitrification in soils and transport from the stratosphere. Although NO_x is emitted mainly as NO, cycling between NO and NO_2 takes place in the troposphere on a time scale of a few minutes during daytime. It is therefore more appropriate to consider the budget of the NO_x family as a whole. The photochemical NO_x cycle is written as:

$$NO + O_3 \rightarrow NO_2 + O_2 \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \textbf{R 11}$$

$$NO_2 + h\nu \rightarrow NO + O \qquad\qquad (\lambda<420nm) \qquad\qquad\qquad \textbf{R 12}$$

$$\underline{O + O_2 + M \rightarrow O_3 + M \qquad\qquad (+\Delta Q)} \qquad\qquad\qquad\quad \textbf{R 2}$$

$$NO_2 + h\nu + O_2 \rightarrow NO + O_3 \qquad\qquad\qquad\qquad\qquad\qquad\quad \textbf{R 13}$$
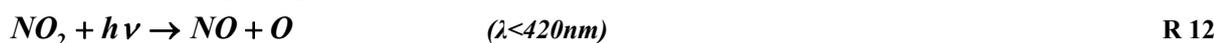
This cycle is null in term of ozone and NO_x production or destruction.

There are two major paths for removal of NO_x from the atmosphere. The principal daytime removal path for NO_x is oxidation to nitric acid HNO_3 (R14). During the night, NO_x is present exclusively as NO_2, which reacts with O_3 (R15). The produced nitrate radical NO_3 reacts also with NO_2 to produce N_2O_5 (R16). Whereas gas-phase reactions of N_2O_5 are quite slow, the heterogeneous (particle-phase) hydrolysis of N_2O_5 produces two molecules of HNO_3 (R17).

Day: $\quad NO_2 + OH + M \rightarrow HNO_3 + M \qquad\qquad\qquad\qquad\qquad \textbf{R 14}$

Night: $NO_2 + O_3 \rightarrow NO_3 + O_2 \qquad\qquad\qquad\qquad\qquad\qquad\qquad \textbf{R 15}$

$$\qquad\quad NO_3 + NO_2 + M \rightarrow N_2O_5 + M \qquad\qquad\qquad\qquad\quad \textbf{R 16}$$

$$\qquad\quad N_2O_5 + H_2O\ (aerosol) \rightarrow 2HNO_3 \qquad\qquad\qquad\qquad \textbf{R 17}$$

The resulting lifetime of NO_x varies from approximatively 6 hours in the boundary layer to several days to a week in the upper troposphere. In the troposphere, HNO_3 is mostly scavenged by precipitation because of its high solubility in water, leading to a lifetime of a few days in the lower-troposphere and a few weeks in the upper troposphere. Therefore, HNO_3 is not an effective reservoir for NO_x. Note that HNO_3 is a pollutant which, together with sulphuric acid, is responsible for the phenomenon of acid rain.

A more efficient mechanism for long-range transport of anthropogenic NO_x to the global troposphere is through the formation of reservoir species called peroxyacyl nitrates (PANs). PANs are produced in the troposphere by photochemical oxidation of organic compounds in the presence of NO_x. The general formulation of PAN formation is:

$$C_xH_yO_3 + NO_2 \rightarrow C_xH_yO_3NO_2 \qquad\qquad\qquad\qquad\qquad\qquad \textbf{R 18}$$

Formation of PAN is generally less efficient as a sink for NO_x than formation of HNO_3. However, in contrast to HNO_3, PAN is only sparingly soluble in water and is not removed by deposition. Its principal loss is by thermal decomposition, regenerating NO_x. The lifetime of PAN is only 1 hour at 295K but several months at 250K. In the lower troposphere, NO_x and PAN are typically near chemical equilibrium but in the middle and upper troposphere, PAN can be transported over long distances and decompose to release NO_x far from their sources. Long-range transport of PAN at high altitude plays a critical role in allowing anthropogenic sources to affect tropospheric NO_x on a larger scale (and hence O_3 and OH). Nitrogen oxides are often grouped into one large family under the denomination NO_y, including NO_x, HNO_3 and PAN.

Figure 2-14 illustrates the photochemical NO_x cycle. The basic NO_x photochemical cycle is null for the production of ozone, unless perturbed by the presence of HO_2 (or RO_2), resulting from the oxidation of carbonyl compounds (see sections 2.5.3 and 2.5.4). Removal processes by reaction with OH or by formation of PANs are also illustrated.
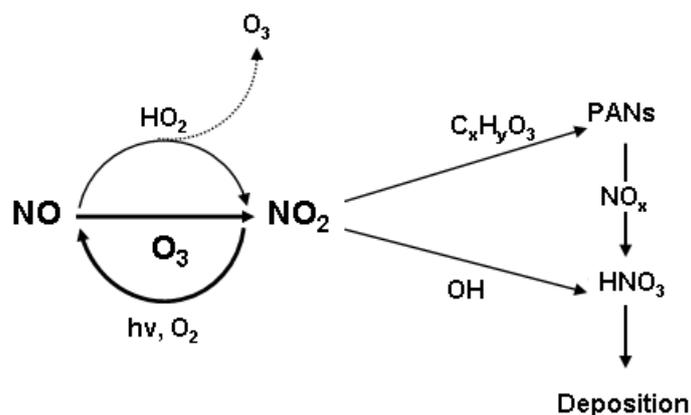
Figure 2-14: Photochemical cycle of NO₂, NO and O₃. Heavy arrows interconnecting NO and NO₂ indicate that this cycle occurs more frequently than either the NO+HO₂ or NO₂+OH reactions.

2.5.3 CO oxidation mechanism

The average mixing ratio of CO in the troposphere is around 120 ppb, but its distribution presents significant spatial variability because of its two months lifetime. CO is the simplest carbon containing compound in the atmosphere but its oxidation mechanism exhibits many of the key features of more complex VOC oxidation. The oxidation of CO by OH produces the H atom (R19), which reacts rapidly with O_2 to produce the hydroperoxyl radical HO_2 (R20).

$$CO + OH \rightarrow CO_2 + H \qquad\qquad\qquad\qquad\qquad\qquad\text{R 19}$$

$$\underline{H + O_2 + M \rightarrow HO_2 + M} \qquad\qquad\qquad\qquad\qquad\text{R 20}$$

$$CO + OH + O_2 \rightarrow CO_2 + HO_2 \qquad\qquad\qquad\qquad\text{R 21}$$

In low NO_x conditions, the resulting HO_2 radical can self-react to produce hydrogen peroxide H_2O_2 (R22). Hydrogen peroxide is highly soluble in water and is removed from the atmosphere by deposition on a time scale of a week. It can also photolyse or react with OH (R23 and R24).

$$HO_2 + HO_2 \rightarrow H_2O_2 + O_2 \qquad\qquad\qquad\qquad\qquad\text{R 22}$$

$$H_2O_2 + hv \rightarrow 2OH \qquad\qquad\qquad\qquad\qquad\qquad\text{R 23}$$

$$H_2O_2 + OH \rightarrow HO_2 + H_2O \qquad\qquad\qquad\qquad\quad\text{R 24}$$

In high NO_x conditions, the most important atmospheric reaction undergone by the HO_2 radical is:

$$NO + HO_2 \rightarrow OH + NO_2 \qquad\qquad\qquad\qquad\qquad\text{R 25}$$

Reaction R25 regenerates OH and also produces NO_2 which goes on to photolyse (R12), allowing the production of O_3 and the regeneration of NO in a catalytic cycle (R13). The resulting sequence of reactions summarized below (R21-R25-R13) is a chain mechanism for O_3 production in which the oxidation of CO is catalysed by the HO_x chemical family and by NO_x. The chain is initiated by the source of HO_x from reaction R10 and is terminated by the loss of the HO_x radicals through R22 or R14. This last reaction removes both HO_x and NO_x from the cycle. The propagation efficiency of the chain (the chain length) is determined by the abundance of NO_x. In this theoretical high-NO_x scenario, a maximum of one molecule of O_3 is produced per oxidized molecule of CO (R26).

$$CO + OH + O_2 \rightarrow CO_2 + HO_2 \qquad \text{R 21}$$

$$NO + HO_2 \rightarrow OH + NO_2 \qquad \text{R 25}$$

$$\underline{NO_2 + hv + O_2 \rightarrow NO + O_3} \qquad \text{R 13}$$

$$CO + 2O_2 \rightarrow CO_2 + O_3 \qquad \text{R 26}$$
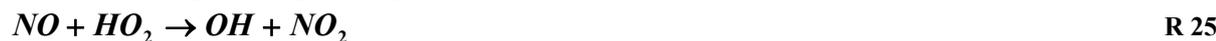
In remote regions of the troposphere, ozone loss by the HO_x catalysed mechanism (R8-R9) can be important, when NO concentrations are sufficiently low for R9 ($HO_2 + O_3 \rightarrow OH + 2O_2$) to compete with R25 ($NO + HO_2 \rightarrow OH + NO_2$).

2.5.4 CH$_4$ oxidation mechanism

Because of its long lifetime (about nine years), CH_4 is the most important carbon-containing compound after CO_2. Despite the relatively slowness of its oxidation by OH, CH_4 exerts a dominant effect on the chemistry of the background troposphere because of its high concentration (about 1770 ppb). The mechanism for oxidation of CH_4 involves many more steps than the oxidation of CO but follows the same schematic.
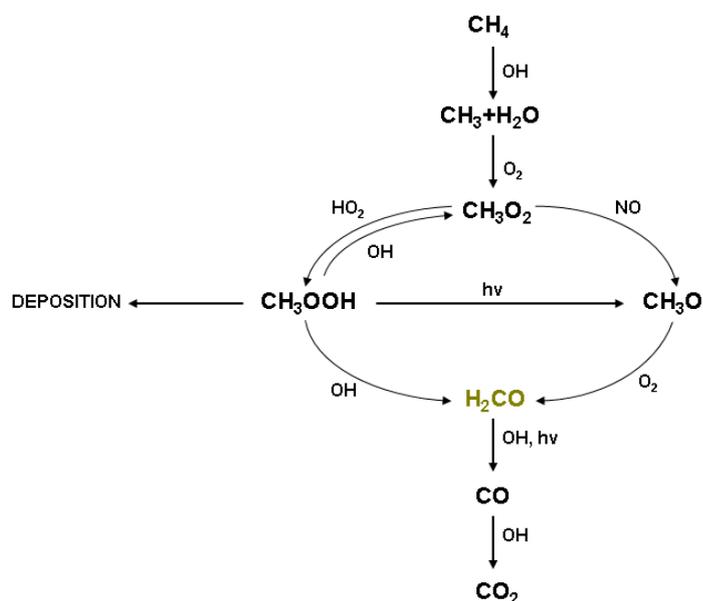
Figure 2-15: Atmospheric methane oxidation chain [*Seinfeld and Pandis*, 2006]

The oxidation of methane is initiated by OH, yielding the methyl radical CH_3 (R27), which reacts almost instantaneously with O_2 (R28). The produced methyl peroxy radical CH_3O_2 is analogous to HO_2 and is considered part of the HO_x family. Its main sinks are reaction with HO_2 and NO (R30 and R31). Such as H_2O_2, CH_3OOH may either photolyse or react with OH (R32, R33 or R34). The reaction with OH has two branches because the H abstraction can take place either at the methyl or at the hydroperoxy group. The first branch leads to formaldehyde (H_2CO) and OH production (R32). The second branch leads to the peroxy radical CH_3O_2 which will reacts either with HO_2 or NO (R30 or R31). The methoxy radical CH_3O either produced by reactions R31 or R34, also leads to formaldehyde production (R35). Figure 2-15 schematizes the different branches of the CH_4 oxidation. Ultimately, every CH_4 molecule is converted to formaldehyde (H_2CO).

$$CH_4 + OH \rightarrow CH_3 + H_2O \qquad \text{R 27}$$

$$CH_3 + O_2 + M \rightarrow CH_3O_2 + M \qquad \text{R 28}$$

$$\overline{CH_4 + OH + O_2 \rightarrow CH_3O_2 + H_2O} \qquad \text{R 29}$$

$$CH_3O_2 + HO_2 \rightarrow CH_3OOH + O_2 \qquad \text{R 30}$$

$$CH_3O_2 + NO \rightarrow CH_3O + NO_2 \qquad \text{R 31}$$

$$CH_3OOH + OH \rightarrow H_2CO + OH + H_2O \qquad \text{R 32}$$

$$CH_3OOH + OH \rightarrow CH_3O_2 + H_2O \qquad \text{R 33}$$

$$CH_3OOH + h\nu \rightarrow CH_3O + OH \qquad \text{R 34}$$

$$CH_3O + O_2 \rightarrow H_2CO + HO_2 \qquad \text{R 35}$$

Formaldehyde can either react with OH (R36) or photolyse (two photolysis branches, R37 and R38). Reactions R36 and R37 are very important for tropospheric chemistry because they produce the HCO radical, which reacts rapidly with O_2 to yield CO and HO_2 (R38). The lifetime of formaldehyde against the two photolysis reactions is respectively 9 and 7 hours, while it is 31 hours against OH reaction. The overall H_2CO lifetime is therefore relatively short (3.6 hours). Both H_2CO photolysis and OH reaction lead to the formation of CO. Therefore, the CO oxidation chemistry becomes part of the CH_4 oxidation chain. Eventually, CO is converted into CO_2 on a several-month timescale (as illustrated in Figure 2-15).

$$H_2CO + OH \rightarrow HCO + H_2O \qquad \text{R 36}$$

$$H_2CO + h\nu + O_2 \rightarrow HCO + HO_2 \qquad \text{(\textit{λ<330nm})} \qquad \text{R 37}$$

$$H_2CO + h\nu \rightarrow CO + H_2 \qquad \text{(\textit{λ<370nm})} \qquad \text{R 38}$$

$$HCO + O_2 \rightarrow CO + HO_2 \qquad \text{R 39}$$
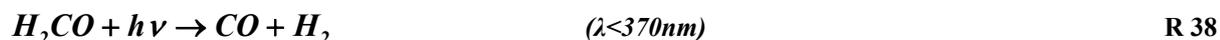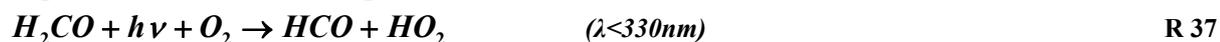
Formaldehyde is a first-generation oxidation product of CH_4 and of many other hydrocarbons. Indeed, reaction of OH with organic compounds (RH) leads to alkyl peroxy radicals (RO_2) and, although the complexity stepping in the oxidation chain arising from the multiple fates of the RO_2 radicals as well as of the organic compounds, the basic elements of ozone production are similar to those of CO and CH_4, with H_2CO as a central chain product.

As an illustration, the O_3 and HO_x yields from CH_4 oxidation can be calculated in two extreme:

- *High-NO_x regime*: Consider first a situation where HO_2 and CH_3O_2 react only by R25 and R31 respectively and H_2CO is removed solely by R37. By summing all reactions in the mechanism we come to the following net reaction for conversion of CH_4 to CO_2:

$$CH_4 + 10O_2 \rightarrow CO_2 + H_2O + 5O_3 + 2OH \qquad \text{R 40}$$
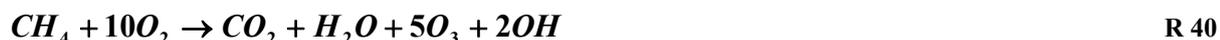
With an overall yield of five O_3 molecules and two HO_x molecules per molecule of CH_4 oxidised. Such a maximum yield is not realised in the actual atmosphere because of all the competing reactions that have been neglected. Similarly to CO, the oxidation of CH_4 in this high-NO_x case is a chain mechanism for O_3 production where HO_x and NO_x serve as catalysts. Formaldehyde photolysis, which provides the extra source of HO_x as part of the propagation sequence, branches the chain.

- *Low-NO$_x$ regime*: In contrast, consider an atmosphere devoid of NO$_x$ so that CH$_3$O$_2$ reacts by R30. Further assume that CH$_3$OOH reacts by R32 and H$_2$CO reacts by R36. Summing all reactions in the mechanism yields the net reaction:

$$CH_4 + 3OH + 2O_2 \rightarrow CO_2 + 3H_2O + HO_2 \qquad \textbf{R 41}$$

So that no O$_3$ is produced and two HO$_x$ molecules are consumed. This result emphasises the critical role of NO$_x$ for maintaining O$_3$ and OH concentrations in the troposphere.

2.5.5 Global budget of tropospheric ozone

The tropospheric budget of ozone can be calculated by global chemistry-transport models that integrate HO$_x$-NO$_x$-VOCs chemical mechanisms in a three-dimensional framework. They can be used to estimate the importance of the different sources and sinks in the tropospheric O$_3$ budget. Table 2-2 gives the present tropospheric ozone budget, as derived from a study based on 25 tropospheric chemical transport models [*Stevenson et al.*, 2006 and *Denman et al.*, *IPCC*, 2007]. It is fairly well established that the abundance of tropospheric O$_3$ is largely controlled by chemical production and loss within the troposphere, while transport from the stratosphere and dry deposition are relatively minor terms. The reactions HO$_2$+NO (R25) and CH$_3$O$_2$+NO (R31) account respectively for 70% and 20% of the global chemical production of tropospheric ozone. The global tropospheric ozone burden is estimated at 340±40 TgO$_3$, which corresponds to a mean lifetime of 22±2 days, using equation (2-3) and the estimated total sink in Table 2-2.

Table 2-2: Global budget for tropospheric ozone, based on simulations performed with 25 chemical transport model [*Stevenson et al.*, 2006 and *Denman et al.*, *IPCC*, 2007].

Global tropospheric ozone budget [$TgO_3\,yr^{-1}$]	
Sources	
In situ chemical production	5060±570
Transport from stratosphere	520±200
Total	**5580**
Sinks	
In situ chemical loss	4560±720
Dry deposition	1010±220
Total	**5570**

Figure 2-16 illustrates typical distribution of surface ozone mixing ratio in January and in July for pre-industrial time conditions and present time conditions, simulated with the global chemical transport model LMDZ-INCA [*Hauglustaine et al.*, 2004]. Higher O$_3$ concentrations are found in the Northern than in the Southern hemisphere, reflecting the highest concentrations of NO$_x$ and CO at Northern midlatitudes, due to the large source from fossil fuel combustion. O$_3$ concentrations are also higher in summer time, when photochemical production is most active. Model results suggest that anthropogenic emissions have increased NO$_x$ and CO concentrations in most of the troposphere by factors 2-8 (NO$_x$) and 3-4 (CO) from pre-industrial times to today. Ozone concentrations have increased by 50-100% in most of the troposphere, the largest increases being at low altitudes in the Northern hemisphere. The pre-industrial distribution shows values around 10-30 ppb during all year, while at present, summer ozone maxima up to 60 ppb are found in polluted regions of Northern hemisphere.
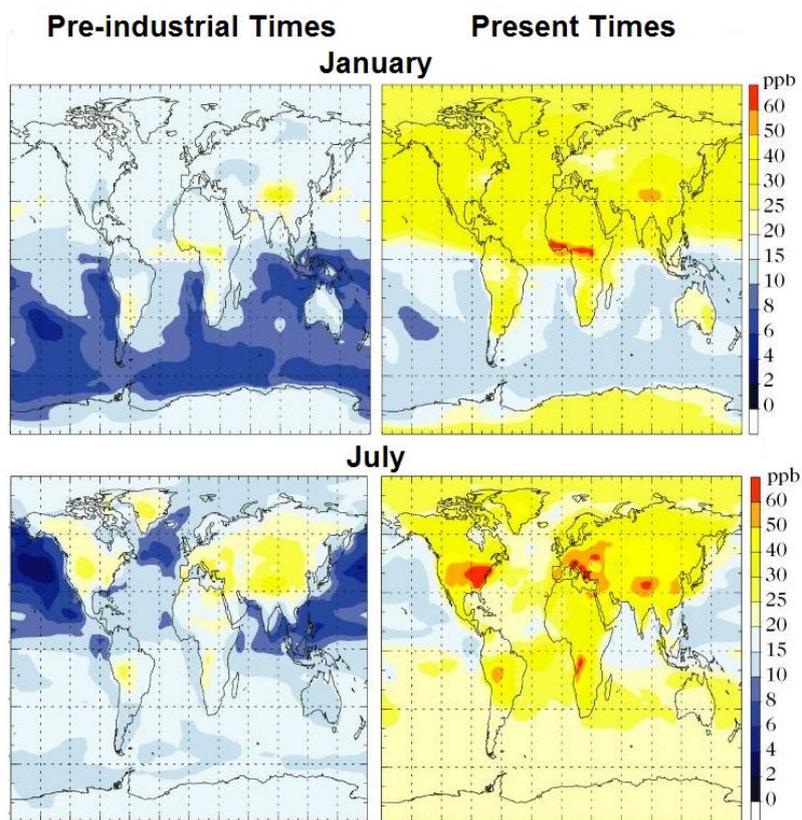
Figure 2-16: Distribution of surface ozone mixing ratio [ppb] simulated with the 3D-CTM model LMDz-INCA in January (upper panels) and in July (lower panels) for pre-industrial time conditions (left) and present time conditions (right) [*Hauglustaine et al.*, 2004].

2.5.6 Air pollution: relative roles of NO$_x$ and NMVOCs in ozone formation

The chemistry of the background troposphere is dominated by that of CO and CH$_4$. However, in continental regions where both biogenic and human emissions significantly influence the composition of the atmosphere, NMVOCs are critical for rapid production of O$_3$ and play also an important role in the long-range transport of NO$_x$, as mentioned with PANs. Fuel combustion and other activities of our industrial society release into the atmosphere a large number of pollutants which can, by direct exposure, be harmful to human or vegetation. This air pollution is called "smog" because it is associated with reduced visibility, due to the diffusion of light by high concentrations of anthropogenic aerosols. The health hazards of smog are caused in part by the aerosol particles but also by gases either emitted as primary pollutants (NO$_x$, SO$_2$, CO, VOCs) or produced as secondary pollutant (O$_3$). In surface air, O$_3$ is toxic to humans and vegetation because it oxidizes biological tissue. In most industrialised countries, air quality standards have been imposed to protect the population against exposure to air pollutants. National air pollution index includes measurements of O$_3$, other pollutants (SO$_2$, NO$_2$, CO) and particulate matter (PM$_{10}$). When the standards are exceeded, emission controls must be enacted. Concentrations of O$_3$ in clean surface air are in the range of 10-30 ppb. In Europe, the National Emission Ceilings directive (NEC directive) for reaching air quality standards is that 8-hour average O$_3$ concentration cannot exceed 120 μg·m^{-3} (about 60 ppb) more than 25 days per year. Authorities must alert the population when O$_3$ concentration exceeds 180 μg·m^{-3} over one hour. In the US, the Environmental Protection Agency (EPA) recommends a maximum 8-hour average O$_3$ concentration of 75 ppb not to be

exceeded more than three days per year. In China, the first grade of air quality is limited to a maximum 1-hour average O_3 concentration of 160 $\mu g \cdot m^{-3}$ and the second grade is set to 200 $\mu g \cdot m^{-3}$ (~100 ppb).

O_3 is known to be the most difficult pollutant to bring into compliance with air quality standards because the response of photochemical ozone production to changes in NO_x or in NMVOC emissions is not linear. It is important to understand the factors controlling ozone concentrations to set up efficient emission control. Figure 2-17 shows the results of chemical model calculations where simulated O_3 concentrations over the Eastern United States are displayed as a function of NO_x and hydrocarbons emissions. The thick line on the figure separates the NO_x- and hydrocarbon-limited regimes. To the left of the line is the NO_x-limited regime: O_3 concentrations increase or decrease with increasing or decreasing NO_x and are insensitive to hydrocarbons. To the right of the line is the hydrocarbon-limited regime (also called NO_x-saturated regime): O_3 concentrations increase with increasing hydrocarbons while O_3 decrease with increasing NO_x. The nonlinear dependence of O_3 on precursor emissions is apparent from these simulations. In the NO_x-limited regime, hydrocarbon emission controls are of no benefit for decreasing O_3. In the hydrocarbon-limited regime, NO_x emission controls can even cause an increase in O_3. Formulation of a successful strategy against O_3 pollution evidently requires knowledge of the chemical regime for O_3 production, and also accurate values of NO_x and hydrocarbon emissions. Underestimating emissions of hydrocarbons (from human or natural origin) is misleading because an upward of hydrocarbon emissions can shift the chemical regime from hydrocarbon to NO_x limited. For example, emissions of biogenic isoprene (C_5H_8) in the Eastern United-States are larger than the sum of all anthropogenic hydrocarbon emissions and it is therefore important to know with a good spatial and temporal precision the amount of isoprene emitted by the vegetation.
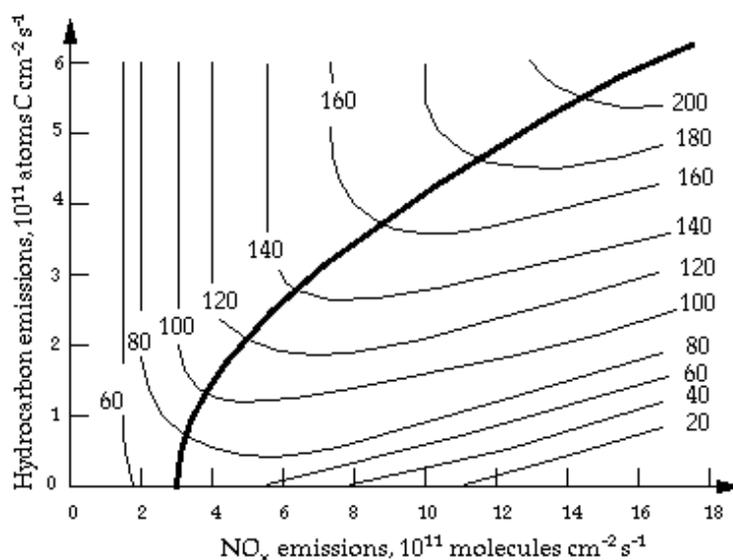
Figure 2-17: Ozone concentrations (ppb) simulated by the GEOS-CHEM photochemical model as a function of NO_x and hydrocarbon emissions. The thick line separates the NO_x-limited (top left) and hydrocarbon-limited (bottom right) regimes [*Jacob et al.*, 1999].

High O_3 concentration episodes occur solely in spring or in summer because its photochemical formation needs sufficient solar radiation. In summer, the lifetime of ozone in the boundary layer is a few days. The time needed to form a significant ozone amount in an agglomeration is limited by the hydrocarbon degradation time, which goes from one to several hours for the most reactive ones.

Therefore, maximum concentrations of ozone are expected outside the agglomerations because of wind effects. As illustrated in Figure 2-18, up-wind an agglomeration, ozone and NO_x concentrations are at their background level. When an air mass crosses the agglomeration, nitrogen oxide emissions transform a part of the ozone into NO_2. Down-wind the agglomeration, dispersion decreases NO_x concentrations and, in presence of strong solar radiation, radical cycle initiated by the hydrocarbon oxidation leads to efficient ozone production. Events of strong ozone concentrations often occur over several days period associated with high temperatures, strong solar radiation and low winds. When temperature is elevated, biogenic emissions of NMVOCs also increase and further contribute to ozone formation.
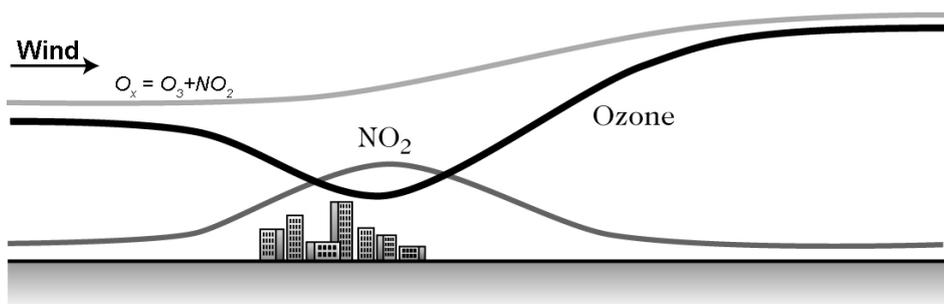
Figure 2-18: Schematic variation of concentrations in an air mass crossing an agglomeration following wind direction [*Delmas et al.*, 2005].

The perturbation of the ozone amounts by man-made and natural emissions of NO_x and VOCs can be illustrated with two examples. Figure 2-19 (left panel) shows ozone concentration over Western Europe during a large scale ozone event associated with the heat wave of Augustus 2003, simulated with the regional chemical transport model CHIMERE [*Bessagnet et al.*, 2004]. The event was characterised by exceeding of alert levels in France and other Western European countries between the first and the 15th Augustus. During the first days, only the vicinity of large agglomerations were polluted but the ozone plumes grew up during several days, agglomerated to finally form a very large polluted area covering almost entirely France, Belgium, Netherlands and Germany. This ozone event was exceptional by its geographical extend, its duration and the O_3 concentration levels that were attained (up to 400 $\mu g \cdot m^{-3}$ near Marseille). For comparison, regular ozone levels calculated with the same model on 30th Augustus 2010 are shown on the right panel of Figure 2-19.
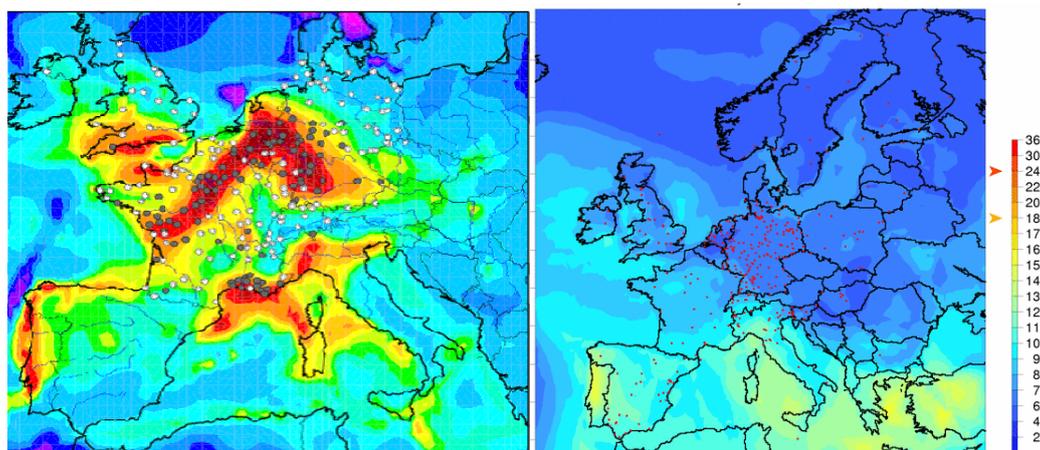
Figure 2-19: Ozone surface concentrations in Europe calculated with the regional CTM CHIMERE [in $\mu g \cdot m^{-3}$]. Left panel: ozone levels on the 8th Aug. 2003. White dots indicate operating ground-based observation stations while grey dots indicate the stations where the threshold of 180 $\mu g \cdot m^{-3}$ was exceeded. Right panel: ozone concentrations simulated on 30th Aug. [© PREV'AIR, http://www.prevair.org]

In China, the second grade of Chinese air quality standards is regularly exceeded in a large area Eastern and Southern of Beijing. This is illustrated in Figure 2-20 that presents 8-hour average O_3 concentrations between Aug.30 and Sept.2 2010 simulated with the CHIMERE model in the framework of the AMFIC project [www.amfic.eu]. The effect of meteorological conditions is clearly visible on these simulations. Furthermore, it illustrates the fact that O_3 is a regional rather than urban pollution problem. The distribution of O_3 concentrations in continental surface air reflects roughly the distribution of population. However, within a given region, there is little difference between cities, suburbs and nearby rural areas. The broad spatial extent of the O_3 pollution problem has important implications not only for population exposure but also for effects on crops and forests. There is accumulating evidences that sensitive crops are damaged by O_3 concentrations as low as 40 ppb, well below the current air quality standards.
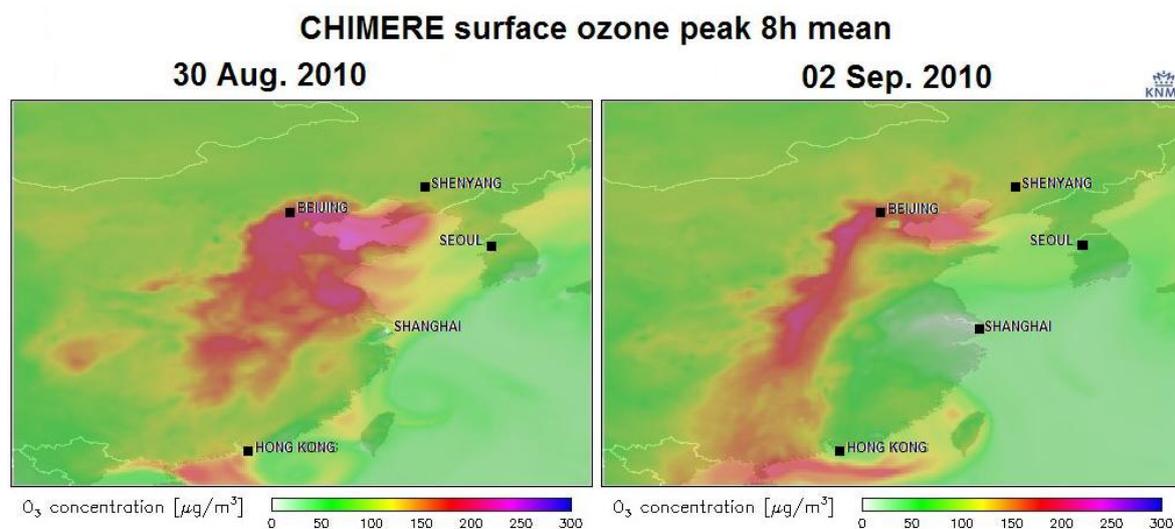
Figure 2-20: 8-hour average O_3 concentration in eastern China for the 30th Aug. and 02sd Sept. 2010 simulated by the regional CTM CHIMERE for the AMFIC project [http://www.amfic.eu/bulletin/index.php].

Quantitative understanding of tropospheric ozone production is largely dictated by the extent of knowledge of the amounts and distribution of tropospheric NO_x and VOCs. As will be explained in the next chapter, on the global and regional scale, large uncertainties exist in the biogenic and anthropogenic NMVOC emission inventories because of the diversity and the spatial and temporal variability of NMVOCs. Satellite remote sensing presents the capability to improve the knowledge on NO_x and NMVOCs emissions through the observation of nitrogen dioxide and formaldehyde, with implications for air quality control strategy. H_2CO being a central product in the oxidation chain of many VOCs and its lifetime being only a few hours, its satellite observations can be directly related to NMVOC emissions and help to constrain emission inventories. Furthermore, it has been demonstrated that the ratio of H_2CO to tropospheric NO_2 columns is an indicator of the relative sensitivity of ozone production to emissions of NO_x and VOCs [*Martin et al.,* 2004a, *Duncan et al.*, 2010]. Indeed, the H_2CO/NO_2 ratio is an efficient indicator to diagnose NO_x sensitive versus NO_x saturated conditions for several reasons: (1) the bulks of tropospheric NO_2 and H_2CO columns over polluted regions are within the boundary layer, (2) H_2CO is positively correlated with the source of HO_x in polluted environments while tropospheric NO_2 is closely related to NO_x emissions, and (3) the H_2CO/NO_2 column ratio largely reflects the ratio of HO_x to NO_x sources near the surface.

3   Volatile Organic Compounds

This chapter gives an overview of the global budget and the emission sources of the most important carbon-containing compounds for atmospheric chemistry, namely the greenhouse gases CO_2 and CH_4 and the reactive gases CO and NMVOCs. Because they are long-lived gases, CO_2 and CH_4 are well mixed in the atmosphere and their concentrations are known with a good accuracy. However, uncertainties remain in the spatial distribution of the sources and in their type repartition, mainly for CH_4. Because of their shorter lifetimes, CO and NMVOC concentrations are much more variable and nonlinearly related to their sources, which generally include direct surface emissions and *in situ* chemical production. As a consequence, emission inventories of CO and NMVOCs present very large uncertainties. After being emitted in the atmosphere, CH_4, CO and NMVOCs are photochemically oxidized through a succession of steps leading ultimately to CO_2. Formaldehyde is a high-yield intermediate product in these oxidation chains. Direct H_2CO emissions are negligible compared to in-situ production. Furthermore, the H_2CO lifetime is short enough that a measure of its abundance can be related to the emission strength of its parent VOCs.

This chapter starts with a simplified inventory of the different surface sources and the general principles of global budgets and emission inventories of atmospheric compounds (section 3.1). The current knowledge on the budgets and emissions of CO_2, CH_4, CO and NMVOCs will be presented in section 3.2. Finally in section 3.3, the global budget of formaldehyde estimated with a chemistry-transport model will be presented. We discuss the potential of H_2CO satellite observations to provide quantitative information on NMVOC emissions by inverse modelling. Note that satellite observations in the troposphere will be presented in the next chapter. Here we focus on the global budgets and emissions of volatile organic compounds.

3.1 Global budgets and emission inventories

The global budget of an atmospheric compound consists of three quantities: the global source and sink and the atmospheric burden, defined as the total mass of the compound integrated over the whole atmosphere. The global burden and its growth rate (*i.e.* the difference between sources and sinks) can be determined from atmospheric measurements. For long-lived gases, such as CO_2 and CH_4, they are usually the best-known quantities in the budgets. Global source and sink estimates can be deduced from mass balance calculations (Equations 2-2 and 2-3). In contrast, the atmospheric burdens of reactive gases such as CO and NMVOCs cannot be measured with great accuracy.

The global source strength is the sum of all sources, including emissions and *in situ* chemical production. Likewise, the sink strength (or global loss rate) can have several independent components. The main natural sources are biogenic emissions (gases produced during biological processes) in oceans, soils and terrestrial vegetation, volcanic emissions of gases and aerosols and wind erosion which lead to aerosol emissions over continents and oceans. The number of anthropogenic sources is larger. Regarding the intensity and the variety of the emitted pollutants, the main anthropogenic sources are combustion processes. They include fossil fuel combustion (coal, gas, oil), essentially for energy production and its diverse uses (thermal regulation, transports, industries …), and biomass burning. These sources are comparable in terms of emitted compounds (carbon, nitrogen and sulphur compounds with various oxidation levels) because in both cases, the organic matter that is burned (oxidized) is of similar chemical composition. They are also

comparable in terms of the quantity of carbon burned annually, even if fossil fuel combustion dominates biomass burning. The principal natural and anthropogenic sources are summarized in Table 3-1. It should be mentioned that the distinction between natural and anthropogenic sources is sometimes difficult to make, as some sources classified as anthropogenic, like biomass burning or agricultural activities, are actually anthropogenic forcing of natural processes (started fires, soil fertilization, livestock farming). From Table 3-1, one can conclude that carbon-containing compounds are emitted by all types of sources. Since a few decades however, the anthropogenic sources of several species (*e.g.* CH_4 and CO) tend to become preponderant.

Table 3-1: Principal natural and anthropogenic sources of atmospheric compounds [*Delmas et al.*, 2005].

Type of source	Emitted compounds
Natural sources	
Ocean (wind erosion and biogenic emissions)	CO_2, DMS[1], N_2O, primary and secondary marine aerosols[2]
Continental wind erosion	Primary inorganic aerosols
Biogenic emissions in terrestrial ecosystems	CO_2, NO_x, N_2O, CH_4, NMVOCs, secondary organic aerorols
Volcans	CO_2, SO_x, primary and secondary aerosols
Anthropogenic sources	
Fossil fuel combustion (oil, coal, natural gas -energy, transport, heating, industry)	CO_2, CO, CH_4, NMVOCs, NO_x, N_2O, SO_x, O_3, primary and secondary aerosols
Biomass burning (heating, firewood, deforestation, savanna fires, agricultural wastes)	CO_2, CO, NMVOCs, NO_x, N_2O, SO_x, O_3, primary and secondary aerosols
Industries, mines, gas and petroleum exploitations	CO_2, CO, NMVOCs, NO_x, N_2O, SO_x, O_3, primary and secondary aerosols
Agriculture, farming, fertilization, rice growing	CH_4, NH_3, NO_x, N_2O
Landfills and burning of domestic and industrial wastes	CO_2, CO, CH_4, NMVOCs, NO_x, H_2S, SO_x

Emissions can be quantified at different scales. On the global scale, the budget of the main atmospheric compounds allows to explain the average chemical composition of the atmosphere and to understand its evolution by comparing anthropogenic and natural emissions. However, the global knowledge of sources and sinks is inadequate for any detailed study of atmospheric chemistry, like three-dimensional chemical-transport modelling, for which it is essential to know their spatial and temporal distribution. Emission inventories are therefore needed, at different spatial and temporal scales, depending on the application (from global chemistry-climate coupling model to regional and local air quality model). Trace gas emission inventories are compiled within the Global Emissions Inventory Activity (GEIA) [http://www.geiacenter.org]. This data portal provides gases and aerosols emission gridded data, aggregated in several categories (anthropogenic, biomass burning, biogenic and oceans) at global or regional scales. Widely used inventories are the Emission Database for Global Atmospheric Research (EDGAR) for anthropogenic emissions [*Olivier et al.*, 2005], the Global Fire Emissions Database (GFED) for biomass burning [*van der Werf et al.*, 2006] and the Model of Emissions of Gases and Aerosols from Nature (MEGAN) for biogenic emissions [*Guenther et al.*, 2006]. Regional anthropogenic emissions inventories are also developed for Europe (European Monitoring and Evaluation Program (EMEP) Centre on Emission Inventories and Projections (CEIP)), Asia (Regional Emission Inventory in Asia (REAS)) or the United-States (National Emissions Inventories (NEI)).

Emission inventories are generally based on bottom-up methodologies, in which ground-based and laboratory measurements are extrapolated to estimate regional and global emissions by applying the following general equation:

---

[1] DMS: dimethyl sulfide is an organosulfur compound with the formula $(CH_3)_2S$.
[2] Primary aerosols are directly emitted as particles, while secondary aerosols are formed in the atmosphere by gas-to-particle conversion processes.

$$E = A \cdot (EF) \cdot P_1 \cdot P_2 \ldots \hspace{4cm} \text{(3-1)}$$

where E is the emission (typically expressed in kg m^{-2} s^{-1}), A is the activity rate for a given source (or group of sources, for example, kg of coal burned in a power plant) and (EF) is the emission factor which is based on source-specific emission measurements (for example, kg carbon emitted per kg coal burned). P_1, P_2 ... are parameters that apply to the specified source types and species to modulate the calculated emission factors in function, for example, of the meteorological conditions or the spatiotemporal variability of the sources. The bottom-up approach suffers from several limitations. Emission factors are developed from only a limited sampling of the emission source population for any given category, and the reported values are an average of those limited samples. Furthermore, geographical and statistical data are used to extrapolate measurements of emission factors, typically available only on a sparse spatial and temporal network. Because of the high spatial and temporal variability of the emission fluxes, this leads to important errors in emission inventories. The uncertainties in emission estimates may be reduced by accumulating direct flux measurements near the surface or above the canopy (for biogenic emissions) and by improving the representation of the dependences to environmental variables. An alternative methodology, called the top-down or inverse modelling approach, consists in using chemistry-transport models in an inverse mode, *i.e.* in adjusting the surface emission fluxes to minimize the differences between the model predictions and a set of atmospheric observations. In this method, it is implicitly assumed that the relationship between surface fluxes and atmospheric abundances is reasonably well predicted by the model, so that the biases between the model and the data are mostly due to errors in the emission inventories [*Müller and Stavrakou*, 2005].

3.2 Sources, sinks and global budgets of carbon-containing compounds

As mentioned in chapter 2, on a limited temporal scale, concentrations of species in the atmosphere are controlled by emissions, transport, chemistry and deposition. However, from an Earth system perspective, the composition of the atmosphere is ultimately controlled by the exchange of elements between the different reservoirs of the Earth.
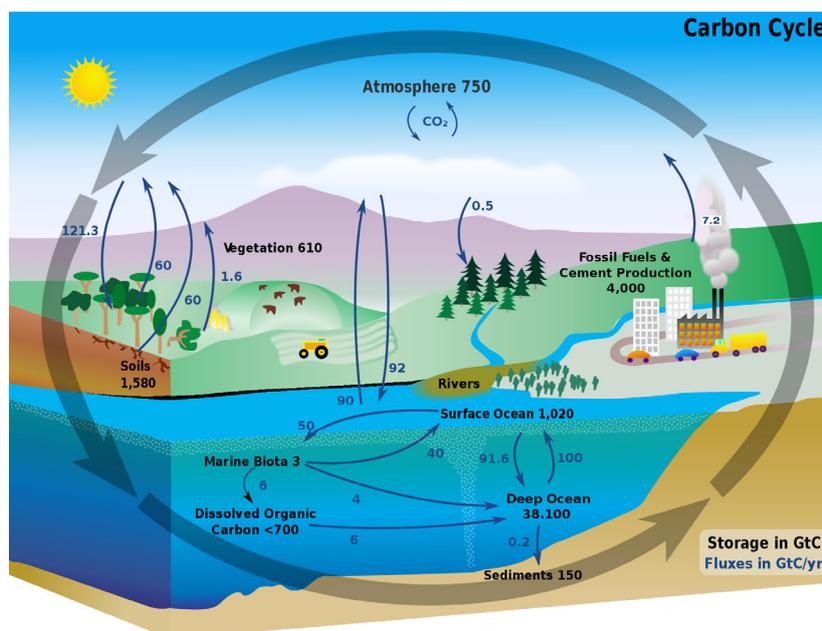
Figure 3-1: Diagram showing the storage and annual exchange of carbon between the atmosphere, hydrosphere and geosphere [© PD-USGov-NASA, based on *Denman et al.*, *IPCC*, 2007].

As illustrated in Figure 3-1, the stock of carbon in the main reservoirs is accounted in billion tons (Gt) of carbon and the fluxes in billion tons per year. The oceanic reservoir is by far the most important (~40000 GtC), followed by the continental biosphere (2200 Gt) and the atmosphere (750 GtC). These reservoirs naturally exchange important quantities of carbon, predominantly under the form of CO_2 fluxes, but also, to a lesser extent, through CH_4, CO and NMVOC fluxes.

3.2.1 Carbon dioxide

Over continents, carbon is converted to plant biomass by photosynthesis, which captures atmospheric CO_2. Plant, soil and animal respiration (including decomposition of dead biomass) returns carbon to the atmosphere as CO_2 (or as CH_4 under anaerobic conditions). CO_2 is also continuously exchanged between the atmosphere and the ocean. CO_2 entering the ocean surface immediately reacts with water and is transformed into dissolved carbon[3]. A fraction of the dissolved carbon is used by phytoplankton and transformed into oceanic biomass, a small fraction of which sinks into the deep ocean as biogenic particles. The loss through sedimentation is compensated by transport from rivers and net exchange from ocean to the atmosphere. Terrestrial and oceanic exchange fluxes with the atmosphere are estimated to respectively 120 and 90 GtC yr^{-1}. These processed are illustrated in Figure 3-1. Over the last 1000 years, atmospheric CO_2 level had been relatively stable around 280 ppm, but since industrialization started, CO_2 concentrations have strongly increased and by 2005, the global mean atmospheric CO_2 level has reached almost 380 ppm [*Forster et al.*, *IPCC*, 2007, see also Figure 2-11]. Continuous atmospheric measurements made since 1958 at Mauna Loa Observatory in Hawaii, shown in Figure 3-2, demonstrate the continuous increase of CO_2. Superimposed on the trend is a seasonal oscillation that reflects the seasonality of the natural carbon cycle, due the vegetation seasonal cycle. Each year during the Northern hemisphere growing season, atmospheric CO_2 mixing ratios decrease as carbon is incorporated into leafy plants. From October through January, photosynthesis is largely confined to the tropics and the relatively small land mass of the Southern hemisphere continents. At this time of year, plant respiration and decay dominate and CO_2 levels increase.
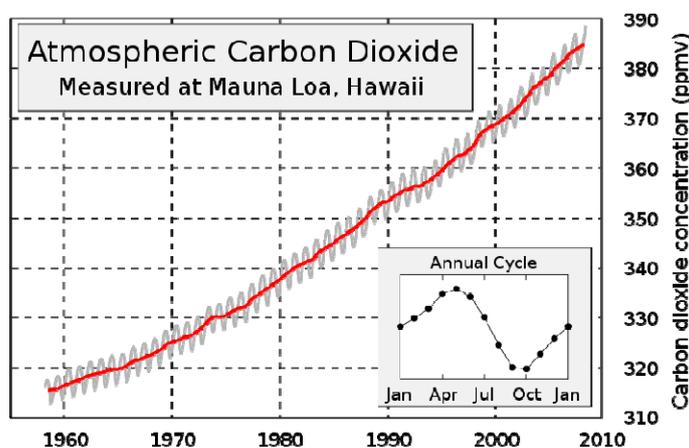
Figure 3-2: Monthly (in grey) and annual (in red) atmospheric CO_2 concentrations as directly measured at Mauna Loa, Hawaii. The annual cycle CO_2 is shown in the inset figure [© R. Rohde/global warming art].

In terms of the emitted total mass, carbon dioxide is the most important waste product of our industrialized society. Anthropogenic emissions consist in two fractions: CO_2 from fossil fuel burning and cement production, recently released from hundreds of millions of years of geological

---

[3] dissolved carbon: bicarbonate (HCO_3^-) and carbonate (CO_3^{2-}) ions.

storage, and CO_2 from deforestation and agricultural development, which has been stored for decades to centuries. About 80% of anthropogenic CO_2 emissions result from fossil fuel burning and about 20% from land use change. Fossil fuel and cement emissions rose up to 7.2 GtC yr^{-1} in 2005 [*Denman et al.*, *IPCC*, 2007]. These numbers are estimated based upon international energy statistics. There are large regional variations in emission trends. Since industrialization, the United States and Western Europe have been the main carbon emitters worldwide. From 1990, the base year of the Kyoto protocol, emissions of developing countries (mostly China) have become equivalently important and since 2005, China has become the first carbon emitter [*Olivier et al.*, 2008]. Although anthropogenic CO_2 fluxes between the atmosphere and both the land and ocean are just a few percent of the gross natural fluxes, they resulted in measurable changes in the carbon content of the reservoirs since pre-industrial times. It is estimated that about 40% of the combined anthropogenic CO_2 emissions actually accumulated in the atmosphere. The other part has been transferred to other geochemical reservoirs (oceans, biosphere and soils). As both fossil fuel emissions and atmospheric CO_2 levels are know with good precision, the total CO_2 sink is relatively well known. There are larger uncertainties in the repartition between the uptake of oceans and lands. The factors controlling these sinks as well as the coupling between climate change and the carbon cycle need to be better understood in order to predict future trends in atmospheric CO_2 [*Denman et al.*, *IPCC*, 2007].

Exchanges of carbon between the biosphere and the atmosphere are not limited to CO_2. Natural and anthropogenic sources of CH_4, CO and NMVOCs are numerous. Even if they represent only about 2.5 Gt of carbon emitted per year, these compounds participate to the carbon cycle after oxidation in the atmosphere, essentially by OH.

3.2.2 Methane

Methane is the most abundant hydrocarbon in the atmosphere. As displayed in Figure 2-11, its atmospheric concentration has nearly tripled since pre-industrial times, reaching 1774 ppb in 2005. Anthropogenic sources of CH_4 include rice agriculture, livestock, landfills and waste treatment, some biomass burning and fossil fuel combustion (natural gas, petroleum and coal). Natural CH_4 is emitted from biogenic sources such as wetlands, oceans, forests, termites and non biogenic sources such as fires and geological sources. Table 3-2 summarizes the global sources and sinks of CH_4, which were estimated to be around 580 Tg(CH_4) yr^{-1} (or 435 TgC yr^{-1}) for the period 2000-2005 [*Denman et al.*, *IPCC*, 2007]. Total pre-industrial emissions of CH_4 are estimated at 200 to 250 Tg(CH_4) yr^{-1}, of which at least 80% was emitted by natural sources [*Houweling et al.*, 2000]. In contrast, anthropogenic emissions dominate current CH_4 budgets, accounting for more than 60% of the global budget. The sink of methane through reaction with OH in the troposphere is estimated at 511 Tg(CH_4) yr^{-1}, and the loss to the stratosphere is estimated at 40 Tg(CH_4) yr^{-1}. Microbial uptake in soils contributes further to an estimated 30 Tg(CH_4) yr^{-1} removal rate [*Denman et al.*, *IPCC*, 2007]. As illustrated by Table 3-2, significant differences exist between the available CH_4 inventories, depending mainly of the considered period. Several sources of uncertainties need to be considered for the assessment of the global CH_4 budget. Uncertainties on atmospheric concentrations, measured to 1774±1.8 ppb in 2005, are small. The main sources of uncertainties come from the sink estimates and the lifetime used in the mass balance equation. The total global CH_4 source is relatively well known but the strength and trend of each contributing source are not. Figure 3-3 shows gridded methane emissions in 2005 as provided by the EDGARv4 emission inventory. It must be pointed out that large uncertainties exist on the location and intensity of the emission sources [*Forster et al.*, *IPCC*, 2007]. National inventories based on bottom-up studies can

grossly underestimate emissions and top-down assessments of reported emissions are required for verification [*Bergamaschi et al.*, 2005].

Table 3-2: Sources, sinks and atmospheric budget of CH$_4$ in Tg(CH$_4$) yr^{-1} [*Denman et al., IPCC*, 2007].

Reference	*Houweling et al.*, 2000	*Mikaloff Fletcher et al.*, 2004	*Chen and Prinn*, 2006	*Ehhalt et al., IPCC*, 2001	*Denman et al., IPCC*, 2007
Base years	Pre-industrial	1999	1996-2001	1998	2000-2005
Natural sources	**222**	**260**	**168**		
Wetlands	163	231	145		
Termite	20	29	23		
Others	39				
Anthropogenic Sources		**350**	**428**		
Coal mining		30	48		
Gas, oil, industry		52	36		
Landfills & waste		35			
Ruminants		91	189		
Rice agriculture		54	112		
Biomass burning		88	43		
Total sources	**610**	**596**	**598**		**582**
Sinks					
Soils		30		30	30
Tropospheric OH		507		506	511
Stratospheric loss		40		40	40
Total sink		**577**		**576**	**581**
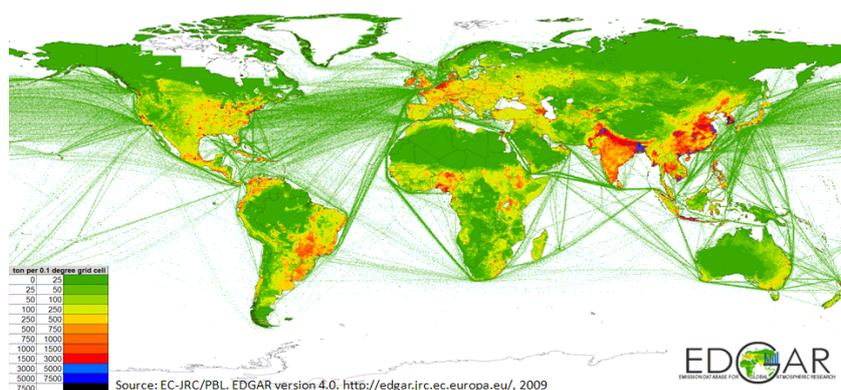
Figure 3-3: Global gridded methane emissions in 2005 [http://edgar.jrc.ec.europa.eu/part_CH4.php].

Figure 3-4 shows recent measurements of CH$_4$ concentrations derived from measurements of an international network of monitoring stations [*Forster et al., IPCC*, 2007; *Dlugokencky et al.*, 2003]. The atmospheric CH$_4$ growth rate has decreased markedly since the early 1990s and the level of methane has remained relatively constant during 1999 to 2005. However, large fluctuations are observed from year-to-year, although their causes remain uncertain [*Bousquet et al.*, 2006]. The decrease in atmospheric methane growth during the 1990s is thought to have been caused by a decline in anthropogenic emissions. However, since 1999, anthropogenic emissions of methane have risen again. It is thought that the reduction of the CH$_4$ growth rate is a consequence of source stabilisation, the effect of the increase in anthropogenic emissions being balanced by a coincident decrease in wetland emissions. Atmospheric methane levels may increase in the near future if wetland emissions return to their mean 1990s levels [*Bousquet et al.*, 2006; *Denman et al., IPCC*, 2007]. Since biogenic CH$_4$ production and emission from major sources are influenced by climate variables such as temperature and moisture, the effect of climate on emissions from these sources could be significant. Observations indicate substantial increases in CH$_4$ released from Northern peat lands that are experiencing permafrost belt [*Denman et al., IPCC*, 2007]. For that matter, from 2007, a renewed growth in CH$_4$ concentrations seems to be observed [*Rigby et al.*, 2008].
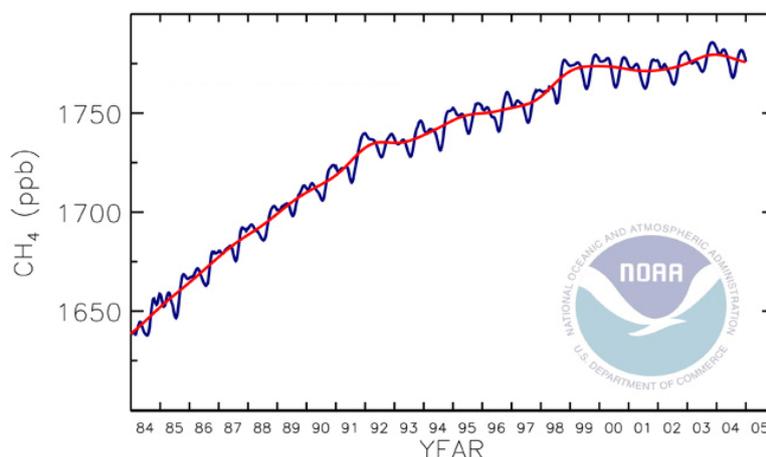
Figure 3-4: Global average atmospheric CH$_4$ mixing ratios (blue line) determined from surface sites operated by NOAA/GMD. The red line represents the de-seasonalised long-term trend [© Ed Dlugokencky, NOAA].

3.2.3 Carbon monoxide

As a key precursor of tropospheric ozone, carbon monoxide is an essential trace gas for the understanding of both air quality and climate forcing. CO is directly emitted at the Earth's surface, mostly from incomplete combustion processes. It is also produced chemically by the oxidation of hydrocarbons in the troposphere. CO is removed by OH oxidation and it is the largest global sink of the OH radical. Hence, it plays an important role in the CH$_4$ atmospheric concentrations [*Ehhalt et al., IPCC*, 2001; *George et al.*, 2009]. Because of its lifetime of the order of a month, CO provides an integrated view of regional emissions and transport. For this reason, CO has been extensively used to evaluate global chemistry transport models [*Shindell et al.*, 2006]. In the case of long-lived gases like CO$_2$ and CH$_4$, atmospheric concentrations are linearly dependent on emissions and global mass balance equation can be used to constrain emission inventories. It is not the case for atmospheric abundances of reactive gases like CO and NMVOCs, which are nonlinearly related to surface emissions [*Müller and Stavrakou*, 2005]. CO has large gradients in the atmosphere, and its global burden is more uncertain than those of CO$_2$ or CH$_4$ [*Ehhalt et al., IPCC*, 2001].
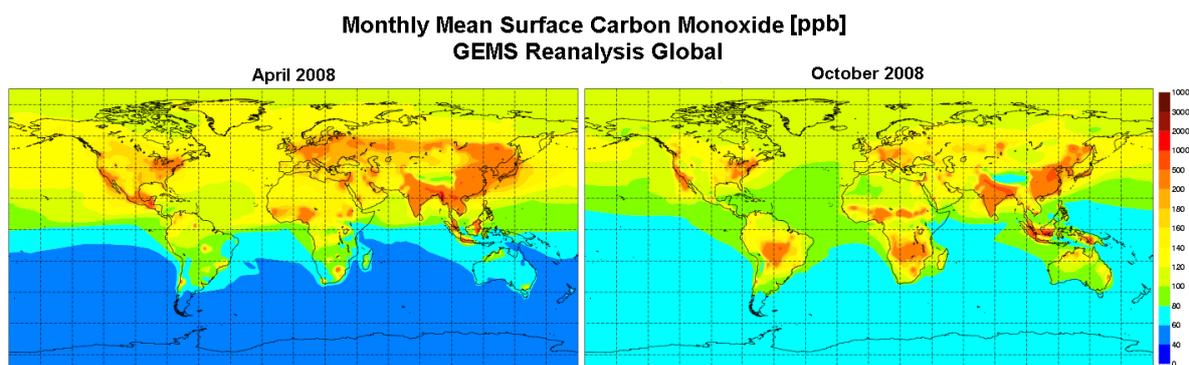
Figure 3-5: Surface CO mixing ratio in April and October 2008 simulated within the MACC/GEMS project [http://www.gmes-atmosphere.eu/d/services/gac/reanalysis/gems].

As illustrated in Figure 3-5, tropospheric CO mixing ratios range from 40 to 200 ppb. The abundance of CO in the Northern Hemisphere is about twice that in the Southern Hemisphere and has increased in the second half of the 20th century along with industrialisation and population. The maximum values are found near the surface at Northern midlatitudes. In general, the CO mixing ratio decreases with altitude in the Northern Hemisphere to a free tropospheric average value of

about 120 ppb near 45°N. In the Southern Hemisphere, CO tends to be more uniformly mixed vertically with a mixing ratio of about 60 ppb near 45°S, and the maximum concentration is observed to occur during the biomass burning season [*Seinfeld and Pandis*, 2006].

Estimates of the global sources and sinks of CO, taken from the third IPCC assessment report [*Ehhalt et al., IPCC*, 2001] and from more recent inverse modelling studies based on satellite data [*Stavrakou and Müller*, 2006; *Kopacz et al.*, 2010], are given in Table 3-3. Direct sources account for about 1550 Tg(CO) yr^{-1} (or 664 TgC yr^{-1}), while in-situ production accounts for about 1230 Tg(CO) yr^{-1}, for a total source of 2780±150 Tg(CO) yr^{-1}. Both direct and indirect sources of CO include contributions from natural and anthropogenic activities. Anthropogenic activities (deforestation, savanna and waste burning, fossil and domestic fuel use) accounts for more than 80% of the direct emissions of CO. About half of the *in-situ* source can also be attributed to anthropogenic emissions. A source of 800 Tg(CO) yr^{-1} is estimated from CH$_4$ oxidation, while the estimated source from other hydrocarbons is more uncertain (from 430 to 798 Tg(CO) yr^{-1}). The major sink for CO is reaction with OH radical, soil uptake and diffusion into the stratosphere being minor sinks. The understanding of the CO budget remains inadequate, as illustrated by a recent CTM comparison exercise showing large disagreements between models and observations [*Shindell et al.*, 2006]. Modelling of the spatial, seasonal, and interannual variability of CO involves a complex interplay of sources, transport, and chemistry. Errors in sources can exceed a factor of two on continental scales [*Kopacz et al.*, 2010]. Satellite-based studies have provided important new constraints for CO emissions, revealing in particular an underestimation in current inventories of the Asian anthropogenic sources, the Tropical biogenic source and the African vegetation fire source. Furthermore, indirect sources of CO calculated by chemistry-transport models depend to a large extent on the representation of NMVOC emissions and oxidation, which are still more uncertain. It is unclear how global CO emissions will evolve in the future and various scenarios predict either a decrease or a continued increase depending mostly on the development in Asia [*Shindell et al.*, 2006].

Table 3-3: Sources, sinks and atmospheric budget of CO in Tg(CO) yr^{-1}

Reference	POET [*Olivier et al., 2003*]	*Ehhalt et al., IPCC*, 2001	*Stavrakou and Müller*, 2006	*Kopacz et al.*, 2010
Base year	**2000**	**2000**	**2000-2001**	**2004-2005**
Oxidation of CH$_4$		800	755	
Oxidation of NMVOCs		430	798	
Sub-total *in situ* oxidation		**1230**	**1553**	**1500**
Vegetation	160	150	195	
Oceans	20	50	20	
Biomass burning	364	700	404	
Fossil & domestic fuel	577	650	688	
Sub-total direct emissions	**1121**	**1550**	**1308**	**1350**
Total sources		**2780**	**2860**	**2630**

3.2.4 Non-methane volatile organic compounds

Under the common name of NMVOCs are included a large number and variety of compounds emitted by natural sources, fires and anthropogenic activities. They influence the climate system through their production of secondary organic aerosols (SOA)[4], and their impact on ozone and on the oxidising capacity of the atmosphere [*Ehhalt et al., IPCC*, 2001; *Denman et al., IPCC*, 2007]. Unlike CH$_4$ and CO, NMVOC emissions are still largely dominated by natural emission from

---

[4] Secondary organic aerosols are organic aerosols formed by condensation or nucleation of low-volatility products of hydrocarbons photo-oxidation.

vegetation, representing 85% of the total emissions, half of it being isoprene emissions. Emissions of biogenic VOCs (BVOCs) other than isoprene are still poorly known and largely remain to be quantified. Anthropogenic sources of VOC include fuel production, distribution, and combustion, with the largest source being emissions from motor vehicles and from biomass burning. Thousands of different NMVOCs with varying lifetimes (hours to months) and chemical behaviour have been observed in the atmosphere. Given their variety and their variability, it is very difficult to derive a global atmospheric burden for most NMVOCs, and their speciation, *i.e.* the repartition between each NMVOC species, is still highly uncertain. Due to their importance in tropospheric chemistry, systematic measurements and analysis of their budgets are essential for understanding the chemistry-climate coupling [*Ehhalt et al., IPCC*, 2001].

Biogenic NMVOCs

Efforts are made to compile inventories of biogenic hydrocarbon emissions [*Guenther et al.*, 1995 and 2006]. Measurements in wooded and agricultural areas coupled with emission studies from selected individual trees and agricultural crops have demonstrated the ubiquitous nature of biogenic emissions and the variety of organic compounds that can be emitted. Biogenic VOCs can be grouped into four categories: isoprene, monoterpenes (for example α-pinene), reactive BVOCs with lifetimes shorter than 1 day (for example 2-methyl-3-buten-2-ol), and other BVOCs with lifetimes longer than a day (for example methanol) [*Guenther et al.*, 1995]. Isoprene shows the largest emission rate, followed by monoterpenes. They are the most important biogenic VOCs for tropospheric photochemistry. Furthermore, monoterpenes play an important role in secondary organic aerosol formation [*Kanakidou et al.*, 2005]. Vegetation releases other NMVOCs at relatively small rates, and the oceans also emit small amounts of NMVOCs. Isoprene (C_5H_8) is a species-dependent by-product of plant photosynthesis and photorespiration. It is emitted from a wide variety of mostly deciduous vegetation. It is not stored within or metabolized by plants and therefore, no isoprene is emitted without illumination. Isoprene emissions are temperature and light-dependent. By contrast, monoterpenes are stored in plant reservoirs, so they are emitted throughout the day and night, mostly by conifer trees, yet their emission varies with ambient temperature. The dependence of natural isoprene and terpenoid emissions on temperature results in a large variation in the rate of production of biogenic VOCs over the course of a growing season [*Seinfeld and Pandis*, 2006]. Biogenic hydrocarbon emission rates from individual plant species are estimated experimentally (mostly for isoprene and the dominant monoterpenes) and are used with empirical algorithms accounting for temperature and light effects to scale up to entire geographic regions, based on land-use data and biomass density factors. As an illustration, different estimates of global biogenic NMVOC emissions are provided in Table 3-4 and maps of MEGAN isoprene fluxes in January and July are shown in Figure 3-6. The total biogenic NMVOC emissions are estimated at 1150 TgC yr^{-1}. Isoprene and monoterpenes account respectively for 503 and 127 TgC yr^{-1}, while other biogenic VOCs represent an additional 520 TgC yr^{-1} [*Guenther et al.*, 1995]. For example, biogenic emissions of methanol are estimated to be 57 TgC yr^{-1} [*Jacob et al.*, 2005]. Values adopted in the third IPCC assessment report were significantly lower, 377 TgC yr^{-1} for the total biogenic NMVOC emissions, including 220 TgC yr^{-1} of isoprene and 127 TgC yr^{-1} of terpene (taken from GEIA) [*Ehhalt et al., IPCC*, 2001]. Large uncertainties remain in the estimates, at least 30% for isoprene [*Guenther et al.*, 1995 and 2006; *Müller et al.*, 2008] and much more for other biogenic NMVOCs, mainly because of the scarcity of direct measurements, in particular in the tropical ecosystems, which represent more than half of the global isoprene emissions, resulting from a combination of high temperatures and large biomass densities [*Guenther et al.*, 1995]. Smaller, but still important, amounts are emitted in the Northern mid-latitudes and boreal regions mainly in the warmer seasons. During summer months, the maximum flux of biogenic hydrocarbon emissions in

the Southeastern US is predicted to be as large as that in the Tropics. Since they are sensitive to temperature changes, BVOC emissions are likely to change in response to evolving chemical and climate environment over the century, but the projection of such changes is currently very uncertain [*Ehhalt et al., IPCC*, 2001, *Müller et al.*, 2008].

Table 3-4: Annual emission of isoprene and other biogenic hydrocarbons (in TgC yr^{-1}).

Reference	GEIA [*Guenther et al.*, 1995]	IPCC TAR [*Ehhalt et al.*, 2001]	MEGAN [*Guenther et al.*, 2006]	[*Jacob et al.*, 2005]	MEGAN-ECMWF [*Müller et al.*, 2008]
Isoprene	503	220	529		362
Monoterpenes	127	127	97		-
Methanol				57	-
Other VOCs	~520	~30 (acetone)	~237		-
Total	**1150**	**377**	**863**		
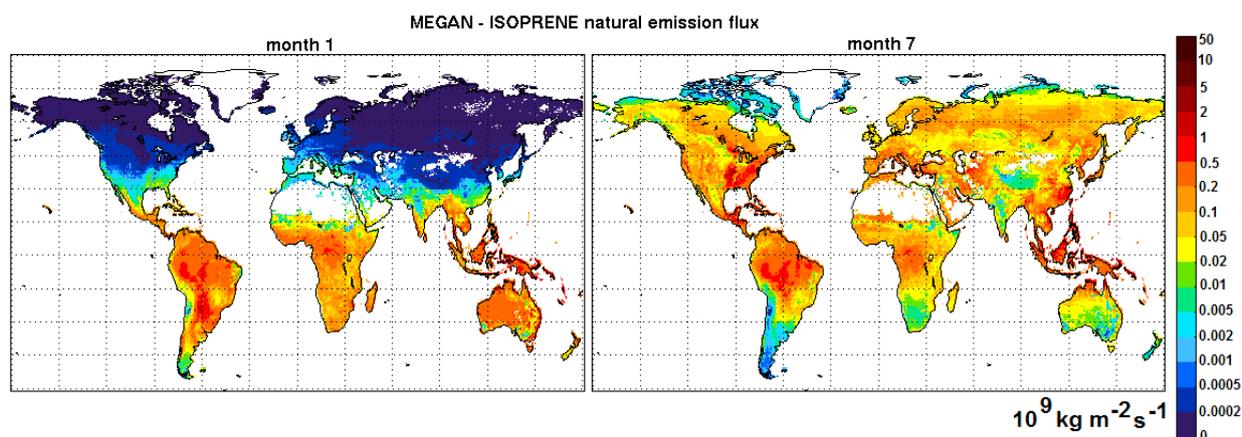
Figure 3-6: Isoprene emission fluxes in January and July estimated by the MEGAN model [*Guenther et al.*, 2006].

Anthropogenic NMVOCs

Anthropogenic NMVOC sources include emissions from fossil fuel combustion (oil, gas, and coal), industrial activities, waste disposal, biofuel use and agricultural waste burning. Biomass burning accounts for forest and savanna fires. In the EDGAR inventory, anthropogenic emissions are based on national activity data, emission factors, and grid maps for the spatial distribution of emissions within a country (e.g. population maps). Biomass burning emissions are estimated with a climatology, and satellite active fire maps are used for the spatial and temporal distribution of the emissions. Values adopted in the third IPCC assessment report are based on the EDGAR database. Anthropogenic sources are estimated to account for 161 TgC yr^{-1}, corresponding to about 210 Tg(NMVOC), depending on the adopted speciation [*Olivier et al.*, 2001; *Ehhalt et al., IPCC*, 2001]. The vegetation fire source of NMVOCs is of the order of 30 to 60 TgC yr^{-1}, but biomass burning emissions are highly variable from year to year [*Ehhalt et al., IPCC*, 2001; *Andreae and Merlet*, 2001; *Stavrakou et al.*, 2009a]. Globally, largest emissions are found to be road transport (about 20%), oil production/handling, residential biofuel use and solvent use (each about 15%), whereas savannah burning and temperate vegetation fires contribute another 5% each [*Olivier et al.*, 2003]. However, these sources are not distributed equally over world regions, let alone individual countries. Figure 3-7 shows NMVOC emission fluxes from anthropogenic and biomass burning sources estimated in the EDGAR inventory for the year 2000. In economically developed countries, fossil-fuel use and industrial processes tend to dominate, while in other regions, biomass burning is often the largest source. Anthropogenic emissions occur mostly where natural emissions are relatively low, so they have significant impacts on regional air quality and photochemistry despite smaller

global emissions. Some NMVOCs like benzene, toluene and other aromatic hydrocarbons are toxic and carcinogen pollutants. A few VOC, such as ethane and acetone, are longer-lived and impact tropospheric chemistry on hemispheric scales.
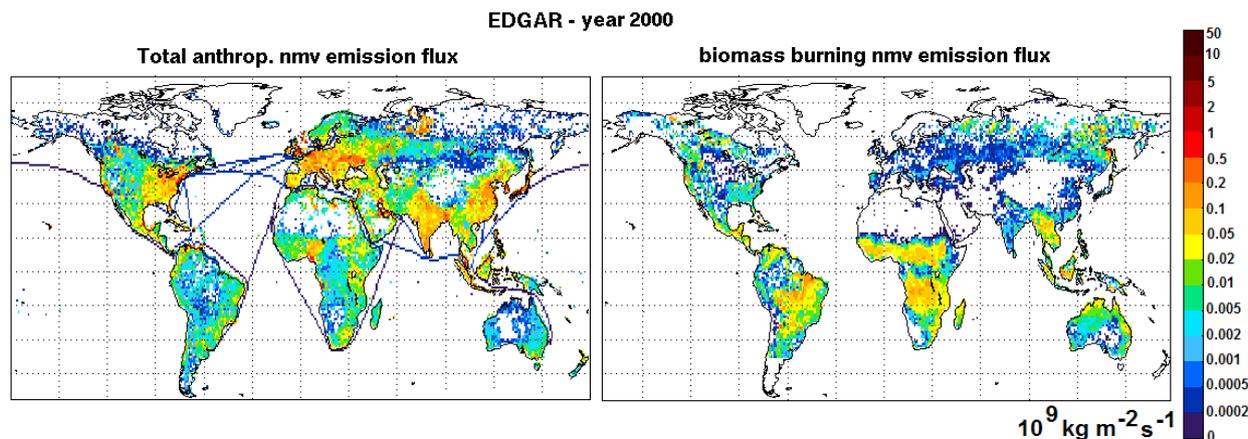
Figure 3-7: Anthropogenic and biomass burning NMVOC emission fluxes in 2000 estimated by the EDGAR inventory [*Olivier et al.*, 2001].

Models dealing with NMVOC chemistry require an explicit speciation of NMVOC emissions. A major shortcoming in any global set of NMVOC emissions is the fact that this detail is not available. Indeed, due to the variety of different compounds it is a very difficult task to compile a global inventory of NMVOC including all species and only emissions by aggregated classes are provided. Table 3-5 gives the speciation of NMVOCs provided in the POET emission inventory [*Olivier et al.*, 2003], and the corresponding emissions for the year 2000. Discrepancies between the different emission inventories and their speciation are very important, yet it appears that the most emitted NMVOCs by anthropogenic activities are butane and higher alkanes, and aromatic compounds (toluene, benzene, xylene). During vegetation fires, the strongest carbon emitters are ethene, alcohols (methanol) and aromatics.

Table 3-5: NMVOC speciation and estimated emissions in 2000 by the POET inventory [*Olivier et al.*, 2003].

NMVOC	Chemical Formula (Molecular Weight)	Anthropogenic emission Tg(species) yr^{-1} (TgC yr^{-1})	Biomass burning emission Tg(species) yr^{-1} (TgC yr^{-1})
Ethene	C_2H_4 (28)	04.24 (03.63)	4.49 (3.85)
Ethane	C_2H_6 (30)	05.70 (04.56)	2.19 (1.75)
Propene	C_3H_6 (42)	02.09 (01.79)	1.42 (1.22)
Propane	C_3H_8 (44)	06.51 (05.33)	0.41 (0.34)
Butene*	C_4H_8 (56)	05.88 (05.04)	1.05 (0.90)
Butane*	C_4H_{10} (58)	54.10 (44.77)	0.83 (0.69)
Formaldehyde	H_2CO (30)	01.24 (00.50)	2.54 (1.02)
Acetaldehyde*	CH_3CHO* (44)	01.84 (01.00)	2.71 (1.48)
Methanol	CH_3OH (32)	00.97 (00.36)	6.95 (2.61)
Ethanol*	C_2H_5OH* (46)	02.82 (01.47)	0.20 (0.10)
Acetone	$(CH_3)_2CO$ (58)	00.31 (00.19)	1.39 (0.86)
Mek*	C_4H_8O* (72)	01.13 (00.75)	3.20 (2.13)
Toluene*	$C_6H_5CH_3$* (92)	28.31 (25.85)	2.22 (2.03)

*   Butene and higher alkenes
*   Butane and higher alkanes
* CH_3CHO is a lumped species for all non-H_2CO aldehydes
* C_2H_5OH is a lumped species for all non-CH_3OH alcohols
*   Mek (Methyl-ethyl-ketone) is a lumped species for all non-acetone ketones
*   Toluene is a lumped species for all aromatic species

In conclusion, uncertainties on NMVOC emissions and their spatial and temporal variability remain important, though the quality of emission inventories is a determining factor for the quality of atmospheric model simulations. Local, regional and global anthropogenic emission inventories are calculated with emission factors and algorithms which are based on a limited number of observations. Another important uncertainty is the application of source category specific NMVOC emission profiles. In general NMVOC profiles are available for Northern America and Europe, while for other regions no information is available. A lot of hypotheses need to be made for the estimation of emissions in other countries. These uncertainties can have strong effects on global inventories, because of the rapid growing of anthropogenic emissions in certain parts of the world, like in Asia. The lack of temporal resolution in the dataset is also an important shortcoming. In order to model present atmospheric conditions, emission estimates should be updated for the most recent years. Anthropogenic NMVOC emissions could drastically increase in the next years/decades due to emissions in developing countries for which an increasing economic activity is accompanied with less emission control than in developed countries. In that respect, spaceborne observations of NMVOCs, like formaldehyde, have the potential to provide quantitative information about biogenic, biomass burning and anthropogenic NMVOC emissions.

3.3 Formaldehyde

Formaldehyde is the most abundant aldehyde in the atmosphere. A very small fraction is due to direct emissions by vegetation fires [*Andrea and Merlet*, 2001] and fossil fuel combustion [*Olivier et al.*, 2003] (see Table 3-5), but its main source is the chemical degradation of methane and non-methane volatile organic compounds [*Stavrakou et al.*, 2009a]. Therefore, formaldehyde constitutes an important indicator of NMVOC emissions and photochemical activity. In the continental boundary layer, H_2CO concentrations of the order of a few ppb are reported and can reach values greater than 10 ppb in tropical regions and large cities. The H_2CO concentration falls below 0.5 ppb far from the sources in the background atmosphere. Its abundance decreases with altitude and reach values as low as a few tens of ppt in the upper troposphere [*Dufour et al.*, 2009a and references therein] and in the stratosphere [*Steck et al.*, 2008]. Globally, methane oxidation represents the main source, accounting for more than half the H_2CO global production, the remainder being produced by NMVOC oxidation. Over continents, due to its short lifetime (a few hours), formaldehyde is directly related to the emissions of biogenic isoprene and many other short-lived NMVOCs emitted by biomass burning and anthropogenic sources. Therefore, satellite measurements of H_2CO are more and more used to constrain NMVOC emission inventories used in current state-of-the-art chemical transport models [*Palmer et al.*, 2003; *Abbot et al.*, 2003; *Palmer et al.*, 2006; *Shim et al.*, 2005; *Fu et al.*, 2007; *Stavrakou et al.*, 2009b].

3.3.1 Simulations of the global formaldehyde budget

Four categories of sources contributing to the atmospheric formaldehyde concentration can be considered: methane oxidation, anthropogenic emissions, vegetation fires and biogenic emissions. To estimate the contribution of each emission source to the global H_2CO distribution, simulations have been conducted using the global chemistry-transport model IMAGES [*Müller and Stavrakou*, 2005]. The latest version of IMAGES includes an optimized chemical scheme with respect to H_2CO production [*Stavrakou et al.*, 2009a]. Vegetation fires are obtained from the GFEDv2 biomass burning inventory [*van der Werf et al.*, 2006] and the MEGAN-ECMWF database is used for isoprene emissions [*Müller et al.*, 2008]. Anthropogenic CO and total NMVOC emissions are taken from the EDGAR v3.3 inventory for 1997 [*Olivier et al.*, 2001]. Speciation of VOC emissions is obtained from the POET database [*Olivier et al.*, 2003]. The contribution of each

source category to the total H_2CO column is calculated by including only emissions from that source. The resulting annual mean H_2CO columns are shown in Figure 3-8 [*Stavrakou et al.*, 2009a].
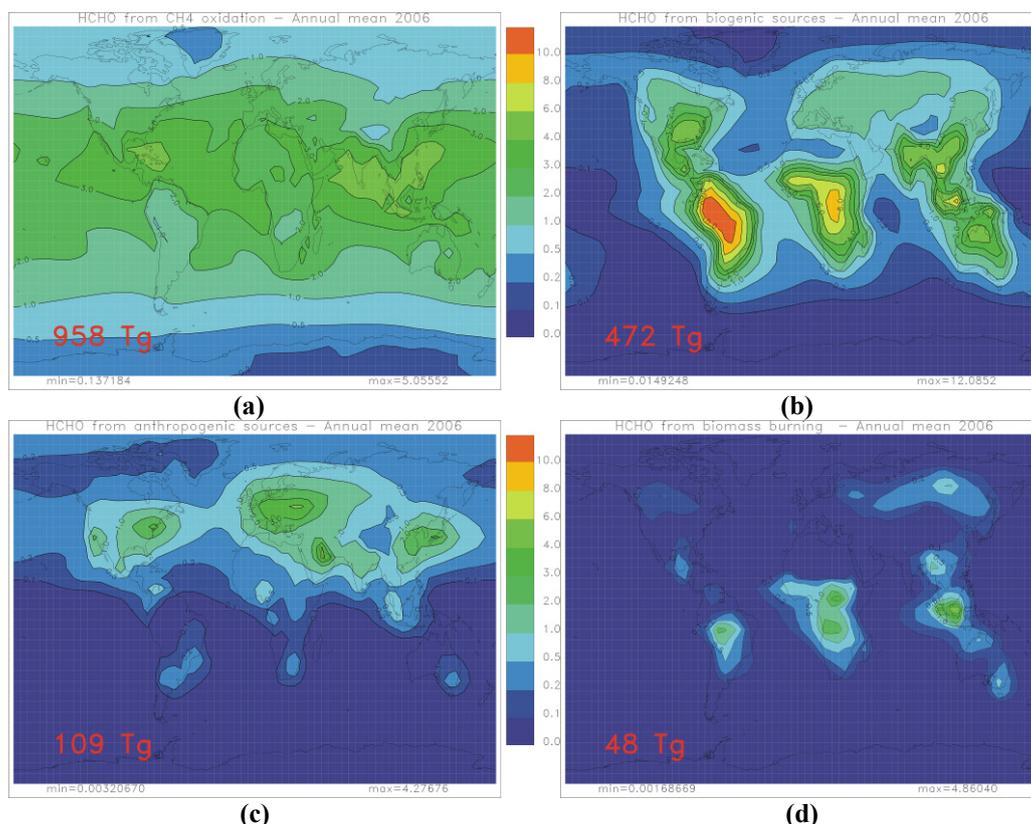
Figure 3-8: Contributions to the total modelled H_2CO columns (at a resolution of 5°x5° and averaged over 2006) from different emission sources: (a) CH_4 oxidation, (b) biogenic, (c) anthropogenic, and (d) pyrogenic NMVOC oxidation. Units are 10^{16} molec.cm^{-2}. The global amount of H_2CO produced annually is indicated inset [*Stavrakou et al.*, 2009a].

The total production of formaldehyde is estimated to be 1600 Tg(H_2CO) yr^{-1} (or 640 TgC yr^{-1}). The global amount of H_2CO produced annually by each source is provided inset each plot. The oxidation of methane in the background troposphere represents 60% of the H_2CO source on the global scale. However, the oxidation of locally emitted NMVOCs generally dominates over continents. Among them, biogenic compounds (mostly isoprene) are dominant and contribute up to 30%. If anthropogenic sources are significant in populated and industrialized areas, they are responsible for only 7% of the global source. The smallest contribution (3%) comes from biomass burning. Although insignificant on the global scale, fire episodes still lead to locally enhanced H_2CO columns detectable by satellite, especially in tropical regions [*Stavrakou et al.*, 2009a and 2009b]. The contribution of directly emitted H_2CO is small (less than 1% of the global total), yet it can also be locally important, especially during fire events. The total H_2CO signal is the strongest in tropical regions because of the combined effect of biogenic and biomass burning emissions.

3.3.2 Use of satellite formaldehyde data to constrain NMVOC emissions

Satellite observations of CH_4 and CO are used to constrain their surface emissions in a top-down approach. The potential of H_2CO observations to constrain NMVOC emissions in an inverse modelling framework is the main motivation of this thesis, which focuses on the retrieval of H_2CO columns from different satellite instruments. Formaldehyde is a key component of NMVOC

oxidation chains and its lifetime is short enough to be directly related to its precursor emissions. Satellite H_2CO column measurements offer the unique opportunity to constrain the emission of several NMVOCs from biogenic, pyrogenic and anthropogenic origin. Furthermore, H_2CO satellite measurements provide a global view of the NMVOC distribution, with a relatively good spatial resolution and regular observations over a long period of time. Multi-annual comparisons between simulated and measured formaldehyde columns are a mean to test the ability of chemistry transport model to reproduce the observed features of the H_2CO distribution in different regions of the world. These comparisons also provide a unique opportunity to asses the performance of emission inventories, highlighting their potential strengths but also point to possible weaknesses.

Emissions of a given NMVOC can be constrained by H_2CO satellites observations provided that this NMVOC is characterized by significant emissions, an elevated H_2CO yield and a short lifetime. Indeed, the lifetimes of VOCs against conversion to H_2CO may range from less than one hour for isoprene to ten years for methane. Transport will thus smear the spatial relationship between H_2CO column and VOC emission. VOCs of very long lifetime, such as methane, yield a global H_2CO background with no detectable localized signal. Short-chain alkanes, benzene, acetone and methanol have lifetime in excess of days and are also largely smeared. Biogenic monoterpenes and anthropogenic toluene have lifetimes typically less than a day, but their H_2CO production is delayed because of long-lived intermediates (leading to the formation of organic aerosols for monoterpenes) and the resulting signals are also largely smeared. Alkenes, isoprene, and xylenes have lifetimes shorter than a day and H_2CO is produced promptly upon oxidation, so that smearing is minimal. These are considered as "reactive" NMVOCs since their emission should produce a local H_2CO signal [*Fu et al.*, 2007]. For example, North American NMVOC emissions in summer are dominated by isoprene. Isoprene does have a short lifetime (of the order of an hour) and a large regional variability in emission, so that the variability of H_2CO columns over North America in summer is expected to reflect that of isoprene emissions [*Palmer et al.*, 2003]. Estimates of isoprene emissions have been derived from H_2CO columns retrieved from satellite, in particular over the Eastern US during summertime [*Palmer et al.*, 2003; *Abbott et al.*, 2003; *Palmer et al.*, 2006], but also over China and Southeast Asia [*Fu et al.*, 2007], over Europe [*Dufour et al.*, 2009b], and on the global scale [*Shim et al.*, 2005]. In Tropical regions, the derivation of emissions from satellite data is made more difficult, to a large extent because of the strong contribution of biomass burning to the observed H_2CO signal, difficult to separate from the biogenic VOC contribution [*Müller et al.*, 2008]. Anthropogenic NMVOC emissions are also expected to produce a detectable signal in industrialized areas, but they are difficult to detect in summertime because of the strong biogenic emissions. Their contribution in wintertime is expected to be more significant [*Fu et al.*, 2007; *Millet et al.*, 2008].

One of the main current limitations to NMVOC inversion is the accuracy of the H_2CO production schemes in the models. Indeed, although the reaction pathways leading from methane to H_2CO are well established, the degradation mechanisms of the many NMVOCs are still very uncertain [*Stavrakou et al.*, 2009a]. The other main limitation is the accuracy of the satellite H_2CO observations. Inverse modelling depends vitally on the quality of the retrieved columns. To this purpose, we aim at retrieving high quality H_2CO satellite observations, consistent over long time periods, using several satellite instruments, and properly quantifying the errors on the derived columns. This will be the subject of the chapters 6, 7 and 8. In parallel to this work, developments of the representation of NMVOC chemistry in the IMAGESv2 model have been conducted by the BIRA-IASB tropospheric modelling team (J.-F. Müller and co-workers), in order to improve modelling of biogenic, biomass burning and anthropogenic NMVOC degradation schemes. A summary of their results obtained using our satellite data product will be given in chapter 9.

4   Satellite Remote Sensing of Tropospheric Composition

During the last three decades, satellite platforms dedicated to the observation of atmospheric chemistry have considerably developed. As a result, remote regions over continents and oceans, where ground-based stations are usually rare, can now be probed regularly from space and variations of geophysical parameters on small and large scales can be studied. The section 4.1 of this chapter describes the general principles of spatial observations. In section 4.2, a selection of past, present and future satellite instruments designed for sensing atmospheric composition is presented, while in section 4.3, examples of satellite observations in the lower troposphere are shown. Finally, the three UV-Vis satellite instruments that have been used during this thesis to retrieve formaldehyde tropospheric columns, namely the global ozone monitoring experiment 1 and 2 (GOME-1 and GOME-2) and the scanning imaging absorption spectrometer for atmospheric chartography (SCIAMACHY), are presented in more details in section 4.4.

4.1 General principles of passive remote sensing from space

Spaceborne observations depend on the instrument specifications but also on the spatial platform which carries the instrument, particularly on its orbital parameters. The choice of the satellite platform and the instrument will depend on the scientific objectives (study of air pollution, ozone layer, trends, climate), on the atmospheric layers (troposphere, stratosphere), the geographical regions (tropics, mid-latitudes, polar regions) and the compounds that are to be observed.

| (a) | (b) |
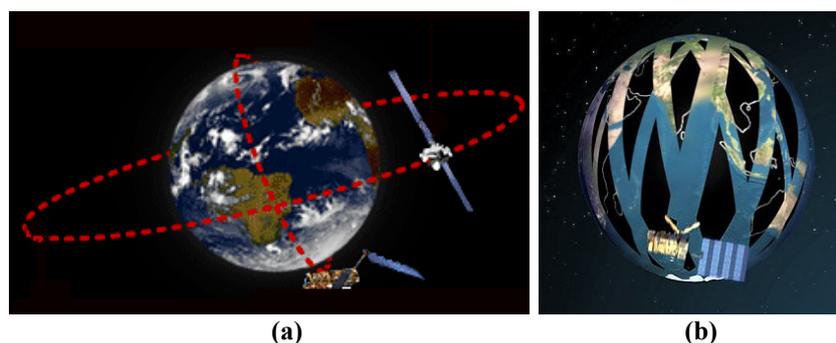
Figure 4-1: (a) Illustration of geostationary and low-Earth orbits around the Earth. (b) Earth coverage after several orbits of a near-polar satellite in LEO [© ESA - AOES Medialab].

Artificial satellites can be classified into two main categories depending on their orbit, as illustrated in Figure 4-1 (a): geostationary satellites (GEO) and low-Earth orbiting satellites (LEO). GEO satellites fly at an altitude of about 36000 km in the equatorial plane, in synchronisation with the Earth, so that they always cover the same region of the planet. So far, this kind of platform has mainly been used for meteorological applications. However, their high temporal resolution makes them also interesting for the monitoring of atmospheric compounds. LEO satellites fly at altitudes ranging from 600 to 2000 km. Depending on the orbit inclination (the angle between the orbit and the equator plane), the satellite coverage can be concentrated on tropical or mid-latitude regions, or extend from pole to pole. This last kind of orbit is called polar (or near-polar) and is inclined at (or nearly at) 90° to the equatorial plane. The orbit is said to be sun-synchronous when the angle between the orbital plane and the sun remains constant, so that the satellite always crosses a particular latitude at the same local time. The main satellites yielding information on atmospheric trace constituents fly in such sun-synchronous, near-polar orbits. Their coverage of the Earth is illustrated in Figure 4-1 (b). The common altitude for this type of satellite is ~800 km, yielding 100

minute orbits with approximately 14 orbits per day. Because of the Earth's rotation, each new orbit begins west of the previous one. The orbit crosses the equator once in the descending direction and once in the ascending direction, so that at low and mid latitudes, observation of the same location can be performed maximum twice a day. At higher latitudes, such orbits have a relatively higher sampling frequency. The actual time to achieve global coverage of the Earth varies from a day to a week, and is determined by the swath width of the instrument and its ability to measure during the night.

Satellite instruments exploit nadir and/or limb viewing geometries (both represented in Figure 4-2). Nadir viewing instruments observe a given field of view centred around the point directly beneath the satellite. Spatial coverage is performed by horizontal scanning and/or imaging systems. Limb viewing instruments look at the edge of the atmosphere and scan the atmosphere vertically, observing large horizontal paths at different altitudes. Occultation measurements are a particular case of limb viewing, but with the sun or the moon (or stars) in the field of view of the instrument. Limb viewing generally yields good vertical resolution, but but do not reach down to the planetary boundary layer. Therefore, limb observations are primarily used for sounding the mesosphere and stratosphere down to the upper troposphere. In contrast with the limb geometry, the nadir-viewing geometry observes the entire atmospheric column and is the only geometry that probes the boundary layer.
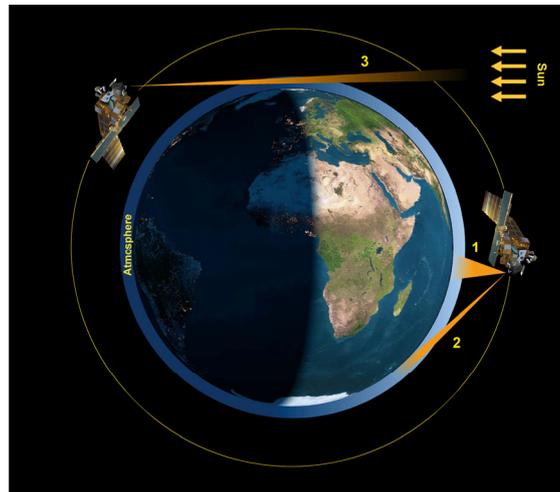
Figure 4-2: Observation geometries from space (1: nadir, 2: limb, 3: occultation) [© DLR-IMF, *Gottwald et al.*, 2006].

Passive remote sensing of atmospheric composition uses radiation from natural source, either direct (sun, moon, star) or indirect (atmosphere, surface). The measured radiation at the top of the atmosphere contains information about the sounded atmospheric layers. Passive remote sensing may be subdivided into two categories depending on the nature of the measured radiation: (1) radiation emitted by the Earth and followed by absorption/emission by the atmospheric compounds and (2) solar radiation after absorption and scattering in the atmosphere.

Retrieval methods based on Earth emission measurements are used in the thermal infrared (TIR: λ from 1 mm to 3 µm) and micro-wave domains ($\lambda > 1$ mm). At these wavelengths, the radiation emitted by the Earth dominates the signal. TIR measurements can be performed both during day and night. The radiance signal (I) recorded by the instrument is a composite of the initial emission of the Earth (I_0), modified by gaseous absorption and emission processes along the path. Figure 4-3 shows individual molecular transmittances (I/I_0) contributing to the recorded spectrum at the top of the atmosphere. Concentrations can be retrieved for each absorbing molecule present on the

atmospheric path using iterative inversion methods. Thanks to the information contained in the pressure broadening of molecular lines, TIR measurements yield vertical profile information on O_3, CO, CH_4, CO_2, H_2O, N_2O, HNO_3 and CFCs. Tropospheric total column amounts of PAN, NH_3, SO_2, C_2H_4, CH_3OH and HCOOH can also be derived in enhanced conditions. Nadir TIR techniques offer a sensitivity rapidly decreasing in the lowest troposphere and boundary layer and the capability of TIR sounders to probe the troposphere varies according to the temperature difference between the surface and the atmospheric layers [*Clerbaux et al.*, 2009].
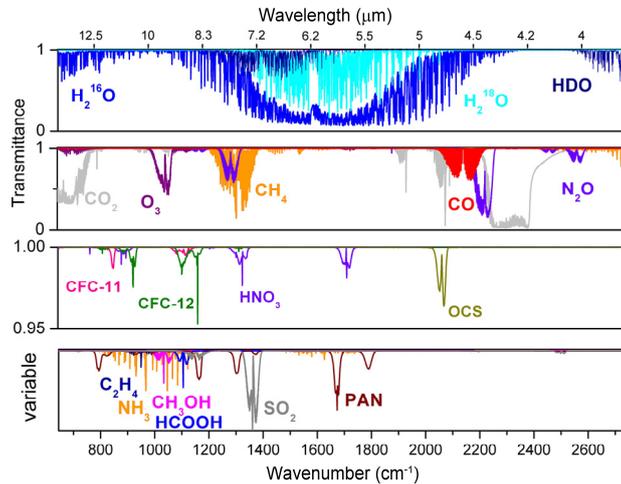
Figure 4-3: Radiative transfer simulations showing the transmittances of the main absorbing gases and weaker absorbers in the thermal infrared [*Clerbaux et al.*, 2009].

In the ultra-violet (λ < 400 nm), visible (400-700 nm), near infrared (NIR: 700-1400 nm) and short-wave infrared (SWIR: 1.4-3 µm) regions, the Earth radiation is too weak to be detected and an extraterrestrial source is used (the sun, the moon or stars). The measurements consist in detecting both the direct radiation from the source (I_0), and the radiation after its travel through the Earth atmosphere (I). While nadir and limb instruments measure light scattered by the atmosphere, occultation instruments look directly at the source trough the atmosphere. Figure 4-4 shows optical depths (proportional to the absorbance = $ln(I_0/I)$) of the main atmospheric absorbers in the UV-Vis-NIR-SWIR regions. Nadir solar backscatter measurements are sensitive to the troposphere, including the boundary layer, but height resolved information is typically not directly derived. No measurement can be made during the night or polar winters. The method to retrieve abundances of atmospheric compounds in the UV-Vis wavelength range will be the subject of chapter 5.
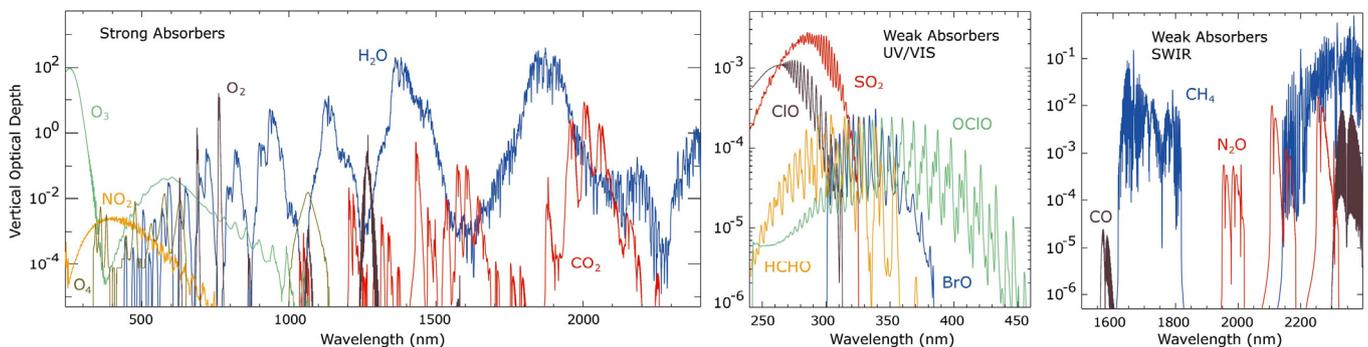
Figure 4-4: Simulated vertical optical depth of atmospheric constituents [© IUP-IFE, *Gottwald et al.*, 2006].

4.2 Spatial observations of atmospheric chemical composition: past, present and future missions
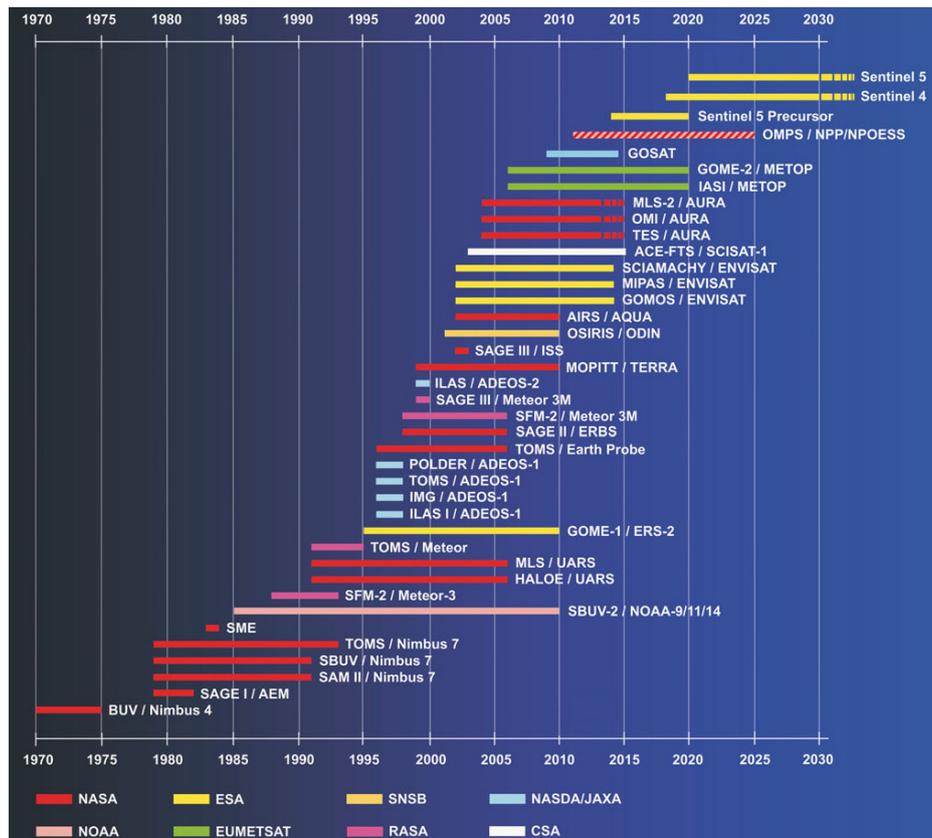
Figure 4-5: Atmospheric science spaceborne instruments and missions since 1970 [© DLR-IMF].

An overview of the past, present and future atmospheric science spaceborne instruments since 1970 is given in Figure 4-5 [*Gottwald et al.*, 2011]. The list of missions is not intended to be complete but rather to illustrate the progress in spaceborne instrumentation for atmospheric composition monitoring, with a focus on the troposphere. The reader is referred to Table 4-1 for details about the instrument characteristics. The past three decades can be viewed as the pioneering golden age of Earth observation with research missions generally having been much more successful than originally expected. In the early 1970s NASA started global measurements of atmospheric ozone with the BUV instrument, followed by SBUV and a series of SBUV-2 instruments operated from 1985 up to now. While early missions were limited to the observation of the stratosphere, satellite remote sensing of the troposphere began in 1978 with the TOMS instrument onboard the Nimbus 7 satellite. Although TOMS was primarily aimed at determining global knowledge of stratospheric ozone, it was later recognised to also yield information about the troposphere [*Fishman et al.*, 1990]. TOMS observes solar backscattered radiation through a single monochromator at few UV wavelengths where ozone has strong absorption features. It also provides important information on tropospheric ozone, volcanic SO_2 and aerosols [*IGAC*, 2007]. NASA operated several TOMS instruments on a variety of satellites until 2006.

In 1995, the European contribution to ozone and atmospheric composition measurements started with GOME, which was launched onboard the ESA ERS-2 satellite. This instrument was designed primarily to provide global observation of stratospheric ozone and related species, but it was also designed to investigate the potential of nadir remote sounding of backscattered UV-Vis radiation for

the retrieval of tropospheric trace gases. The broad spectral coverage of GOME from UV to NIR, coupled with spectroscopic inversions, has been enabling to probe several key species in tropospheric chemistry like O_3, NO_2, H_2CO, BrO, SO_2 and H_2O [*Burrows et al.*, 1999b and references therein]. The next ESA ENVISAT platform was launched in 2002. It carries ten optical and radar instruments with the aim to provide continuous observation of the Earth. Among them, SCIAMACHY measures backscattered solar radiation in the UV-Vis-NIR-SWIR regions. This wide range of wavelengths allows the detection of all the species observed with GOME, as well as aerosols and clouds, and in addition CO_2, CH_4 and CO. SCIAMACHY combines nadir and limb viewing. Also onboard ENVISAT, MIPAS performs TIR limb measurements from the mesosphere to the upper troposphere. These are used to make observations such as O_3, CO, PAN, H_2O and HNO_3. GOMOS performs also limb measurements from the mesosphere to the upper troposphere, but using the UV-VIS radiation of stars in occultation mode.

Since its launch in December 1999, MOPPIT has also significantly expanded scientific perspective on the scale of tropospheric pollution. It is a nadir-viewing radiometer operating in the 4.7 μm band of carbon monoxide. It flies on the TERRA satellite, which is part of the NASA Earth Observing System (EOS). EOS is completed by the AQUA and AURA satellites, respectively launched in 2002 and 2004. AQUA carries the AIRS instrument, which measures also in the thermal infra-red to retrieve temperature, water vapour, cloud properties and greenhouse gases. With the launch of AURA, began a period of tropospheric trace gases monitoring with unprecedented spatial coverage based on three separate instruments: MLS, OMI and TES. MLS is a limb sounder that operates in the microwave during both day and night, allowing the retrieval of O_3, CO and H_2O profiles in the stratosphere and the upper troposphere. TES is a Fourier transform spectrometer observing thermal emission with both nadir and limb capability. OMI is a nadir viewing UV-Vis spectrometer in the heritage of GOME and SCIAMACHY. OMI's technology is advanced by employing two-dimensional CCD detector arrays for the first time, improving spatial resolution in combination with a daily global coverage.

The first METOP satellite was launched by EUMETSAT and ESA in October 2006 as part of the EUMETSAT Polar System (EPS), including a series of three satellites. METOP is the first meteorological platform to have instrumentation dedicated to trace gas measurements. Each platform will contain two instruments for remote sensing of tropospheric chemistry: GOME-2 and IASI. GOME-2 is an improved version of GOME, while IASI is a nadir viewing Fourier transform spectrometer measuring thermal emission. This series of satellites will provide continuous monitoring capability to the year 2020 from polar sun-synchronous low-Earth orbit.

In parallel, initiatives emerged to realise a comprehensive global Earth observation system. Global Monitoring for Environment and Security (GMES) is a joint initiative of the European Commission and ESA which aims at bridging the gap between scientific data and the provision of reliable and timely information to government and data users. The GMES space component includes a series of five Sentinel satellites. Sentinels 4 and 5 are devoted to monitor the atmosphere and are planned for launch respectively in 2018 and 2020. Furthermore, in order to bridge parts of the data gap between the SCIAMACHY-OMI period and Sentinel 4 and 5, a new imaging spectrometer, TROPOMI, is planned for launch in 2014 as the GMES Sentinel-5 precursor mission. TROPOMI will combine the large swath of OMI and the extended wavelength range of SCIAMACHY [*Gottwald et al.*, 2006].

With the exception of Sentinel 4, the missions presented so far refer to space platforms in low-Earth sun-synchronous polar orbits which cannot capture the diurnal variation of photochemically active trace gases. However, air pollution quantification and forecast require measurement of diurnal variability. To address these aspects, the GeoSCIA (Geostationary Scanning Imaging Absorption

Spectrometer) concept has been adopted by EUMETSAT and ESA to form the Sentinel 4 UVN instrument. It will be launched on the next generation of European geostationary meteorological satellite MTG (METEOSAT Third Generation). It shall perform daytime observations of solar backscattered spectra to determine the total and tropospheric column amounts of ozone, ozone precursors and pollutants over Europe with temporal and spatial resolutions better than 1h and 10km, increasing the frequency of cloud-free measurements and therefore improving the observation of the boundary layer. On the same MTG platform, EUMETSAT will install the IRS sensor, which will focus on meteorological parameters but shall also obtain vertical distributions of O_3 and CO during day and night with temporal and spatial resolutions of about 20–60 min and 4 km [*Gottwald et al.*, 2011].

Table 4-1: Passive satellite instruments designed to determine trace gas distributions in the atmosphere. The list of sensors refers to Figure 4-5 [*Gottwald et al.*, 2011].

Instrument	Name	Atm. layer (1)			View. Geom. (2)			Target species
		T	S	M	N	L	O	
ACE-FTS	Atmospheric chemistry experiment		x				x	O_3, H_2O, CO, CFCs, HNO_3, NO
AIRS	Atmospheric IR Sounder	x			x			O_3, H_2O, CO_2, CH_4
BUV	Backscatter UV ozone experiment		x		x			O_3
GOME 1, 2	Global ozone monitoring experiment	x	x		x			O_3, NO_2, H_2O, BrO, OClO, SO_2, H_2CO, CHOCHO, IO; clouds and aerosols
GOMOS	Global ozone monitoring by occultation of stars		x	x			x	O_3, NO_2, H_2O, NO_3; aerosols, T
HALOE	Halogen occultation experiment		x				x	CO_2, H_2O, O_3, NO_2, HF, HCl, CH_4, NO
IASI	IR atmospheric sounding interferometer	x	x		x			O_3, H_2O, CO, CH_4, N2O; T
ILAS I, II	Improved limb atmospheric spectrometer		x				x	O_3, NO_2, N_2O, H_2O, CFC-11, CH_4; aerosols
IMG	Interferometric monitor for greenhouse gases	x	x		x			O_3, N_2O, H_2O, CH_4, CO, CO_2
MIPAS	Michelson inferometer for passive atmospheric sounding		x	x		x		O_3, NO_x, N_2O_5, $ClONO_2$, CH_4, CFCs,HNO_3; T, P
MLS 1-2	MW limb sounder		x	x		x		ClO, O_3, H_2O, HNO_3; T, P (1) + CO, HCL, BrO, N_2O, SO_2, HCN, CH_3CN (2)
MOPITT	Measurement of pollution in the troposphere	x			x			CO, CH_4
OMI	Ozone monitoring instrument	x	x		x			O_3, SO_2, NO_2, BrO, H_2CO; aerosols
OMPS	Ozone monitoring and profiling suite	x	x	x	x	x		O_3, SO_2; aerosols
OSIRIS	Optical spectrograph and IR imaging system		x	x		x		NO, OClO, O_3, NO_2; aerosols
SAGE I-II-III	Stratospheric aerosol and gas experiment		x			x		O_3, NO_2; aerosols (I) + H_2O (II) + OClO, BrO, NO_3 (III)
SAM II	Stratospheric aerosol measurement		x			x		Aerosols
SBUV 1, 2	Solar backscatter UV ozone experiment	x	x		x			O_3
SCIAMACHY	Scanning imaging absorption spectrometer for atmospheric chartography	x	x	x	x	x	x	O_3, O_2, O_4, NO, NO_2, N_2O, BrO, OClO, H_2O, SO_2, H_2CO, CHOCHO, IO, CO, CO_2, CH_4; cloud, aerosols
SENTINEL 4 UVN	UV-Vis-NIR	x	x	x	x			O_3, NO_2, SO_2, H_2O, CHOCHO; cloud, aerosols
SENTINEL 5 UVNS	UV-Vis-NIR-SWIR	x	x	x	x			O_3, NO_2, BrO, SO_2, H_2CO, CHOCHO, H_2O, CO, CO_2, CH_4; cloud, aerosols
TROPOMI	Tropospheric Ozone Monitoring Instrument	x	x	x	x			O_3, NO_2, BrO, SO_2, H_2CO, CHOCHO, H_2O, CO, CH_4; cloud, aerosols
SFM-2	Spectrophotometer		x				x	O_3, aerosols
TANSO (GOSAT)	Thermal and NIR sensor for carbon observation	x			x			CO_2, H_2O, CH_4
TES	Tropospheric emission spectrometer	x	x		x	x		HNO_3, O_3, NO, H_2O, CH_4, CO, SO_2
TOMS	Total ozone monitoring spectrometer	x	x		x			O_3, SO_2, aerosols

(1) Atmospheric layer: T=Troposphere, S=Stratosphere, M=Mesosphere
(2) Viewing geometry: N=Nadir, L=Limb, O=Occultation

4.3 Remote sensing of air quality from space

Air quality monitoring from space relies primarily on a combination of UV-Vis, SWIR and TIR nadir-viewing sensors. Table 4-2 lists the main current satellite instruments allowing the observation of trace gas concentrations in the lower troposphere. All instruments in Table 4-2 are in low-Earth, sun-synchronous polar orbits and in nadir (or limb/nadir) viewing geometry. Global observations of tropospheric trace gas columns provide unique information on the spatial distribution of pollutants on a regional and continental scale. The spatial resolution and Earth coverage is increasing with instrumental developments. Furthermore, as tropospheric trace gas data records extend to a decade or longer, important trend information about some emissions in specific regions of the world is now becoming available [*Martin et al.*, 2008; *Fishman et al.*, 2008]. However, satellite remote sensing in the lower troposphere is substantially more difficult than in the stratosphere. Measurements show a decreasing sensitivity towards the surface, particularly in the UV because of molecular scattering, and in the TIR due to the reduced thermal contrast between the atmospheric layers and the surface. Moreover, passive nadir viewing techniques are characterised by their limited vertical resolution, by difficulties in separating tropospheric concentration from the total column and in accounting for the effects of clouds and water vapour [*Martin et al.*, 2008].

Table 4-2: Nadir viewing satellite instruments used for the remote sensing of tropospheric chemistry. Instruments measuring aerosols properties are not included [*Martin et al.*, 2008; *IGAC*, 2007].

Instrument	GOME	MOPITT	AIRS	SCIAMACHY	OMI	TES	GOME-2	IASI
Platform	ERS-2	TERRA	AQUA	ENVISAT	AURA		METOP	
Measurement period	1996-2003	2000-	2002-	2002-	2004-		2006-	
Equator crossing time	10:30 am	10:30 am	01:30 pm	10:00 am	01:45 pm		09:30 am	
Nadir resolution [km^2]	320x40	22x22	14x14	60x30	24x13	8x5	80x40	12x12
Spectral region	UV-Vis-NIR	SWIR-TIR	Vis-NIR-TIR	UV-Vis-NIR-SWIR	UV-Vis	TIR	UV-Vis	TIR

4.3.1 Carbon monoxide

The first global CO measurements from satellite were obtained in 1996 by IMG on the Japanese ADEOS platform, confirming that the contribution of anthropogenic sources to the CO column was highest than the contribution of biogenic sources [*Barret et al.*, 2005]. Unfortunately, IMG operated for less than a year because of a failure of ADEOS in June 1997. Since 1999, MOPITT has been continuously measuring vertically-resolved CO concentrations. Global observations of CO from MOPITT have revealed important processes in emissions and transport of industrial and biomass burning events and have provided new data for studying pollution sources, chemistry and transport in details [*Fishmann et al.*, 2008; *Stavrakou and Müller*, 2006]. MOPITT long-term observational record is used to investigate links between chemistry, climate and other components of the Earth system. For example, strong correlations are found between Southern Hemispheric CO fire activity in Indonesia and El Niño conditions [*Edwards et al.*, 2006]. CO is also retrieved from SCIAMACHY in the 2.3 μm absorption band [*Buchwitz et al.*, 2007], and from AIRS, TES and IASI in the 4.7 μm absorption band [*McMillan et al.*, 2005; *Lopez et al.*, 2008; *George et al.*, 2009; *Turquety et al.*, 2009]. MOPITT, AIRS, TES and IASI observe thermal emission and are thus more sensitive to the mid-troposphere, while SCIAMACHY observes backscattered solar radiation, offering high measurement sensitivity down to the surface, though with no vertical resolution [*Kopacz et al.,* 2010]. Figure 4-6 shows an example of CO total columns retrieved from SCIAMACHY measurements in Apr.-May-Jun. and Oct.-Nov.-Dec. 2003. Between April and June,

CO concentrations are higher in the Northern Hemisphere, where human impact is greater than in the Southern Hemisphere. However, between October and December, the huge plumes emitted from forest and grassland fires burning in South America and Southern Africa dominate the emissions, travelling as far as Australia.
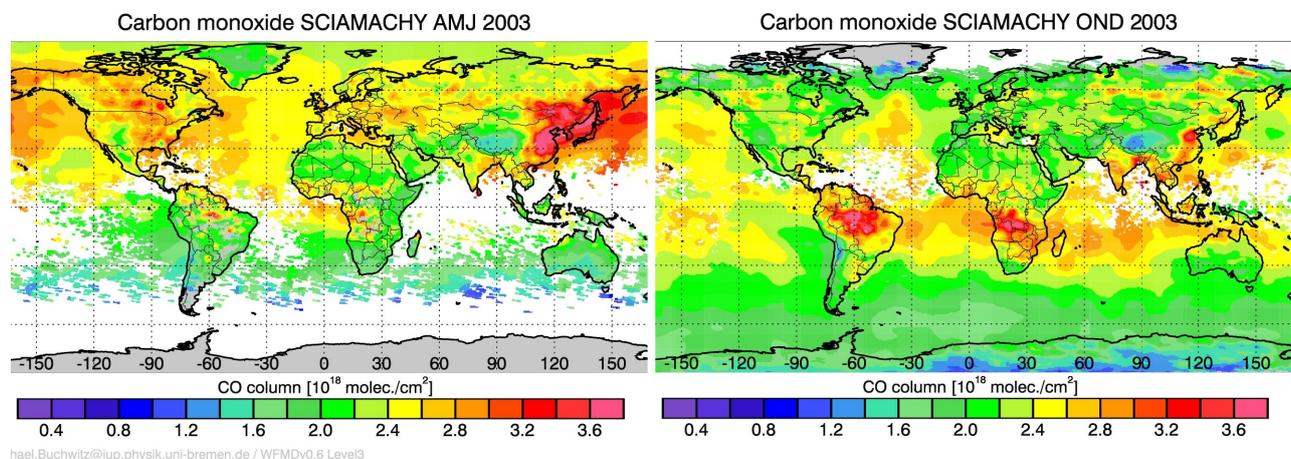
Figure 4-6: Seasonally averaged CO columns retrieved from SCIAMACHY/ENVISAT in Apr.-May-Jun. (left panel) and Oct.-Nov.-Dec. (right panel) 2003 [*Buchwitz et al.*, 2007].

4.3.2 Tropospheric ozone

Since the launch of the first TOMS instrument, ozone total columns primarily provide information about the distribution of ozone in the stratosphere where 90% of the atmospheric ozone is found, the remaining 10% being located in the troposphere. However, enhancements in total ozone at low latitudes over the South Atlantic and adjacent continents have been regularly observed during austral spring, with generally a positive correlation with CO measurements and with El Niño Southern oscillations. Those persistent ozone sources in the tropics were the first indications that satellite total ozone columns could be used to identify ozone pollution [*Fishman et al.*, 2008]. In order to retrieve the amount of tropospheric ozone from space-based observations, various techniques have been developed for separating the tropospheric and stratospheric contributions. For example, by assuming that ozone variability in the stratosphere is defined on relatively large spatial scales compared to the troposphere, the stratospheric ozone distribution can be isolated. Tropospheric ozone obtained with this method revealed elevated amounts at Northern temperate latitudes during the summer and in the tropics and subtropics in the Southern Hemisphere during austral spring [*Fishman et al.*, 1996]. The stratospheric contribution can also be separated using ozone profiles retrieved from limb measurements, or with the cloud slicing method [*Ziemke et al.*, 2001]. Thanks to the broad UV-Vis wavelength range of GOME, SCIAMACHY and OMI, direct determination of ozone profiles, including tropospheric information, has been developed by using information from different wavelength ranges and the temperature dependence of the ozone absorption [*Munro et al.*, 1998; *Liu et al.*, 2009]. Another approach for direct retrieval of tropospheric ozone is derived from nadir measurements of TIR emission at high spectral resolution. For example, Figure 4-7 shows monthly mean TES ozone mixing ratio at 681 hPa in May and September 2005 [*Beer et al.*, 2006]. Tropospheric O_3 concentrations are consistently higher in the more populous and industrialized Northern Hemisphere and during biomass burning events in Africa and South America. Tropospheric ozone retrievals in the TIR are also performed from AIRS [*Susskind et al.*, 2003] and IASI [*Boynard et al.,* 2009, *Dufour et al.*, 2010], but these instruments have lower spectral resolution than TES and are thus slightly less sensitive to the troposphere.
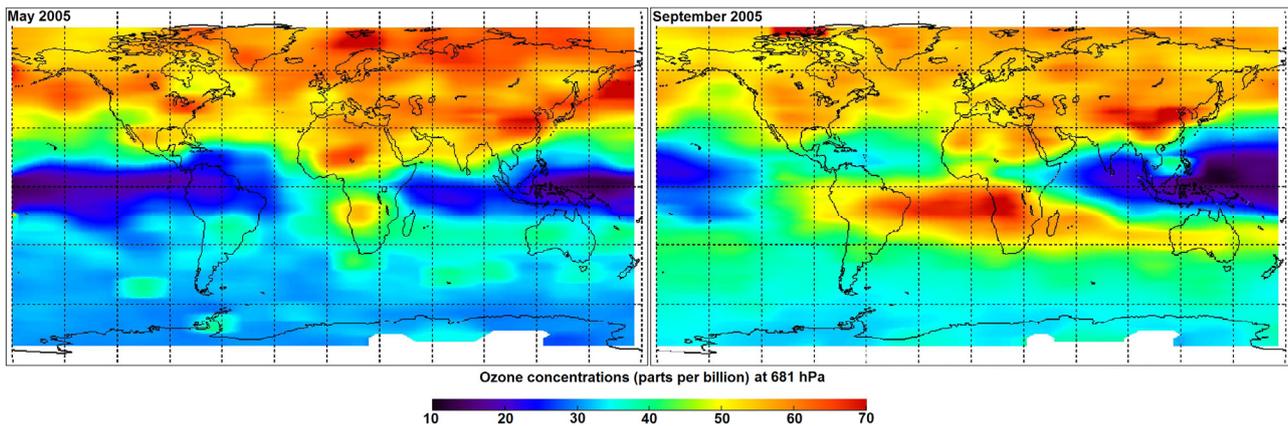
Figure 4-7: Monthly mean ozone concentrations (ppb) at 681 hPa for May (left) and September (right) retrieved from TES measurements in 2005 [*Beer et al.*, 2006].

4.3.3 Tropospheric nitrogen dioxide
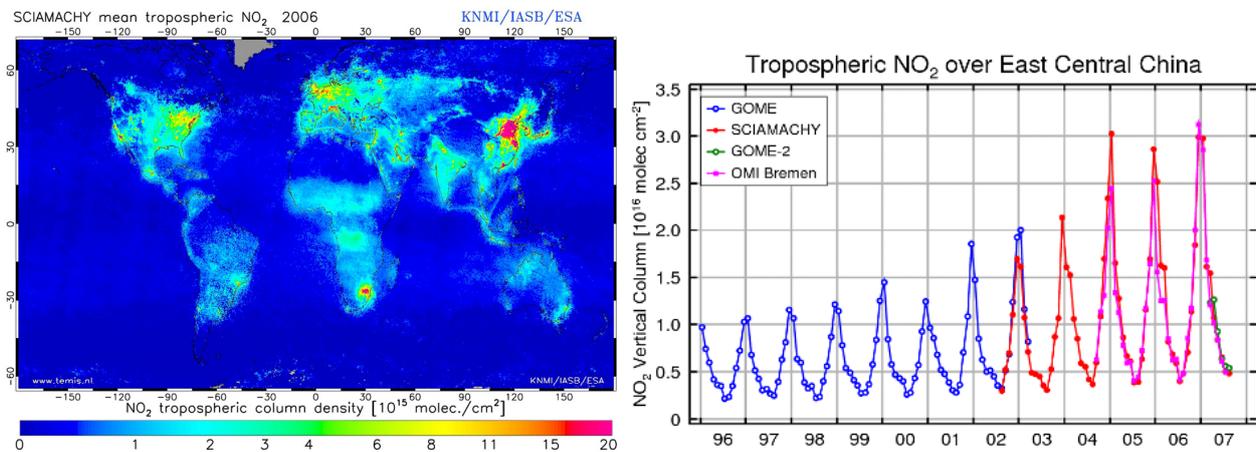
Figure 4-8: Tropospheric NO$_2$ in 2006 retrieved from SCIAMACHY [© www.TEMIS.nl] and tropospheric NO$_2$ above China retrieved from GOME, SCIAMACHY, OMI and GOME-2 [© *A. Richter*, Bremen-IUP].

Tropospheric NO$_2$ columns are retrieved independently from GOME, SCIAMACHY, GOME-2 and OMI by several groups [*Martin et al.*, 2002; *Richter and Burrows.*, 2002; *Boersma et al.*, 2004; *Bucsela et al.*, 2006]. As for ozone, the stratospheric contribution to the total NO$_2$ column needs to be estimated in order to retrieve the tropospheric NO$_2$ distribution. This is done assuming that the NO$_2$ variability in the stratosphere is defined on much larger spatial scales than in the troposphere, or by using assimilation in stratospheric chemistry models. Figure 4-8 shows a map of tropospheric NO$_2$ columns averaged over the year 2006, retrieved from SCIAMACHY in the framework of the TEMIS project [*Boersma et al.*, 2004]. Pronounced enhancements are obvious over major urban and industrial regions. Furthermore, using combined observations of GOME, SCIAMACHY, OMI and GOME-2, trends in emission can be detected. For example, the right panel of Figure 4-8 shows an increase in tropospheric NO$_2$ columns over industrial China during the last decade. The same dataset also reveals that emissions have decreased in Europe and the United States [*Richter et al.*, 2005; *van der A et al.*, 2006 and 2008; *Stavrakou et al.*, 2008].

4.3.4 Non-methane volatile organic compounds

Formaldehyde is used as a major proxy to detect and constrain NMVOC emissions. In contrast to CO and NO$_2$, the H$_2$CO distribution is dominated by natural sources, even if anthropogenic NMVOC emissions contribute significantly to the H$_2$CO columns in populated and industrial areas. H$_2$CO was first retrieved from GOME spectra during intense biomass burning events over Asia in 1997, a year associated with strong El Niño events [*Thomas et al.*, 1998] and over the Eastern United States where densely populated deciduous forests emit large quantities of isoprene during the summer [*Chance et al.*, 2000; *Wittrock et al.*, 2000]. Although the H$_2$CO column is to a very large extent concentrated in the lower troposphere, significant upper tropospheric H$_2$CO mixing ratios have been observed during the growing season of the terrestrial biosphere in the Northern Hemisphere and during the biomass burning season in the Southern Hemisphere. Seasonally-resolved upper tropospheric profiles of formaldehyde have been retrieved from ACE-FTS [*Coheur et al.*, 2007; *Dufour et al.*, 2009a] and from MIPAS [*Steck et al.*, 2008]. Global observations of H$_2$CO total columns in the lower troposphere are retrieved from GOME, SCIAMACHY [*Wittrock et al.*, 2006; *De Smedt et al.*, 2008], OMI [*Millet et al.*, 2008] and GOME-2 [*De Smedt et al.*, 2009; *Vrekoussis et al.*, 2010]. Another proxy for NMVOC sources is glyoxal, which can provide further constrains on NMVOC emissions. Glyoxal is retrieved from SCIAMACHY, OMI and GOME-2 measurements [*Wittrock et al.*, 2006; *Vrekoussis et al.*, 2010; *Lerot et al.*, 2010b]. IASI measurements allow detecting other NMVOCs such as formic acid (HCOOH), methanol (CH$_3$OH), ethene (C$_2$H$_4$) and PANs above intense emission sources originating from biomass burning events [*Coheur et al.*, 2009], but also at the global scale for methanol and formic acid [*Razavi et al.*, 2011]. Note that H$_2$CO has not been identified in IASI spectra because of the weak H$_2$CO lines in the near infrared above 2500 cm^{-1}, where the noise performances of IASI are significantly reduced [*Coheur et al.*, 2009].

GOME H$_2$CO VCD [x10^{15} molec.cm^{-2}]
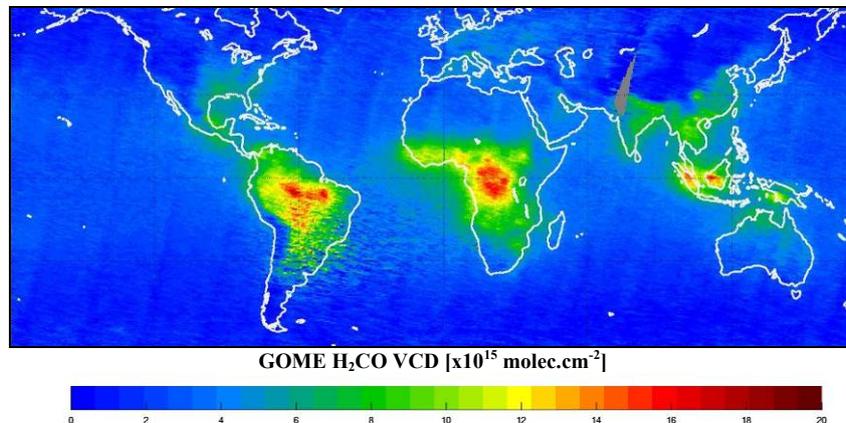
Figure 4-9: H$_2$CO tropospheric columns in 1997 retrieved from GOME measurements [*De Smedt et al.*, 2008].

The retrieval of formaldehyde columns from earthshine backscattered spectra remains a challenge because of the relatively weak H$_2$CO absorption signatures compared to other UV-Vis absorbers like O$_3$, SO$_2$ or NO$_2$. During this thesis, efforts have been made to create a consistent dataset of H$_2$CO observations composed by GOME, SCIAMACHY and GOME-2 measurements [*De Smedt et al.*, 2008]. Figure 4-9 shows for example a map of tropospheric H$_2$CO columns averaged over 1997, as retrieved from GOME. This dataset is used to constrain biogenic and biomass burning NMVOC emissions [*Stavrakou et al.*, 2009 a and b; *Barkley et al.*, 2008], and to a lesser extent, anthropogenic NMVOCs emissions. Thanks to the long period of observations covering more than

14 years, it was possible to detect trends in anthropogenic emissions of NMVOCs in Asia [*De Smedt et al.*, 2010].

4.4 UV-Visible satellite instruments: GOME, SCIAMACHY and GOME-2

This section gives a description of the three instruments that have been used to retrieve H_2CO columns during this thesis, namely GOME, SCIAMACHY and GOME-2. They are three spectrometers measuring the Earth backscattered radiance as well as direct solar irradiance spectra. They fly on polar, sun-synchronous orbits with a morning descending node crossing time. The GOME instrument, flying onboard ERS-2 since 1995, was proposed as a precursor to SCIAMACHY, flying on ENVISAT since 2002. Both instruments were developed by ESA. The observations started by GOME are now continued by the GOME-2 instrument, operated by EUMETSAT on the METOP platforms. METOP-A was launched in 2006 and is the second largest European Earth-observation satellite after ENVISAT. It is planned that subsequent METOP satellites will be launched at approximately five year intervals, maintaining the service until at least 2020, extending the European atmospheric chemistry measurement series in the UV-Vis into the next two decades.

As GOME and GOME-2 instruments are almost identical, their characteristics are discussed together, while SCIAMACHY is presented in a separated section. At the end of the chapter, the main features of the three instruments are compared in Table 4-3.

Figure 4-10: Artist impressions of (1) ERS-2 (followed by ENVISAT), (2) ENVISAT, and (3) METOP [© ESA].

4.4.1 GOME/ERS-2 and GOME-2/METOP-A

The GOME instrument is a four channel grating spectrometer measuring radiation in the ultraviolet, visible and near-infrared. The optical design of GOME-2 is almost identical to GOME. A schematic diagram of their layout is shown in Figure 4-11. The main improvements of GOME-2 consist in ground pixels four times smaller than GOME-1 with a swath width twice as large. Furthermore, the GOME-2 instrument includes a more complex polarisation monitoring unit [*Callies et al.*, 2000].

At the nadir view port, a scan mirror is used in order to change the instrument line-of-sight angle (LoS). The scan mirror directs the light emitted from the atmosphere into a telescope, designed to match the field of view of the instrument to the dimensions of the entrance slit. Except for the scan mirror, all spectrometer parts are fixed. The spectra are recorded simultaneously from 240 nm to 790 nm. Light entering the instrument is split into four spectral bands by a pre-disperser prism and a channel separator prism. In each of the four spectral channels, the light is dispersed by a diffraction grating and focused onto a silicon linear detector array comprising 1024 individual detector pixels. The spectral resolution varies between 0.2 nm (UV, channel 1) and 0.4 nm (Vis, channel 4) for

GOME and between 0.25 nm and 0.55 nm for GOME-2 (see Table 4-3). Due to the large dynamic range of the signal below 307 nm, resulting from the increasing ozone absorption, the channel 1 diode array is divided into two virtual bands (1a/1b) with different integration times in order to optimize the signal-to-noise ratio. To achieve maximum sensitivity, the detectors are cooled to 273K.
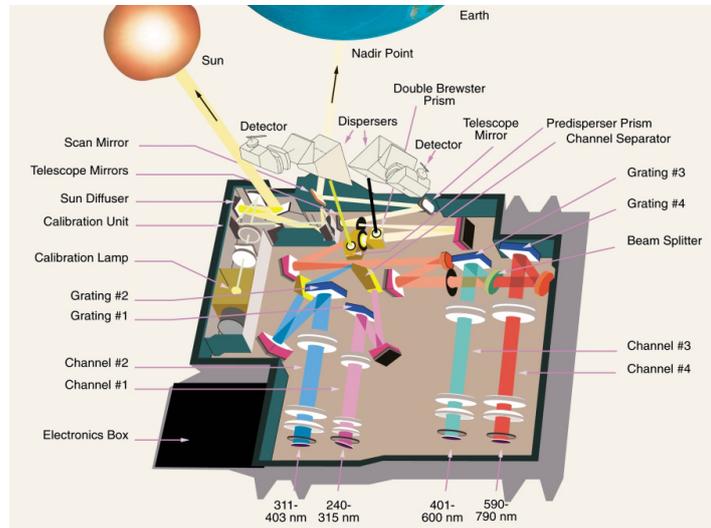
Figure 4-11: Schematic view of the GOME and GOME-2 optical layout. The double Brewster prism and the dispersers/detector are new PMD features in GOME-2 (not present in GOME) [© ESA].

Reflected light from the Earth's atmosphere is, in general, partially polarized. To correct for effects caused by the polarization sensitivity of the instrument, a small fraction of light is reflected toward dedicated polarization measurement devices (PMDs). Compared to the main channels, the PMD measurements are performed at lower spectral resolution, but at higher spatial resolution. This makes the PMD measurements useful for detecting rapid changes in the observed surface reflectivity and cloudiness during scanning. In GOME, the PMDs are three fast broadband silicon diodes whose spectral range covers the optical channels 2, 3 and 4, from 300 to 750 nm. The new polarization unit of GOME-2 monitors the 312 to 790 nm range using 200 silicon detector pixels with a spectral resolution that varies from 2.8 nm at 312 nm to about 40 nm at 790 nm.

The scan mirror can also point to two internal calibration light sources (a white light source to monitor the pixel-to-pixel variability and a spectral light source for wavelength calibration) and to the Sun view port. Once a day, solar irradiance measurements are performed when the satellite crosses the terminator in the North Polar Region coming from the night side. The solar radiation is attenuated by a mesh and directed via a diffuser plate onto the entrance slit of the spectrometer. Unfortunately, in GOME/ERS-2, as a result of residual periodic structures on the diffuser, spectral artefacts are present in the recorded solar irradiances. These structures depend on the illumination angle and therefore on the seasonally varying position of ERS-2 relative to the sun [*Richter and Wagner*, 2001].

During the illuminated part of the orbit, the instruments perform nadir observation of the atmosphere by scanning the surface from east to west (forward scan) and back (backward scan), while the satellite moves on the descending node. The mean local equator crossing times of GOME and GOME-2 are respectively 10:30 and 09:30 am. The maximum line-of-sight angle of GOME/ERS-2 is 32°, corresponding to a swath width of 960 km. As illustrated in Figure 4-12 (a), an across track scan sequence of GOME consists of four ground pixel types called east, nadir, west, and one backscan pixel. The nominal integration time is 1.5 seconds, resulting in a resolution of

40×320 km^2 for the forward pixels [*Burrows et al.*, 1999b]. The GOME-2 line-of-sight angle has been extended up to 57°, corresponding to a larger swath width of 1920 km. As for GOME, the along-track dimension of the instantaneous field-of-view is ~40 km. The scan mirror sweeps in 4.5 seconds from negative to positive LoS angle, followed by a backward scan of 1.5 seconds. As illustrated in Figure 4-12 (b), the integration time of 187.5 milliseconds for the main channels corresponds to a ground pixel resolution of 40×80 km^2 in the forward scan. The default integration time of 187.5 milliseconds is used in all channels except in the backward scan mode or when longer integration times are needed because of low light intensity, for example at high solar zenith angles.

(a)   (b)
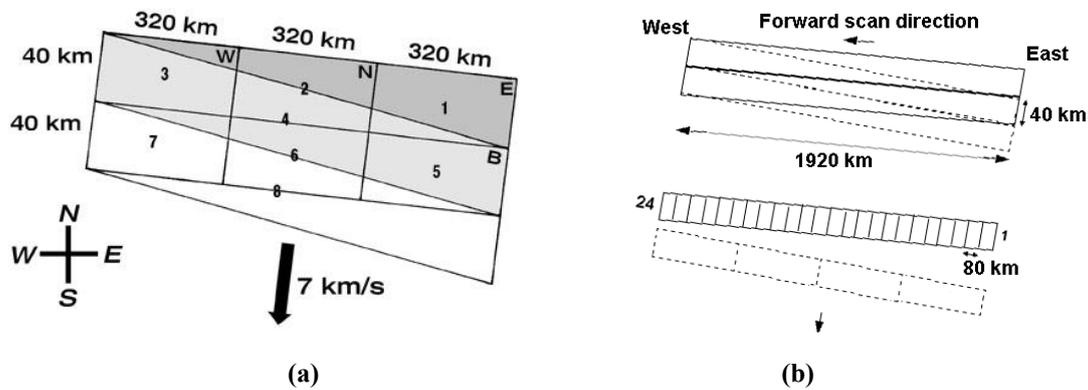
Figure 4-12: GOME (a) and GOME-2 (b) scan geometry in their default nadir viewing mode. Two successive scan sequences are shown. The GOME forward scan consists of 3 pixels (E, N, W), followed by a backscan pixel (B), each with an integration time of 1.5 sec. The GOME-2 forward scan consists of 24 pixels, each with an integration time of 187.5 msec., followed by 4 backscan pixels (375 msec.). Note that the figure is not drawn to scale [© ESA/EUMETSAT].

As illustrated in Figure 4-13, GOME achieves global coverage at the equator within three days, while the larger swath width of GOME-2 enables global coverage within 1.5 days. In June 2003, the tape recorder on ERS-2 permanently failed. Since then, only those GOME measurements are available that can be directly recorded by a ground receiving station, reducing the global coverage by about 70 percent. GOME is still in operation but with limited Earth coverage over the Northern Hemisphere and part of Antarctica.
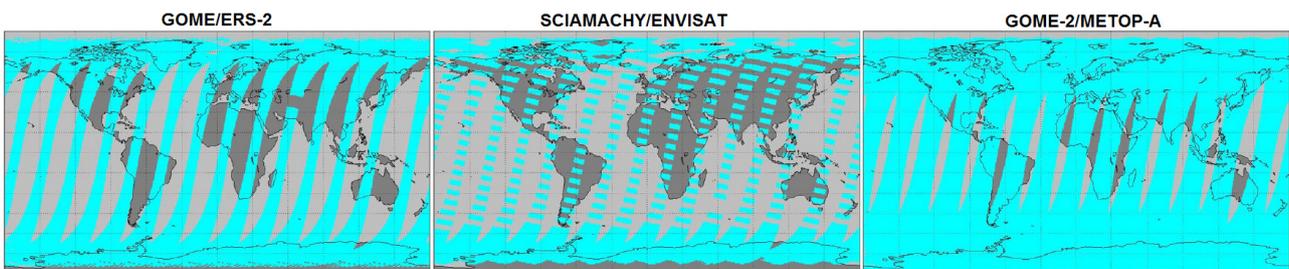
Figure 4-13: Illustration of the maximum spatial coverage achieved in one day with the GOME, SCIAMACHY and GOME-2 satellite instruments (in nadir viewing mode) [*Theys et al.*, 2010].

4.4.2 SCIAMACHY/ENVISAT

The SCIAMACHY instrument is a height channel grating spectrometer covering the spectral range from ultraviolet to short-wave infrared at moderate spectral resolution. The spectra are recorded between 240 and 1700 nm and in selected regions of SWIR between 1900 and 2400 nm (with a resolution comprised between 0.2 nm and 1.5 nm), allowing the observation of a very large number of atmospheric trace gases [*Bovensmann et al.*, 1999]. The large wavelength range of

SCIAMACHY is also ideally suited for the determination of aerosols and clouds. The optical system is represented in Figure 4-14.
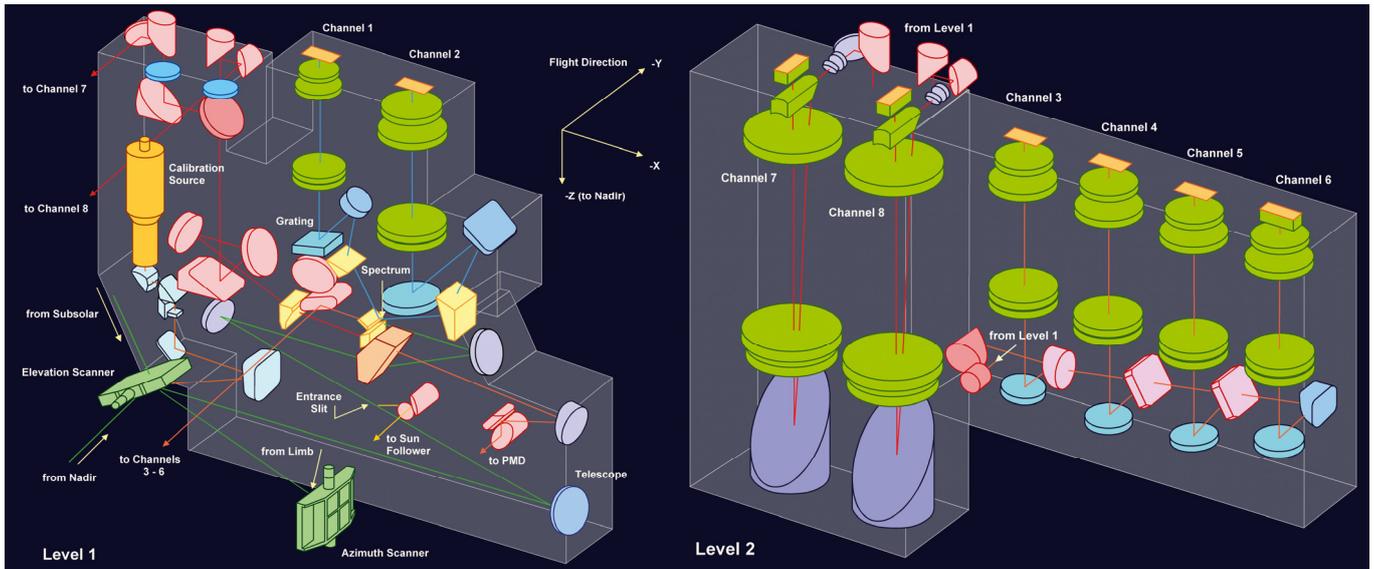
Figure 4-14: SCIAMACHY optical configuration: level 1 and level 2 [DLR-IMF, *Gottwald et al.*, 2006]

The spectrometer contains a pre-disperser prism which separates the light into three spectral bands followed by a series of dichromatic mirrors which further divide the light into a total of eight channels. The optical unit is organised in two levels. Entrance optics, pre-disperser prism, calibration unit and channels 1-2 are contained in level 1. Channels 3-8 are located in level 2. A grating is located in each channel to diffract the light into a high-resolution spectrum which is then focused onto eight detectors. Two types of detectors are used. For the UV-Vis-NIR range covered by channels 1-5, standard silicon photodiodes with 1024 pixels are used. Additionally, UV channels 1 and 2 are electronically divided into two virtual bands 1a/1b and 2a/2b, which can be configured separately. The SWIR channels 6-8 employ InGaAs detectors specifically developed for SCIAMACHY. All channels are cooled to achieve the specified signal-to-noise performance. The operational temperature range is channel dependent and lowest for the SWIR channels.

The pre-disperser is also used to separate polarised light, a part of which being sent to the PMDs. The output of the PMDs is later used to correct for the polarisation effects. The onboard calibration unit consists of two calibration lamps, one for white light and one for spectral lines. During nominal measurements, light enters the instrument via the azimuth (ASM) or elevation (ESM) scan mechanisms. Whilst the ASM captures radiation coming from regions ahead of the spacecraft (limb observation), the ESM either views the ASM or the region directly underneath the spacecraft (nadir observation). For obtaining the solar irradiances, the sun is measured via two aluminium diffusers, one on the backside of the ESM mirror, and one on the backside of the ASM mirror.

The ENVISAT local crossing time at the equator is 10:00 am. As illustrated in Figure 4-15, a special feature of SCIAMACHY is the alternation between limb and nadir measurements, facilitating the measurement of both vertical profiles from the mesosphere down to the upper troposphere at low spatial resolution and the measurement of total columns at high spatial resolution. The limb and nadir scans can be combined to provide tropospheric columns [*Beirle et al.*, 2010]. When possible, SCIAMACHY is also performing solar and lunar occultation measurements. In its nadir view the LoS is limited to an angle of 32° across track. The SCIAMACHY swath width is therefore equivalent to GOME/ERS-2 (960 km). It takes 4 seconds to scan the swath from east to

west (forward), and 1 second in backward scan. Within a ground scene, the individual pixel size is defined by the integration time. Depending on wavelength and solar elevation, it varies from 0.125 to 1 second. In the UV-Vis wavelength range, the nominal integration time is 0.25 second, corresponding to 16 ground pixels of 30x60 km^2 (see Figure 4-16). In this case, the ground resolution is seven times better than GOME, but with an equivalently lower signal to noise ratio due to the shorter integration time. In its limb view, SCIAMACHY performs vertical scanning over an altitude range of 100 km, with a sampling of 3 km. The integration time determines the area covered by the limb observation (from 1060x3 to 230x3 km^2). After each azimuth scan, the elevation is increased until the maximum altitude of 100 km is reached. In nominal operation, limb and nadir viewing is alternated in such a way that an air mass observed during limb observation is a few minutes later observed in nadir. Projected on the Earth surface, this leads to footprints where both limb and nadir are measured, alternated with footprints where no measurements are available. In order to retain a global coverage, the location of gaps and measurements is alternated every orbit. In this way, global coverage is achieved in six days (see Figure 4-13).
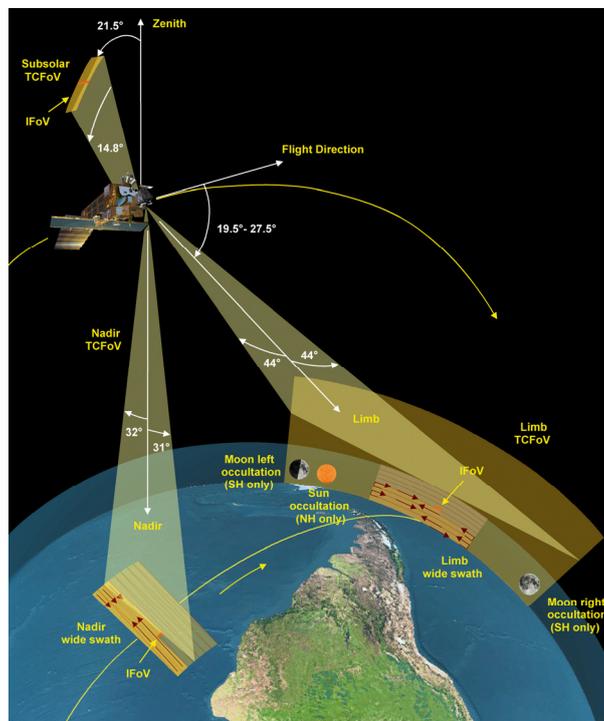
Figure 4-15: Sketch of SCIAMACHY total clear field of view (TC FoV) and observation geometries. The spectrometer entrance slit defines the instantaneous field of view (IFoV), which is about 25x0.6 km² in the nadir mode and 105x2.5 km² in the limb mode [© DLR-IMF, *Gottwald et al.*, 2006].

Figure 4-16: Pattern of nadir ground pixels for integration times of 1 sec (left) and 0.25 sec (right). Only the forward scans are shown [© DLR-IMF].

Table 4-3: Main features of GOME, SCIAMACHY and GOME-2 spectrometers [*Burrows et al.*, 1999b; *Bovensmann et al.*, 1999; *Munro et al.*, 2006].

Instrument	GOME		SCIAMACHY		GOME-2	
Satellite	**ERS-2**		**ENVISAT**		**METOP-A**	
Orbital parameters	**Near polar, sun-synchronous, descending node**					
inclination	98.5°		98.5°		98.7°	
equator local crossing time	10:30am		10:00am		09:30am	
altitude	785 km		800 km		817 km	
orbital period	100 min		100.6 min		101 min	
orbit velocity	7 km/s		7.45 km/s		7.4 km/s	
orbits per day	14		14		14	
repeat cycle	35		35		35	
Instruments	4 channels grating spectrometer		8 channels grating spectrometer		4 channels grating spectrometer	
Channels		Wavelength [nm]	Spectral res. [nm]		Wavelength [nm]	Spectral res. [nm]		Wavelength [nm]	Spectral res. [nm]

Channels		Wavelength [nm]	Spectral res. [nm]		Wavelength [nm]	Spectral res. [nm]		Wavelength [nm]	Spectral res. [nm]
	1	237-316	0.20	1	240-314	0.24	1	240-315	0.24-0.29
	2	311-405	0.17	2	309-405	0.26	2	311-403	0.26-0.28
	3	405-611	0.29	3	394-620	0.44	3	401-600	0.44-0.53
	4	595-793	0.33	4	604-805	0.48	4	590-790	0.44-0.53
				5	785-1050	0.54			
				6	1000-1750	1.48			
				7	1940-2040	0.22			
				8	2265-2380	0.26			

Polarisation measurement device	PMD1	295-397	broadband	PMD1	310-377	broadband	PMDs	312-790	
	PMD2	397-580	broadband	PMD2	450-525	broadband			
	PMD3	580-745	broadband	PMD3	617-705	broadband	2 polarisation channels with 15 programmable bands	2.8@312 to 40@790	
				PMD4	805-900	broadband			
				PMD5	1508-1645	broadband			
				PMD6	2290-2405	broadband			

Nominal nadir resolution	Integration time [s]	Area [km²]	Integration time [s]	Area [km²]	Integration time [s]	Area [km²]
high SZA and band-1a	12	~ 100×960	1	30x240	1.5	40x640
backscan pixel	4.5	40x960	0.5	30x120	0.375	40x160
Forward pixel	1.5	40×320	0.25	30x60	0.1875	40x80
Nominal PMD resolution	0.09375	40×20	0.125	30x30	0.023	40x 10
Scanning time	4.5 s forward + 1.5 s backscan		4 s forward + 1 s backscan		4.5 s forward + 1.5 backscan	
LoS angle	±32°		+32°/-31°		±57°	
Swath width	960 km		960 km		1920 km	
Global coverage (equator)	3 days		6 days		1.5 days	

<div style="background-color:green; padding:10px;">

5   Differential Optical Absorption Spectroscopy

</div>

Satellite-based UV-Visible nadir sensors provide measurements of the top-of-atmosphere (TOA) spectral radiances and of the extraterrestrial solar irradiance. Absorption and scattering properties of the Earth's atmosphere can be determined by comparing the radiance reflected by the surface and scattered by the atmosphere to the solar irradiance entering the top of the atmosphere. To derive quantitative information on atmospheric composition from these measurements, an accurate modelling of the transport of the solar radiation through the atmosphere is required (forward model), as well as mathematical methods to invert the constituent properties from the measured top-of-atmosphere radiance (inversion method). The Differential Optical Absorption Spectroscopy technique (DOAS) is a widely used inversion method in the UV-Vis for the measurement of atmospheric trace gas concentrations.

In the first part of this chapter, the transport of the radiation and its interactions with the atmosphere are introduced. We then concentrate on the principles of absorption spectroscopy and describe in details the method used in this thesis to retrieve tropospheric trace gas columns from nadir satellite measurements using the DOAS approach.

5.1 *Transport of radiation in the atmosphere*
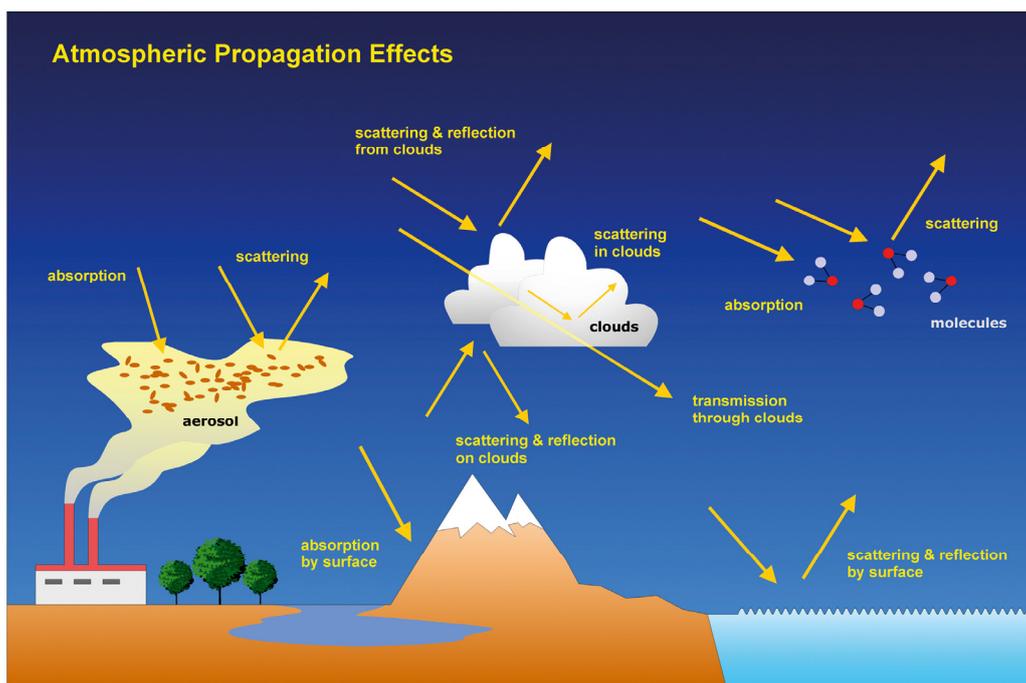
Figure 5-1: Scheme of the relevant interactions of solar light with the Earth's atmosphere and surface in the UV to SWIR spectral ranges [© DLR-IMF, *Gottwald et al.*, 2006].

The retrieval of information on atmospheric trace gases relies on the knowledge of the absorption, emission and scattering of electromagnetic radiation in the atmosphere. From the ultraviolet to the short-wave infrared spectral ranges, absorption-based measurements consist in quantifying the attenuation of radiation in the atmosphere, which is caused by absorption and scattering processes. Indeed, as illustrated in Figure 5-1, the transport of solar radiation through the atmosphere is driven by trace gas absorptions, scattering by air molecules, scattering and absorption by aerosol and cloud

particles, and surface reflection. All these processes must be taken into account when retrieving atmospheric geophysical parameters from satellite measurements. This section describes the basic quantities and the fundamental laws governing radiation transport in the atmosphere.

5.1.1 Absorption

When solar radiation is absorbed by molecules or aerosols, photons are removed from the radiation field and usually re-emitted in the form of radiation at longer wavelength. The absorption and emission of radiation by molecules is accompanied by a change in the internal energy of the molecule.

Absorption lines

Atoms and molecules are quantum mechanical systems which can exist in many different energetic states. Besides the electronic energy levels, molecules present lower energy excitation levels which are not found in atoms: the rotation of the entire molecule and the vibration of atoms relative to each other within the molecule. In addition to the kinetic energy, the energy state of a molecule is a combination of its electronic, vibrational and rotational energies, described by a set of quantum numbers according to the laws of quantum mechanics[1]. Transitions between energy states are governed by selection rules and are associated to the absorption or emission of electromagnetic radiation. The energy of a photon absorbed or emitted by a molecule corresponds to the difference of energy between two particular states i and j of this molecule:

$\Delta E = E_j - E_i = hc\tilde{\nu}$, where $\tilde{\nu} = 1/\lambda$ is the wavenumber of the photon[2]. Depending on the energy of the radiation, its interaction with matter will take different forms (see Figure 5-2):

- Transitions between two rotational states correspond to wavelengths in the sub-millimeter or microwave range. As the rotational energy levels of most molecules are rather closely spaced, those transitions involve relatively low-energy photons.
- Transitions between vibrational states correspond to wavelengths in the infrared spectral range. In addition to vibrational excitation, a molecule is generally also rotationally excited, so that each vibrational transition splits into a series of rotational lines. This coupling gives rise to a group of lines known as the ro-vibrational band.
- UV and visible radiation lead to the excitation of the molecule to higher electronic states. Electronic transitions are usually accompanied by changes in the vibrational and rotational state of the molecule. At atmospheric temperatures, most molecules are in the ground electronic and vibrational state.

The total absorption line spectrum of a molecule is determined by all transitions that are allowed between pairs of energy levels [*Petty*, 2006]. The molecular absorption is usually expressed in terms of an absorption cross-section, which represents the absorption probability of a photon at a certain wavelength [in unit of $cm^2 \cdot molec^{-1}$]. Each molecule is characterized by a unique absorption spectrum, which makes possible selective analysis by spectroscopic techniques.

The natural width of a spectral line depends on the lifetime of the excited state. For atmospheric conditions, the natural width is generally small compared to Doppler and pressure broadening, which are respectively caused by thermal motion and collisions between molecules. In the infrared

---

[1] The vibrational and rotational quantum numbers are denoted *v* and *J*, respectively.
[2] *h*: Planck constant (6.626×10^{-34} J·s), *c*: speed of light (2.998×10^8 m·s^{-1}).

region, a combination of pressure and Doppler broadenings is found in the upper atmosphere, whereas in the lower atmosphere, below about 20 km, pressure broadening prevails because of the higher air density. In the UV-Vis spectral region, pressure and Doppler broadenings are roughly equal (typically 0.001 nm) [*Platt and Stutz*, 2008]. While in the microwave and infrared region it is possible to resolve the individual lines, it is usually not the case in the UV-Vis spectral range because of the complexity of the rovibronic spectrum. Moreover, the typical resolution of grating used in satellite instruments (about 0.2 nm) is too coarse to allow resolving individual lines. As a result, for our application, the absorption cross-section of a molecule in the UV-Vis can be considered as quasi-continuous within a vibrational band.
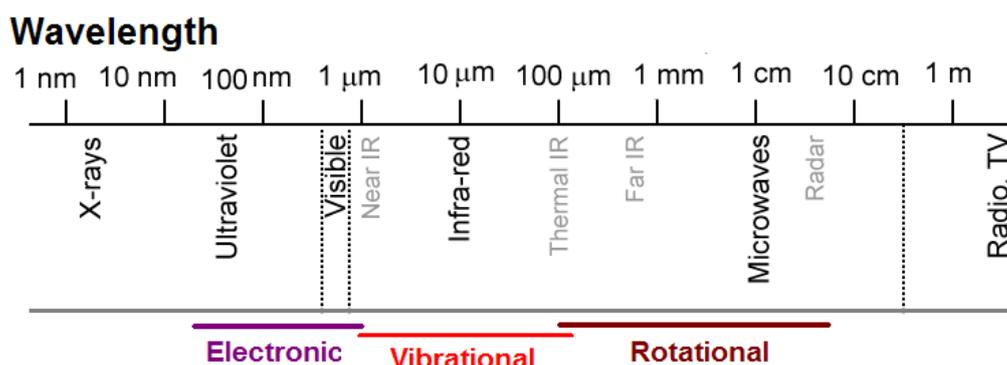
Figure 5-2: Electromagnetic spectrum from radio-waves to X-rays [© *Petty*, 2006].

Formaldehyde absorption cross-section

Figure 5-3 shows the absorption cross-section of formaldehyde in the ultra-violet and thermal infrared (TIR: 1 mm<λ< 3 μm) spectral ranges. In the UV spectral region, formaldehyde presents strong and structured absorption bands between 240 and 360 nm. This corresponds to an electronic transition from the ground state to the first excited state, giving rise to an extended progression of ro-vibrational bands [*Brauers et al.*, 2007]. For satellite remote sensing of formaldehyde in the atmosphere, only the bands above 320 nm can be used, because of the strong increase of ozone absorption below 320 nm [*Chance et al.*, 2000].
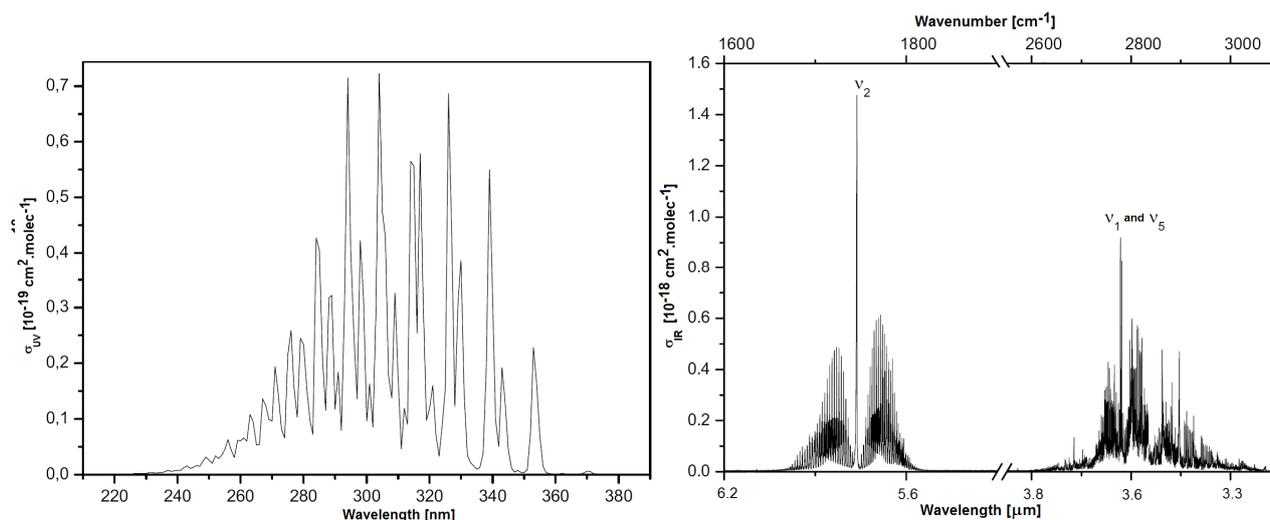
Figure 5-3: Ultra-violet and infrared formaldehyde absorption spectra at 298K [*Meller and Moorgat*, 2000; *Gratien et al.*, 2007]

In the infrared, the main absorption regions of formaldehyde are the 3.6 μm and the 5.7 μm bands [*Gratien et al.*, 2007; *Perrin et al.*, 2009]. At 298K, the 3.6 μm region corresponds to the strong ν_1 and ν_5 vibrational bands of H_2CO. This region is relatively transparent to water vapour absorption and is therefore the best suitable for the detection of tropospheric H_2CO by infrared satellite instruments such as ACE [*Coheur et al.*, 2007; *Dufour et al.*, 2009a]. The 5.7 μm region corresponds principally to the ν_2 band, which appears in a less transparent region of the atmosphere. This band has been used to retrieve concentration profiles of formaldehyde in the upper troposphere and lower stratosphere with the MIPAS instrument [*Steck et al.*, 2008].

5.1.2 Scattering

Scattering is a physical process by which a particle continuously abstracts energy from a given incident electromagnetic wave and re-radiates that energy in all directions. The probability of scattering depends on the number of scattering particles present in the atmosphere, on the size of molecules or particles, and on the wavelength of the incident radiation. In the atmosphere, the incident photon can be scattered only once (single scattering) or several times (multiple scattering). Multiple scattering is an important process for the transfer of radiant energy in the lowermost atmosphere, especially when aerosols and clouds are involved.

Three different types of scattering are commonly defined: Rayleigh, Mie and Raman scattering. The different regimes of scattering can be distinguished depending on the ratio of the size of the scattering particle (particle radius *r*) to the radiation wavelength $\lambda$, the so-called size parameter *$x=2\pi r/\lambda$* (see the left panel of Figure 5-4). The size parameter ranges widely for different common atmospheric constituents. In the UV-Vis region, it is much less than 1 for gas molecules, about 1 for haze and smoke particles and much larger for raindrops and ice crystals. Rayleigh and Raman scattering of solar radiation occurs on atmospheric molecules (*x<<1*), while Mie scattering occur on large particles such as aerosols and droplets (*x≥1*).
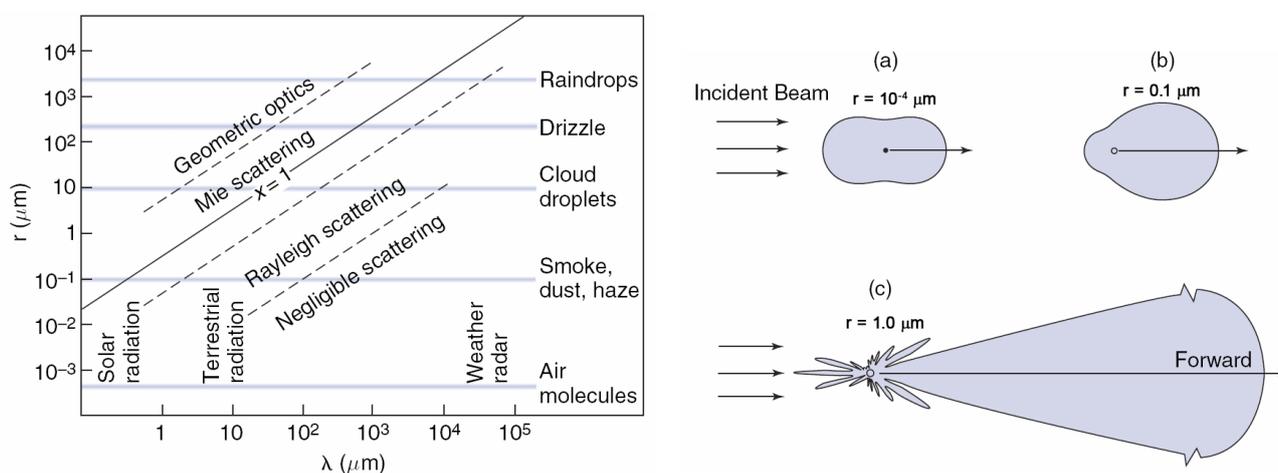
Figure 5-4: Left panel: scattering regimes as a function of the wavelength of the incident radiation and of the size of the scattering molecule or particle. Right panel: scattering phase functions of spherical particles of radius (a) 10^{-4} μm, (b) 0.1 μm, and (c) 1 μm, illuminated by visible light of 0.5 μm. The forward scattering pattern of case (c) is extremely large and has been scaled for the figure [*Liou*, 2002; *Petty*, 2006].

When the scattered photon has the same wavelength as the original photon, the scattering is said elastic. In contrast, when the scattered photon has lower or higher energy than the incident photon, one talks about inelastic scattering. To a very large extent Rayleigh and Mie scattering are elastic

processes. Raman scattering is the part of molecular scattering that is inelastic (about 4% of all molecular scattering processes).

Rayleigh scattering

Rayleigh scattering occurs on particles that are small relative to the wavelength of incoming radiation. In practice, Rayleigh scattering involves the scattering of UV-visible light by molecules having a radius of up to a tenth of the radiation wavelength. The probability of scattering can be described with a scattering cross-section. The Rayleigh scattering cross-section is inversely proportional to the fourth power of the wavelength: $\sigma_R(\lambda) \approx \lambda^{-4}$. As a result, in the atmosphere, ultraviolet radiation is usually scattered before reaching the surface. The phase function gives the distribution of the scattered intensity as a function of the scattering angle Θ: $P(\Theta) = 3/4 \cdot (1 + \cos^2 \Theta)$. The angular distribution of Rayleigh scattering is symmetrical in the plane normal to the incident radiation, so that the forward scatter equals the backward scatter (see the right panel of Figure 5-4, case (a)) [*Liou*, 2002].

Mie scattering

Mie scattering occurs on particles with dimensions comparable or larger than the wavelength of the incident radiation, *i.e* large particles such as aerosols, droplets and suspended matter in liquids. Mie scattering is elastic. The most noticeable difference compared to Rayleigh scattering is the typically much weaker wavelength dependence of the cross-section and a strong dominance of the forward direction in the scattered light. The wavelength dependence of the Mie scattering cross-section can be approximated by: $\sigma_M(\lambda) \cong \lambda^{-\alpha}$ with the Angström exponent α being inversely related to the mean particle radius. Typically, α varies from 0 to 2. The angular distribution of Mie scattering is a complex function of particle size, but it is strongly dependent on the size parameter x, the forward peak being stronger for larger particles (right panel of Figure 5-4, cases (b) and (c)) [*Liou*, 2002].

Raman scattering
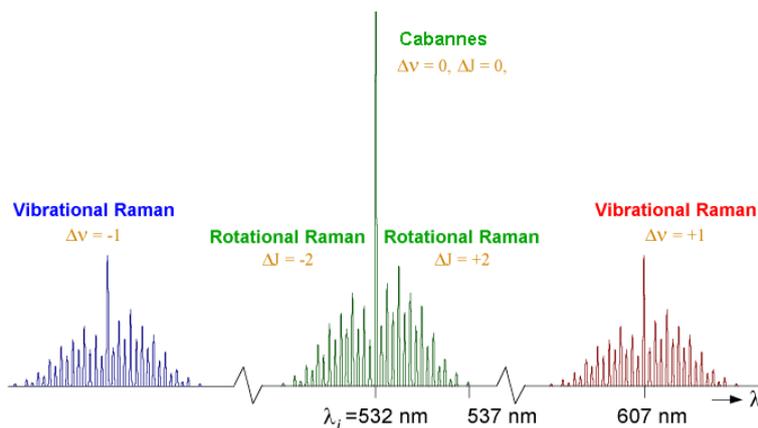
Figure 5-5: Sketch of a typical rotational Raman spectrum for an incident beam at 532 nm on a molecule of N_2 [© http://www.phys.tue.nl].

Raman scattering refers to the inelastic contribution of molecular scattering. Atmospheric Raman scattering mainly involves N_2 and O_2. Its cross-section is proportional to λ^{-4} and its phase function is almost isotropic. Figure 5-5 illustrates the scattering cross-section of N_2 around 532 nm. Molecular

scattering is a two-steps absorption-emission process involving rotational transitions such as $\Delta J = 0,$
± 2. Elastic scattering corresponds to transitions such as $\Delta v=0$ and $\Delta J=0$ (Cabannes scattering, ~96%
of molecular scattering), while inelastic scattering involves transitions such as $\Delta v=0, \pm 1$ and $\Delta J =\pm 2$
(Raman scattering, ~4% of molecular scattering). The term rotational Raman scattering (RRS) is
used if the rotational excitation state is the only energy level to be affected. If the vibrational state
also changes, the term vibrational Raman scattering (VRS) is used. In the air, nearly all inelastically
scattered light is due to rotational Raman scattering. Vibrational Raman scattering by air molecules
is an order of magnitude smaller. VRS is mainly significant in water. In the air, RRS gives rise to
spectral shifts of a few nm, while VRS gives rise to larger wavelength shifts (up to $\pm$ 50 nm).

5.1.3 Reflection at the surface

When the direct or diffused solar radiation reaches the surface, a fraction of it is absorbed, whereas
the other part is reflected in various directions and contributes to the upwelling radiation. The
fraction of incident radiation that is reflected by the surface is called reflectivity or albedo. The
planetary albedo is the average ratio of reflected light from the Earth, including clouds
(approximately 30%). The reflectivity of the ground depends on the nature of the surface. It
depends also on the wavelength and on the direction of the incident radiation. Figure 5-6 displays
examples of spectral reflectivity for a few natural surface types. Fresh snow has a very high
reflectivity (85-95%) in the visible region, whereas the reflectivity of fresh vegetation, soil, and
bodies of water is generally much lower (<15%). There may be relatively little relation between the
reflectivity of a surface type in the visible part of the spectrum and that in the near IR band.
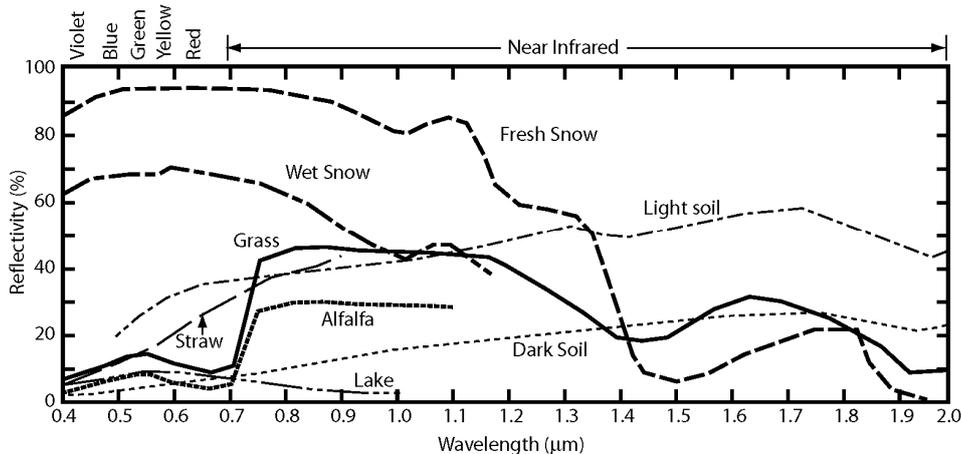
Figure 5-6: Typical shortwave reflectivity of various surface types as a function of wavelength [© *Petty*, 2006].

Reflectance of light from terrestrial homogeneous surface is generally anisotropic and is described
by its bidirectional reflectance distribution function (BRDF), which depends on the surface
properties and on the direction of the incident radiation. Depending on the given viewing and
illumination geometry, surfaces may appear brighter or darker. In practice however, a widely used
approximation for spaceborne nadir observations is to assume that the surface reflectivity is
isotropic. This corresponds to the Lambertian reflection approximation [*Koelemeijer et al.*, 2003].
This approximation is justified for satellites flying on sun-synchronous polar orbits whose
instruments always sample a given region at the same local time, so that the solar zenith angle
slowly varies with the season. Moreover, the albedo datasets used in satellite applications are
generally derived from instruments flying on the same platform, for the same solar zenith angle.

However, in principle and for more refined evaluations, the directionality of the surface reflectance distribution can be considered [*Zhou et al.*, 2010].

5.1.4 The radiative transfer equation

In this section, we present the basic equations describing the radiation transport in absorbing and scattering media. We will first consider the change in intensity radiation while it traverses an infinitesimal distance and then extend the formalism to the whole atmosphere. The propagation of radiation in the Earth atmosphere is a complex process. This is mostly due to the fact that scattered light can propagate in any direction and can undergo further scattering and absorption processes. Both absorption and scattering by molecules and particles lead to the extinction of radiation traversing an atmospheric layer. In addition, one should also consider source terms enhancing the radiation intensity, *i.e.* the radiation reflected and scattered from other directions into the direction of propagation, plus eventual thermal emission from the Earth-atmosphere system.

Basic quantities related to radiation transport

The extinction cross-section of a molecule (or particle) is the sum of its absorption and scattering cross-sections [$cm^2.molec^{-1}$]. When multiplied by the number density, it is referred to as the extinction coefficient [cm^{-1}]. The combined losses by absorption and scattering are thus described by the extinction coefficient: $k(\lambda) = k_a(\lambda) + k_s(\lambda)$.

Several quantities are used to describe the radiation field, *i.e* the angular and spatial distribution of the radiation in the atmosphere [*Liou*, 2002]:

-   The radiant flux $\boldsymbol{\Phi}$ is the energy of radiation per unit time, regardless the direction of its emission [$J \cdot s^{-1}$, W].
-   The irradiance (or flux density) $\boldsymbol{F}$ is the radiant flux $\boldsymbol{\Phi}$ received by an illuminated area $\boldsymbol{A_e}$: $\boldsymbol{F = \Phi / A_e}$ [$W \cdot m^{-2}$].
-   The radiance (or intensity) $\boldsymbol{I}$ is the radiant flux $\boldsymbol{\Phi}$ emitted by the radiating area $\boldsymbol{A_e}$ in the solid angle $\boldsymbol{\Omega}$: $\boldsymbol{I = \Phi / (A_e \cdot \Omega)}$ [$W \cdot m^{-2} \cdot sr^{-1}$].

Variation of radiation intensity over an infinitesimal path
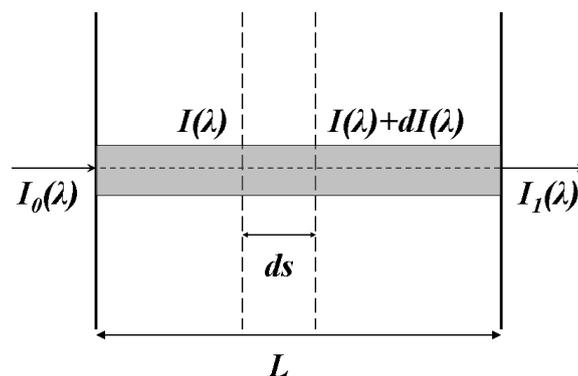
Figure 5-7: Sketch of the radiation traversing an absorbing and/or scattering medium over an infinitesimal path.

Let us consider the radiance $\boldsymbol{I(\lambda)}$ [$W \cdot m^{-2} \cdot sr^{-1}$] along its direction of propagation, traversing a medium of thickness $\boldsymbol{ds}$, small enough so that the extinction coefficient $\boldsymbol{k(\lambda,s)}$ is constant within the

interval. The intensity attenuation is proportional to $k(\lambda,s)$, ds and $I(\lambda)$. The intensity of the radiation may also be strengthened by light emitted from the material, as well as by scattering from all other directions into the direction under consideration. By analogy with the extinction coefficient, the source function coefficient $j(\lambda)$ is defined as the increase in intensity due to emission and scattering. The change in the radiation intensity is then given by:

$$dI(\lambda) = I(\lambda, s+ds) - I(\lambda, s) = -k(\lambda, s)I(\lambda, s)ds + j(\lambda, s)ds \tag{5-1}$$

Moreover, it is convenient to define the source function $J(\lambda,s)$ such that $J(\lambda,s) \equiv j(\lambda,s)/k(\lambda,s)$. In this manner, the source function has units of radiant intensity. It follows that (5-1) becomes:

$$\frac{dI(\lambda)}{k(\lambda,s)ds} = -I(\lambda,s) + J(\lambda,s) \tag{5-2}$$

This is the general form of the energy-conservative radiative transfer equation. It includes all processes affecting the radiation field as a result of its interaction with a medium, without any coordinate system imposed.

Simplification: the Beer-Lambert's law

Frequently, simplifications of the radiative transfer equation are possible. For instance, for wavelengths below 3 µm, the contribution due to thermal emissions can generally be neglected. Moreover, if the radiance added by scattering is also neglected, equation (5-2) reduces to:

$$\frac{dI(\lambda)}{k(\lambda,s)ds} = -I(\lambda,s) \tag{5-3}$$

The emergent intensity at a distance L ($I_1(\lambda)$), can be obtained by integrating (5-3) over the light path:

$$I_1(\lambda) = I_0(\lambda)\exp(-\int_0^L k(\lambda,s)ds) \tag{5-4}$$

$I_0(\lambda)$ being the incident intensity at $s=0$ (see Figure 5-7). The optical density (or optical depth) of the layer is defined as:

$$\tau(\lambda,L) = \int_0^L k(\lambda,s)\cdot ds \tag{5-5}$$

If the medium is homogeneous, $k(\lambda,s)$ is independent of the distance, and (5-4) can be expressed by:

$$I_1(\lambda) = I_0(\lambda)e^{-k(\lambda)L} \tag{5-6}$$

This is the well known Beer-Lambert's law, which states that the decrease in the radiant intensity traversing a homogeneous extinction medium follows the simple exponential function whose argument is the product of the extinction coefficient and the path length. It is the basis of many present quantitative trace gas analytical retrieval methods.

Plane parallel approximation of the Earth atmosphere

In general, a planetary atmosphere is inhomogeneous in both vertical and horizontal directions. However, for many atmospheric radiative transfer applications, it is physically appropriate to consider that the atmosphere is plan-parallel in localized portions such that variations in the intensity and atmospheric parameters (density, temperature, and gaseous profiles) are considered only in the vertical direction. The radiation field will then depend on a single vertical coordinate (the altitude z or the vertical optical depth τ) and the two directional variables μ (cosine of the zenith angle θ) and φ (azimuth angle with respect to a given direction) (see Figure 5-8). Equation (5-2) describes the transmission of radiation over a finite distance along an arbitrary path in the propagation direction of the light. In the vertical coordinates of the plane-parallel atmosphere, the radiative transfer equation (RTE) for the diffuse radiance $I(\tau,\Omega)$ in the direction $\Omega=(\mu,\varphi)$ and at optical density $\tau(z)$ becomes [*Chandrasekhar*, 1960; *Liou*, 2002]:

$$\mu \frac{dI(\tau,\Omega)}{d\tau} = -I(\tau,\Omega) + J(\tau,\Omega) \qquad (5\text{-}7)$$

For simplicity, the wavelength dependence has been omitted. Equation (5-7) is thus valid for a monochromatic intensity field, and inelastic scattering is neglected here.
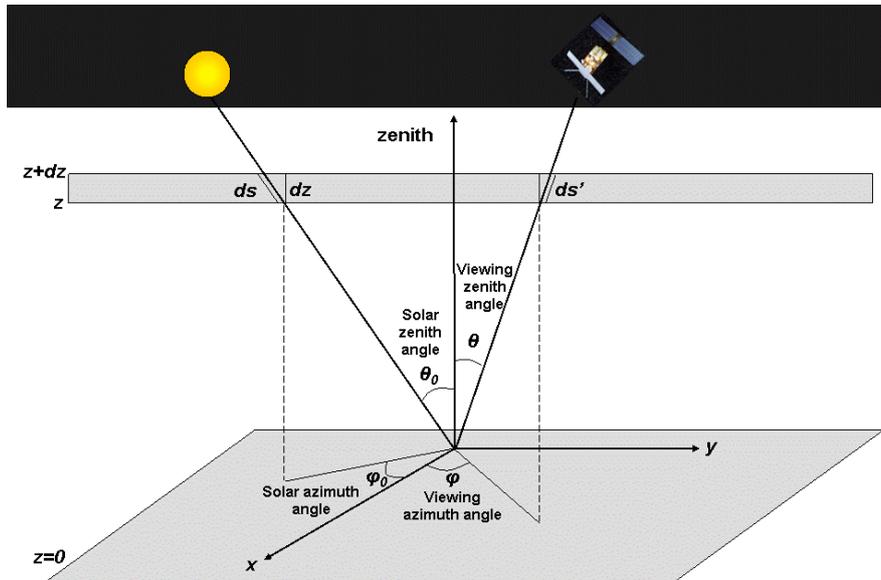
Figure 5-8: Sketch of satellite radiation measurement and geometry in a plane parallel atmosphere.

The source function $J(\tau,\Omega)$ for the surface-atmosphere system is given by Equation (5-8) and composed by [*Liou*, 2002]:

1. The term of scattered light from direct solar beam (5-8 a): the gain of radiation by single scattering of sunlight is proportional to the amount of incident solar radiation, attenuated from TOA to the altitude z, and by the probability of scattering from the direction Ω_0 to the direction Ω.

2. The scattered diffuse light term (5-8 b): the gain of radiation by multiple scattering is given by the amount of light coming from a given direction Ω', already scattered at least once, multiplied by the scattering probability into the direction of interest Ω, and integrated over all possible directions.

3. Reflected light from the surface (5-8 c): The gain of radiation by reflection of light from the surface is proportional to the incident solar radiation (attenuated to the bottom of atmosphere) and to the surface albedo a. The reflected light propagates (and is attenuated by $e^{-(\tau_0-\tau)/\mu'}$) from the surface to $\tau(z)$ over all directions $\boldsymbol{\Omega'}$. The gain of radiation is given by the probability for this reflected light to be scattered in the direction $\boldsymbol{\Omega}$.

4. The thermal emission contribution (5-8 d).

$$J(\tau,\Omega) = \frac{\omega(\tau)}{4\pi} P(\tau,\Omega,\Omega_0)F_0 e^{-\tau/\mu_0} \tag{5-8 a}$$

$$+ \frac{\omega(\tau)}{4\pi} \int_{4\pi} P(\tau,\Omega,\Omega')I(\tau,\Omega')d\Omega' \tag{5-8 b}$$

$$+ \frac{\omega(\tau)}{4\pi} F_0 e^{-\tau_0/\mu_0} \int_{2\pi} P(\tau,\Omega')ae^{-(\tau_0-\tau)/\mu'}d\Omega' \tag{5-8 c}$$

$$+ (1-\omega(\tau))B(T(\tau)) \tag{5-8 d}$$

- F_0 is the direct solar irradiance[3].
- ω is the single scattering albedo, defined as $\omega = k_s/(k_a + k_s)$.
- $P(\Omega,\Omega')$ is the scattering phase function from the direction (μ,φ) to the direction (μ',φ')[4].
- B is the Planck radiation at temperature T (which is a function of $\tau(z)$) and wavelength λ[5].
- τ_0 is the total vertical optical density of the atmosphere[6].

The radiative transfer equation as formulated by (5-7) and (5-8) is a multi-dimensional integro-differential equation which depends on several atmospheric and optical variables. A complete simulation of the solar light transmission requires considering, at each altitude, the loss of radiation by absorption and scattering and the internal atmospheric contribution to the radiation which is due to single and multiple scattering (the contribution of atmospheric emissions can be neglected in the UV-Vis spectral range). The plane-parallel (or stratification) assumption allows RTE solutions to be determined in each layer. The complete radiation field is then put together by means of a set of boundary conditions. In particular, the vertical optical density at top-of-atmosphere (τ_{TOA}) is 0, while it is τ_0 at the surface [6]. The total radiation field measured at TOA can be split into the direct radiation, which is has not been scattered in the atmosphere, and the diffuse radiation, which has been scattered at least once. The solution of the radiative transfer equation for the direct part can be easily derived:

$$I_{dir}(\tau_{TOA},\Omega) = I_0(\Omega_0).a.e^{-(\frac{\tau_0}{\mu_0}+\frac{\tau_0}{\mu})} \tag{5-9}$$

---

[3] The solar irradiance [W·m^{-2}] is the solar radiance component perpendicular to a unit area integrated over an entire hemisphere: $F_0 = \int_\Omega I_0\cos\theta_0 d\Omega = \int_0^{2\pi}d\varphi\int_0^{\pi/2}I_0\cos\theta_0\sin\theta_0 d\theta = \pi I_0$

[4] The scattering angle Θ_s is given by the relation: $\cos\Theta_s = -\mu\mu'+\sqrt{(1-\mu^2)(1-\mu'^2)}\cos(\varphi-\varphi')$.

[5] $B(\lambda,T) = \frac{2hc^2}{\lambda^5(e^{hc/kT\lambda}-1)}$

[6] $\tau(z) = \int_z^{TOA}(k_a(z')+k_s(z'))dz'$

The solution of the diffuse part must take into account single and multiple scattering at all altitudes. In general, solutions have to be calculated by numerical methods.

There are a number of different methods of solving RTE in a scattering atmosphere. They can be divided into two groups [*Liou*, 2002]. First, the radiation field can be derived from Monte Carlo models that simulate physical processes on individual photons using probability functions. The paths of an ensemble of photons (with randomly chosen parameters) through the atmosphere are calculated for the desired conditions and the radiation field can be derived from the photon density. These models have the advantage of a transparent implementation of the physical processes and a true 3-dimentional treatment, but they are very slow. The second type of radiative transfer models compute numerical solutions of the RTE using discretization methods. The implementation of such models is often complex and necessarily based on approximations, but they have the great advantage to be fast. State-of-the-art radiative transfer models are not only able to simulate radiances, but they can also deliver the derivative of the modelled radiance with respect to selected parameters. Such derivations are called weighting functions. They describe the sensitivity of the radiance to a particular parameter and they are an essential input for inversion methods.

In this thesis, we have used the radiative transfer model LIDORT v3.0 (Linearized Discrete Ordinate Radiative Transfer) [*Spurr*, 2008]. LIDORT is an advanced multiple-scattering discrete code which allows for the calculation of both radiances and weighting functions. Since, the plane-parallel assumption becomes invalid for large solar zenith angles (SZA >75°) and for wide off-nadir viewing angles, the model uses the pseudo-spherical approximation which treats the direct light attenuation in a curved atmosphere while the plane-parallel approximation is maintained for the multiple scattering terms. This approach has the advantages of the plane-parallel formalism (flexibility and speed) while providing sufficiently accurate solution for large solar zenith angles and moderate viewing angles (up to 30°). For larger viewing angles, a dedicated sphericity correction is applied for scattering along the line-of-sight [*Spurr*, 2008].

5.2 *Differential optical absorption spectroscopy*

Differential Optical Absorption Spectroscopy (DOAS) is a widely used inversion method for the retrieval of atmospheric trace gas abundances from multi-wavelength light measurements. The basic idea of DOAS is to make use of the structured absorption of many trace gases in the UV, visible and near-infrared spectral ranges. The DOAS method was originally developed for ground-based measurements [*Platt et al.*, 1994; *Platt and Stutz*, 2008; and references therein] and has been successfully adapted to measurements from UV-Vis spaceborne spectrometers [*Gottwald et al.*, 2006]. Here, we restrict the discussion to the retrieval of atmospheric trace gas abundances from spaceborne nadir measurements.

The DOAS method consists of two independent steps. Firstly, molecular absorption cross-sections are fitted to the logarithm of the ratio of the measured radiance and direct solar irradiance without atmospheric absorptions. The resulting fit coefficients are the integrated number of molecules per unit area along the atmospheric light path for each trace gas, also called slant column density (*SC*). The slant column depends on the observation geometry, the position of the sun and also on parameters such as the presence of clouds, aerosol load and surface reflectance. Therefore, in a second step, the slant column is converted to a vertical column (*VC*) through division by an air mass factor (*AMF*) which is computed with a radiative transfer model and accounts for the average light path through the atmosphere.

5.2.1 Solar irradiance and Earth backscattered radiance

Spaceborne UV-Vis sensors measure the sunlight scattered by the atmosphere and/or reflected by the Earth surface. They also measure the direct solar irradiance, which is free of atmospheric spectral signatures. For satellite instruments such as GOME (1 and 2) and SCIAMACHY, earthshine spectra are measured for successive ground pixels across the track of the satellite, the scanning being performed by mirrors. The solar irradiance ($F_0 = \pi I_0$) is recorded once a day, using a diffuser plate (see chapter 4). The left panel of Figure 5-9 illustrates the measurement geometry of solar and backscattered radiation by nadir spaceborne spectrometers, while the right panel shows an example of solar irradiance (F_0) and earthshine radiance (I) spectra measured by SCIAMACHY between 250 and 800 nm. Both spectra show strong absorption features, most of them being Fraunhofer lines caused by absorption occurring in the solar photosphere. They are the dominant spectral features in the UV-Vis region (above 300 nm). To detect small trace gases absorptions, the Fraunhofer lines must be accurately removed (see next section). Note that several strong atmospheric absorption features are apparent in the earthshine spectrum, such as ozone below 320 nm, oxygen at 685 and 765 nm, and water vapour at 725 nm. Note also that compared to the solar irradiance, the extinction of the earthshine spectrum is more pronounced in the visible and NIR spectral regions than in the UV. This is explained by the source function due to Rayleigh scattering, which is more efficient in the UV.
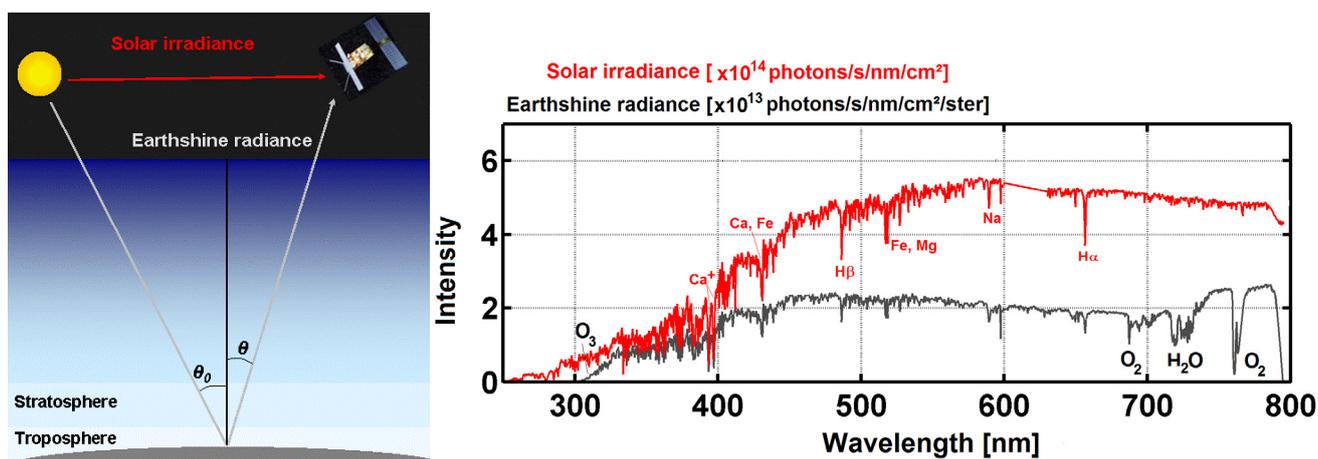
Figure 5-9: Left panel: Measurement geometry of direct solar irradiance and Earth backscattered radiance by nadir spaceborne spectrometers (the figure is not to scale). Right panel: Example of solar irradiance and earthshine radiance spectra measured by SCIAMACHY in channels 1 to 4 [both figures are adapted from http://www.iup.uni-bremen.de/E-Learning/at2-els_NO2/index.htm].

5.2.2 General principle of DOAS for nadir satellite measurements

As illustrated by the left panel of Figure 5-9, the backscattered earthshine spectrum as measured by the satellite spectrometer is a result of the complex solar spectrum with additional features due to interactions in the atmosphere during the incoming and outgoing paths of the radiation. The DOAS method relies on a simplification of the radiative transfer equation, leading to the application of the Beer-Lambert law to the whole atmosphere. The key idea of the DOAS method is to separate broad and narrowband spectral structures of the absorption spectra in order to isolate the narrow trace gas absorption features. In order to do this, the following approximations are made:

1. One considers the mean path followed by the photons through the atmosphere up to the instrument. This allows expressing light attenuation in term of the Beer-Lambert's law and separating spectroscopic retrievals from radiative transfer calculations.

2. The absorption cross-sections are supposed to be independent of temperature and pressure, which allows introducing the concept of slant column densities.

3. Broadband variations are approximated by a common low-order polynomial to compensate for the effects of loss and gain from scattering and reflections by clouds and/or at the Earth surface.

Simplification of the light path

Photons collected by the satellite instrument may have followed very different light paths through the atmosphere depending on their scattering history (see Figure 5-1). However, for most situations we can assume a single "effective" light path which represents an average over a limited wavelength interval of the complex paths of scattered photons in the atmosphere. This simplification is valid if the effective light path is reasonably constant over the considered wavelength range, *i.e* if the atmosphere can be considered as optically thin over the wavelength interval. The length of the effective light path is dealt with separately using radiative transfer calculations.

By expanding the Beer-Lambert's law, one can consider the various factors that influence the light intensity at the top-of-atmosphere by an equation that includes the effect of all absorbers present over the light path as well as Rayleigh and Mie scattering particles. The starting point of DOAS retrieval is the following expression:

$$I(\lambda, \mu_0) = I_0(\lambda) \cdot \exp\left(-\int_{Atm} k(\lambda, s) \cdot ds\right) \cdot A(\lambda, \mu_0) \tag{5-10}$$

$I(\lambda, \mu_0)$ and $I_0(\lambda)$ are respectively the earthshine and solar intensity spectra. The factor $A(\lambda, \mu_0)$ accounts for the fact that the overall intensity of the observed scattered light depends on the scattering efficiency of the Earth/atmosphere system. It takes also into account instrumental effects such as wavelength dependence output. The measured intensity depends on the wavelength and on the extinction coefficient $k(\lambda, s)$. The integral is taken along the effective path followed by the photons through the atmosphere.

The next step is the normalisation of the earthshine spectrum with the solar spectrum in order to remove the Fraunhofer lines. The relation between I and I_0 can be written as:

$$\frac{I(\lambda, \mu)}{I_0(\lambda)} = \exp\left(-\int\left(\sum_j \sigma_j(\lambda, s) \cdot n_j(s) + \sigma_R(\lambda, s) \cdot n_R(s) + \sigma_M(\lambda, s) \cdot n_M(s)\right) \cdot ds\right) \cdot A(\lambda, \mu) \tag{5-11}$$

The summation is performed over all relevant species with absorption cross-sections σ_j and densities n_j. σ_R and σ_M are the Rayleigh and Mie extinction cross-sections, while n_R and n_M are respectively the Rayleigh and Mie scattering particle densities. The densities of the absorbers and scattering particles vary with altitude along the light path. To a lesser extent, it is in principle also the case of the absorption cross-sections because of their dependence on temperature and pressure.

Substitution of slant column densities

The notion of slant column density (SC_j) is defined as the particle number density n_j integrated along the light path s:

$$SC_j = \int n_j(s)ds \tag{5-12}$$

If we now assume that the absorption and scattering cross-sections σ do not depend on s, then the sum over the individual terms and the integral can be exchanged and equation (5-11) can be expressed in terms of slant column densities:

$$\frac{I(\lambda,\mu)}{I_0(\lambda)} = \exp\left(-\left(\sum_j \sigma_j(\lambda)\cdot SC_j + \sigma_R(\lambda)\cdot SC_R + \sigma_M(\lambda)\cdot SC_M\right)\right)\cdot A(\lambda,\mu) \tag{5-13}$$

This assumption is justified by the fact that the pressure dependence of absorption cross-sections can in most cases be neglected in the UV-Vis region. However, the temperature dependence of absorption cross-sections can be significant (for example that of O_3 and NO_2). This can be corrected in first approximation by introducing correction factors during the *AMF* calculation [*Boersma et al.*, 2004] or by fitting two absorptions cross-sections at different temperatures assuming linear dependence on temperature [*Van Roozendael et al.*, 2002b].

Substitution of a scattering polynomial

In the DOAS technique, we use the fact that molecular absorption cross-sections $\sigma_j(\lambda)$ are usually highly spectrally structured, while scattering by molecules and particles, as well as reflection at the surface, have broadband dependences that can be approximated by a low order polynomial:

$$A(\lambda,\mu) = \sum_p a_p \lambda^p \tag{5-14}$$

$$\sigma_R(\lambda)\cdot SC_R + \sigma_M(\lambda)\cdot SC_M = \sum_p b_p \lambda^p \tag{5-15}$$

To obtain the final DOAS equation, the logarithm of (5-13) is taken and all broadband variations are included in a single polynomial:

$$\ln I(\lambda,\mu) - \ln I_0(\lambda) = -\sum_j \sigma'_j(\lambda)\cdot SC_j + \sum_p c_p \lambda^p$$
$$\tag{5-16}$$
$$\tau_{meas}(\lambda) = \tau'(\lambda) + \tau_p(\lambda)$$

In this way, the measured optical density τ_{meas} is separated into its highly structured part τ' and its broadband variation τ_p. The separation into high frequency and low frequency parts of the signal is the "differential" part of DOAS. Note that the absorption cross-sections σ_j are also separated into the so-called "differential absorption cross-sections" σ'_j, which represent the high-pass filtered parts of the absorption cross-sections, and the "low frequency absorption cross-sections" $\sigma_{jp}(\lambda)$:

$$\sigma_j(\lambda) = \sigma_{jp}(\lambda) + \sigma_j'(\lambda) \tag{5-17}$$

Figure 5-10 illustrates the separation of the narrow- and broadband structures, both for the O_3 absorption cross-section and the optical density. The meaning of narrow and broad wavelength variation of the absorption cross-section is of course dependant on the wavelength interval used and

on the width of the absorption bands to be detected. Often, the polynomial is used to perform the entire high pass filtering essential for the DOAS technique and the non high-pass filtered cross-sections σ_j can be used in the fitting process instead of the differential ones.

Equation (5-16) is a linear equation between the logarithm of the measured quantities (**I** and **I$_0$**), the slant column densities of all relevant absorbers (**SC$_j$**) and a low order polynomial with coefficients **c_p**. The separation of broad and narrow spectral structures and the separation of the various absorbers require the measurement of the radiation intensity at multiple wavelengths. DOAS retrievals consist in solving an over-determined set of linear equations, which can be done by standard methods of linear least squares fit [*Platt and Stutz*, 2008].
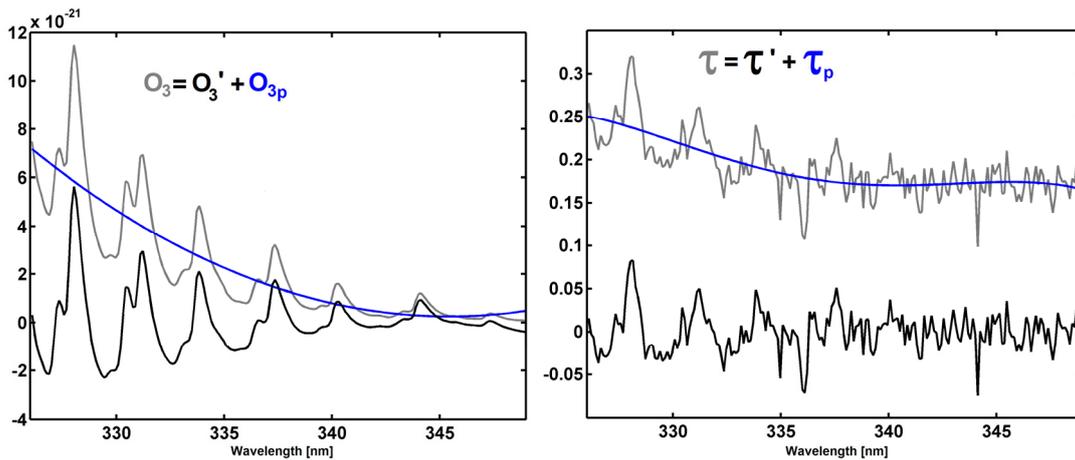
Figure 5-10: DOAS principle: σ and τ are separated by an adequate filtering procedure into a narrow (σ' and τ') and a broad band part (σ_p and τ_p).

5.2.3 Choice of the wavelength interval and trace gases measured by DOAS

DOAS measurements are in principle applicable to all gases having suitable narrow absorption bands in the UV, visible, or near IR regions. However, the generally low concentrations of these compounds in the atmosphere, and the limited signal-to-noise ratio of the spectrometers, restrict the number of trace gases that can be detected. Figure 5-11 shows the absorption cross-sections of a number of trace gases that have been regularly measured by DOAS. Many spectral regions contain a large number of interfering absorbers. In principle, because of their unique spectral structure, a separation of the absorption is possible. However, correlations between absorber cross-sections can sometimes lead to systematic biases in the retrieved slant columns. In general, the correlation between cross-sections decrease if the wavelength interval is extended, but the assumption of a single effective light path defined for the entire wavelength interval may not be fully satisfied, leading to systematic misfit effects that may also introduce biases in the retrieved slant columns. To optimize DOAS retrieval settings, a trade-off has to be found between these effects. A basic limitation of the classical DOAS technique is the assumption that the atmosphere is optically thin in the wavelength region of interest. At shorter wavelengths, the usable spectral range of DOAS is limited by rapidly increasing Rayleigh scattering and O_3 absorption. In addition, 'line-absorbers' such as H_2O, O_2, CO, CO_2 and CH_4 usually cannot be retrieved precisely by standard DOAS algorithms because their strong absorption also depends on pressure and temperature [*Gottwald et al.*, 2006].
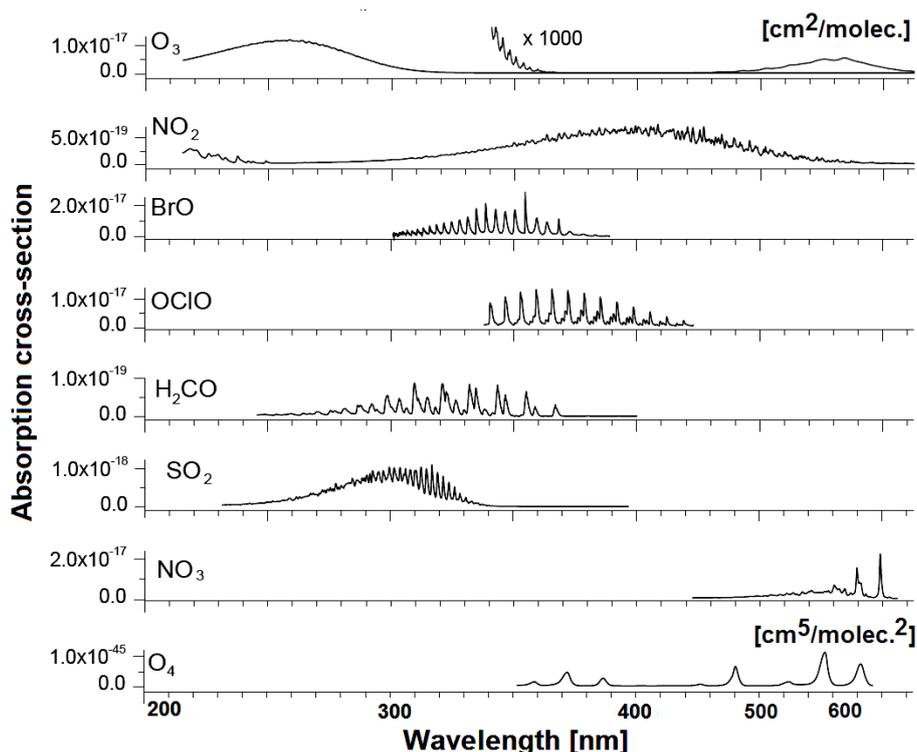
Figure 5-11: Absorption cross-sections of various absorbers retrieved with the DOAS method [*Platt and Stutz*, 2008].

5.2.4 Practical aspects of DOAS retrieval

The spectral fitting of the slant columns is generally a challenge for weak absorbers as very tiny spectral structures (optical density of a few 10^{-4}) need to be quantitatively separated from interfering spectral structures of larger amplitude. In order to optimize the fitting procedure, spectral calibration, knowledge of the instrumental slit function and structured spectral effects (like Ring or solar I_0 effects) have to be considered carefully. Furthermore, the linearity of Equation (5-16) may be broken down by instrumental aspects such as straylight and small wavelength shifts.

In the next chapter, we will present the work performed during this thesis to optimize the retrieval of formaldehyde slant columns, which is made difficult by the relatively low H_2CO optical depth, spectral interferences with O_3 and BrO and some particular instrumental effects. The algorithms for the retrieval of H_2CO slant columns were developed using the QDOAS software, a multi-purpose DOAS analysis software developed at BIRA-IASB [*Fayt and Van Roozendael*, 2001; *Van Roozendael et al.*, 2002a]. Here, we focus on general aspects that need to be dealt with in order to optimise the quality of the DOAS fits.

Wavelength Calibration and instrumental slit function

Fraunhofer absorption lines must be removed from the differential **lnI(λ)-lnI$_0$(λ)** spectrum to avoid interference in the retrieval. Perfect alignment of the two spectra is therefore needed. The wavelength calibration procedure is based on the alignment of the Fraunhofer structures of *I$_0$(λ)* with those of an accurately calibrated high-resolution solar reference atlas, degraded to the resolution of the instrument, *i.e* convoluted with the instrumental slit function. The reference atlas used for this purpose is usually the solar flux atlas from 296 to 1300 nm measured in the National

Solar Observatory at Kitt Peak (Arizona) [*Kurucz et al.*, 1984] spectrum. During the calibration process, the instrumental slit function can also be characterized by convolving the highly resolved solar atlas until best match is found with the recorded satellite solar spectra. QDOAS allows fitting the parameters of different line shapes, as well as their wavelength dependence. This is useful when the slit function provided with the instrument is not described precisely enough in the wavelength interval used for the retrieval. In the same way, to account for the moderate resolution of satellite instruments (about a few tenths nanometres), the absorption cross-sections of the trace gases have to be convoluted by the instrumental slit function (and interpolated on the final I_0 wavelength grid). A good knowledge of the instrumental slit function and of its eventual wavelength variation is important to avoid systematic errors in the retrieved slant columns due to spectral shape mismatch between reference data and atmospheric spectra [*Van Roozendael et al.*, 2002b].

Filling-in of the Fraunhofer lines: the Ring effect

Because of rotational Raman scattering, a small fraction of the incident photons undergo a wavelength change of a few nanometres. This causes an intensity loss at their incident wavelength and a gain at the neighbouring wavelengths to which they are redistributed. RRS causes the so-called "filling-in" of Fraunhofer lines, which have a slightly different shape in the earthshine radiance than in the direct solar light. This effect was first discovered by *Grainger and Ring* [1962] and referred to as the Ring effect. The atmospheric absorption lines are also broadened by RRS events occurring after absorption (molecular Ring effect). Although RRS accounts for only a few percent of the measured TOA intensity, it significantly affects DOAS measurements of scattered radiation since typical trace gas absorptions are of the order of a percent or less. If not properly corrected, the Ring effect produces strongly structured residuals in the differential optical density, due to the fact that Fraunhofer lines do not cancel perfectly between I and I_0. Especially in the UV spectral range the remaining spectral structures can by far exceed the structures of weak atmospheric absorbers.

Usually the Ring effect is taken into account by including an additional 'absorber' in the DOAS fit. Ring cross-sections $\sigma_{Ring}(\lambda)$ can be measured [*Solomon et al.*, 1987] or calculated [*Chance and Spurr*, 1997]. The Ring effect can be approximated using the following development for an optically thin atmosphere [*Wagner*, 1999; *Van Roozendael et al.*, 2002b]. One can consider that in any scattered light observation, the light detected by the satellite instrument (I_{meas}) is the sum of elastic and inelastic scattering processes:

$$I_{meas} = I_{elastic} + I_{RRS} = I_0 e^{-\tau} + I_{RRS} \qquad\qquad \textbf{(5-18)}$$

To analyze a measured spectrum, the logarithm of I_{meas} is taken. Since I_{RRS} is very small compared to $I_{elastic}$, the logarithm can be approximated by the first two terms of the Taylor expansion:

$$\ln I_{meas} = \ln(I_{elastic} + I_{RRS}) \approx \ln I_{elastic} + \frac{I_{RRS}}{I_{elastic}} = \ln I_0 - \tau + \frac{I_{RRS}}{I_{elastic}} \qquad\qquad \textbf{(5-19)}$$

The last term $I_{RRS}/I_{elastic}$ is the Ring term which can be approximated by the product of a Ring coefficient (α_{Ring}) and a Ring cross-section (σ_{Ring}). α_{Ring} can be fitted together with other absorbers in the DOAS procedure. A good approximation of the Ring cross-section can be obtained from calculation of a source term for Raman scattering (I_0^{Raman}) derived by simple convolution of the solar spectrum with Raman cross-sections, *i.e* with calculated N_2 and O_2 RRS cross-sections. The term $\alpha_{Ring} . \sigma_{Ring}$ (where $\sigma_{Ring} = I_0^{Raman}/I_0$) is then substituted to $I_{RRS}/I_{elastic}$ in (5-19), assuming that

the molecular Ring effect can be neglected [*Chance and Spurr*, 1997]. A more rigorous approach to compute $I_{RRS}/I_{elastic}$ relies on radiative transfer simulations of nadir radiances including and excluding Raman scattering. Ideally, modelled Ring correction spectra should be calculated using realistic atmospheric parameters (O_3 profile, solar zenith angle, albedo etc.). However due to the complexity of this approach for operational processing, an approximation is to use two Ring cross-section eigenvectors derived by principal component analysis from a set of Ring cross-sections calculated for different atmospheric conditions and solar zenith angles [*Vountas et al.*, 1998].

Solar I_0 effect

To account for the resolution of the instrument, the cross-sections are convolved with the instrumental slit function (F):

$$\sigma*(\lambda) = F \otimes \sigma(\lambda) = \int \sigma(\lambda') \cdot f(\lambda - \lambda') d\lambda' \tag{5-20}$$

However, it should be noted that in 'reality' this convolution occurs in the spectrometer where the incoming intensity $I(\lambda)$ is convoluted with the instrument slit function:

$$I*(\lambda) = F \otimes I(\lambda) = F \otimes \left[I_0(\lambda) \exp(-\sigma(\lambda) \cdot SC) \right] \tag{5-21}$$

Applying (5-21) to the definition of the optical density leads to:

$$\tau_{meas} = -\ln\left(\frac{I*(\lambda)}{I_0*(\lambda)} \right) = -\ln\left(\frac{F \otimes \left[I_0(\lambda) \exp(-\sigma(\lambda) \cdot SC) \right]}{F \otimes I_0(\lambda)} \right)$$
$$\neq F \otimes -\ln\left(\frac{I_0(\lambda) \exp(-\sigma(\lambda) \cdot SC)}{I_0(\lambda)} \right) = F \otimes \sigma(\lambda) \cdot SC \tag{5-22}$$

Thus the fitting of convoluted cross-sections to the optical depth measured by a low resolving instrument is not strictly correct. Only for an optically thin absorber ($\tau' \ll 1$) the exponential function and the logarithm can be approximated by the first two terms of the Taylor expansions. Furthermore, if the I_0 spectrum can be approximated by a constant on the considered wavelength interval it follows:

$$\tau_{meas} \approx F \otimes \sigma(\lambda) \cdot SC \tag{5-23}$$

However, the solar spectrum is not constant with wavelength; it is highly structured due to the Fraunhofer lines. In some cases this leads to retrieval errors since Equation (5-23) is no longer valid. This effect is often referred to as the solar I_0 effect. For most of the atmospheric absorbers the I_0 effect is negligible but in some cases it has to be considered. This is particularly the case for strong atmospheric absorbers like O_3. The solar I_0 effect can be accounted for using so called "I_0 corrected" cross-sections: for a particular absorber, a highly resolved solar spectrum $I_0(\lambda)$ and an absorption spectrum calculated with a known slant column SC are both convolved with the instrument slit function:

$$I_0*(\lambda) = F \otimes I_0(\lambda) \tag{5-24}$$

$$I*(\lambda) = F \otimes (I_0(\lambda) \cdot \exp(\sigma(\lambda) \cdot SC)) \tag{5-25}$$

The I_0 corrected cross-section is then derived:

$$\sigma*(\lambda, SC) = -\ln\left(\frac{I*(\lambda)}{I_0*(\lambda)}\right)$$ (5-26)

The I_0-corrected cross-sections derived in this manner perfectly matches the absorptions in the measured atmospheric spectrum provided that the **SC** used for the calculation is equal to the true atmospheric **SC**. However, this is not a critical point and the I_0 corrected cross-sections can be used for a large range of atmospheric slant columns [*Wagner*, 1999].

Wavelength shifts and intensity offset

The linearity of (5-16) is broken down by the need to account for possible instrumental straylight requiring an additional intensity offset parameter (**$I_{meas}(\lambda)=I(\lambda)+offset(\lambda)$**) or small wavelength shifts or stretch (**$shift(\lambda)$**) between the various spectra involved in the data evaluation, namely the radiance and irradiance spectra as well as the absorption cross-sections. These parameters are included in Equation (5-16):

$$\ln\left[I(\lambda + shift(\lambda)) - offset(\lambda)\right] = \ln I_0(\lambda) - \sum_j \sigma'_j(\lambda) \cdot SC_j + \sum_p c_p \lambda^p$$ (5-27)

In QDOAS, both offset and shift parameters can be fitted using wavelength polynomials up to the second order:

$$offset(\lambda) = a_{off} + b_{off} \cdot (\lambda - \lambda_0) + c_{off} \cdot (\lambda - \lambda_0)^2$$

$$shift(\lambda) = a_{shift} + b_{stretch} \cdot (\lambda - \lambda_0) + c_{stretch} \cdot (\lambda - \lambda_0)^2$$ (5-28)

where λ is the calibrated wavelength and λ_0 is the centre of the retrieval interval. a_{off}, b_{off}, c_{off} are the fitted offset parameters, while a_{shift} is the fitted wavelength shift and $b_{stretch}$ and $c_{stretch}$ are the first and second orders of the fitted stretch.

Slant columns fitting and residual

Since the spectra are recorded with diode arrays of typically 1024 discrete elements (see chapter 4), the signal obtained by a single diode represents the integrated intensity over the wavelength range covered by this diode. The aim of the fitting process is to minimize the χ^2 function derived from Equation (5-27):

$$\chi^2 = \frac{\sum_{i \in W_{int}}\left(\ln\left[I(\lambda_i + shift(\lambda_i)) - offset(\lambda_i)\right] - \ln I_0(\lambda_i) + \sum_j \sigma'_j(\lambda_i) \cdot SC_j + \sigma_{Ring}(\lambda_i) \cdot \alpha_{Ring} - \sum_p c_p \lambda_i^p\right)^2}{M - N}$$ (5-29)

where the summation is made over the individual diode elements included in the selected wavelength range W_{int}. **M-N** represents the number of degrees of freedom, **M** being the number of selected diode elements and **N** the number of parameters to fit. The QDOAS fitting algorithm uses a combination of Singular Value Decomposition (SVD) and the Levenberg-Marquard methods to solve the linear and non-linear parts of the DOAS equation [*Fayt and Van Roozendael*, 2001]. The difference between the measured optical density and the result of the fitting routine is called the residual spectrum. It is a measure of the quality of the spectral fitting. From the random structures of the residual, the statistical error on the derived slant columns can be determined [*Platt and Stutz*, 2008].

5.3 Effective atmospheric light path: air mass factor calculation

The measured slant column density (*SC*) is the integrated trace gas concentration along the effective light path between the sun and the satellite. This quantity must be converted into a more physical variable, namely the trace gas concentration integrated vertically or vertical column density (*VC*). This conversion is achieved by means of a so-called air mass factor (*AMF*) which describes the sensitivity of the measurement to a specific absorber considering the state of the atmosphere along the light path. The *AMF* is calculated with a radiative transfer model.

5.3.1 The air mass factor formulation

In the DOAS approach, an optically thin atmosphere is assumed. The mean optical path of scattered photons can therefore be considered as independent of the wavelength within the relatively small spectral interval selected for the fit and a single wavelength effective air mass factor can be defined as the ratio of the slant column to the vertical column:

$$AMF = \frac{SC}{VC} \tag{5-30}$$

When a trace gas resides only in the stratosphere, scattering (which occurs mostly in the troposphere) does not need to be taken into account for the *AMF* computation. Indeed, the light paths across the stratosphere have all the same length, completely determined by the solar zenith angle θ_0 and the viewing angle θ. The air mass factor can be well approximated by a simple geometrical form (see Figure 5-9), valid for angles up to 70°:

$$AMF_{geo} = \frac{1}{\cos\theta_0} + \frac{1}{\cos\theta} = \frac{1}{\mu_0} + \frac{1}{\mu} \tag{5-31}$$

In the troposphere, however, photon paths are highly variable and taking into account atmospheric scattering is critical for the *AMF* calculation, particularly in the UV-visible wavelength ranges. This requires full radiative transfer calculations combined with knowledge of surface albedo, cloud properties, and aerosols. Moreover, the *AMF* depends on the vertical distribution of atmospheric absorbers. The concentration profiles have therefore to be known *a priori*. An important limitation of the *AMF* approach lies in the fact that for most applications, some of these parameters are poorly known or with inappropriate spatial and temporal resolution.

The general approach for the *AMF* calculation relies on the following considerations. The extinction due to a particular absorbing species is given by:

$$I^w = I^{wo}e^{-\tau_a} = I^{wo}e^{-\sigma_a \cdot SC} \tag{5-32}$$

where I^w and I^{wo} are respectively the radiances with and without the presence of the absorber, while σ_a and *SC* are the absorption cross-section and slant column of the absorber of interest. Therefore, the *AMF* is generally calculated by performing two simulations, one with and one without the absorber, and is given by:

$$AMF = \frac{\ln I^{wo} - \ln I^w}{\sigma_a \cdot VC} \tag{5-33}$$

where **VC** is the total column of the *a priori* absorber profile used in the simulation. Starting from this definition, *Palmer et al.* [2001] provided a convenient and widely used **AMF** formulation for optically thin absorbers that decouples the measurement vertical sensitivity to atmospheric species from the vertical profile shape of the species of interest:

$$AMF = \int_0^{TOA} w(z)S(z)dz \tag{5-34}$$

where the scattering weighting functions **w(z)** and the vertical shape factors **S(z)** are respectively defined as:

$$w(z) = -\frac{\partial \ln I^w}{\partial \tau} \tag{5-35}$$

$$S(z) = \frac{x_a(z)}{\int_0^{TOA} x_a(z)dz} \tag{5-36}$$

The strength of this formulation is that both components can be studied individually. The scattering weighting functions represent the sensitivity of the measurement to a change of optical density at a certain altitude. In a scattering atmosphere, **w(z)** depends on the viewing angles, the wavelength, the surface albedo, and the surface altitude, but not on the vertical distribution of the considered absorber (optically thin approximation). Radiative transfer models such as LIDORT allow to derive **w(z)**. The vertical profile shape corresponds to the *a priori* vertical profile of the absorber number density $x_a(z)$ [molec.cm^{-3}], normalized with the total *a priori* vertical column. Frequently, $x_a(z)$ is taken from an atmospheric chemistry transport model or from some other prior knowledge of the vertical distribution of the absorber. The final **AMF** formulation is thus given by:

$$AMF = \frac{\int_0^{TOA} w(z)x_a(z)dz}{\int_0^{TOA} x_a(z)dz} \tag{5-37}$$

This **AMF** formulation represents an integral over the entire atmosphere of the vertical distribution of the absorber weighted by sensitivity coefficients derived from the radiative transfer model. It means that the **AMF** does not depend on any multiplicative factor applied to the *a priori* profile $x_a(z)$, but only on its shape. The decoupling of the **AMF** calculation has two main advantages. First, it increases flexibility by allowing radiative transfer and atmospheric chemistry to be studied separately. Values of **w(z)** can be tabulated and then used in combination with any vertical profile shape for local calculations of the **AMF**. Second, using shape factors from an atmospheric chemistry model to formulate the **AMF** ensures consistency for subsequent evaluation of the model with the retrieved vertical columns.

Figure 5-12 (a) illustrates typical vertical distributions of **w(z)** and H_2CO **$S_z(z)$**. Here, **w(z)** has been simulated for three solar zenith angles (θ_0 = 20°, 50°, 80°) with the LIDORT RTM at 340 nm, for an albedo of 0.1 and for a cloud-free, U.S. standard atmosphere [1976]. **S(z)** is taken from a summer time tropospheric formaldehyde profile in a remote region, simulated by the IMAGES model [*Stavrakou et al.*, 2009a]. At high altitude, scattering is negligible and the weighting function equals the geometric **AMF**. Below 5 km, the sensitivity of the measurement decreases because the

penetration of UV radiation is inhibited by scattering. The amount by which the sensitivity decreases with altitude depends on the solar position. If the sun is low (high SZA), less light reaches the surface because the scattering point is higher, and the measurement are very little sensitive to the boundary layer. In the meantime, large SZA increase the light path in the stratosphere and therefore also the sensitivity towards stratospheric absorbers. Figure 5-12 (b) compares the variation of the geometric (calculated with (5-31)) and the tropospheric *AMF* (calculated with (5-37)) with θ_0. It shows that the tropospheric *AMF* values are approximately half of the geometric *AMF* for typical conditions, up to SZA of ~60°. Scattering thus decreases the sensitivity of UV-Vis nadir measurements to tropospheric H_2CO by about a factor of 2.
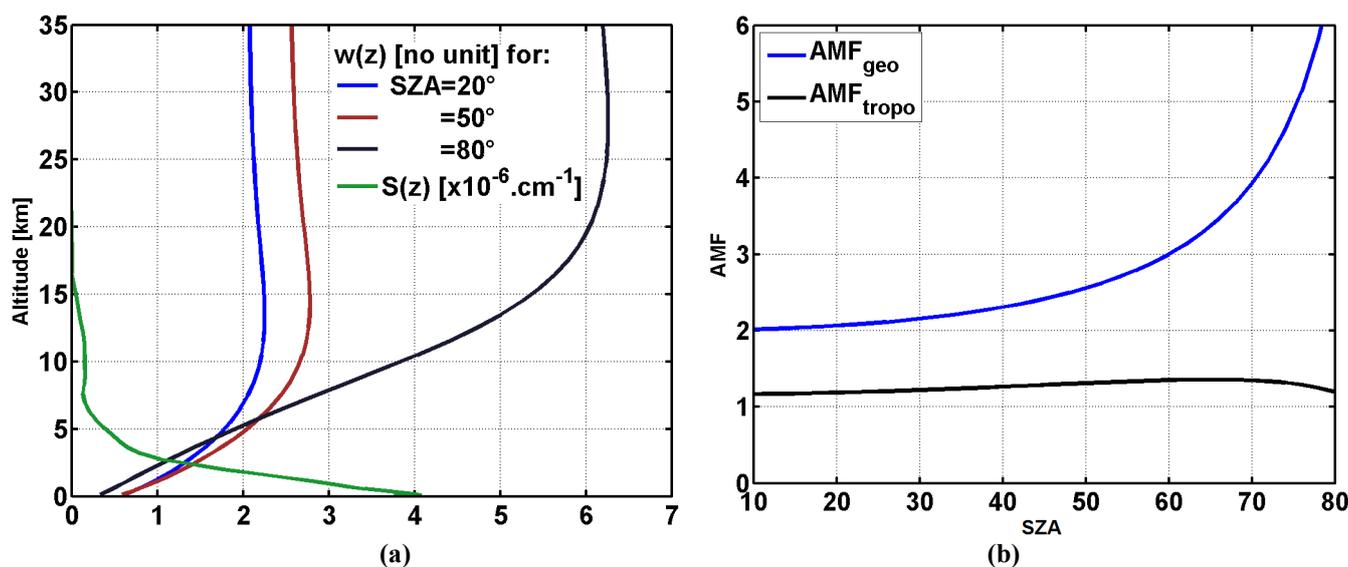
Figure 5-12: (a) Examples of one H_2CO profile shape $S(z)$ and three scattering weighting function $w(z)$ for SZA=20°, 50° and 80°. (b) Variation of the geometric and tropospheric AMFs with solar zenith angle (LOS=0°, azimuth=0°, albedo=0.1, λ=340nm).

More generally, the sensitivity of the measurement in the lower troposphere depends strongly on all the conditions of observation, particularly on the surface albedo, the surface elevation and clouds that modify drastically the light path through the atmosphere. We will present in the next chapter more details on the *w(z)* variations and on the *AMF* calculation for tropospheric H_2CO observations taking clouds into account. When calculating the *AMF*, accuracy depends on the knowledge of the main parameters determining the radiative transfer in the atmosphere. The impact of parameter uncertainties and the resulting implications for the interpretation of the DOAS results will be also studied in the chapter 6.

6 Global Observation of Formaldehyde in the Troposphere Using GOME and SCIAMACHY Sensors

This work presents global tropospheric formaldehyde columns retrieved from UV radiance measurements performed by GOME and SCIAMACHY. A special effort has been made to ensure the coherence and quality of the H_2CO dataset for the entire period 1996-2010. Optimised DOAS settings are proposed in order to reduce the impact of two important sources of error in the derivation of slant columns, namely, the polarisation anomaly affecting the SCIAMACHY spectra around 350 nm, and a major absorption band of the O_4 collision complex centred near 360 nm. The air mass factors are determined from scattering weighting functions generated using radiative transfer calculations taking into account the cloud fraction, the cloud height and the ground albedo. Vertical profile shapes of H_2CO are provided by the global chemistry transport model IMAGES based on an up-to-date representation of emissions, atmospheric transport and photochemistry. A comprehensive error analysis is presented. This includes errors on the slant columns retrieval and errors on the air mass factors which are mainly due to uncertainties in the *a priori* profile and in the cloud properties. The major features of the retrieved formaldehyde column distribution are discussed and compared with previous H_2CO datasets over the major emission regions.

This chapter is based on *De Smedt et al.* [2008][1], although several improvements have been introduced since then. First, the consistency of the combined dataset has been further improved trough adoption of a unique set of high-resolution cross-sections used for both instruments. Secondly, to reduce the errors associated to topography, scattering weighting functions have been calculated with a finer altitude grid near the ground and the *a priori* profiles are rescaled to the ground altitude following *Zhou et al.* [2009]. The resulting AMFs are generally larger by less than 5% except over high altitude plateaus where the differences can reach 20%. Finally, the SCIAMACHY time series has been extended up to 2010.

6.1 Introduction

Satellite observations of formaldehyde started with the launch of GOME on the ERS-2 satellite in April 1995. Early retrievals of formaldehyde were obtained from GOME measurements by application of the Differential Optical Absorption Spectroscopy (DOAS) method, using three absorption bands of H_2CO between 337 and 359 nm [*Thomas et al.*, 1998; *Chance et al.*, 2000; *Wittrock et al.*, 2000; *Marbach et al.*, 2004; *De Smedt et al.*, 2006]. These studies demonstrated the usefulness of satellite observations of H_2CO columns as a means for providing constraints on the emissions of NMVOCs [*Palmer et al.*, 2003 and 2006; *Fu et al.*, 2007; *Shim et al.*, 2005]. However, GOME H_2CO columns retrieved using this fitting interval exhibit anomalous features like for example, excessively low (sometimes negative) columns above desert regions like Sahara and Australia [*Wittrock et al.*, 2000; *De Smedt et al.*, 2006], or systematic structures above oceans in apparent contradiction with tropospheric models [*Abbot et al.*, 2003; *Palmer et al.*, 2006]. As an

---

[1] De Smedt, I., J.-F. Müller, T. Stavrakou, R. van der A, H. Eskes and M. Van Roozendael (2008), Twelve years of global observation of formaldehyde in the troposphere using GOME and SCIAMACHY sensors, *Atmos. Chem. Phys., 8*, 4947-4963.

illustration of these anomalies, an example of "classic" H$_2$CO retrieval from GOME measurements is shown in Figure 6-1, for the period 1997-2001.
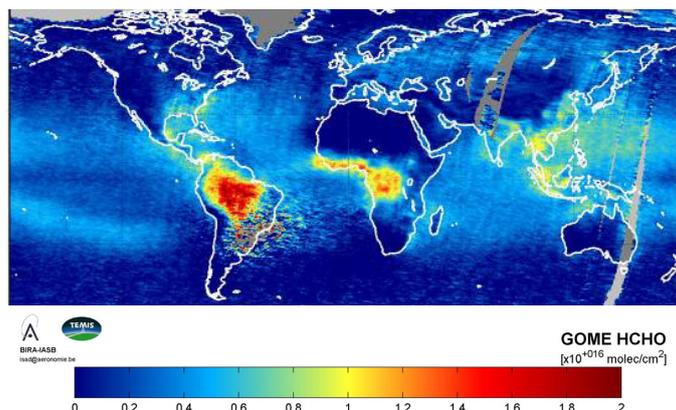
Figure 6-1: Example of early formaldehyde retrieval, obtained from GOME observations between 1997 and 2001, in the commonly used wavelength interval of 337-359 nm [*De Smedt et al.*, 2006].

Launched on ENVISAT in March 2002, SCIAMACHY has been providing nadir UV-visible radiance measurements suitable to extend GOME formaldehyde observations. However, until this study, its use for H$_2$CO retrieval had only been reported once [*Wittrock et al.*, 2006]. This is mainly due to the existence of a strong polarisation anomaly affecting the SCIAMACHY spectra around 350 nm [*Tilstra et al.*, 2005; *Slijkhuis and Lichtenberg*, 2008], preventing the straightforward application of the H$_2$CO retrieval settings derived from GOME [*Afe et al.*, 2005, *Wittrock et al.*, 2006]. One way to sidestep this problem is to fit H$_2$CO at shorter UV wavelengths, hence avoiding the instrument grating polarization signature. A reduced spectral region including only two absorption bands of formaldehyde (334–348 nm) was therefore selected by *Wittrock et al.* [2006]. The resulting H$_2$CO columns retrieved from SCIAMACHY measurements during the year 2005 is shown in Figure 6-2. Such as for GOME, the abnormally low H$_2$CO columns above desert regions are noticeable. Furthermore the data are very noisy because (1) the SCIAMACHY Earth sampling of is half the one of GOME, because of the alternate limb/nadir views of SCIAMACHY, (2) the integration time of the SCIAMACHY measurements is shorter than the GOME measurements and consequently the signal to noise ratio of the spectra are lower, and (3) the number of spectral measurements used for the DOAS fit is smaller than for GOME, increasing the noise in the retrieval.
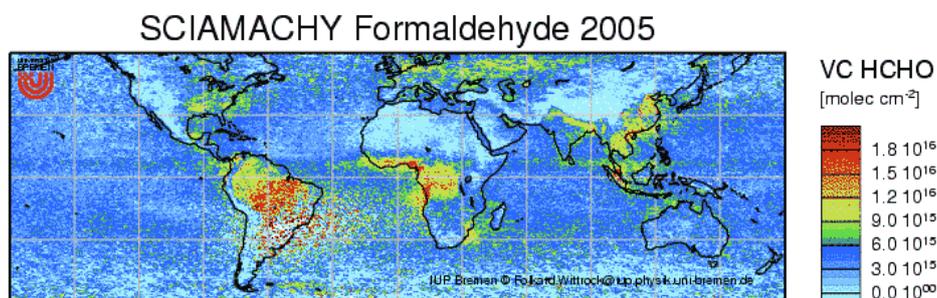
Figure 6-2: Yearly mean formaldehyde derived from SCIAMACHY observations in 2005, using the reduced fitting interval of 334-348 nm [*Wittrock et al.*, 2006].

Until this work, no attempt had been made to create a unique dataset by combining GOME and SCIAMACHY observations. Furthermore, the errors on the retrieved vertical columns were most of the time poorly quantified. A single value was generally provided, corresponding to the typical random error on the slant columns. Here, we have proposed a new fitting window including three absorption bands of formaldehyde and applicable to both GOME and SCIAMACHY. Optimized retrieval settings were developed for the fit of the slant columns and for the estimation of the air mass factors in order to generate a 15-year long dataset of H₂CO columns that consistently combines the two instruments. Moreover, a detailed error budget was drawn up for each individual pixel, separating random from systematic effects. This H₂CO product was developed in the framework of the TEMIS [http://www.temis.nl] and PROMOTE [http://www.gse-promote.org] projects and is available on their websites.

Section 6.2 presents the H₂CO retrieval method, more specifically, the slant column fitting and the determination of the air mass factors. Section 6.3 provides a detailed description of the error budget. Section 6.4 presents the global distribution and temporal evolution of the H₂CO vertical columns derived in this study, as well as a brief comparison with previously retrieved datasets. A description of the two spaceborne spectrometers used in this work is provided in chapter 4 (section 4.4).

6.2 Retrieval method

6.2.1 Slant Column Retrieval

Owing to the sharp absorption bands of formaldehyde in the UV, the DOAS technique can be used to derive H₂CO concentrations. As presented in chapter 5, this method consists of two independent steps. First, the species concentration integrated along the mean light path between the sun and the satellite (the slant column density *SC*) is derived from the satellite radiance using the Beer-Lambert law applied to an optically thin atmosphere [e.g. *Platt and Stutz*, 2008].

$$\ln \frac{I(\lambda, \mu)}{I_0(\lambda)} = -\sum_j \sigma_j(\lambda) \cdot SC_j + \sum_p c_p \lambda^p \tag{6-1}$$

In this expression, *I* and *I₀* are the measured earthshine radiance and solar irradiance spectra. *SCⱼ* and *σᵢ* are respectively the slant column and the absorption cross section of each molecule and *cₚ* are the coefficients of the polynomial function accounting for terms varying slowly with wavelength, like Rayleigh and Mie scattering. The slant column densities are deduced by solving this system of linear equations. In practice, additional effects have to be taken into account like for example, the Ring effect [*Grainger and Ring*, 1962], wavelength shifts between *I* and *I₀* or the possible contamination of measured radiances by spectral straylight, which requires the introduction of an intensity offset parameter (see Chapter 5).

Despite the relatively large abundance of formaldehyde in the atmosphere (of the order of 10^{16} molec.cm^{-2}) and its well-defined absorption bands, the fitting of H₂CO slant columns in earthshine radiances is a challenge because of the low optical density of H₂CO compared to other UV absorbers. As displayed in Figure 6-3, the typical H₂CO optical density is about 2.5 times smaller than BrO, up to 50 times smaller than NO₂ and up to 500 times smaller than O₃. Therefore, the detection of H₂CO is limited by the signal to noise ratio of the measured radiance and by possible spectral interferences due to other molecules absorbing in the same fitting interval, mainly ozone.
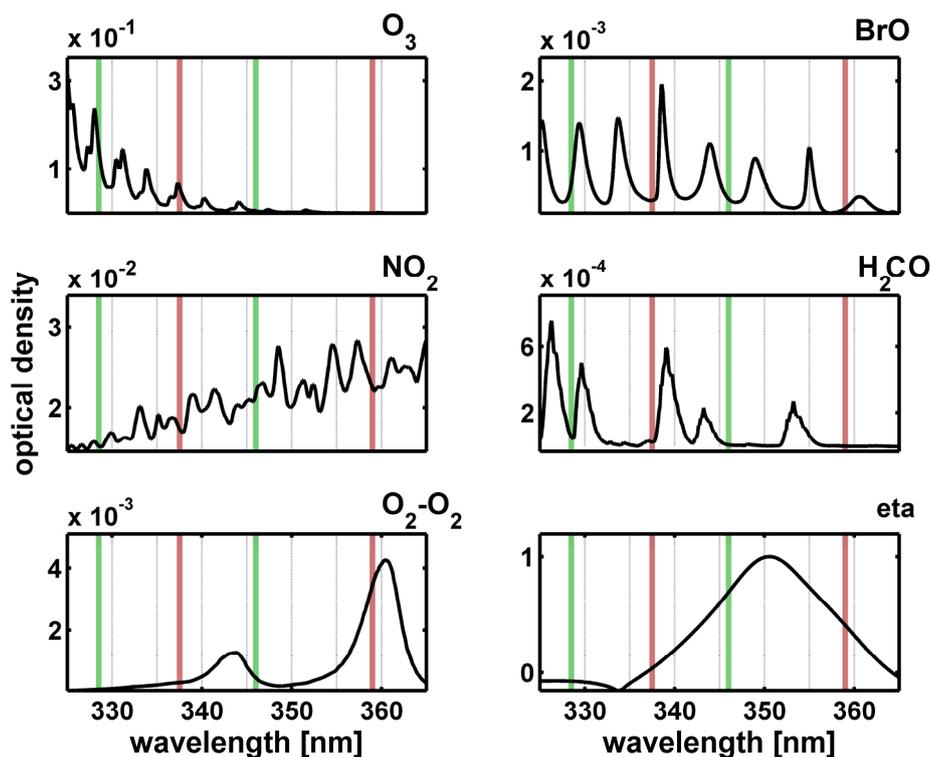
Figure 6-3: Typical optical densities of O_3, NO_2, O_2-O_2, BrO, H_2CO and SCIAMACHY polarisation response in the near-UV [eta, *Slijkhuis and Lichtenberg*, 2008]. The slant columns have been taken as 10^{19} molec.cm^{-2} for O_3, 5×10^{16} molec.cm^{-2} for NO_2, 10^{43} molec^2.cm^{-5} for O_2-O_2, 10^{14} molec.cm^{-2} for BrO and 10^{16} molec.cm^{-2} for H_2CO. The vertical lines indicate the fitting intervals discussed in the text: 328.5-346 nm (green) and 337.5-359 nm (red).

Choice of H_2CO cross-section dataset

The UV formaldehyde absorption spectrum between 300 nm and 360 nm (Figure 6-4) has been measured by several research groups at different spectral resolutions [*e.g. Cantrell et al.*, 1990; *Meller and Moortgat*, 2000; *Bogumil et al.*, 2003; *Gratien et al.*, 2007; *Brauers et al.*, 2007]. The most recent studies covering the wavelength range of interest for the retrieval of tropospheric formaldehyde (320-360 nm) are listed in Table 6-1. Medium-resolution measurements (0.15 to 0.2 nm) were reported by *Gratien* [2007] and *Bogumil* [2003]. The spectrum reported by *Cantrell* [1990] was measured with the highest resolution (0.01 nm) for the 300-358-nm region using Fourier transform spectroscopy. It is the H_2CO cross-section recommended by the HITRAN database [*Orphal and Chance*, 2003]. *Meller and Moortgat* [2000] provided a high-resolution spectrum in the complete UV absorption band (224-373 nm). Their dataset is now recommended by IUPAC [*Atkinson et al.*, 2006] and NASA [*Sander et al.*, 2006]. As illustrated by Figure 6-4, differences exist between the cross-section amplitudes. These differences directly impact the retrieved H_2CO slant column densities. After degradation to the resolution of GOME or SCIAMACHY (~0.2-0.3 nm), the *Cantrell* [1990] cross-section dataset is found to be 9% lower than the *Meller and Moortgat* [2000] dataset. Still compared to *Meller and Moortgat* [2000], the dataset of *Bogumil* [2003] is in very good agreement, while the *Gratien* [2007] dataset is 12% higher. In order to compare H_2CO measurements performed by different satellite instruments or with ground-based instruments (for example MAX-DOAS or FTIR), it is important to use consistent laboratory absorption datasets. Recent intercomparison studies showed that the *Meller and Moortgat* [2000]

cross-section dataset best reproduce the experimental H$_2$CO concentrations in simulation chamber [*Brauers et al.*, 2007; *Hak et al.*, 2005]. Furthermore, *Gratien et al.* [2007] presented an intercomparison of UV (300-360 nm) and infrared (1660-1820 cm^{-1}) absorption coefficients of formaldehyde using simultaneous measurements in both spectral regions. This intercomparison showed good agreement between the UV data published by *Meller and Moortgat* [2000] and a set of four infrared data. In contrast, a rather large discrepancy was observed with the UV data of *Cantrell* [1990]. For these reasons, the *Meller and Moortgat* [2000] H$_2$CO absorption cross-section dataset was used in this work.

Table 6-1: Comparison of high resolution H₂CO cross-section datasets covering the 300-360 wavelength range [*MPI-Mainz-UV-VIS Spectral Atlas of Gaseous Molecules*, www.atmosphere.mpg.de/spectral-atlas-mainz].

H$_2$CO cross-section datasets	Range (Resolution)	Temperature
Cantrell [1990]	300.3 - 385.8 nm (0.01 nm)	293 K
Meller and Moortgat [2000]	224.6 - 376.0 nm (0.02 nm)	298 K
Bogumil [2003]	247.3 - 400.0 nm (0.21 nm)	293 K
Gratien [2007]	240.0 - 364.5 nm (0.15 nm)	298 K
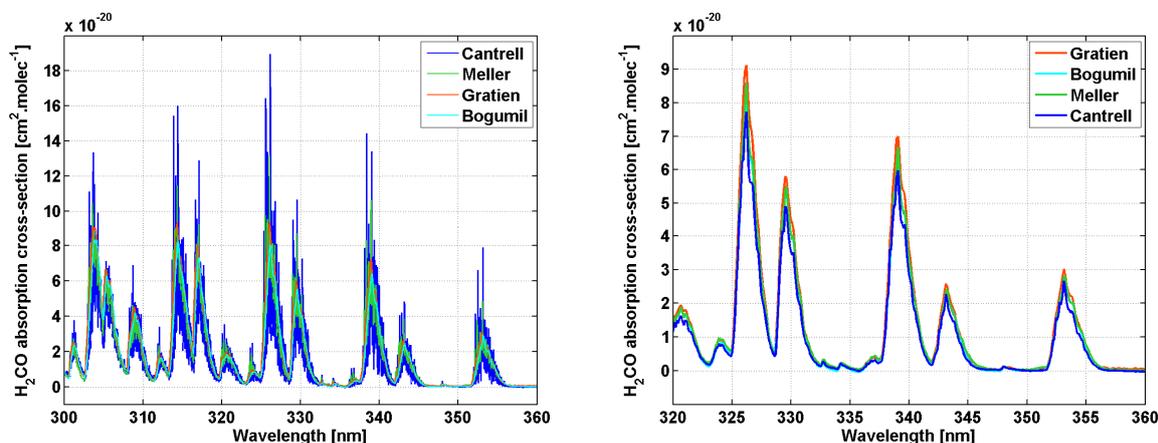
Figure 6-4: Comparison of the H₂CO cross-section datasets at high resolution (left) and convolved with the GOME slit function (right).

Some laboratory measurements additionaly report the variation of the formaldehyde absorption cross-section with temperature [*Cantrell et al.*, 1990; *Meller and Moortgat*, 2000]. Since atmospheric temperatures may change by several tens of degrees in regions where H$_2$CO is observed, we have to evaluate the sensitivity of the retrieved slant column densities to the choice of the temperature cross-section. The *Meller and Moortgat* [2000] dataset is provided at 298 K and 223 K (see the left panel of Figure 6-5). Temperature effects are the strongest at the maximum of the absorption bands, where lower temperature results in larger absorption. In first approximation, we can evaluate the temperature impact on the slant columns assuming a linear relation between the cross-sections throughout the range of atmospheric temperatures (at GOME resolution and in the window 325-360 nm) [*Van Roozendael et al.*, 2002b]. As shown in the right panel of Figure 6-5, this evaluation results in a decrease of the absorption cross-section of 0.05% for each temperature increase of 1° K (with an inversely proportional effect on the retrieved *SC*). This weak temperature dependence can be neglected for our applications. As the bulk of formaldehyde lies in the troposphere, we use the cross-section at 298 K.
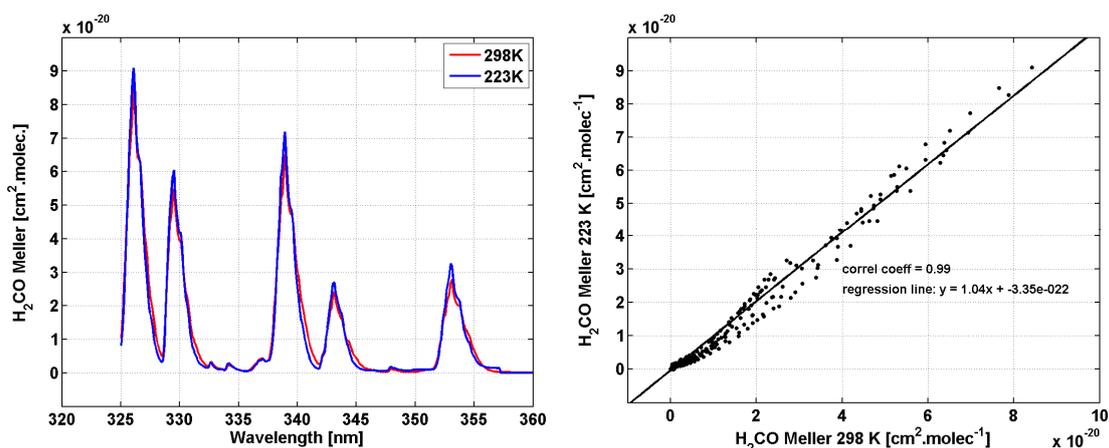
Figure 6-5: Temperature variation of the H₂CO *Meller and Moortgat* absorption cross-section dataset degraded to the GOME resolution [*Meller and Moorgat*, 2000].

Polarisation correction

The signals measured by the GOME and SCIAMACHY instruments are not only dependent on the intensity of the incoming light, but also on its polarization. The sensitivity of the instrument to polarised radiance is described by a set of calibration constants which are dependent on wavelength and on scan mirror angles. Primarily, this sensitivity is changing slowly with wavelength and is not of concern for the DOAS retrieval. However, some instrumental components exhibit sharp changes in sensitivity, and introduce wavelength dependent signatures in the measured spectra if not corrected. The polarization constants (or key data) have been measured during pre-flight calibration at certain wavelengths and scan angles. In the case of SCIAMACHY, a very prominent signature is present around 350 nm, and related to a strong polarisation feature of the channel 2 grating (see the SCIAMACHY "eta" function in channel 2, shown in the last panel of Figure 6-3) [*Slijkhuis and Lichtenberg*, 2008]. During the flight, the polarisation degree is determined using the polarisation monitoring devices (PMDs). The spectral polarisation effects are then calculated using the calibration constants, interpolated to the exact wavelengths and scan angles of the measurement in process. Unfortunately, the instrumental response has changed over time, and the pre-flight calibration curves do not longer correct completely for the effects of polarisation. Those wavelength regions exhibiting the polarisation structures have to be avoided for the retrieval of weak absorbers. This is the case in the SCIAMACHY channel 2 used to retrieved formaldehyde, but also in the GOME channel 3 used to retrieve glyoxal [*Wittrock et al.*, 2006].

Choice of the fitting window

Taking into account the spectral and instrumental constraints, the fitting window has been chosen in order to (1) maximize the sensitivity to formaldehyde, (2) minimize the fitting residuals and the dispersion of the retrieved H₂CO slant columns and (3) minimize the interferences with other absorbers. From inspection of the H₂CO absorption cross-sections in Figure 6-3, it can be seen that a maximum of five characteristic absorption bands are available in the spectral interval from 325 to 360 nm. A compromise has to be found when defining the DOAS settings, since the fitting accuracy is limited at the shorter wavelengths by the increased O_3 absorption and the resulting interference with the formaldehyde absorption, and at longer wavelengths by the interference with the SCIAMACHY polarisation feature around 350 nm (Figure 6-3). Fifteen fitting intervals have been

tested as for their ability to produce consistent and stable retrievals from both GOME and SCIAMACHY, starting from the reduced interval used by the IUP-Bremen group for SCIAMACHY (334–348 nm) and shifting gradually the lower and upper limits of the interval. The best compromise, which maximizes the number of H₂CO absorption bands and minimizes fitting errors while avoiding the SCIAMACHY polarisation peak around 350 nm, was found to lie in the 328.5-346 nm wavelength region. Despite the stronger interference with O_3 absorption at these wavelengths, the use of the 328.5-346 nm window improves the GOME slant columns compared to retrievals commonly performed in the 337.5-359 nm region. Indeed, the use of the newly proposed 328.5-346 nm interval leads to a reduction of the noise over oceans and brings the slant column values above desert regions at the level of the background (Figure 6-6 (a) and (b)). The GOME H₂CO values retrieved using the two fitting windows are generally consistent within 2×10^{15} molec.cm^{-2}, in particular above regions of enhanced emissions. The largest differences are found above arid and semi arid regions, mainly in Africa, the Middle East and Australia, where they range between 2×10^{15} and 10×10^{15} molec.cm^{-2}. Fitting residuals are lower by 15% in the Tropics, but they increase more rapidly with solar zenith angle and exceed the fitting residuals of the 337.5-359 nm window poleward of 50° because of a stronger interference with O_3 (Figure 6-6 (c) and (d)).
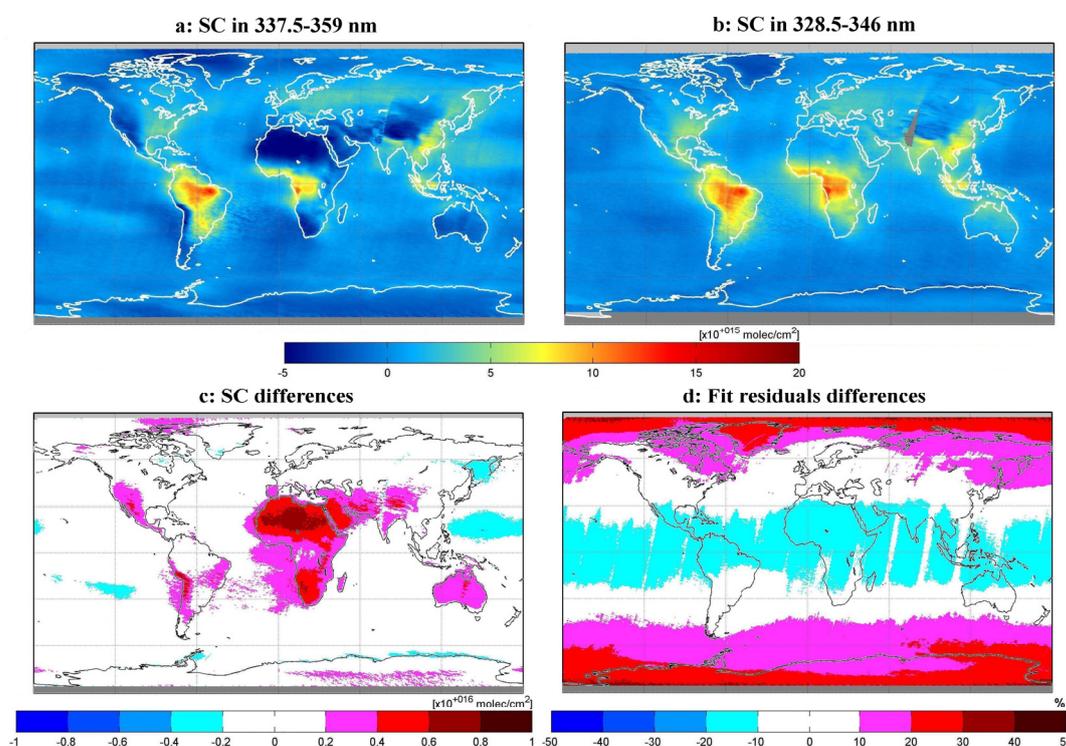
Figure 6-6: Comparison of GOME H₂CO slant columns densities (SC) averaged over 1997-2002 retrieved using (a) the 337.5-359 nm fitting window and (b) the 328.5-346 nm interval introduced in this study. The absolute differences between the slant columns shown in panels (b) and (a) are displayed in panel (c). Panel (d) shows the relative differences between the fitting residuals obtained using the two retrieval windows.

Fitting tests have been performed, showing that the slant column underestimation above desert regions is not observed as long as the upper limit of the fitting window is kept below 350 nm. Furthermore, the magnitude of the negative bias above Sahara, represented as a function of the upper limit chosen for the fitting window (Figure 6-7), correlates with the O_2-O_2 (or O_4) collision complex absorption band peaking at 360 nm. This suggests that the low H₂CO columns observed

above deserts are most likely due to an O_4-related fitting artefact rather than a real particularity of the formaldehyde distribution. Desert regions are generally cloud-free and therefore, they correspond to the highest O_4 columns observed globally. The known uncertainties of the O_4 cross-sections are expected to significantly impact the quality of the H_2CO retrievals in these regions, especially when the strong absorption band of O_4 at 360 nm is included in the fit, even partially [*Greenblatt et al.*, 1990; *Wagner et al.*, 2002].
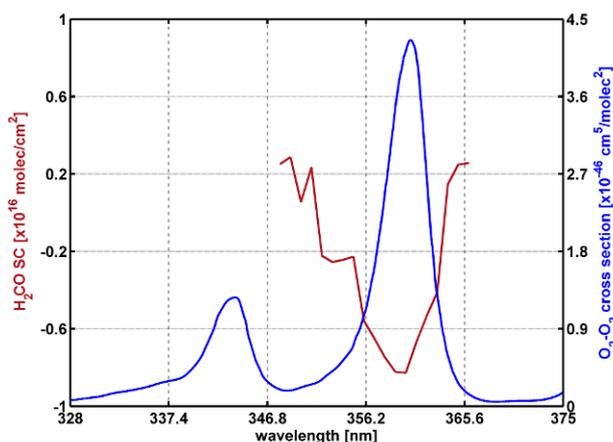
Figure 6-7: In red are plotted the fitted H_2CO slant columns (*SC*) averaged above the Sahara region as a function of the upper limit chosen for the fitting interval (the lower limit is fixed to 337.5 nm). These slant columns have been retrieved for latitudes between 15 and 30°N, for the GOME orbit number 40585 of the 24th of January 2003 (20-30°E). In blue is shown the absorption cross-section of the O_2-O_2 (or O_4) collision complex in the same wavelength region.

GOME and SCIAMACHY retrieval settings and fit results

In this work, most DOAS settings are common to both GOME and SCIAMACHY retrievals in order to optimize the consistency between the two satellite instruments. High–resolution cross-section datasets, convolved with the instrumental slit function of GOME or SCIAMACHY have been used. In addition to H_2CO, the fitting procedure includes reference spectra for other interfering species (O_3 (228K and 243K), NO_2 and BrO) (see Table 6-2 for the cross-section dataset references). The O_3 and NO_2 cross-sections are corrected for the I_0 effect [*Wagner*, 1999; *Aliwell et al.*, 2002]. The Ring effect is corrected according to *Chance and Spurr* [1997]. A linear intensity offset correction is further applied as well as a polynomial closure term of order 5. This quite high polynomial degree was found to reduce substantially the fitting residuals.

Table 6-2: Absorption cross-section dataset included in the fit of formaldehyde and associated uncertainties.

Molecule	Data source	Cross-section error
H_2CO	*Meller and Moorgat* [2000], 298°K	10 %
Ozone	*Brion et al.* [1998], 228°K + 243°K	5 %
BrO	*Fleischmann et al.* [2004], 223°K	8 %
NO_2	*Vandaele et al.* [2002], 220°K	3 %
Ring effect	*Chance and Spurr* [1997]	5 %

GOME solar spectra are affected by spectral artefacts caused by angular dependent spectral features in the reflectivity of the Sun diffuser plate [*Richter and Wagner*, 2001]. In order to minimise the impact of these artefacts, daily radiance spectra are used as reference instead of solar irradiance spectra. These are selected in a region of the equatorial Pacific Ocean where the formaldehyde column is assumed to be low and stable in time [*Stavrakou et al.*, 2008]. This choice of reference spectrum also minimizes fitting residuals because the Ring effect is attenuated. Indeed, the filling-in of the Fraunhofer lines due to inelastic scattering cancels better between two radiance spectra.

Examples of formaldehyde optical density fits are shown in Figure 6-8 for a GOME pixel in September 1997 and a SCIAMACHY pixel in September 2006, both located over Indonesia during fire episodes. The ground resolutions of GOME and SCIAMACHY are respectively 40x320 km^2 and 30x60 km^2. The GOME pixels are thus seven times larger than those of SCIAMACHY. From photon statistical consideration, when these instruments of similar design perform at best, the ratio between the corresponding fitting residuals should be about 2,6 ($\sqrt{7}$). We find a slightly better ratio of 2.4 ($6.5 \times 10^{-4}/2.7 \times 10^{-4}$), demonstrating that the noise in the retrieved columns is consistent with the signal to noise ratio of the instruments. However, it should be noted that degradation effects impacting the fitting residuals are present in the GOME observations starting from 2001 (see section 6.3.2 about the slant column errors), making more difficult a fair comparison between the two instruments.
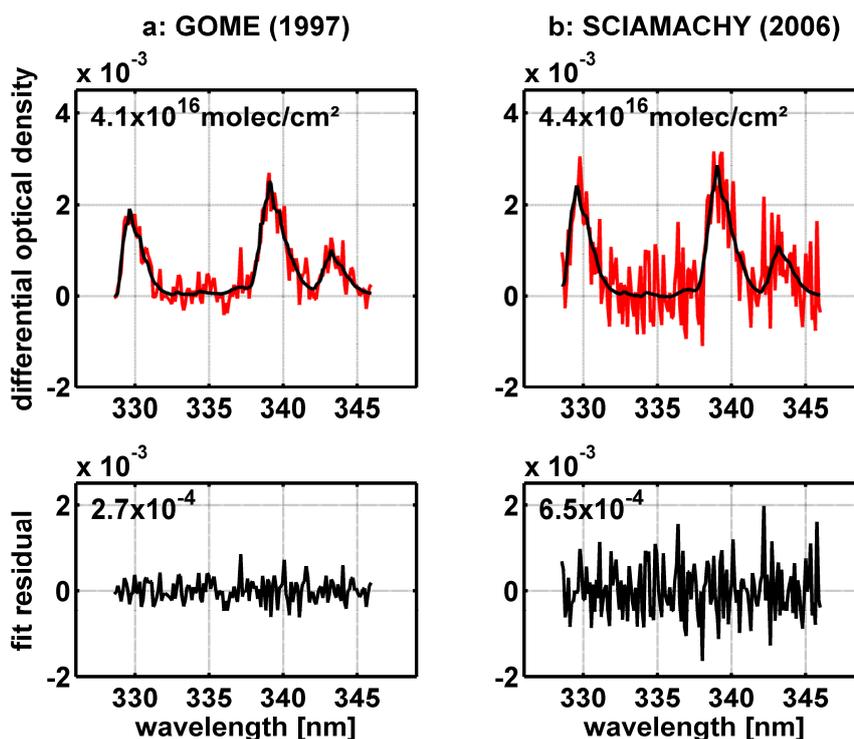
Figure 6-8: DOAS fit of H₂CO optical density and residual of the fit for (a) a GOME pixel (02/09/1997, orbit 12382, SZA: 24.6°) and (b) a SCIAMACHY pixel (24/09/2006, Orbit: 23875, SZA: 28.6°). The fitted slant columns densities are given inset the upper panels and the fit residuals are given inset the lower panels.

Among other fitting windows tested for SCIAMACHY, the 328.5-346 nm interval provides the lowest dispersion in the SCIAMACHY results. This wavelength interval also displays the lowest differences with GOME, as verified over the six first months of 2003, when the two satellite

measurements globally overlap. Figure 6-9 shows the slant columns for two overlapping orbits of GOME and SCIAMACHY over Eastern China, on the 14th of April 2003. The mean H_2CO values agree well and, in this case, the ratio between the standard deviation of the GOME and SCIAMACHY slant columns is only 1.7, as a result of the long term degradation of the GOME instrument.
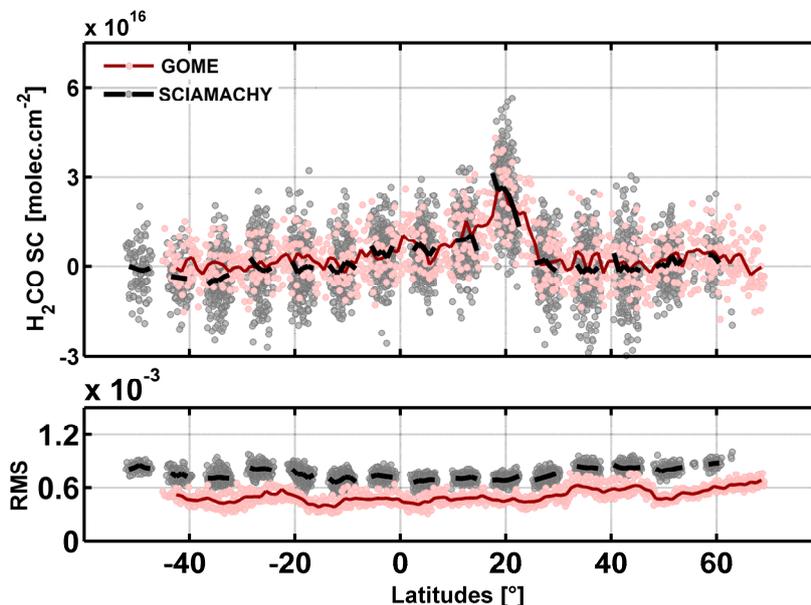
Figure 6-9: Formaldehyde slant columns (*SC*) and fit residuals (*RMS*) for overlapping GOME and SCIAMACHY orbits passing over South-eastern China on the 14th of April 2003. The darker lines are the averaged slant columns on a 1° latitude grid. A selection of 70° has been applied on the solar zenith angles. The standard deviations of GOME and SCIAMACHY slant columns are respectively 6x10[15] molec.cm[-2] and 1x10[16] molec.cm[-2].

6.2.2 Reference sector correction

Because a daily radiance spectrum is used as control spectrum for the DOAS fit (*I_0* in equation (6-1)), the retrieved slant columns actually represent the difference in slant column with respect to the slant column contained in the control spectrum. Hence the quantity of formaldehyde present in the reference spectrum must be estimated on a daily basis, based on suitable external information. Furthermore, an analysis of the retrieved slant column measurements reveals obvious zonally and seasonally dependent artefacts. Indeed, over the remote Pacific Ocean, the formaldehyde columns are often found to differ from the background levels attributed to methane oxidation by more than $1x10^{16}$ molec.cm[-2]. This is especially clear at latitudes higher than 60° and in March. This behaviour is caused by unresolved spectral interferences with ozone and BrO absorptions, which are very important at high solar zenith angles in spring. Fortunately, these interferences are small in regions where the sources of formaldehyde are significant. However, to reduce the impact of these artefacts, an absolute normalisation is applied on a daily basis using the reference sector method [*Khokhar et al.*, 2005]. The reference sector is chosen in the central Pacific Ocean (140°-160°W), where the only significant source of H_2CO is the CH_4 oxidation. On a daily basis, the latitudinal dependency of the H_2CO slant columns in the reference sector is modelled by a polynomial, subtracted from the slant columns and replaced by the H_2CO background taken from the tropospheric 3D-CTM IMAGES [*Müller and Stavrakou*, 2005] in the same region. The value of the

IMAGES H$_2$CO background added to the total column is plotted in Figure 6-21 and Figure 6-22. A similar normalisation approach has been applied in previous H$_2$CO studies [e.g. *Abbot et al.*, 2003].

6.2.3 Air mass factors determination

The second step in the retrieval of tropospheric H$_2$CO total columns is the calculation of the air mass factors necessary to convert slant columns into corresponding vertical columns. In the troposphere, scattering by air molecules, clouds and aerosols leads to complex altitude-dependent air mass factors. Full multiple scattering calculations are required for the determination of the weighting functions, and the vertical distribution of the absorber has to be known *a priori*. The total column air mass factor can be obtained from the formula [*Palmer et al.*, 2001]:

$$AMF = \int_{atm} w(z,\theta_0,\theta,\varphi,a_S,h_S,C_f,C_h) \cdot S(z,lat,long,month) \cdot dz \qquad \text{(6-2)}$$

where the scattering weighting function w represents the sensitivity of the satellite measurements to the molecule concentration as a function of the altitude z, sensitivity which varies with the solar and viewing zenith angles θ_0 and θ, the relative azimuth angle φ, the surface albedo a_S, the surface altitude h_S and the effective cloud fraction C_f and cloud altitude C_h. The shape factor S represents the normalised vertical profile of the absorbing molecule, which depends on the time and location of the measurement.

Shape factor

The shape factor is defined by:

$$S(z) = \frac{x_a(z)}{\int_{atm} x_a(z)dz} \qquad \text{(6-3)}$$

where x_a is the chosen *a priori* vertical concentration profile of formaldehyde.

Formaldehyde concentrations decrease with altitude owing to near-surface sources of short-lived NMVOC precursors, the temperature dependence of CH$_4$ oxidation, and the altitude dependence of photolysis. S varies depending on local NMHC sources, boundary layer depth, photochemical activity, and other factors. To resolve this variability, the monthly output of an updated version of the IMAGES chemical transport model [*Müller and Stavrakou*, 2005] optimized with respect to H$_2$CO production [*Stavrakou et al.*, 2009a] has been used to specify the vertical profile of the H$_2$CO distribution. A more detailed description of the IMAGES model is given in chapter 9. The IMAGES profiles have been validated with various aircraft measurements like the INTEX-A campaign over Northern America in July-August 2004 [*Singh et al.*, 2006]. The measurements were performed by two groups, the National Centre for Atmospheric Research (NCAR) and the University of Rhode Island (URI) and are publicly available at the NASA data centre website [www.air.larc.nasa.gov]. The URI dataset has systematically lower values than the NCAR data, by about 30% in the boundary layer. In Figure 6-10, the IMAGESv2 profiles are compared with the mean observed vertical distribution during this campaign over the Northern American continent (left panel) and over the Atlantic Ocean (right panel). Over the continent, the IMAGES modelled mixing ratios reproduce the boundary layer enhancement of H$_2$CO measured during the INTEX-A campaign, although the decline in the free troposphere is steeper than observed. Over the Atlantic, the IMAGES profile lie mostly between the values defined by the two datasets [*Stavrakou et al.*, 2009a]

and the observations are representative of a marine atmosphere where CH_4 is the dominant H_2CO precursor. The decrease of H_2CO concentration with altitude is less steep than over the continent and reflects principally the temperature dependence of CH_4 oxidation. The comparison of the IMAGES profiles with aircraft observations shows that the model provides useful information on the shape factors of H_2CO, capturing in particular the observed decline of concentrations between the boundary layer and the free troposphere and the difference in the slope of this decline between continental and marine atmospheres [*Palmer et al.*, 2001].
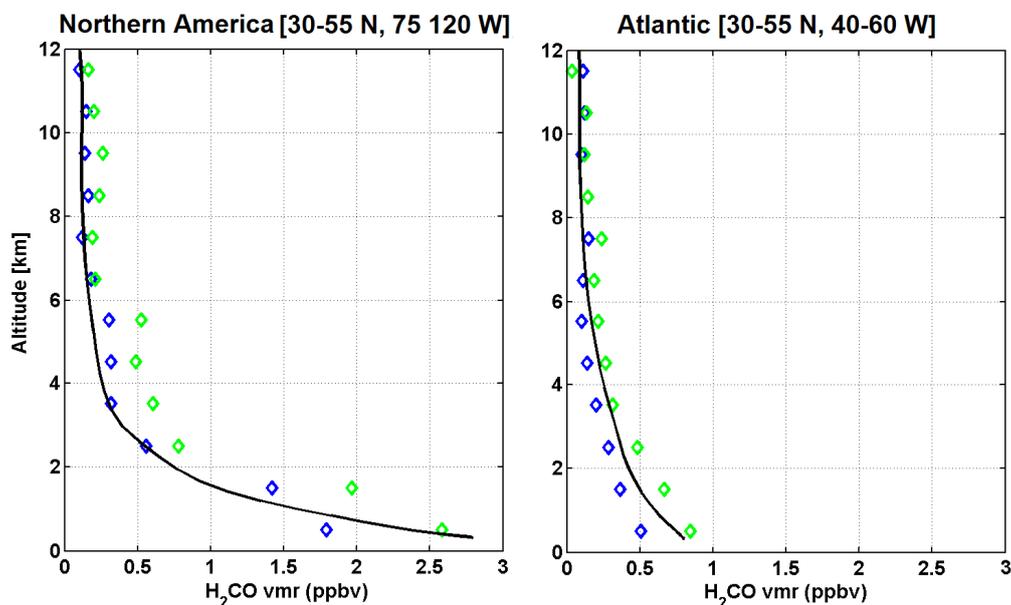
Figure 6-10: Comparison between the formaldehyde concentrations measured during the INTEX-A aircraft campaign (July–August 2004) and the corresponding values simulated with the IMAGES CTM (black line). The green and blue symbols correspond to measurement data from NCAR and Univ. Rhode Island, respectively.

Vertical H_2CO profile shapes are provided on a monthly basis between 1997 and 2008 at the resolution of 5°x5° and interpolated for each satellite geolocation. For the satellite measurements of 2009 and 2010, the IMAGES profiles of 2008 are used. To reduce the errors associated to topography and the rather broad spatial resolution of IMAGES, the *a priori* profiles are rescaled to effective ground altitude of the satellite pixel following *Zhou et al.* [2009].

Scattering weighting function

In general, *w(z)* depends on the wavelength as scattering and absorption processes vary with wavelength. However, the amplitude of the *w* variation is found to be small (less than 5% for SZA lower than 70°) in the 328.5-346 nm fitting window and a single air mass factor representative for the entire wavelength interval is used (at 340 nm). It should be noted that for solar zenith angles larger than 70°, the ozone absorption becomes so large that the fundamental DOAS hypothesis of an optically thin atmosphere is not fully satisfied, leading to tropospheric air mass factor variations of up to 10% within the wavelength interval. A limit for solar zenith angles larger than 70° is therefore applied to ensure the quality of the retrieval.

Figure 6-11 illustrates the dependency of *w(z)* with the viewing angles, *i.e.* θ_0 (a), θ (b), and φ (c), and with scene conditions like a_s (d) and h_s for a weakly (e) or highly reflecting surface (f). The decrease of sensitivity in the boundary layer is more important for large solar zenith angles and

wide instrumental line-of-sight angles. The azimuth angle does have relatively less impact on the measurement sensitivity. In the UV, surface not covered with snow present albedo lower than 0.1, while snow and clouds generally present albedo larger than 0.8. For a weakly reflecting surface, the sensitivity decreases near the ground because photons tend to be scattered before reaching the boundary layer. Increasing values of the surface albedo allow more solar radiation to be reflected to space, thus increasing measurement sensitivity to tropospheric absorbers (see Figure 6-11 (d)).
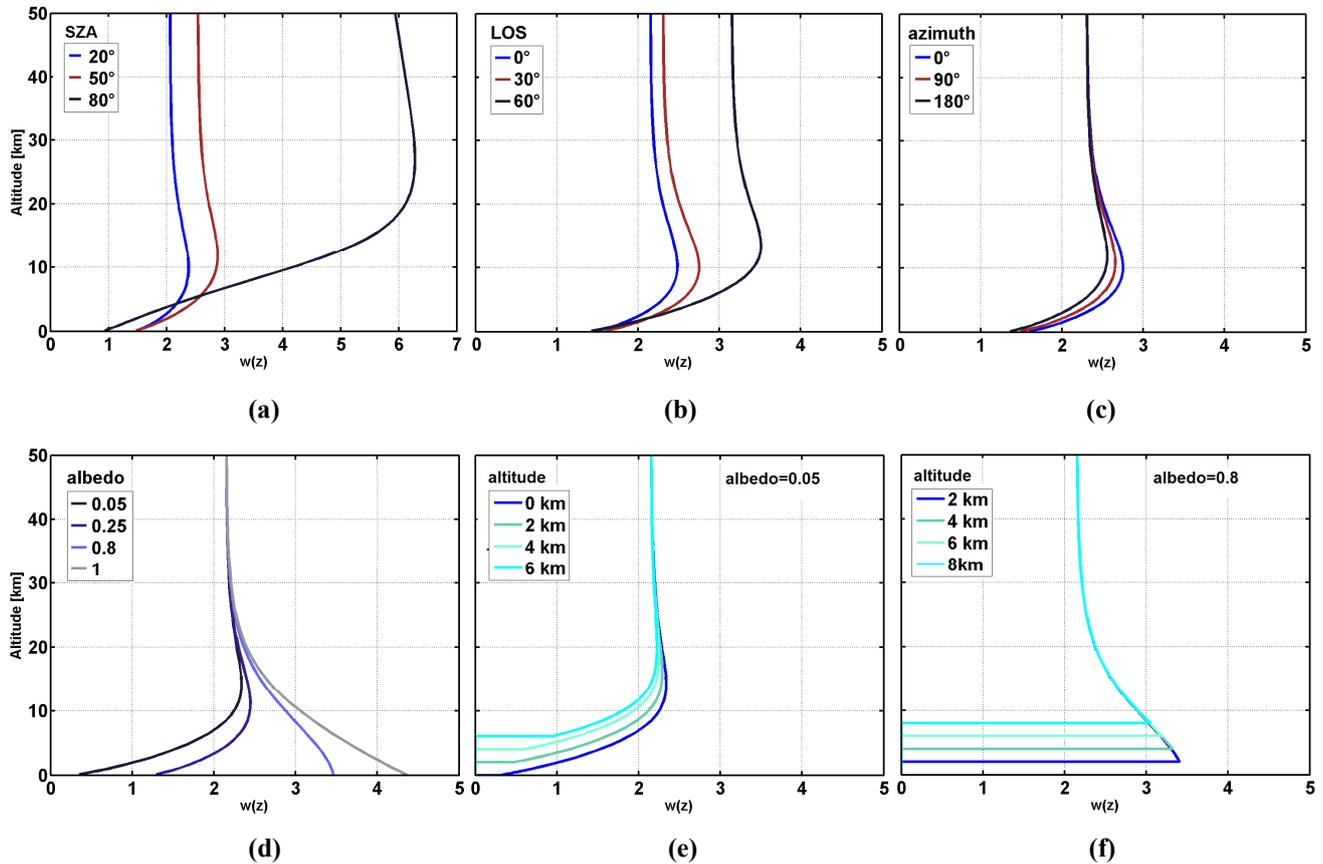
Figure 6-11: Variation of the weighting function with: (a) solar zenith angle, (b) viewing zenith angle or line-of-sight, (c) relative azimuth angle between the sun and the satellite, (d) surface albedo, (e) surface altitude for a weakly reflecting surface, (f) surface altitude for a highly reflecting surface. Unless specified, the parameters chosen for the radiative transfer simulations are: SZA=30°, LOS=0°, azimuth=0°, albedo=0.3, altitude=0km, λ=340nm.

For this work, weighting functions have been evaluated with the LIDORT v3.0 radiative transfer model [*Spurr*, 2008], at 340 nm, for a number of representative viewing geometries, surface albedos and altitudes, and stored in a look-up table. Weighting functions are then interpolated for each particular observation conditions. The parameter values chosen for the look-up table are detailed in Table 6-3. In particular, the grid of surface elevation is very thin below 5 km, in order to minimise interpolation errors caused by the generally low albedo of ground surfaces. Indeed, as illustrated by Figure 6-11 (e) and (f), the variation of the weighting functions is more discontinuous with surface elevation (low reflectivity) than with cloud altitude (high reflectivity). The distribution of surface albedo is taken from the climatology of *Koelemeijer et al.* [2003], which provides monthly Lambert-equivalent reflectivity at 335 nm at a resolution of 1°x1°. Surface elevation is taken from the ETOPO5 global gridded 5-minute database [*ETOPO5*, 1988], degraded to the resolution of 0.25°x0.25°, more comparable to the size of satellite ground pixels.

Table 6-3: Parameters in the weighting function lookup table

Parameter name	Grid of values
Solar zenith angle (SZA)	10, 20, 30, 40, 45, 50, 55, 60, 65, 70, 72, 74, 76, 78, 80, 85 [°]
Line of Sight zenith angle (LoS)	0, 10, 20, 30, 40, 50, 60 [°]
Relative azimuth angle (azim)	0, 45, 90, 135, 180 [°]
Surface albedo (a_S)	0, 0.03, 0.06, 0.1, 0.25, 0.5, 0.8, 1
Surface elevation (h_S)	0, 0.2, 0.4, 0.6, 0.8, 1, 1.2, 1.4, 1.6, 1.8, 2, 2.25, 2.5, 2.75, 3, 3.25, 3.5, 3.75, 4, 4.25, 4.5, 4.75, 5, 6, 9, 12, 15 [km]

Treatment of partly cloudy scenes

The applied cloud correction is based on the independent pixel approximation [*Martin et al.*, 2001], in which a inhomogeneous satellite pixel is considered as made of two independent homogeneous scenes, one completely clear and the other completely cloudy. The intensity measured by the instrument for the entire scene is decomposed into the contributions from the clear-sky and cloudy fractions, respectively weighted by (*1-C_f*) and C_f. Accordingly, the scattering weighting function of a partly cloudy scene is a combination of two weighting functions, calculated respectively for the cloud-free and cloudy fractions of the scene:

$$w = (1 - \Phi) \cdot w_{clear}(a_S^{ground}, h_S) + \Phi \cdot w_{cloud}(a_S^{cloud}, C_h) \tag{6-4}$$

where w_{clear} is the weighting function for a completely cloud-free pixel, w_{cloud} is the weighting function for a completely cloudy scene, and the intensity-weighted cloud fraction Φ is defined as:

$$\Phi = \frac{C_f \cdot I_{cloud}}{(1 - C_f) \cdot I_{clear} + C_f \cdot I_{cloud}} \tag{6-5}$$

I_{clear} and I_{cloud} are respectively the radiance intensities for clear-sky and cloudy scenes whose values are calculated with LIDORT and stored in look-up tables with the same grids as the weighting functions. w_{clear} and I_{clear} are evaluated for a surface albedo a_S^{ground} and a surface elevation h_S, while w_{cloud} and I_{cloud} are estimated for a cloud albedo a_S^{cloud} and at the cloud altitude C_h. In such weighting function calculations, a single cloud-top height (approximated by a Lambertian reflector) is assumed within a given viewing scene. This rather crude cloud description is consistent with assumptions that are often used in cloud retrieval algorithms [*Koelemeijer et al.*, 2002]. For our retrievals, the cloud parameters are obtained from the FRESCO v5 cloud product, which provides an effective cloud fraction and cloud top height assuming a Lambertian-reflecting cloud with an albedo of 0.8 yielding the same top-of-atmosphere radiance as the real cloudy scene [*Wang et al.*, 2008, *Koelemeijer et al.*, 2002].

Figure 6-12 shows three weighting functions calculated respectively for cloud free, half cloudy and totally cloudy pixel. For the totally cloudy pixel, the cloud act as a shield between the satellite and the ground surface and the measurement sensitivity equals zero below the cloud. In the same time, the sensitivity is high above the cloud because of its high reflectivity (albedo=0.8). The half cloudy weighting function is a combination of the cloud free and cloudy weighting functions as defined by (6-4). It should be noted that this formulation of the weighting function for a partly cloudy pixel implicitly includes a correction for the H₂CO column lying below the cloud and therefore not seen

by the satellite, the so-called "ghost column". Indeed, the total ***AMF*** calculation as expressed by (6-2) and (6-3), assumes the same shape factor and implies an integration of the *a priori* profile from the top of atmosphere to the ground, for each fraction of the scene. The ghost column information is thus coming from the IMAGES model. In the case of large cloud fractions, the result of the retrieval is therefore more influenced by the *a priori* profile than by the measurement itself. For this reason, result analysis should include only observations with moderate cloud fractions. It should also be noted that no explicit correction is applied for aerosols, but their impact is partially accounted for by the cloud correction scheme, as further explained in section 6.3.5.
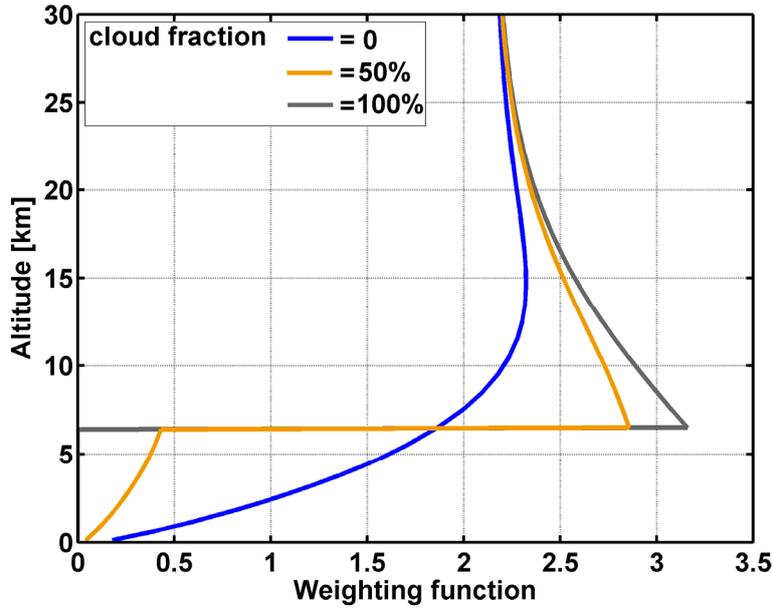
Figure 6-12: Weighting function for a cloud free pixel (in blue), a pixel with a cloud fraction of 0.5 (in orange) and a fully cloudy pixel (in grey). The cloud altitude is 6 km. In all cases, we consider a nadir view, a SZA of 30° and a ground albedo of 0.02. The a priori profile shape is a typical oceanic H₂CO profile. The tropospheric air mass factors are respectively 1.2, 0.69 and 0.53.

6.3 Error analysis

6.3.1 Expression of the total error on the vertical column

A formulation of the total error can be derived analytically starting from the equation of the vertical column which directly results from the different steps detailed in section 6.2:

$$VC = \frac{(SC - SC_0)}{AMF} + VC_0^{CTM} = \frac{\Delta SC}{AMF} + VC_0^{CTM} \tag{6-6}$$

In this expression, ***ΔSC*** represents the difference between the fitted slant column ***SC*** and the mean slant column in the reference sector ***SC₀***. ***VC₀^{CTM}*** is the IMAGES background in this reference region. ***AMF*** is the air mass factor. As these terms are determined independently, they are assumed to be uncorrelated. If we further assume normal probability distributions, the total error on the tropospheric vertical column can be expressed as [*Boersma et al.*, 2004]:

$$\sigma_{VC}^2 = (\frac{\partial VC}{\partial SC})^2 \cdot \sigma_{SC}^2 + (\frac{\partial VC}{\partial VC_0^{CTM}})^2 \cdot \sigma_{VC_0^{CTM}}^2 + (\frac{\partial VC}{\partial AMF})^2 \cdot \sigma_{AMF}^2 \tag{6-7}$$

In our case, it results from (6-6) that the total error can be derived from the following expression:

$$\sigma_{VC}^2 = \frac{1}{AMF^2} \cdot \frac{\sigma_{SC_{rand}}^2}{N} + \frac{1}{AMF^2} \cdot \sigma_{SC_{syst}}^2 + \sigma_{VC_0^{CTM}}^2 + (\frac{\Delta SC}{AMF^2})^2 \cdot \sigma_{AMF}^2 \qquad \text{(6-8)}$$

$\sigma_{SC_{rand}}$ and $\sigma_{SC_{syst}}$ are respectively the random and systematic parts of the error on the slant column, in the sense that the random part of the error can be reduced when the number of measurements increases, which is not the case of the systematic part. In case of averaging, $\sigma_{SC_{rand}}$ can be divided by the square root of the number of satellite pixels taken into the mean (N). σ_{AMF} is the error on the air mass factor evaluation and $\sigma_{VC_0^{CTM}}$ the error on the reference sector correction. Theses two latter sources of uncertainties have systematic but also random components that may average out in space or in time. However, these components can hardly be separated in practice and we will consider these uncertainties as systematic. The total error calculated with equation (6-8) is therefore an upper limit of the real error on the vertical columns.

6.3.2 Error on Slant Columns

Random error

The random error on the slant columns can be expressed by the standard deviation of measured columns around the mean value, but this applies only for regions of weak variability in the H₂CO columns. On the other hand, and provided that fitting residuals are dominated by shot noise in the detection system, the random error on the slant columns can be estimated from the properly weighted root mean square (**RMS**) of the fitting residuals, and that in all regions [*Taylor et al.*, 1982]. Therefore, we use **RMS** values as an estimate of the random error.
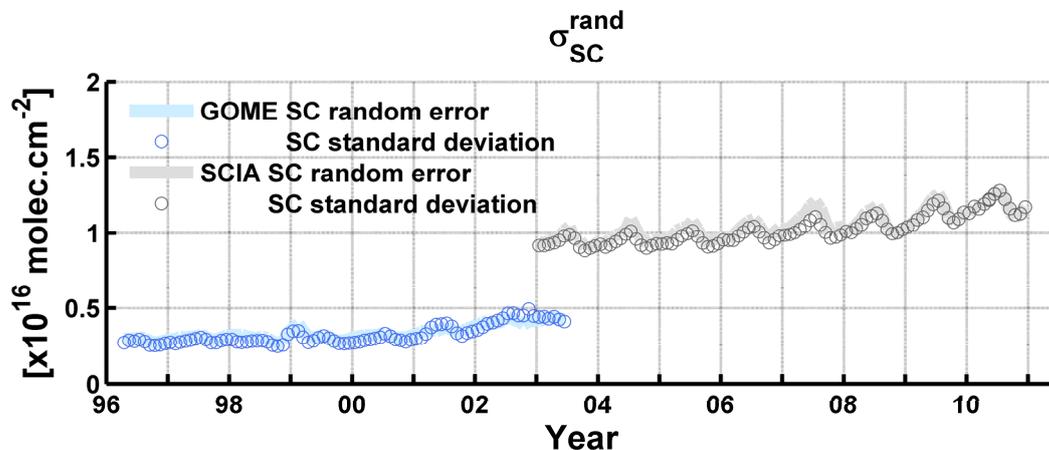
Figure 6-13: Standard deviation and random error of the monthly averaged H₂CO slant columns retrieved from GOME and SCIAMACHY measurements between 1996 and 2010 over a remote region at the equatorial Pacific (15°S-15°N, 160°-240° E).

Figure 6-13 compares the calculated standard deviation of the monthly averaged slant columns with the estimated random errors of GOME and SCIAMACHY retrievals over a remote region at the equatorial Pacific as a function of time. Due to the degradation of the GOME instrument throughout the years and the resulting lower signal to noise ratio of the spectra, $\sigma_{SC_{rand}}$ is found to increase from

4×10^{15} molec.cm^{-2} in 1996 up to 5×10^{15} molec.cm^{-2} in 2003. At the beginning of the SCIAMACHY time series, $\sigma_{SC_{rand}}$ (1×10^{16} molec.cm^{-2}) is twice as large as GOME because of the poorer signal to noise ratio of individual measurements, and increases also with time. In both cases, the degradation of the fitting residuals accelerates after 5-6 years of measurements. For single pixels, the random error on the slant columns is the most important source of error on the total vertical column. It can be reduced by averaging, but of course to the expense of a loss in time and/or spatial resolution.

Systematic error

The systematic errors on the slant columns are largely connected to uncertainties on the absorption cross-sections included in the fit, although other sources of systematic errors also contribute to the error budget, like uncertainties related to instrumental effects (wavelength calibration, spectral stray light, slit function…), or systematic misfit effects when for example, the ozone absorption becomes so strong that the DOAS approximation of an optically thin atmosphere is not fully satisfied anymore. Systematic slant column errors due to absorption cross-section uncertainties and their correlations can be estimated using the Rodgers formalism [*Rodgers et al.*, 2000] for a linear over-constrained problem like the DOAS inversion [*Theys et al.*, 2007; *Theys et al.*, 2010]:

$$\sigma^2_{SC_{syst}} = \sum_{j=1}^{m} SC_j^2 \left[G \cdot S_{bj} \cdot G^T \right] \qquad (6\text{-}9)$$

In this equation, the species index j (=1,…,m) runs over all molecules taken into account in the DOAS fit. G is the contribution matrix of the DOAS retrieval, constructed as $G = \left[K^T \cdot K \right]^{-1} \cdot K^T$ where K ($n\times m$, with n the number of wavelengths in the fitting interval) is a matrix formed by the absorption cross-sections. SC_j is the fitted slant column of the jth molecule and S_{bj} ($n\times n$) is the cross-section error covariance matrix. Any error on the amplitude of the H₂CO cross-section dataset will result in a systematic error on the retrieved slant columns. For the other absorbers included in the fit, only errors affecting the shape of the differential cross-sections will have a significant impact on the formaldehyde column retrieval. Estimates of such uncertainties have been taken from the literature and are summarized in Table 6-2. In order to valid the results obtained with this method of error estimate, retrieval sensitivity tests have been performed by selecting alternative cross-section datasets for each molecule, and assessing the resulting variability of the H₂CO SC. The two methods (sensitivity analysis versus matrix analysis) were found to produce consistent error estimates. However, although the formalism introduced here accounts for correlation effects between the absorption cross-sections, it does not include sources of systematic error other than the cross-section uncertainties. Therefore an additional contribution taken as 12% of the slant columns has been included to account for other sources of error. This value has been derived from retrieval tests, where the fitting windows (328.5-346 nm and 337.5-359 nm), the calibration options (e.g. the type of the slit function), and the polynomial order for the stray light correction were varied respectively.

The total systematic error budget on the H₂CO slant columns, resolved into its different contributions, is represented in Figure 6-14, for SCIAMACHY in March 2003. The total error increases from 2.5×10^{15} molec.cm^{-2} in the Tropics to 8×10^{15} molec. cm^{-2} at high latitudes. In the Tropics, fitting parameters like the retrieval window and the H₂CO absorption cross-section are the most important sources of uncertainty. Ozone, NO₂ and BrO cross-section errors have comparatively little impact in these regions. At higher latitudes, errors related to ozone absorption and the Ring effect are largely dominant, their contributions respectively reaching 7 and 3×10^{15}

molec.cm^{-2}. The error due to spectral interference with BrO absorption structures increases also at these latitudes because, as for ozone, BrO is more abundant poleward. The contributions from the H_2CO absorption cross-section as well as other fitting parameters are relatively constant meridianwise.
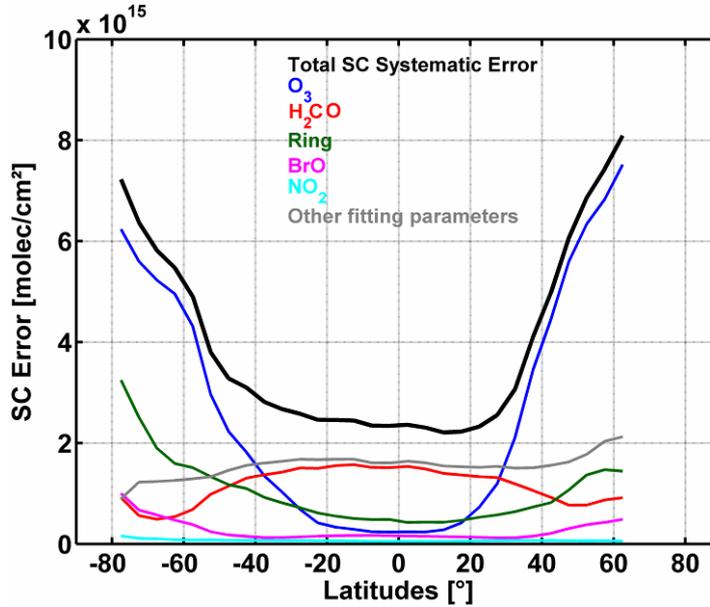
Figure 6-14: Systematic error on the SCIAMACHY H₂CO slant column (*SC*) averaged in March 2003 and on a latitude grid of 2°. The total error is shown in black, while the corresponding contributions from cross-sections uncertainties and other fitting parameters are plotted in colour.

6.3.3 Error on the reference sector correction

In equation (6-8), $\sigma^2_{VC_0^{CTM}}$ represents the uncertainty due to the reference sector correction. It has been evaluated as the monthly averaged differences between the IMAGES columns and the columns estimated using another tropospheric model, namely the TM model [*Eskes et al.*, 2003a], above the reference sector from 1997 to 2002. The errors range between 0.5 and 2×10^{15} molec.cm^{-2}. Therefore, $\sigma_{VC_0^{CTM}}$ is small compared to other error sources (see Figure 6-17).

6.3.4 Error on the air mass factor

Following the definitions in section 6.2.3, the error on the air mass factor depends on input parameters uncertainties and on the sensitivity of the air mass factor to each of them [*Boersma et al.*, 2004]:

$$\sigma^2_{AMF} = (\frac{\partial AMF}{\partial a_S} \cdot \sigma_{a_S})^2 + (\frac{\partial AMF}{\partial C_f} \cdot \sigma_{C_f})^2 + (\frac{\partial AMF}{\partial C_h} \cdot \sigma_{C_h})^2 + (\frac{\partial AMF}{\partial S} \cdot \sigma_S)^2 \qquad \textbf{(6-10)}$$

where $\sigma_{a_S}, \sigma_{C_f}, \sigma_{C_h}, \sigma_S$ represent respectively the uncertainties on the surface albedo, the cloud fraction, the cloud height and the profile shape. They are estimated from the literature or derived from comparisons with independent data, as summarized in Table 6-4. In order to evaluate the uncertainty on the profile shape, we have estimated the uncertainty on the peak altitude and its width, by comparing the model results with aircraft measurements performed during campaigns like

INTEX-A above the United States in 2004 [*Millet et al.*, 2006; *Singh et al.*, 2006] or TRACE-A in the Tropics in 1992 [*Fishman et al.*, 1996; *Emmons et al.*, 2000]. Note that in (6-10), the errors on the solar angles, the viewing angles and the ground altitude are here supposed to be negligible. In this expression, we have not considered the impact of possible correlations between uncertainties on parameters like for example the surface albedo and the cloud top height [*Boersma*, 2004]. We implicitly make the assumption that such uncertainties are random in their relative behaviour.

Table 6-4: Uncertainties on parameters included in the air mass factor evaluation.

AMF parameter	Uncertainty	Source
Albedo	0.02	Koelemeijer et al., 2003
Cloud fraction	0.05	Koelemeijer et al., 2002
Cloud height	1 km	Koelemeijer et al., 2002
Altitude of profile maximum	100 m	Validation
Width of profile	500 m	Validation

The air mass factor sensitivities, *i.e.*, the air mass factor derivatives with respect to the different input parameters, have been evaluated for different representative conditions of observation and for different profile shapes. Figure 6-15 (a) shows the air mass factor dependence on the ground albedo for two different formaldehyde profile shapes (in blue: profile typical of remote area, in red: typical of emission regions). The *AMF* sensitivity (the slope), is almost constant with albedo, being only slightly higher for low albedo values. As expected, the *AMF* sensitivity to albedo is found to be higher for an emission profile peaking near the surface than for a background profile more spread in altitude. As illustrated in Figure 6-15 (b), the sensitivity to the cloud fraction is significant only when the cloud lies below the H₂CO layer. Indeed, for this figure, the cloud altitude (1 km) is located above the formaldehyde maximum in the polluted case (in red) but below the main formaldehyde layer in the background case (in blue). Figure 6-15 (c) shows the air mass factor variation with cloud altitude. The slope is large when the cloud height is located below or at the altitude of the formaldehyde peak. For higher clouds, the sensitivity of the air mass factor to any change in cloud altitude is very weak.

(a)   (b)   (c)
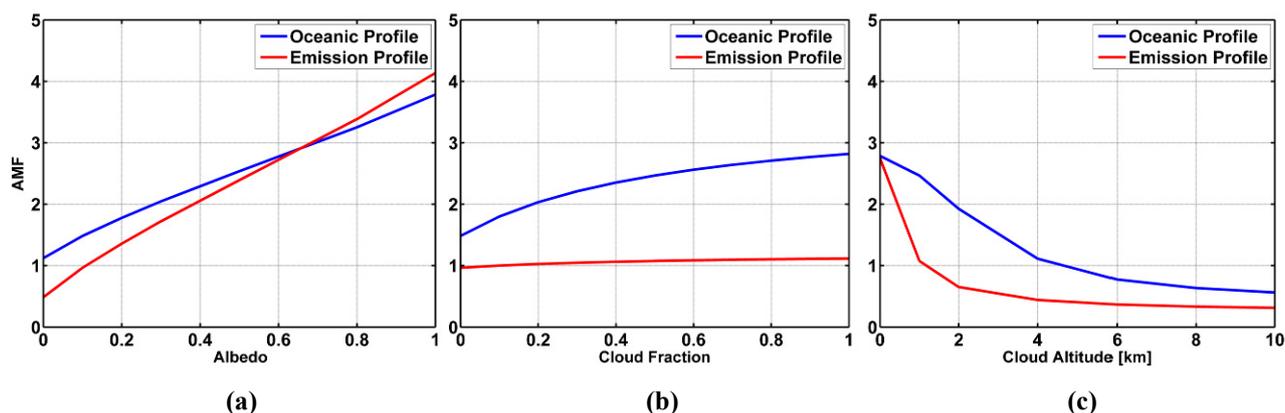
Figure 6-15: Dependence of the tropospheric air mass factor (AMF) on (a) the surface albedo, (b) the cloud fraction and (c) the cloud altitude. The blue lines correspond to a H₂CO profile shape typical of remote oceanic areas and the red lines correspond to a profile peaking at the surface, typical of emission areas. In all cases, we consider a nadir view and a solar zenith angle of 30°. In (a) the pixel is cloud free, in (b) the ground albedo is 0.02 and the cloud altitude is 1 km, in (c) the albedo is 0.02 and the cloud fraction is 0.5.

The contribution of each parameter to the total air mass factor error depends on the observation conditions. Figure 6-16 displays the total error on the air mass factor as a function of the cloud fraction in two cases: when (a) the cloud altitude is at 8 km or (b) at 1 km. The contribution from each parameter to the total error is also displayed. For clear sky conditions, the total air mass factor error is estimated to be 18% with almost equivalent contributions from the profile shape, the albedo and the cloud fraction uncertainties. For a cloud fraction of 0.5, the total error reaches 30% for a high cloud and 50% for a low cloud. The most important error sources are uncertainties on the cloud altitude and on the profile shape for low clouds (< 2 km), particularly over emission regions. Other sources, like the albedo and the cloud fraction uncertainties represent smaller contributions to the total error on the air mass factor. For high clouds, the most important error source is the profile shape uncertainty as their associated measured column is mostly influenced by the *a priori* profile via the ghost column correction. It is therefore recommended not to use the pixels strongly contaminated by clouds. In practice, we filter out all pixels with a cloud fraction larger than 40%, this threshold resulting from a trade-off between number of measurements and limited cloud contamination.
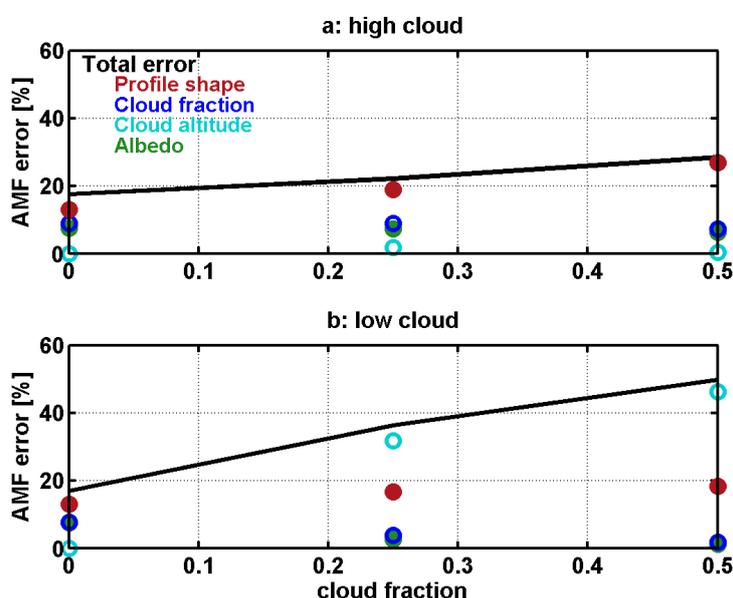
Figure 6-16: Uncertainties on the air mass factor (AMF) for (a) a high cloud (8 km) and (b) a low cloud (1 km). In all cases, the SZA is 30° and the surface albedo is 0.02. The H₂CO vertical distribution is typical of emission areas, peaking near the surface. The total error is represented by black lines and the contributions of the different parameters are represented as coloured circles.

In this study, we have not explicitly considered the effect of aerosols on the air mass factors. To a large extent, however, the effect of the non-absorbing part of the aerosol extinction is implicitly included in the cloud correction [*Boersma et al.*, 2004]. Indeed, in the presence of aerosols, the cloud detection algorithm FRESCO is expected to overestimate the cloud fraction. Since non-absorbing aerosols and clouds have similar effects on the radiation in the UV-visible range, the omission of aerosols is partly compensated by the overestimation of the cloud fraction, and the resulting error on air mass factor is small, typically below 16% [*Boersma et al.*, 2004; *Millet et al.*, 2006]. In some cases however, the effect of clouds and aerosols will not be the same. For example, when the cloud height is significantly above the aerosol layer, clouds will have a shielding effect while the aerosol amplifies the signal through multiple scattering. This will result in an

underestimation of the AMF. Absorbing aerosols have also a different effect on the air mass factors, since they tend to decrease the sensitivity to H_2CO concentration. In this case, the resulting error on the air mass factor can be as high as 30% [*Palmer et al.*, 2001; *Martin et al.*, 2002]. This may, for example, affect significantly our derivation of H_2CO columns in regions dominated by biomass burning as well as over heavily industrialized regions. This is another reason for not using pixels strongly contaminated by clouds.

6.3.5 Total error on the H₂CO vertical column

From equation (6-8), the total error on the vertical column can be calculated for each satellite pixel. As an illustration, Figure 6-17 shows the total error on the monthly and zonally averaged vertical columns, as a function of latitude, calculated for March 2003. The number of SCIAMACHY pixels involved in these averages being generally large, the contribution from slant column random errors is below 2%. Note that, when considering individual pixels, the total error is generally much larger (about 70%) and dominated by the random error. The error due to the reference sector correction ranges from 5 to 12%. At low and mid latitudes, the contributions from the slant columns and from the air mass factor uncertainties are equivalent, with values between 10 and 20%. At higher latitudes, the error due to slant column uncertainties is larger (15-40%) mainly due to the typically larger ozone absorptions. The total error on H_2CO monthly means generally ranges between 20 and 40% for both GOME and SCIAMACHY data. This estimation is in agreement with the value of 25-30% deduced by *Millet et al.* [2006] who compared air mass factors calculated with the GEOS-Chem model and air mass factors derived from aircraft measurements over North America in summer 2004.
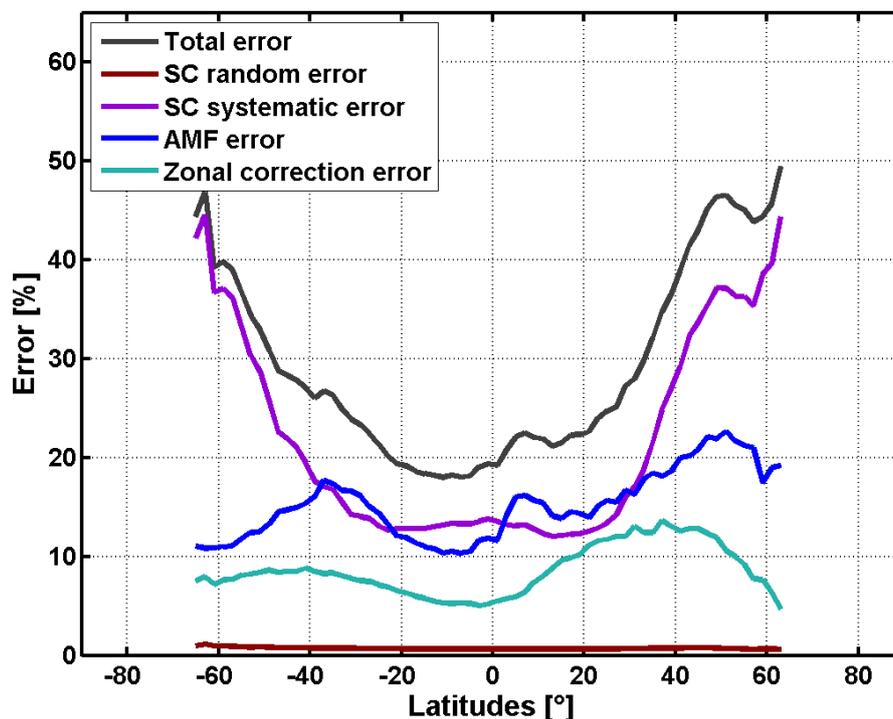

Figure 6-17: Error budget for the monthly and zonally averaged H₂CO vertical columns retrieved from SCIAMACHY in March 2003. The black line represents the total error. The coloured lines represent the contribution of errors on the slant column density (*SC*), the air mass factor (*AMF*), and the zonal correction.

6.3.6 Averaging kernels

Retrieved satellite quantities always represent a weighted average over all parts of the atmosphere that contribute to the signal observed by the satellite instrument. The DOAS total column retrieval is implicitly dependant on the *a priori* tracer profile x_a. Indeed, the radiative transfer calculation accounts for the sensitivity of the measurement to the H₂CO concentrations at all altitudes and the sensitivities are weighted with the assumed *a priori* profile shape to produce the vertical column. The averaging kernel (*AK*) is proportional to this measurement sensitivity profile, and provides the relation between the retrieved column *VC* and the true tracer profile *x* [*Rodgers*, 2000; *Rodgers and Connor*, 2002; *Eskes and Boersma*, 2003b]:

$$VC - VC_a = AK \cdot (x^{pc} - x_a^{pc}) \tag{6-11}$$

where the profiles are expressed in partial columns.

Column averaging kernels are particularly useful when comparing measured columns with *e.g.* model simulations, because they allow removing the effect of the *a priori* profile shape information used in the retrieval [*Boersma et al.*, 2004]. Following the definition for total column observations of optically thin absorbers [*Eskes and Boersma*, 2003b], DOAS averaging kernels are calculated as follows:

$$AK(z) = \frac{w(z)}{AMF} \tag{6-12}$$

where *w* is the weighting function calculated from equation (6-4), and *AMF* is the tropospheric air mass factor calculated from equation (6-2). As discussed above, the air mass factor and therefore the retrieved vertical column, depends on the *a priori* profile shape, in contrast to the weighting function which describes the sensitivity of the slant column to changes in trace gas concentrations at a given altitude and does not depends on the *a priori* profile in an optically thin atmosphere. From the definition of *AK*, we have $VC_a = AK \cdot x_a^{pc}$ and Equation (6-11) simplifies to:

$$VC = AK \cdot x^{pc} \tag{6-13}$$

The averaging kernel varies with the observation conditions. In the H₂CO product created during this work, the *AK* is provided together with the error budget for each individual pixel. The provided H₂CO vertical columns can be used in two ways, each with its own associated error [*Boersma et al.*, 2004]:

1. For independent study and/or comparison with other independent measurements of total column amounts. In this case, the total error related to the column *VC* = *SC/AMF* consists of slant column measurement errors and air mass factor errors. The latter consists of errors related to uncertainties in the assumed profile x_a and errors related to the *w* parameters as developed in section 6.3.1.

2. For comparisons with chemistry transport models or validation with independent profile measurements, if the averaging kernel information is used, the *a priori* profile shape error no longer contributes to the total error. Indeed, the relative difference between the retrieved column *VC* and an independent profile x_i is: $\delta = \dfrac{VC - AK \cdot x_i^{pc}}{VC}$. The total *AMF* cancels

since it appears as the denominator of both **VC** and **AK**. Because only the total **AMF** depends on x_a, the comparison using the averaging kernel is not influenced by the chosen *a priori* profile shape. The comparison error consists of error contributions from the slant column and from the weighting functions. The *a priori* profile error not influences the comparison, but of course, it still does influence the error on the retrieved vertical column.

6.4 Results and discussion

6.4.1 GOME and SCIAMACHY tropospheric vertical columns

The global distribution of the yearly averaged H$_2$CO columns from GOME and SCIAMACHY between 1997 and 2010 are displayed in Figure 6-18 and Figure 6-19. Both instruments capture the regions of elevated NMVOC emission which are mainly located in the Tropics, in particular over Amazonia, Africa and Indonesia. The Eastern United States and Southeastern Asia are also major regions of emission. The interannual variability in the concentrations is related to fire events like, for example in Indonesia in 1997, in Alaska in 2004 or in Russia in 2010, or to particularly high temperatures such as in Europe in summer 2003 or in the US in 2007. The better spatial resolution of SCIAMACHY allows observing more localized hotspots, like for example in the Chinese regions of Beijing and Chongqing and in the Indo-Gangetic plains. In South America, the impact of the South Atlantic Anomaly (SAA)[2], *i.e.* the larger error and scatter in H$_2$CO columns which results from exposure of the satellite instrument to enhanced radiation and high energy particles, is more pronounced in the case of SCIAMACHY. The reason for the difference of behaviour between GOME and SCIAMACHY is currently not fully understood. It might be related to differences in the design of the detectors used in both instruments (mainly the electronic gain and the instrument throughput for photons). Differences in the instruments shielding could also possibly explain the observed different sensitivities to SAA. For these maps, pixels have been selected for solar zenith angles below 60°, cloud fractions lower than 40% and fitting residuals lower than 1.5×10^{-3}.

---

[2] The South Atlantic Anomaly (SAA) refers to the area where the Earth's magnetic field is weakest. This leads to an increased flux of energetic particles in this region and exposes orbiting satellites to higher than usual levels of radiation.
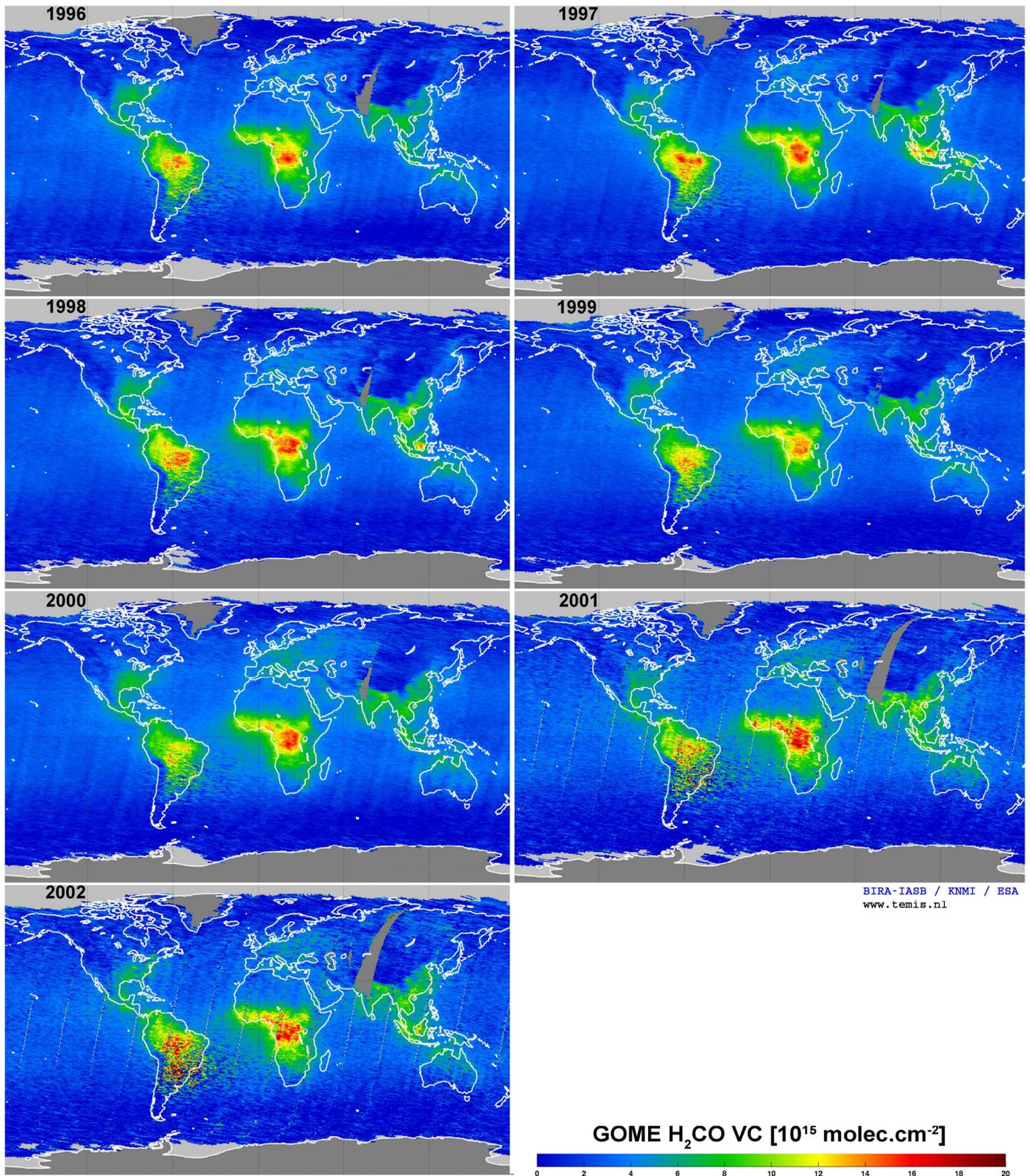
**Figure 6-18: Yearly averaged GOME H₂CO columns from 1996 to 2002.**
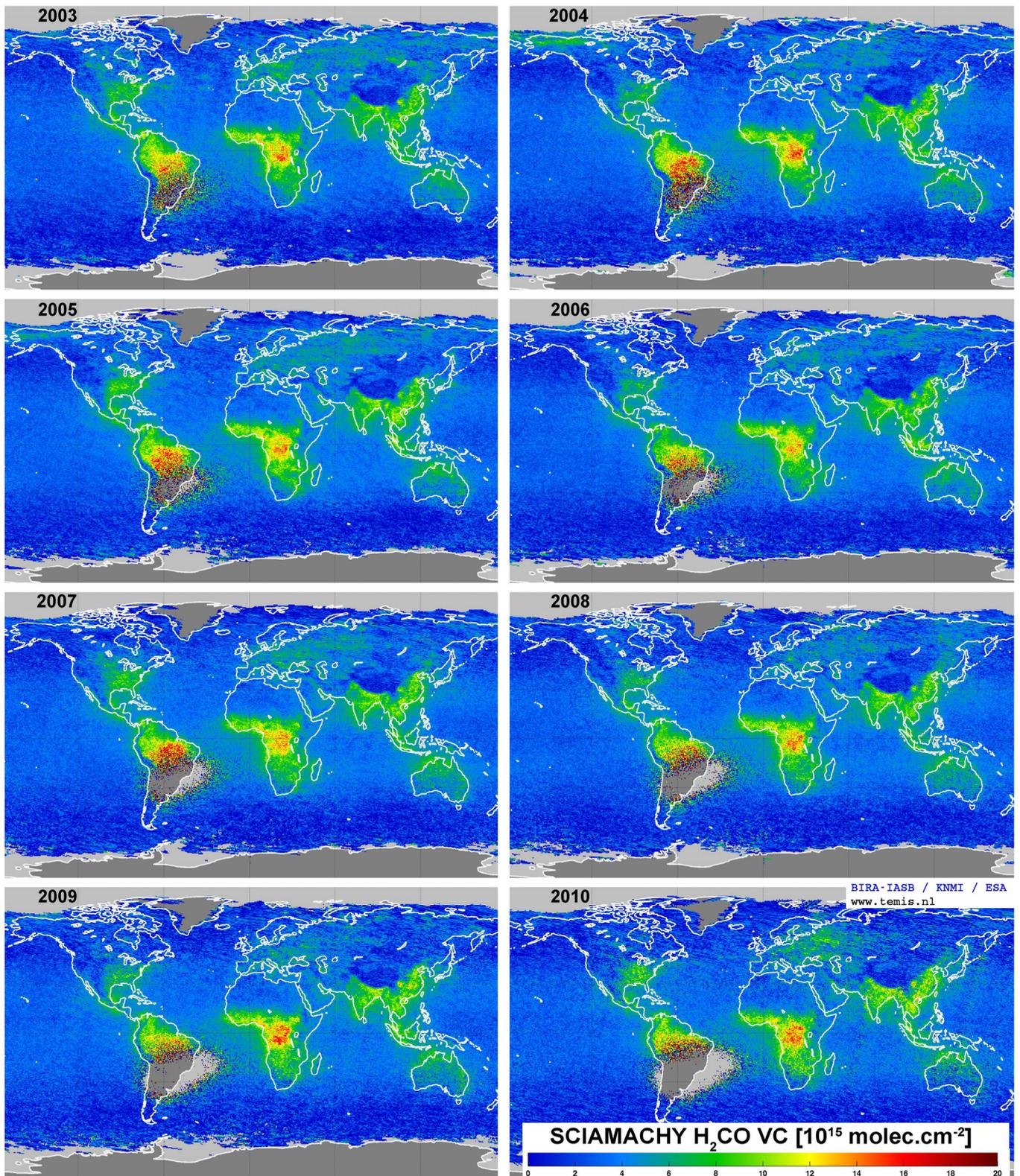
Figure 6-19: Yearly averaged SCIAMACHY H₂CO columns from 2003 to 2010.

Turning from annual global distribution to regional monthly means, H$_2$CO columns over areas generally characterized by large NMVOC emissions are presented in Figure 6-21 (for the American and African continents) and Figure 6-22 (for Asia, Europe and Australia). The region boundaries are displayed in Figure 6-20. The same observation selection criteria as for the maps have been applied. In particular, cloud fractions are lower than 0.4. We find that using alternate cloud fraction thresholds of 0.3 or 0.5 does not alter significantly the mean H$_2$CO slant column statistics over either continental or oceanic regions. The value of the reference sector correction (VC_0^{CTM}) is also shown on the plots, indicating the H$_2$CO detection limit. This allows discriminating the regions (or time periods) where the total column is constrained by the satellite observations themselves rather than by the model background. In the Tropics, the observed H$_2$CO columns are well above the detection limit all over the year. In mid-latitudes regions, the summer H$_2$CO observations are also significant. In winter, however, they are generally close to, or just above, the detection limit defined by the reference sector correction value. This limitation is explained by the low H$_2$CO emissions in mid-latitude regions during wintertime coupled to the fact that the sensitivity of the satellite retrieval in the boundary layer decreases when the solar zenith angle increases.

The general features of the seasonal cycle of formaldehyde columns are generally well reproduced during the whole period, suggesting that SCIAMACHY data consistently extends the GOME time series despite their larger random errors. In all the regions of the American continent (Eastern US, Guatemala, Amazonia) and in Tropical Asia (Thailand and Indonesia), the agreement between the GOME and SCIAMACHY columns averaged over the whole period of measurement is better than 7%. In Australia, China and India, we find positive offsets between the GOME and SCIAMACHY time-series, of respectively 20%, 18% and 14%. In Europe and Africa, negative offsets of about -10% are found between GOME and SCIAMACHY. The reasons for these offsets are not clearly identified (they will be further investigated in chapter 8), but they are well below the estimated errors on the vertical columns in the corresponding regions. The distribution of the formaldehyde columns is detailed in the next subsections for each region of interest. A comprehensive comparison between this dataset and the formaldehyde columns simulated with the IMAGES global chemical transport model over regions dominated by biomass burning and biogenic sources is presented in *Stavrakou et al.* [2009a]. The results of these comparisons will be summarized in chapter 9.
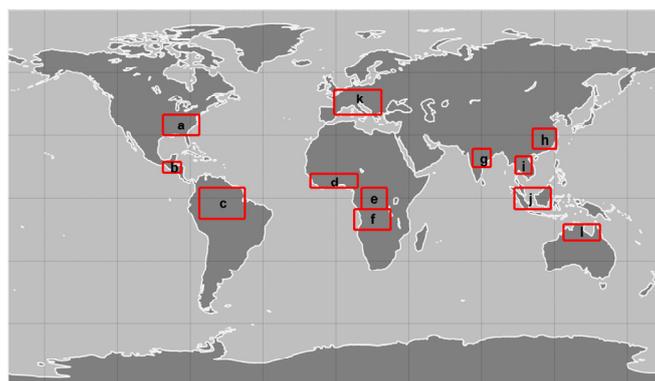
Figure 6-20: Limits of the regions chosen to calculate the H₂CO means in Figure 6-21 and Figure 6-22.
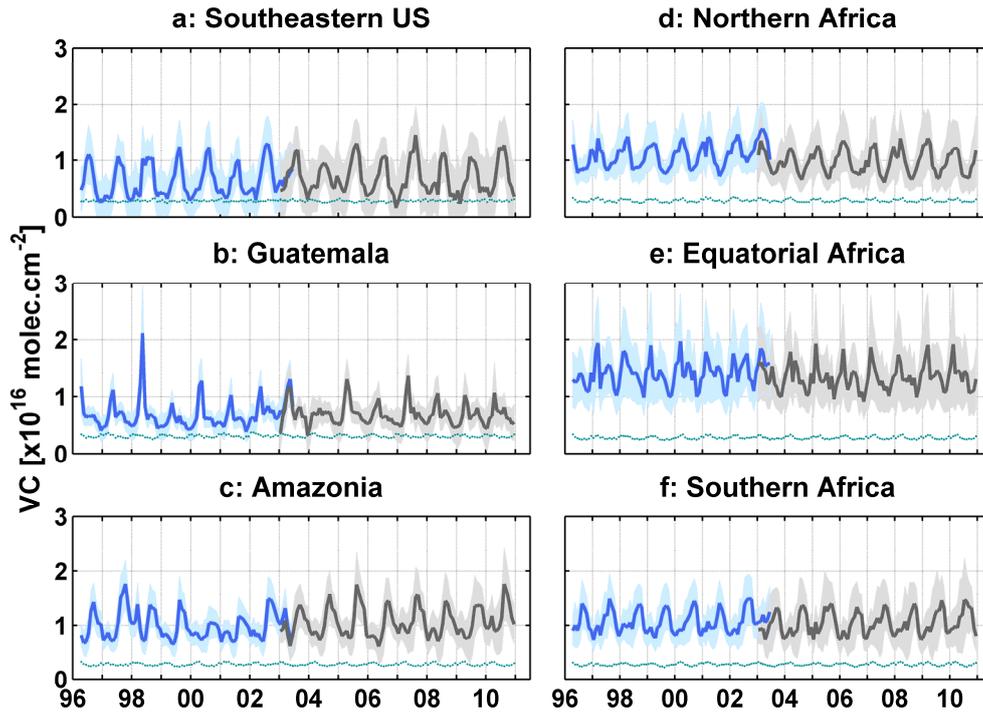
Figure 6-21: Time series of H₂CO vertical columns (*VC*) monthly averaged over regions of the American (a,b,c) and African (d,e,f) continents. GOME data are in blue and SCIAMACHY data are in grey. The monthly total errors on the vertical columns are displayed in light colour. The cyan dotted line is the IMAGES column in the reference sector (VC_0^{CTM}).
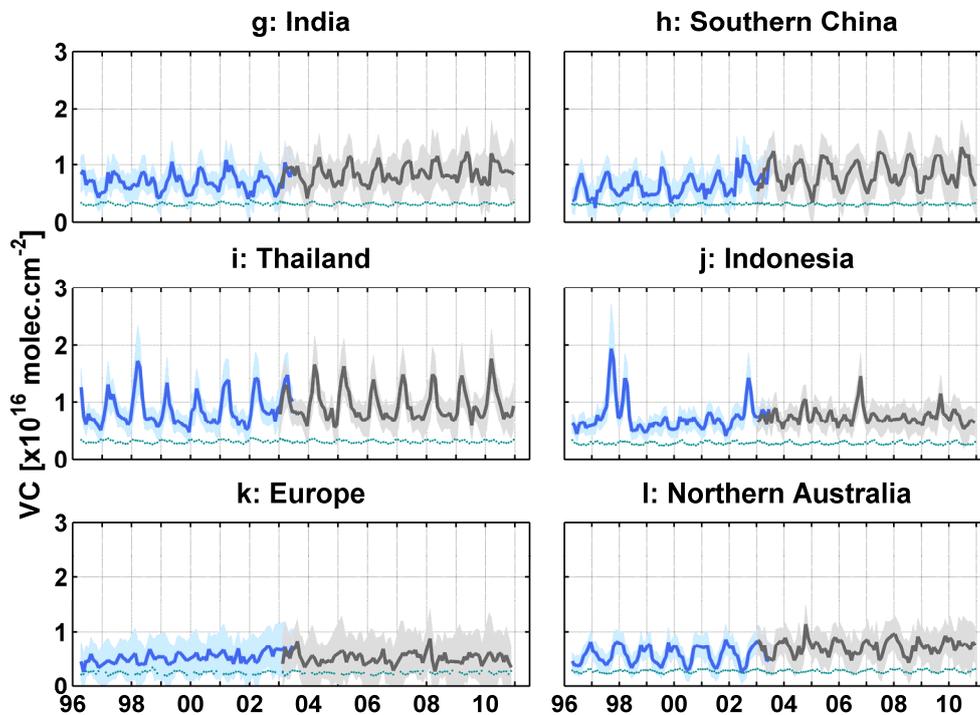
Figure 6-22: same as Figure 6-21 for regions of Asia (g, h, i, j), Europe (k) and Australia (l).

6.4.2 America

Over the Southeastern United States, the formaldehyde columns show a strong seasonal variation primarily related to the oxidation of biogenic VOCs (mainly isoprene) emitted during the summer season [*Abbot et al.*, 2003; *Stavrakou et al.*, 2009a]. The monthly means range from $3x10^{15}$ molec.cm^{-2} in winter (close to the detection limit) to a maximum of about $13x10^{15}$ molec.cm^{-2} in summer (Figure 6-21 (a)). The GOME and SCIAMACHY H$_2$CO columns present a remarkable agreement (better than $0.4x10^{15}$ molec.cm^{-2}). The GOME results agree qualitatively, if not quantitatively, with the series of investigations conducted by the modelling team of Harvard University on the basis of another GOME retrieval [*Palmer et al.*, 2003; *Abbot et al.*, 2003; *Martin et al.*, 2004; *Palmer et al.*, 2006]. The amplitude of the seasonal variation in the H$_2$CO columns derived in our study over North America in 1997 is smaller than in the Harvard retrieval [*Chance et al.*, 2000], with values about $4x10^{15}$ molec.cm^{-2} lower in summer. The interannual variability of the summer columns is very consistent in both datasets between 1996 and 2001, except for the 20% to 30% lower values in our data. It should be noted that the Harvard retrieval has been performed using the *Cantrell* [1990] H$_2$CO cross-section dataset, leading to H$_2$CO columns about 10% higher than those retrieved with the *Meller and Moorgat* [2000] dataset. It is also worth noting that aircraft measurements performed over Texas in the summer of 2000 suggested an overestimation of the Harvard GOME data of about $5.5x10^{15}$ molec.cm^{-2} [*Martin et al.*, 2004]. The GOME-derived isoprene fluxes over North America have been found to be roughly consistent with the flux estimations of the MEGAN model [*Guenther et al.*, 2006], although with GOME emissions typically 25% higher or lower at the beginning or the end of the growing season, respectively [*Palmer et al.*, 2006]. In contrast with this result, the modelling study of *Stavrakou et al.* [2009a] using the IMAGES model and the formaldehyde dataset presented in this study concludes that the MEGAN isoprene fluxes might be largely overestimated in this region. Although the main reason for this discrepancy is the lower H$_2$CO columns derived in the present study, it is also partly due to model differences, in particular regarding the chemical mechanism and the formaldehyde yield in the oxidation of isoprene by OH, about 20-30% higher in IMAGES than in the GEOS-Chem model [*Stavrakou et al.*, 2009a].

Over Tropical America (Figure 6-21 (b) and (c)), and more generally over Tropical forests and savannas, biomass burning emissions contribute significantly, although sporadically, to H$_2$CO abundances. The spectacular enhancement of the monthly averaged vertical column over Guatemala in May 1998 is an extreme case, since the formaldehyde column at this time ($22x10^{15}$ molec.cm^{-2}) was a factor of 4 higher than the usual wet season values ($5x10^{15}$ molec.cm^{-2} between November and February), and about a factor of 2 higher than the dry season peaks of other years. Over Amazonia (Figure 6-21 (c)), the H$_2$CO columns lie between $8x10^{15}$ molec.cm^{-2} during the wet season and $18x10^{15}$ molec.cm^{-2} during the fire season extending generally from August to November. As in the Southeastern United States, no significant offset is found between the GOME and SCIAMACHY datasets in these two regions. Our retrieved columns in these regions agree within 10% with the SCIAMACHY H$_2$CO columns derived by the IUP-Bremen group for the year 2005 [*Wittrock et al.*, 2006]. However, our columns are about 30% lower than the corresponding values of the Harvard dataset used by *Shim et al.* [2005] to constrain isoprene emissions at the global scale.

6.4.3 Africa

The seasonal variations of H$_2$CO abundances in Northern Africa (Figure 6-21 (d)) and in Southern Africa (Figure 6-21 (f)) are quite similar, except for a time shift of 6 months. The amplitude of the seasonal cycle, of about $6x10^{15}$ molec.cm^{-2} between the dry season and the wet season, is of the same order in the two regions. As in other Tropical regions, biomass burning is very probably the main cause of the dry season maximum. Over Tropical forests, however, biogenic isoprene emissions are also expected to peak during the dry season [*Müller et al.*, 2008] and could therefore contribute to the observed dry season enhancement of the formaldehyde columns. In Equatorial Africa (Figure 6-21 (e)), the H$_2$CO columns present two maxima every year, in February and in July-August, corresponding to an equatorial local climate with two dry seasons and two wet seasons [*Stavrakou et al.*, 2009a]. The interannual variability is weak, the columns being always comprised between 10 and $20x10^{15}$ molec.cm^{-2}. In these three regions, negative offsets of about $-1x10^{15}$ molec.cm^{-2} are found between the GOME and SCIAMACHY datasets, the largest being in Northern Africa. Still, these offset values represent less than 10% of the averaged H$_2$CO column in Africa. Our H$_2$CO data are about 25% lower than in the GOME Harvard dataset between September 1996 and August 1998 [*Shim et al.*, 2005] but 15% higher than the SCIAMACHY Bremen values in 2005 [*Wittrock et al.*, 2006]. Above the Sahara, our H$_2$CO values are at the same level as the oceanic background, in contrast with previous studies which found lower values above deserts [*Shim et al.*, 2005; *Wittrock et al.*, 2006]. This difference is seemingly due to differences in the fitting window used in the retrievals (see section 6.2.1). The behaviour observed over Sahara with our data is better in agreement with simulations of tropospheric models like TM [*Wittrock et al.*, 2006], GEOS-Chem [*Shim et al.*, 2005] or IMAGES [*Stavrakou et al.*, 2009a].

6.4.4 Asia

Like over the Eastern United States, the H$_2$CO columns over Southern China (Figure 6-22 (h)) show a strong seasonal variation mainly associated with biogenic VOC emissions, with values ranging from $5x10^{15}$ molec.cm^{-2} in winter to $12x10^{15}$ molec.cm^{-2} in summer. In this region, our data are about 20% lower than the GOME Harvard dataset in summer [*Fu et al.*, 2007]. A positive offset is also noted between the GOME and SCIAMACHY winter values, of the order of $1.2x10^{15}$ molec.cm^{-2}. Furthermore, an apparent trend in the GOME winter minima is found between 1996 and 2000, corresponding to an increase of 20% of the columns during that period. The same apparent trend is present in the SCIAMACHY data. It would be tempting to attribute this increase to anthropogenic emissions, in a country where NO$_x$ emission trends could be detected from spaceborne NO$_2$ measurements [*Richter et al.*, 2005; *van der A et al.*, 2006]. However, as seen in Figure 6-22, a trend is also apparent in the GOME winter values over Europe (where anthropogenic emissions are not expected to increase). The impact of the long-term degradation of GOME might possibly be enlarged over these mid-latitude regions, especially during winter, due to the high solar zenith angles encountered by the satellite and the corresponding measurement noise increase at low photon rates. Further studies based on robust statistical analysis of the time series are needed to discriminate real trends in H$_2$CO columns from artefacts. This will be the subject of chapter 8 [*De Smedt et al.*, 2010].

Over Thailand (Figure 6-22 (i)), the formaldehyde maxima occurring between February and April each year is about $15x10^{15}$ molec.cm^{-2}, more than a factor of two above the levels found during the rest of the year. Over Indonesia (Figure 6-22 (j)), the H$_2$CO levels show an unusually strong interannual variability related to the massive VOC emission associated to forest fires during El Niño

events like in 1997-1998, 2002 and 2006. In those two regions, the offset is negligible between GOME and SCIAMACHY.

In India (Figure 6-22 (g)), maximum values around 10×10^{15} molec.cm^{-2} are found in the period May-July. They are attributed to biomass burning. A second maximum of smaller amplitude (8×10^{15} molec.cm^{-2}) is found in September-October, and is attributed to biogenic emissions [*Fu et al.*, 2007]. The minimum values are of the order of 5×10^{15} molec.cm^{-2}. Like in Southern China, an offset of 0.9×10^{15} molec.cm^{-2} is found between GOME and SCIAMACHY, and a trend is apparent in the two time series. Our yearly averaged columns over India for 2005, about $8\text{-}9 \times 10^{15}$ molec.cm^{-2}, are well above the values ($<6 \times 10^{15}$ molec.cm^{-2}) determined by *Wittrock et al.* [2006]. Our H₂CO values agree with the Harvard dataset [*Fu et al.*, 2007] during winter, but they are again 20% lower in summer.

6.4.5 Australia

As shown in Figure 6-22 (l), the Northern Australia region presents summertime maxima (Nov.-Feb.) typical of biogenic emissions. Their values, around 9×10^{15} molec.cm^{-2}, are somewhat lower than in other regions dominated by biogenic emissions. A significant offset of about 1.3×10^{15} molec.cm^{-2} is found between the GOME and SCIAMACHY values. In this region, we find an important negative bias (70%) with respect to the Harvard dataset and a good agreement with the Bremen results (within 15%) [*Wittrock et al.*, 2006].

6.4.6 Europe

Over Europe, the columns range typically from 2×10^{15} molec.cm^{-2} in winter to 6×10^{15} molec.cm^{-2} in summer, although the seasonal cycle is not clearly defined in the data. The SCIAMACHY data are noisier than GOME, and display an offset of -0.5×10^{15} molec.cm^{-2} (10%). The value of this offset is largely below the estimated errors on the vertical columns, which range between 50 and 100% in Europe. Although there seems to be a positive trend in the GOME columns, it is not present in the SCIAMACHY data.

6.5 Conclusions

A 15-year dataset of H₂CO vertical columns has been developed on the basis of GOME and SCIAMACHY radiance measurements. The retrieval of slant columns relies on the DOAS technique. The measured radiances are fitted in a wavelength window (328.5-346 nm) shifted to shorter wavelengths compared to previously used retrieval settings. Although this window is possibly less appropriate in case of high ozone absorption (*i.e.* at high latitudes), it effectively reduces the impact of the polarisation anomaly affecting the SCIAMACHY spectra and it moves the fit away from an O₄ absorption band which appears to be a significant source of bias in the H₂CO retrievals. The improved quality of the resulting H₂CO retrieval at low latitudes is demonstrated by a reduction of the fitting errors, as well as by the disappearance of disturbing features found in previous retrievals, like very low or negative columns over deserts.

Since the ultimate goal is to make satellite data available to users for quantitative regional air quality monitoring, pollution trend studies and emission constrain via inverse modelling, realistic estimates of the columns, their errors and their information content are essential. In this perspective, the H₂CO vertical columns are provided with their averaging kernels and an error estimate based on a detailed analysis. For individual satellite measurements, the random error on the slant column is

the largest source of uncertainty. However for those applications where spatial and time resolution can be compromised, average in time and over larger regions allows reducing this contribution to a negligible value. At high latitudes, the systematic error due to the interference with ozone is dominant. At lower latitudes, the total error is mainly due to uncertainties on the formaldehyde absorption cross-sections and to the impact of clouds, aerosols and profile shape uncertainties on the air mass factors. The error estimate demonstrates the need for highly accurate cross-section datasets, cloud retrieval schemes and realistic *a priori* H₂CO profile shapes. Pixels with cloud fractions above 40% and low cloud altitudes are characterized by very large errors (>50%) and should not be considered for quantitative analysis. The effect of aerosols on the error has to be better quantified since it may have an important impact on the retrieved H₂CO columns in case of strong biomass burning events. This dataset has been compared with previous retrievals over the major H₂CO source regions. Over the Eastern United States and over Southern China during summertime, as well over most Tropical forests, the columns retrieved in this study are found to be 20-30% lower than in the dataset of *Chance et al.* [2000] based on GOME measurements. This systematic difference can be partly attributed to the H₂CO cross-sections used in the two retrievals. The consequences of this finding on the global budget of reactive organic compounds will be explored in further inverse modelling studies based on the dataset presented in this study.

The differences between the existing satellite datasets stress the need for a detailed comparison of the formaldehyde retrievals in order to ascertain the reliability of the inverse modelling results. Furthermore, validation using independent correlative dataset is needed. Although adequate large scale validation means are currently lacking, ground-based measurements obtained with FTIR and MAX-DOAS instruments are becoming increasingly available [see *e.g. Heckel et al.*, 2005; *Jones et al.*, 2009; *Vigouroux et al.*, 2009] and will be used in future studies to validate the satellite results.

7 Formaldehyde columns retrieved from GOME-2

Here, we present results of formaldehyde retrievals performed using GOME-2 observations between 2007 and 2010, based on an improved version of the algorithms developed for the GOME and SCIAMACHY instruments. Main features of the new retrieval scheme include a two-step fitting procedure that strongly reduces the interferences between H_2CO and BrO spectral structures. Furthermore, a modified DOAS approach has been adopted to better handle the strong O_3 absorption effects, improving the fit at large solar zenith angles. These two corrections lead to a significant reduction of the scatter in the H_2CO slant columns and improve the agreement with coincident SCIAMACHY observations. The correlation coefficient of the linear regression between monthly averaged SCIAMACHY and GOME-2 time series is generally higher than 0.8, and the mean difference between the H_2CO vertical columns is less than 7% in most of the regions. The GOME-2 H_2CO columns are found to be less sensitive to the South Atlantic Anomaly, offering new perspectives for the study of NMVOC emissions in the Amazon. Furthermore, the better Earth sampling of GOME-2 is shown to allow for a significant reduction of the noise on the monthly columns leading to a better identification of the emission sources. We discuss first attempts to validate our satellite data using ground-based MAX-DOAS instruments at two sites, namely Cabauw (Netherlands) and Beijing (China). Taking into account the different vertical sensitivities of the ground-based and satellite measurement techniques, results show a satisfactory level of agreement in Cabauw, while larger discrepancies are found in Beijing. We speculate that this might be related to enhanced pollution levels near the surface, where the satellite measurements have their lowest sensitivity, or to local effects not captured by the satellite measurements.

The newly developed GOME-2 retrieval settings have been published in the proceedings of the 2009 EUMETSAT Meteorological Satellite Conference[1] and an adapted version of this scientific algorithm has been implemented in the GOME-2 trace gas operational processor developed at DLR-IMF, in the framework of the Satellite Application Facility on Ozone and Atmospheric Chemistry Monitoring (O3M SAF).

7.1 Introduction

In this chapter, we focus on formaldehyde retrievals from the second Global Ozone Monitoring Experiment (GOME-2), which was launched in October 2006 onboard the Meteorological Operational satellite-A (METOP-A). METOP-A is the first of a series of three EUMETSAT meteorological operational satellites, planned to be successively launched at approximately five year intervals [*Callies et al.*, 2000]. The measurements made by the three GOME-2 instruments have the potential to extend by more than a decade the successful time-series of global formaldehyde observations established with measurements from the ESA GOME and SCIAMACHY instruments, as well as with OMI on the NASA AURA platform.

---

[1] De Smedt, I., T. Stavrakou, J.-F. Müller, N. Hao, P. Valks, D. Loyola, M. Van Roozendael (2009), H_2CO columns retrieved from GOME-2: first scientific results and progress towards the development of an operational product, proceedings of the 2009 EUMETSAT Meteorological Satellite Conference, *Bath, U.K.*

GOME-2 is a nadir UV-visible spectrometer that covers the 240–790 nm wavelength range, with a mean resolution of 0.27 nm in the UV [*Munro et al.*, 2006]. As shown in chapter 6, the retrieval of formaldehyde from UV radiances remains challenging, especially from space, because of the relative weakness of the H_2CO absorption bands and because the H_2CO bulk lies mostly in the lower troposphere. The noise in the GOME and SCIAMACHY H_2CO measurements is to a large extent defined by the sampling rate of the instrument. Global Earth coverage is achieved in respectively three days with GOME and in six days with SCIAMACHY, because of the alternating limb/nadir viewing modes which introduce further discontinuities in the observations. The full width of a normal GOME-2 scanning swath is 1920 km, twice larger compared to GOME and SCIAMACHY, allowing for global Earth coverage within one day and a half. Although GOME-2 is characterized by a moderate spatial resolution of 80×40 km^2 (four times finer than GOME (320×40 km^2) but 1.8 times coarser than SCIAMACHY (60×30 km^2)), its higher spatial sampling is expected to allow for a better definition of short-lived tropospheric species structures, and for a more precise identification of the emission areas [*Lerot et al.*, 2010b; *Richter et al.*, 2011]. While GOME and SCIAMACHY have been measuring simultaneously (with a global coverage) only during the six first months of 2003, the continuous observations made with SCIAMACHY and GOME-2 since 2007 allow for a more precise comparison of the H_2CO columns. Considering that GOME-2, SCIAMACHY and GOME have differences in overpass times of maximum one hour (between 09:30 and 10:30 am), no significant diurnal variation is expected to impact the column comparison, contrary to the results of OMI, which measures in the early afternoon. Up to now, retrievals of formaldehyde from GOME-2 have been reported in one study of the IUP-Bremen group, for measurements between 2007 and 2008 [*Vrekoussis et al.*, 2010]. The results are consistent with their previous H_2CO observations, although artefacts over deserts and other regions with bare soil are still present [*Wittrock et al.*, 2006; *Vrekoussis et al.*, 2010].

In this work, we concentrate on optimising H_2CO retrieval settings for the GOME-2 instrument, while, at the same time, we assess the consistency with the GOME and SCIAMACHY time-series. In section 7.2, GOME-2 H_2CO columns retrieval issues are discussed and improved DOAS settings are introduced. In section 7.3, results from GOME-2 H_2CO retrievals are presented and the agreement with coincident SCIAMACHY data is investigated in the period from January 2007 to December 2010. Based on case studies, the added-value of the GOME-2 observations is then illustrated. We finally introduce first attempts to validate our H_2CO satellite observations using ground-based MAX-DOAS measurements at two sites: Cabauw (The Netherlands) and Beijing (China).

7.2 Formaldehyde retrieval from GOME-2 measurements

7.2.1 Baseline slant column retrieval

As extensively described in chapter 6, the optimization of DOAS settings for one given species (*e.g.* H_2CO) consists in minimizing the fitting residuals, reducing as much as possible the noise on the slant columns, and avoiding fitting artefacts that can be introduced by interferences with other absorbers or instrumental effects. In the present study, a major additional concern is to optimize the consistency of the H_2CO columns retrieved from GOME-2 with those from previous GOME and SCIAMACHY spectra. Therefore, our initial baseline for GOME-2 retrievals has been to use the settings previously developed for GOME and SCIAMACHY. The details of the H_2CO DOAS settings are given in Table 7-1. Slant columns are retrieved in the 328.5–346 nm wavelength range. All reference cross-section datasets listed in Table 7-1 have been convolved using the accurately characterized GOME-2 slit functions, which depend on the detector wavelength [*Siddans et al.*,

2006]. To correct for the Ring effect, two pseudo absorption cross-sections generated after *Vountas et al.* [1998] using the SCIATRAN radiative transfer model [*Rozanov et al.*, 2001] are included in the fit.

Table 7-1: Baseline DOAS settings used to retrieve H₂CO slant columns from GOME-2 spectra

Parameter	
Fitting interval	328.5-346 nm
Absorption cross-sections	
H₂CO	*Meller and Moortgat* [2000], 298°K
Ozone	*Brion et al.* [*1998*], 228°K + 243°K
BrO	*Fleischmann et al.*[2004], 223°K
NO₂	*Vandaele et al.* [2002], 220°K
Slit function	GOME-2 slit function depending on the detector wavelength [*Siddans et al.*, 2006]
Ring effect	2 vectors generated using SCIATRAN [*Vountas et al.*, 1998; *Rozanov et al.*, 2001]
Polynomial	5th order
Intensity offset correction	Linear offset
Reference spectrum (I₀)	The latest solar spectrum measured by GOME-2

In the case of GOME-1, the I_0 reference spectrum used to calculate the atmospheric optical density (equation 5-16) was a daily radiance spectrum selected over the Pacific Ocean. This choice was made in order to avoid spectral artefacts in the GOME solar spectra caused by the Sun diffuser plate anomaly [*Richter and Wagner*, 2001] and to minimise the fitting residuals. For GOME-2, the results obtained with the solar irradiance spectra lead to equivalent magnitudes and standard deviations of the H₂CO slant columns than those obtained with a radiance reference. Nevertheless, the fitting residuals are 10 to 20% higher because of the larger differences between the filling-in of the Fraunhofer lines in the different spectra; differences which are never perfectly corrected by the implemented Raman scattering correction in the DOAS approach (see section 5.2.4). However, as it does not affect the quality of the slant columns, it has been decided to use solar irradiances as reference for GOME-2. Note that for SCIAMACHY, the solar irradiances also give satisfactory results and are actually used in the present study. A practical advantage of using solar irradiances is the lowest computational cost, as the latest recorded solar spectrum is provided in each orbit file, while the pre-selection of radiance spectra in the reference sector is time consuming. More fundamentally, the use of a solar spectrum as reference (by definition free of atmospheric absorption) allows in principle for the retrieval of an absolute column density, while a radiance reference spectrum only provides a differential slant column. However, as already discussed in the previous chapter, in the case of weak absorbers such as H₂CO, a reference sector normalisation must be applied to the retrieved columns, in order to reduce the impact of fitting artefacts due to unresolved spectral interferences, finally leading to differential slant columns. The method used to calculate the tropospheric *AMF* is the same as the one developed for the GOME and SCIAMACHY H₂CO products and the respective contributions to the total error on individual columns are estimated as presented in chapter 6.

These baseline settings have been applied to GOME-2 measurements between 2007 and 2008. The resulting GOME-2 H₂CO columns are shown in Figure 7-1 (lower panel) and compared with the SCIAMACHY results over the same period (higher panel). As can be seen, the overall agreement is good, although the GOME-2 columns appear slightly lower than those of SCIAMACHY. As expected from the larger number of GOME-2 observations, the 2-years averaged GOME-2 map is smoother than with SCIAMACHY.
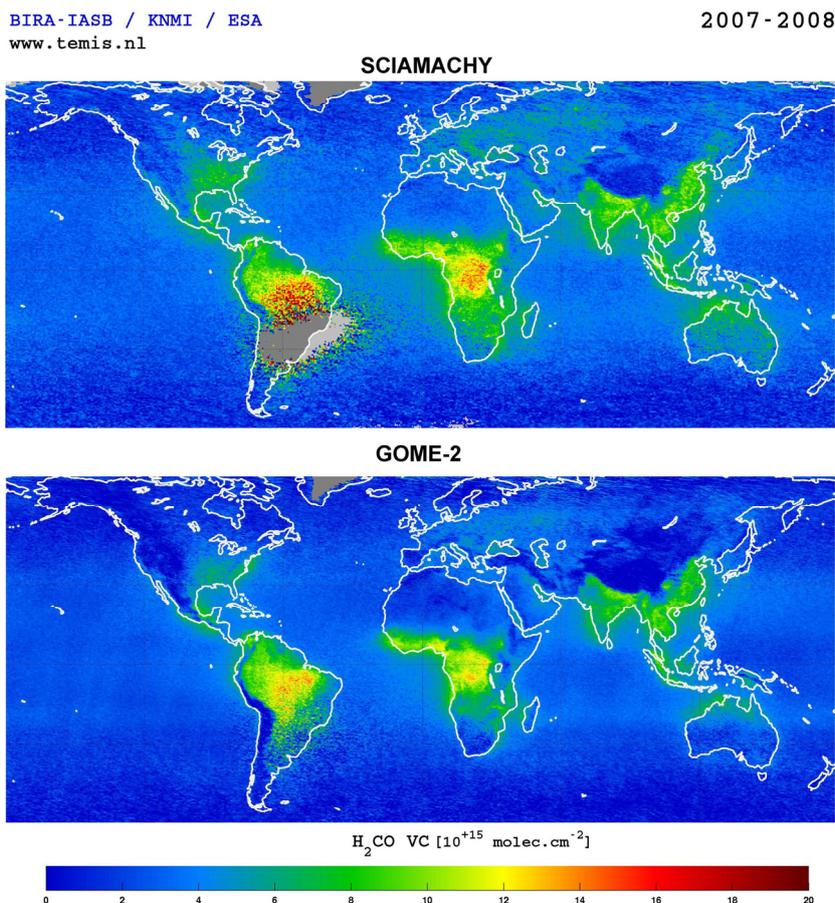
Figure 7-1: H₂CO vertical column retrieved from SCIAMACHY and GOME-2 between 2007 and 2008. The retrieval settings used for GOME-2 are based on the SCIAMACHY settings [*De Smedt et al.*, 2008].

For a more quantitative comparison, the left column of Figure 7-2 compares the SCIAMACHY and GOME-2 (baseline settings) time series of monthly averaged H₂CO columns over five selected regions where emissions are significant. Although the correlation coefficients between the two satellite retrievals are higher than 0.8, the GOME-2 H₂CO columns are generally found to be 15% to 20% lower than those of SCIAMACHY (except in Amazonia). The right panels of Figure 7-2 illustrate the improvements brought by the modified GOME-2 H₂CO retrieval settings that will be presented in the next section.

Additionally, the scatter of the individual GOME-2 measurements has been compared to GOME and SCIAMACHY retrievals. As shown in chapter 6, the H₂CO retrievals are largely limited by the photon-noise in the detection system. Since GOME, SCIAMACHY and GOME-2 have similar optical designs, differences in the standard deviations on the retrieved columns should be related to the different ground pixel sizes of the instruments. In order to evaluate the noise in the retrievals, differential slant columns over the clean equatorial Pacific (15°S–15°N, 150–210°E) have been analyzed for July 2000 (GOME-1) and July 2008 (SCIAMACHY and GOME-2). In this area, one can assume that the tropospheric H₂CO production due to NMVOC oxidation is weak, and that spatial variations over the region can be neglected. The standard deviation in the retrieved H₂CO columns is therefore a measure of the random noise of the measurements. The results of this analysis are shown in Figure 7-3 (together with the results of the same analysis on the improved GOME-2 retrieval that will be discussed later). Only GOME-2 pixels whose line-of-sight angles are lower than 32° were selected in order to avoid optical path length differences. For the figure, a

normalization has been applied in order to account for large differences in the number of observations. The distributions found for GOME, SCIAMACHY and GOME-2 are Gaussian, with σ standard deviations[2] of respectively 0.44, 1.163 and 1.16×10^{16} molec.cm^{-2}. This result is not as good as expected for GOME-2. Indeed, from the different ground pixel sizes, we can deduce that the noise on the GOME-2 slant columns should be only 2 times higher than GOME-1 and smaller than SCIAMACHY (see chapter 6). The reasons for the larger than expected scatter in the GOME-2 H₂CO columns are not understood yet, but are probably related to additional noise in GOME-2 spectra. For that matter, the same effect has been identified in the retrieval of NO₂ in the visible wavelength range [*Richter et al.*, 2011].
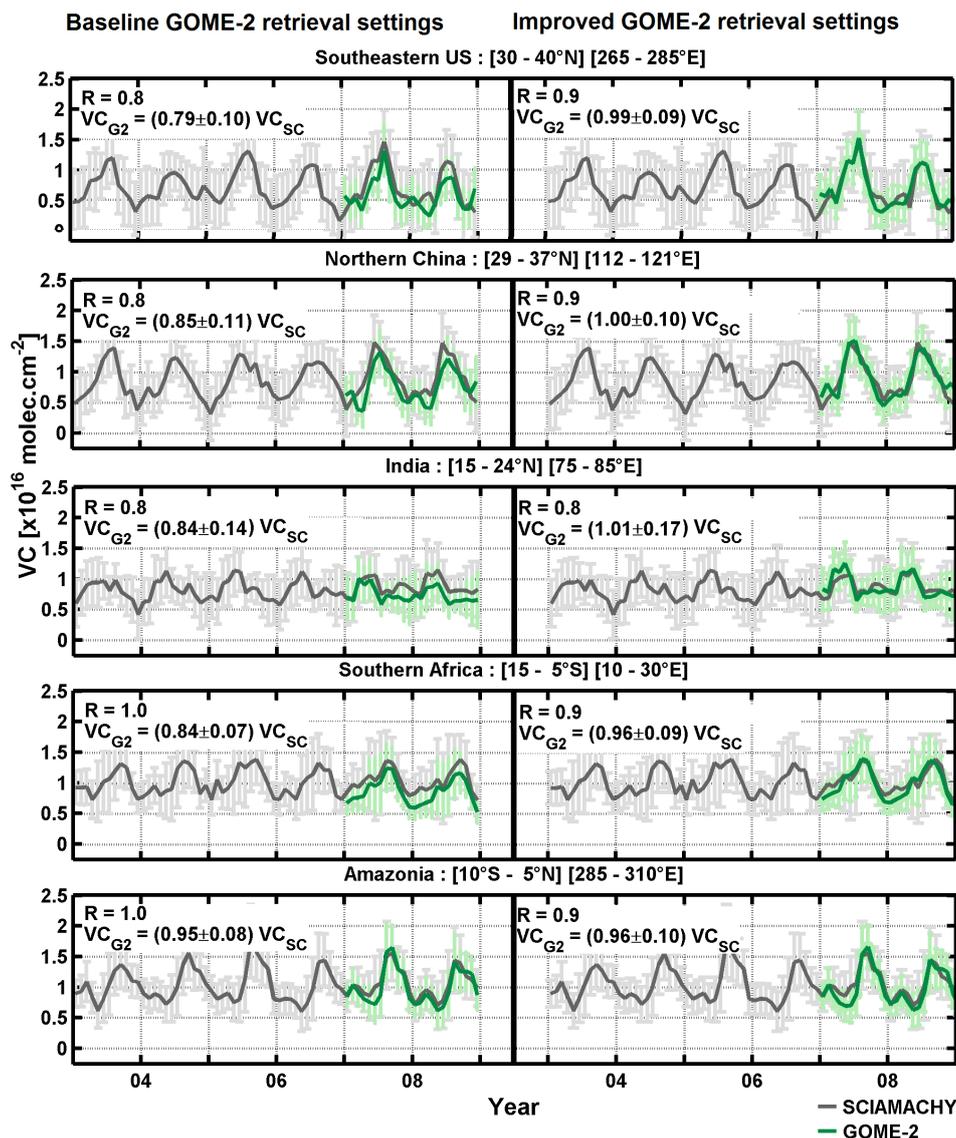
Figure 7-2: Regional monthly averaged H₂CO vertical column retrieved from SCIAMACHY (in grey) and GOME-2 (in green) measurements. The left panels show the GOME-2 results obtained with the baseline retrieval settings between 2007 and 2008, while the right panels show the results obtained with the improved settings over the same period (see text for details).

---

[2] $y = a \cdot e^{\frac{-(x-\mu)^2}{2\sigma^2}}, \quad FWHM = 2\sigma\sqrt{2\ln 2}$
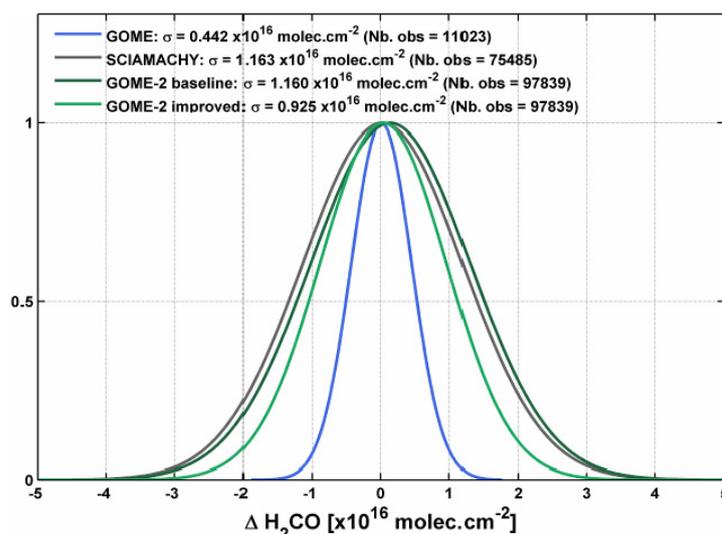
Figure 7-3: Normalised distribution of the retrieved H$_2$CO slant columns from GOME, SCIAMACHY and GOME-2 over the equatorial Pacific for July 2000 (GOME) and 2008 (SCIAMACHY and GOME-2). Only cloud fractions lower than 40% were included. Two versions of GOME-2 results are shown; the baseline retrieval settings, identical to the ones used for GOME and SCIAMACHY, and the improved settings (see text for details).

7.2.2 Improved slant column retrieval

One way to reduce the noise on the retrieved slant columns and to maximize the sensitivity to weak absorbers is to increase the size of the fitting window, including more absorption bands and thereby additional information in the DOAS analysis [*Richter et al.*, 2011].

As the lower limit of our fitting window already goes quite far in the UV with regard to O$_3$ absorption, tests have been performed to extend the fitting interval towards longer wavelengths, which in contrast to the SCIAMACHY case, is possible with GOME-2 spectra since no strong and structured polarization features have been reported in channel 2. Different fitting windows have been tested with GOME-2 spectra, including the interval originally used for GOME-1 (337-359 nm), or starting from the newer 328.5-346 nm window and increasing the upper limit of the interval by steps of 2 nm up to 360 nm. However, as previously shown with the GOME-1 H$_2$CO retrievals, artefacts above desert regions and Oceans, which were found to be related to a strong O$_4$ absorption band peaking at 360 nm [*De Smedt et al.*, 2008], have also been identified in the GOME-2 results. For this reason, we decided to restrict the upper limit of the fitting interval under 350 nm (see the Figure 6-7 in chapter 6). Nevertheless, as illustrated in Figure 7-4, increasing the upper limit of the fitting interval allows for a reduction of the correlation between the H$_2$CO and BrO absorption features (from 0.54 in 328.5-346 to 0.45 in 328.5-359 nm), and our tests have shown that using a larger fitting interval allows for the reduction of the noise of the BrO slant column densities. For that matter, the improvements brought by the fit of BrO in a larger interval for the global observations of BrO with GOME-2 have been reported in *Theys et al.* [2011].

For these reasons, a two-step DOAS fitting approach has been designed that effectively leads to a reduction of the noise on the H$_2$CO columns. The BrO slant columns are first fitted in a wider wavelength interval including six absorption bands of BrO (328.5-359 nm). In a second step, the fit of H$_2$CO is performed in the original 328.5-346 nm interval, but fixing the BrO slant column to the value determined in the first step. This approach allows the decorrelation of the BrO and H$_2$CO retrievals while at the same time, the O$_4$-related bias in oceanic and desert regions is avoided.
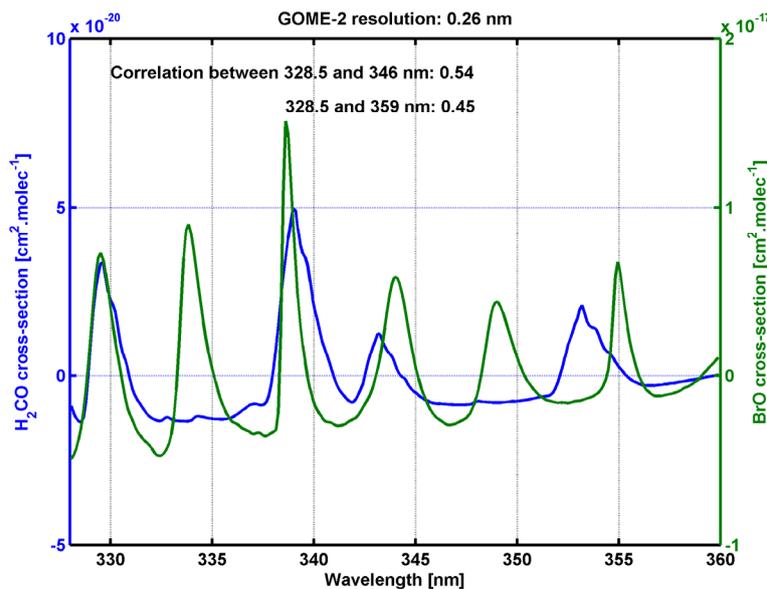
Figure 7-4: H₂CO and BrO differential cross-section datasets, convolved with the GOME-2 slit function. The correlation coefficients between the 2 datasets are given, in the two wavelength intervals (328.5-346 and 328.5-359 nm).

The DOAS settings used in both windows are almost identical (and detailed in Table 7-1). However, in the 328.5-359 nm interval, two additional polarisation vectors are included in the analysis (taken from the GOME-2 calibration key data [*EUMETSAT*, 2009]), to correct for structures appearing at large scanning angles. Figure 7-5 illustrates this two-steps procedure in the case of a GOME-2 spectrum measured on the 9th August 2007 and overpassing a strong H₂CO signal due to isoprene oxidation in the Southeastern United States. Residuals of the fit are lower in the second window, and the H₂CO fitted slant column is slightly enhanced. Note the fact that the optically stronger absorbers like O_3 and the Ring effect are consistently retrieved in both windows, bringing confidence in the quality of the fitting process. Note also that the parameters included in the fit (Table 7-1) are not all shown in the Figure 7-5.

A second correction has been applied to the fitting process. In order to better cope with the strong ozone absorption at wavelengths shorter than 336 nm, an AMF-modified DOAS approach [*Lee et al.*, 2008*; Theys at al.*, 2011] has been adopted. It consists in replacing the ozone cross-section in the fitting procedure by the same ozone cross-section weighted by a wavelength-dependent *AMF*. The fitted ozone columns are therefore "AMF-weighted" columns rather than slant columns. Wavelength dependent *AMF*s have been calculated off-line with a radiative transfer model, for several solar zenith angles and using a typical mid-latitude stratospheric ozone profile [*US standard atmosphere*, 1976], corresponding to a total column of 9.4×10^{18} molec.cm^{-2} (or ~350 Dobson units). This correction allows reducing the fitting residuals and the H₂CO slant columns underestimation at large solar zenith angles, when the ozone absorption becomes optically thicker because of the increased light path in the stratosphere.
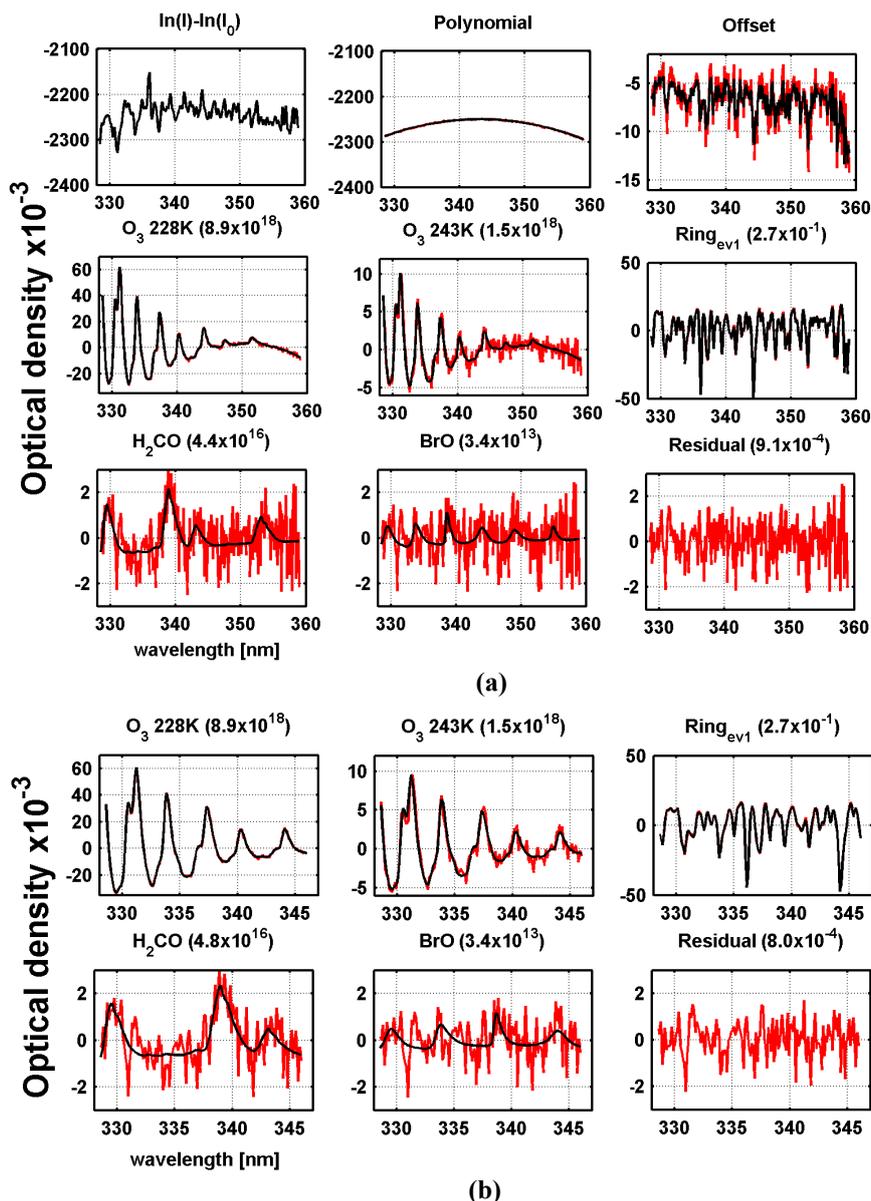
Figure 7-5: DOAS 2-steps fit of a GOME-2 spectrum overpassing a region of large biogenic emissions in South-Eastern US (09/08/2007, orbit 41762, SZA: 28°). (a) First step in the wavelength interval 328.-5-359 nm, (b) second step in the wavelength interval 328.-5-346 nm. The BrO slant column is not fitted in the second step. Each plot shows the optical density (in black) and the residual (in red), while the fitted slant columns are given above.

Figure 7-6 illustrates the effects of the two corrections for one particular GOME-2 orbit of the 13th August 2007, which is typical of background-level H₂CO concentrations. Figure 7-6 (a) shows the differential H₂CO slant columns and the fitting errors before and after the corrections (in black and in green), while Figure 7-6 (b) shows the BrO and O₃ vertical columns. For this figure, the O₃ and BrO slant columns retrieved with the baseline settings (in black) have been divided by a stratospheric AMF, while the green curves are the directly fitted "AMF-weighted" vertical columns. Owing to the first correction (the BrO pre-fit procedure), the standard deviation on the H₂CO slant columns is reduced by 30% (just like the standard deviation on the BrO columns) and the H₂CO slant columns errors are globally reduced by 10%. The second correction (the AMF-modified DOAS approach) allows for a reduction of the H₂CO slant column error and under-estimation at

mid-latitudes by more than 25% (below 40°S). This is directly related to the larger O_3 columns in the Southern Hemisphere in August.

(a)   (b)
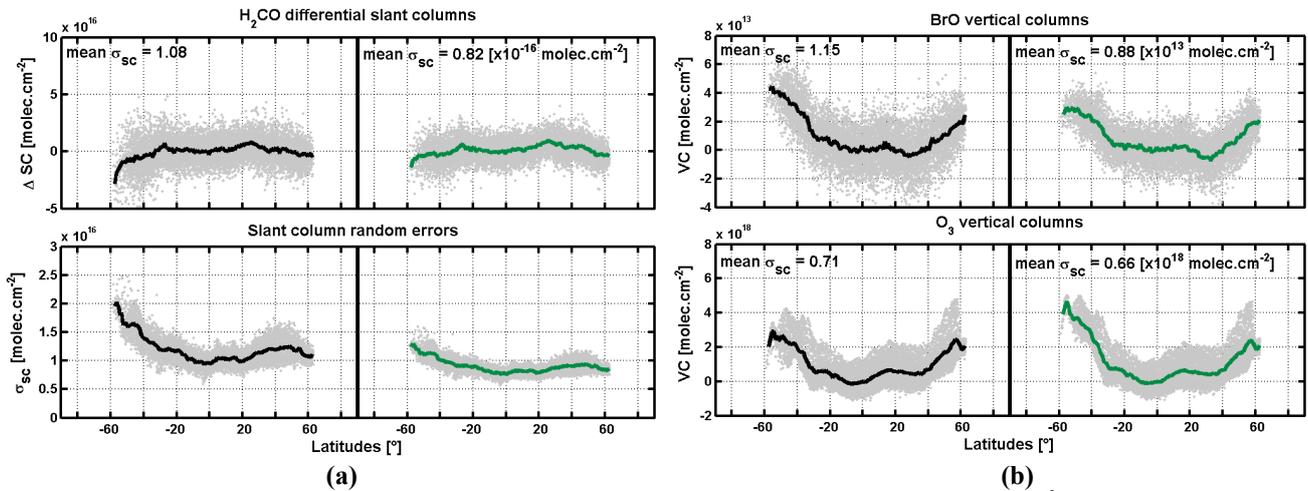
Figure 7-6: Results of the DOAS fit in the 328.5-346 nm interval for a GOME-2 orbit (13th August 2007, orbit number 4227). (a) H₂CO differential slant columns, errors and standard deviation. The mean standard deviations on the H₂CO slant columns are given inset. (b) BrO (upper panels) and O₃ (lower panels) vertical columns. The mean standard deviations on the BrO and O₃ columns are given as inset. For (a) and (b), the left panels correspond to the baseline retrieval (in black), while the right panels show the effect of the BrO and O₃ corrections (in green).
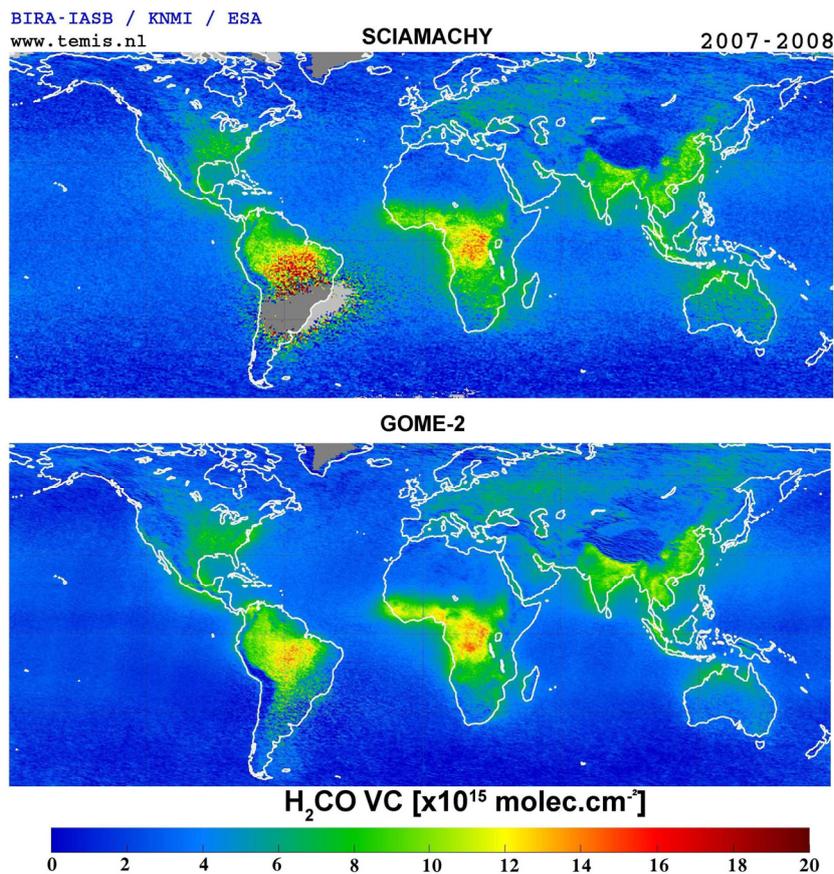
Figure 7-7: Averaged SCIAMACHY and GOME-2 H₂CO vertical columns for the 2007-2008 period. The modified retrieval settings are used to improve the GOME-2 results.

Note that, in the case of SCIAMACHY, it is not possible to fit BrO in a larger window because of the strong polarization anomaly in the spectra around 350 nm (see chapter 6). The two-steps approach has been tested on GOME-1 spectra. However, the fit results were found to be degraded compared to the baseline settings. Although the reasons are difficult to identify, one possibility could be the existence of a residual large-band polarisation feature, not perfectly corrected by the DOAS polynomial in the larger window. This assumption is supported by the enhancement of the misfits at the edges of the scan (East and West pixels). The AMF-modified DOAS approach has not been tested yet with GOME and SCIAMACHY data. However, this type of correction accounting for strong ozone absorption will be further investigated in the future as it could potentially improve the retrievals at mid-latitudes in winter, where tropospheric pollution is to be monitored.

7.3 Results

7.3.1 Comparison with GOME and SCIAMACHY time series

Yearly averaged formaldehyde columns retrieved from GOME-2 measurements between 2007 and 2010 are presented in Figure 7-8. These maps can be compared with the SCIAMACHY observations for the same years investigated in chapter 6 (Figure 6-19). As observed with previous satellite measurements, the highest annual columns are found over Tropical regions in Africa and Amazon, where biogenic and biomass burning sources dominate the emissions [*Müller et al.*, 2008; *Barkley et al.*, 2008; *Stavrakou et al.*, 2009a], and in Southeastern Asia, where biogenic, biomass burning and anthropogenic sources all contribute to the observed H₂CO column [*Fu et al.*, 2007; *Stavrakou et al.*, 2009a].
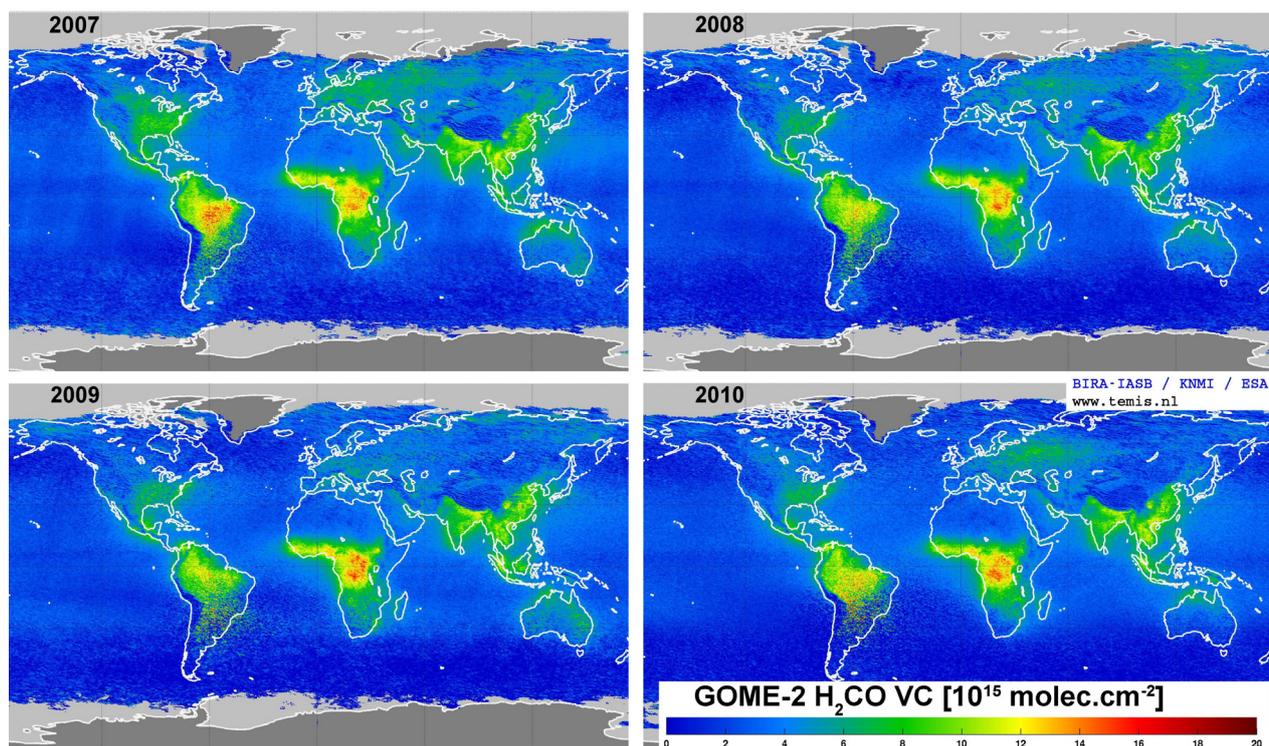
Figure 7-8: Yearly averaged GOME-2 H₂CO columns from 2007 to 2010.

Interestingly, over the Eastern part of South America, the GOME-2 retrievals are less affected by the South Atlantic anomaly than SCIAMACHY or even GOME. This is important since the Amazonian biogenic emissions of NMVOCs are particularly uncertain and need to be better

constrained [*Müller et al.*, 2008, *Stavrakou et al.*, 2009a and b, *Barkley et al.*, 2009]. On the contrary, the GOME-2 retrievals in the channel 3 are found to be more sensitive than SCIAMACHY to the SAA, pointing to different behaviours of the detectors in the UV and in the visible [*Richter et al.*, 2011]. In general, our H₂CO columns agree reasonably well with the annual distribution reported by the IUP-Bremen group, although the latter data display persistent anomalies over arid and semi-arid regions extending over large parts of North and South Africa, the Arabian Peninsula, the Australian continent and the Western part of the North-American continent, preventing their use over these regions [*Vrekoussis et al.,* 2010].

In order to get a more precise picture of the formaldehyde spatial distribution and its annual cycle, maps of seasonally averaged GOME-2 H₂CO columns are shown in Figure 7-9. These seasonal variations can be related to the global land cover map shown in Figure 7-10, which was generated using data collected by the Medium Resolution Imaging Spectrometer (MERIS) instrument onboard ENVISAT [*ESA GlobCover Project*]. The formaldehyde seasonal variations are primarily related to the increase of biogenic emissions during local summer months at mid-latitudes over deciduous forests (in America, in Europe, in Northern Asia or in Australia), and during the dry season over evergreen Tropical forests (in Amazon and Africa). In the Tropics, biomass burning also contributes significantly to the H₂CO columns. Over the Amazon, burning is more widespread in the dry season (September–November), but over Africa, a dipole pattern exists owing to different seasonal burning (and emissions) either side of the Equator. Over East Asia, biogenic emissions dominate the H₂CO signal, coupled with seasonal agricultural burning on the North China Plain and forest fires in the Russian Far East [*Fu et al.*, 2007]. In Southeastern Asia, important biomass burning events take place in India and in Indochina from March to May. In highly populated regions of Asia, the anthropogenic signal of H₂CO is significant, in particular in the regions of Beijing and North of India. However, even there, the biogenic contribution to the total H₂CO column largely dominates. For this reason, the anthropogenic signal should be easier to isolate during winter time. This is not the most favourable situation since the quality of the H₂CO satellite retrieval decreases in the case of low sun observations.

BIRA-IASB / KNMI / ESA
www.temis.nl
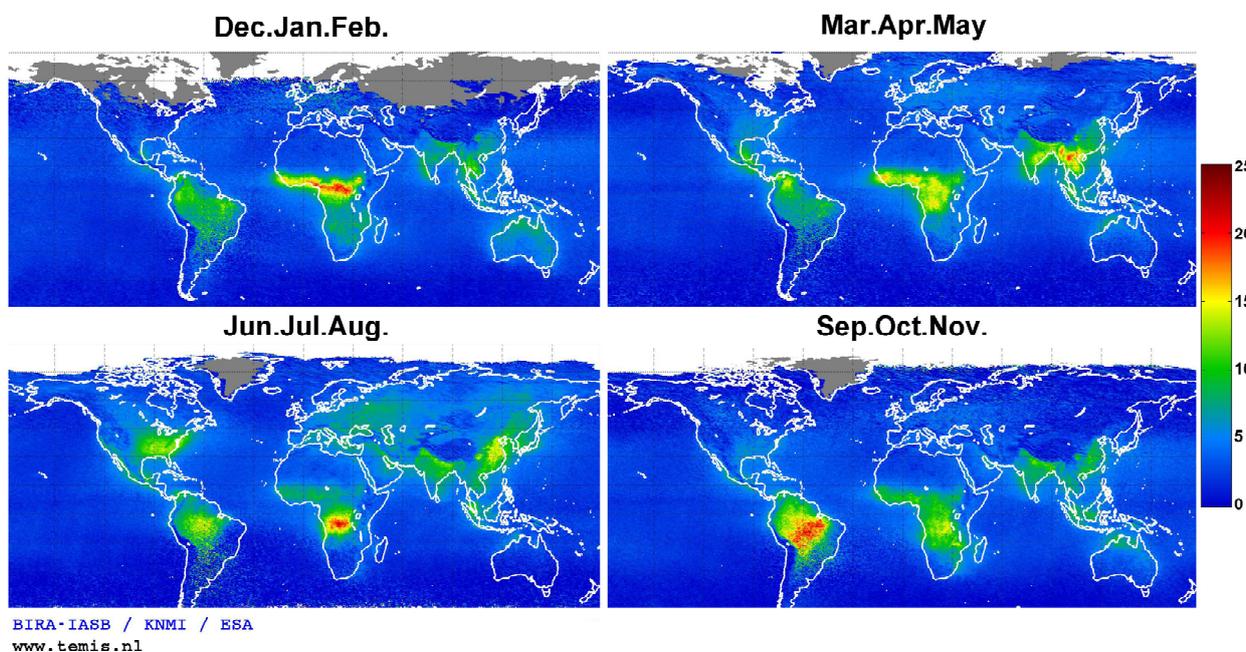
Figure 7-9: Seasonal variations of the formaldehyde distribution as retrieved from GOME-2 measurements between 2007 and 2010.
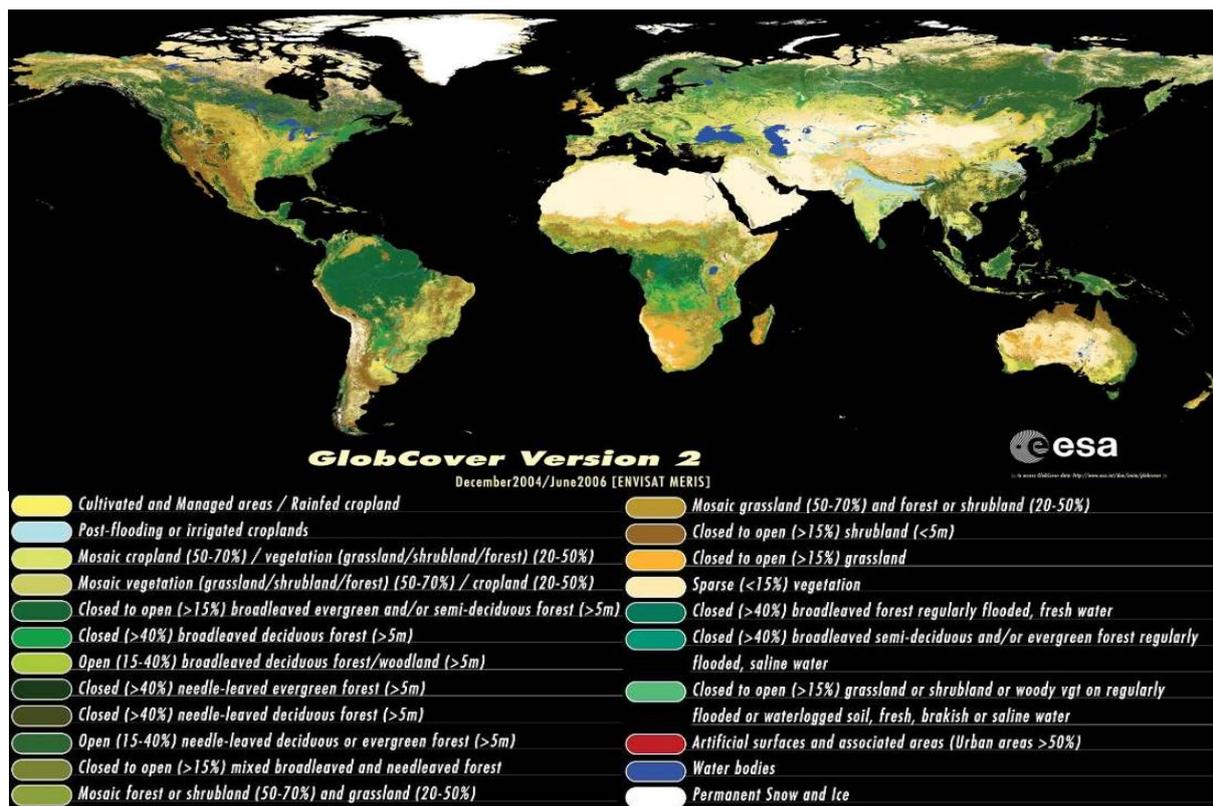
Figure 7-10: Land cover over North America and Asia [© *ESA GlobCover Project*, MEDIAS France/Postel]

In Figure 7-11 and Figure 7-12, the GOME-2 observations are compared with the GOME (Apr. 1996 - Jun. 2003) and SCIAMACHY (2003 - 2010) time series of monthly averaged formaldehyde columns retrieved over a selection of regions (defined in Figure 6-20). For each region, the correlation with coincident SCIAMACHY observations is given within the plot, as well as the slope of the regression line for GOME-2 versus SCIAMACHY. The correlation coefficients are higher than 0.8 in almost all regions. They are slightly lower (0.7) in Europe and in Indonesia where the amplitude of the H_2CO variations is the weakest between 2007 and 2010. The agreement between the SCIAMACHY and GOME-2 H_2CO columns is found to be better than 3% in Southeastern US, India, Southern China, Thailand, Indonesia, Equatorial Africa and Europe. In Northern Africa, the GOME-2 columns are 7% higher, while in Southern Africa, they are 6% lower than SCIAMACHY. In Guatemala and Amazonia, GOME-2 values are also found to be 5 to 7% lower than SCIAMACHY. Note that these differences are mostly due to winter or wet season values, during which the GOME-2 H_2CO columns are lower than those of SCIAMACHY, while the H_2CO values during summer or dry seasons are equivalent. This general level of agreement is very satisfactory in regard to the estimated uncertainties on the total columns, which range between 20% and 40% for monthly and spatially averaged data. Largest differences are however found in Australia, where the GOME-2 columns are 20% lower than SCIAMACHY and appear more consistent with the GOME time-series. The discrepancy between SCIAMACHY and GOME-2 is mostly important during the winters of 2009 and 2010, when the amplitude of the SCIAMACHY data significantly reduces. An instrumental degradation effect in SCIAMACHY time-series, after height years of measurements, cannot be excluded.
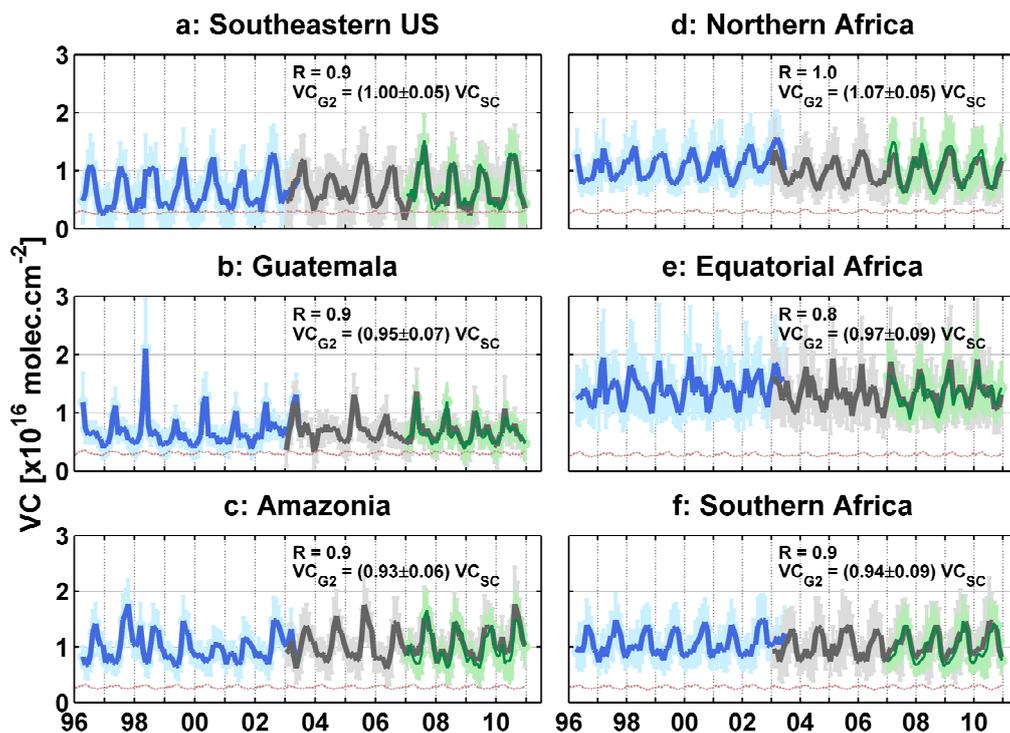
Figure 7-11: Time series of H₂CO monthly averaged vertical columns (*VC*) over regions of the American (a, b, c) and African (d, e, f) continents retrieved from GOME (in blue), SCIAMACHY (in grey) and GOME-2 (in green) measurements. The total errors on the vertical columns are displayed in light colour and value of the reference sector correction in pink. The correlation between the SCIAMACHY and GOME-2 columns are given inset each plot, as well as the slope of the regression line fitted between the two datasets.
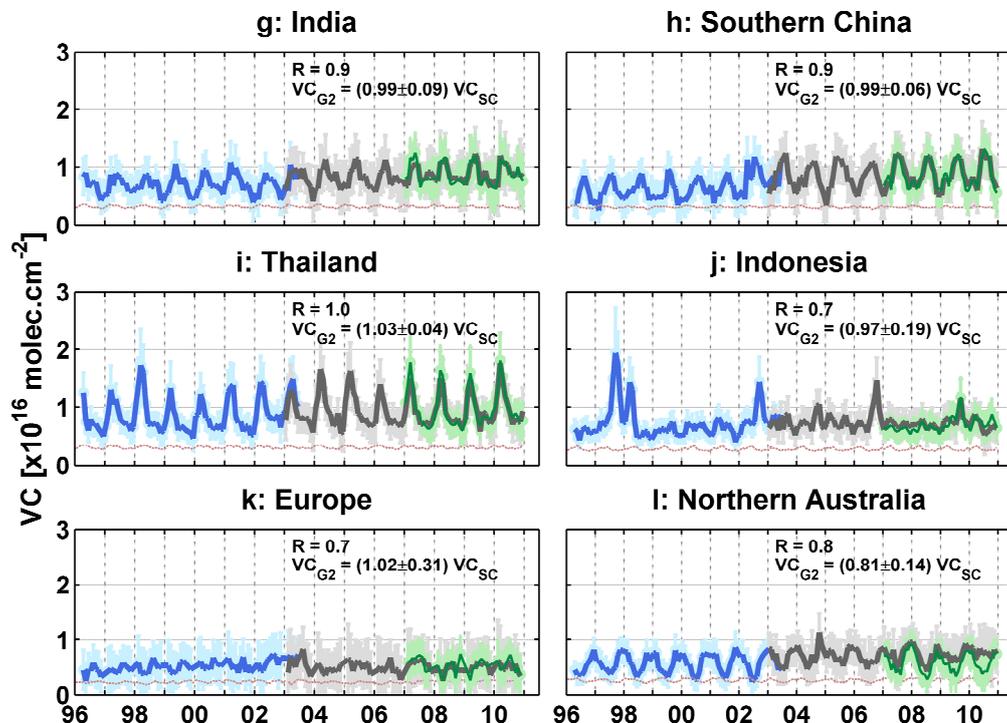
Figure 7-12: Time series of H₂CO vertical columns over regions of Asia (g, h, i, j), Europe (k) and Australia (l) retrieved from GOME, SCIAMACHY and GOME-2 measurements. See Figure 7-11 for legend details.

The effect of instrumental degradation of the SCIAMACHY and GOME-2 instruments is also illustrated in Figure 7-13, which compares the random errors retrievals over the equatorial Pacific. As already mentioned, SCIAMACHY has started showing signs of degradation since 2007, but with no significant impact on the H₂CO columns (except over Australia). It must be noted that the stability of the H₂CO vertical columns is partly due to the applied normalization in the reference sector, which corrects for long term global degradation effects possibly present in the retrieved slant columns. The GOME-2 H₂CO retrievals are degrading noticeably faster than SCIAMACHY, so that the noise of the retrieved slant columns has reached the SCIAMACHY level by the end of 2009. Tests have been preformed on the GOME-2 instrument since September 2009 in order to assess if substantial and sustainable recovery of the signal throughput could be achieved by operating the instrument detectors at higher temperatures [*Eumetsat Product and Service News*, August 12, 2009]. GOME-2 degradation appears to be stabilized since the beginning of 2010, but it is premature to judge on the evolution of weak absorber retrievals from GOME-2.
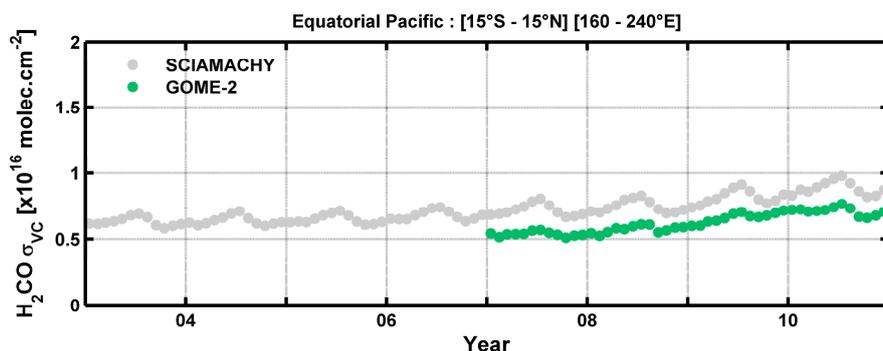
Figure 7-13: Random error on the monthly averaged H₂CO slant columns retrieved from SCIAMACHY and GOME-2 measurements between 2003 and 2010 over a remote region at the equatorial Pacific.

7.3.2 The added value of GOME-2 observations

For NMVOC emission studies and inverse modelling purposes, the precision of the current GOME and SCIAMACHY H₂CO datasets does not allow for a better time resolution than one month. In addition, to further reduce the noise level, the monthly observations need to be averaged in space, for example over large regions as shown in Figure 7-11 and Figure 7-12. GOME-2 offers substantial advantages for resolving spatial and temporal variability and reducing measurement uncertainty through improved sampling statistics. Indeed, the frequency of the GOME-2 observations is two/four times higher than GOME/SCIAMACHY, as illustrated by Figure 7-14, which shows one day of H₂CO measurements from GOME on the 16th August 1997, and from SCIAMACHY and GOME-2 both on the 16th August 2007. Beyond the observational sampling characteristics of each instrument, the gaps in the measurements correspond to cloudy scenes, poor quality fits and/or large solar zenith angles. SCIAMACHY offers the finer ground resolution but with the lowest Earth coverage. On the other hand, the much better GOME-2 coverage allows for a better capture of the H₂CO variations, even with ground pixels larger than SCIAMACHY. For this particular day, it is for example the case over North America, where H₂CO emissions plumes along the Eastern and Western coasts are only visible in the GOME-2 map.
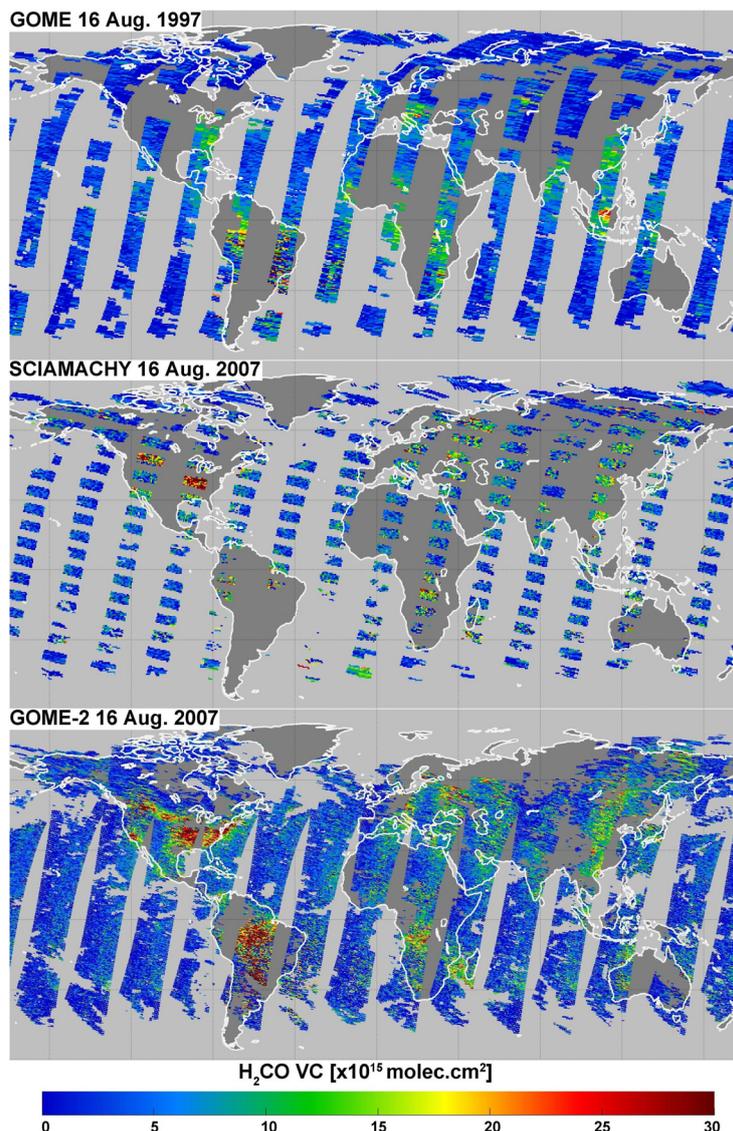
Figure 7-14: Examples of daily observations of H₂CO columns retrieved from GOME, SCIAMACHY and GOME-2 measurements.

In order to further illustrate the added value of GOME-2 in comparison to SCIAMACHY for the identification of regional patterns of the emissions, Figure 7-15 and Figure 7-16 compare GOME-2 and SCIAMACHY seasonal maps (summer and fall seasons, averaged over 2007-2010) in Asia and in North America. During the fall season, biogenic emissions decrease, so that the anthropogenic signal can be more easily detected, while the solar zenith angles are below 60°. Generally speaking, the GOME-2 maps are clearly smoother and show better defined spatial patterns than SCIAMACHY. In Asia, anthropogenic emissions of NMVOCs have been reported over major cities, and over industrial Eastern China, the Ganges valley and Central India where biofuel is extensively used [*Fu et al.*, 2007]. In the GOME-2 map for the September-November period, anthropogenic H₂CO signals are clearly visible in urban areas of Central and Southern China including Chongqing, Wuhan and Changsha, Xi'an, Hong-Kong and the Pearl river delta, and in major Indian cities including New Delhi and Kolkata. These signals are also visible in the SCIAMACHY map, but the noise level makes them more difficult to identify. It should be noted however that the finer ground resolution of SCIAMACHY results in more structured tropospheric

signals that could effectively allow for a better identification of the anthropogenic sources, if accompanied by a better Earth sampling.
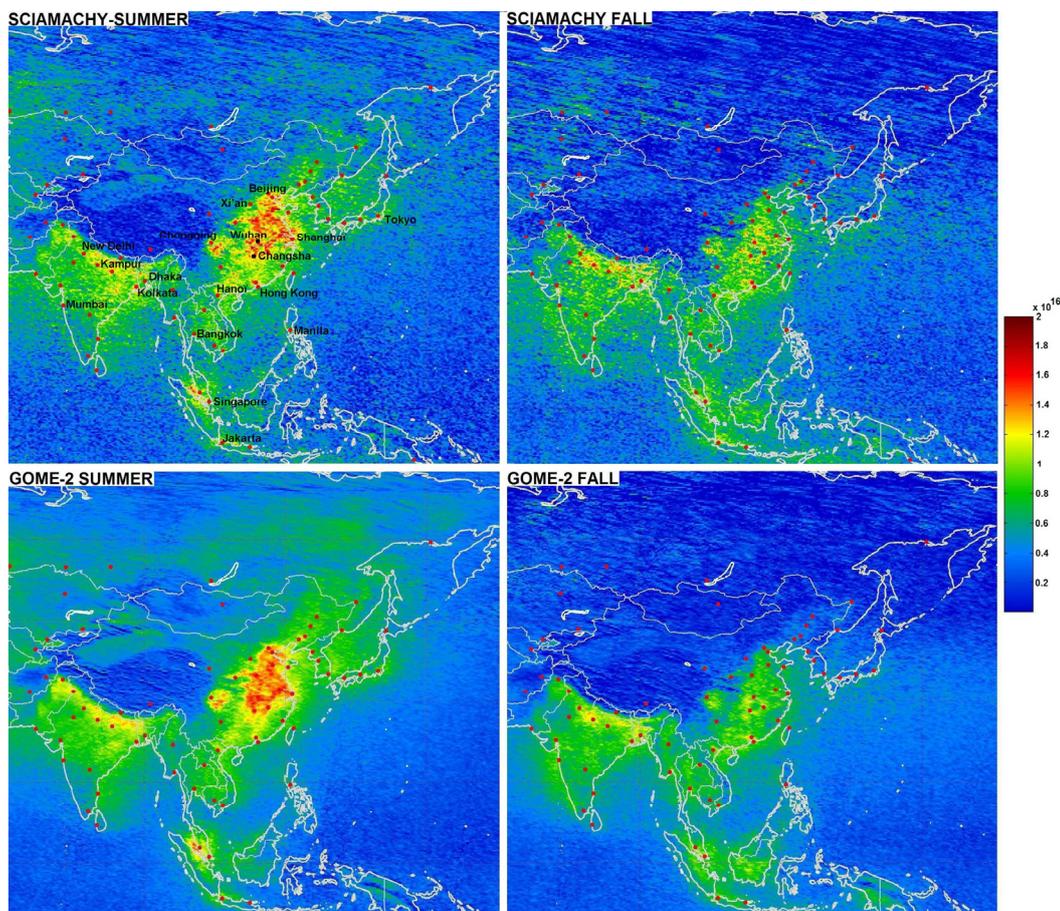
Figure 7-15: Comparison between H₂CO columns in Asia retrieved from SCIAMACHY (upper panel) and GOME-2 (lower pane) during summer (left) and fall (right) seasons over the 2007-2010 period.

In America, biogenic NMVOCs dominate the H₂CO variability. During summer months, elevated H₂CO extend from Eastern Texas to the mid-Atlantic coast. Again, the spatial distribution of the biogenic emissions seen by GOME-2 is better defined. For example, the summertime H₂CO enhanced columns over the Californian coast are clearly visible in the GOME-2 maps, while it was rarely the case with the GOME or SCIAMACHY observations. The lower H₂CO columns over the Appalachian Mountains, right in the middle of the strongest biogenic emissions, are also captured by GOME-2. Similar patterns are observed with the OMI instruments, offering the best current ground resolution and temporal coverage of formaldehyde observations [*Millet et al.*, 2008]. During fall, weaker H₂CO signals are detectable (in both SCIAMACHY and GOME-2 maps) in the vicinity of some urban areas, like for example around Los Angeles and Phoenix and the South of San-Francisco, or between Houston and Dallas in the Southeastern part of the country, were local petrochemical industry is important [*Millet al.,* 2008].
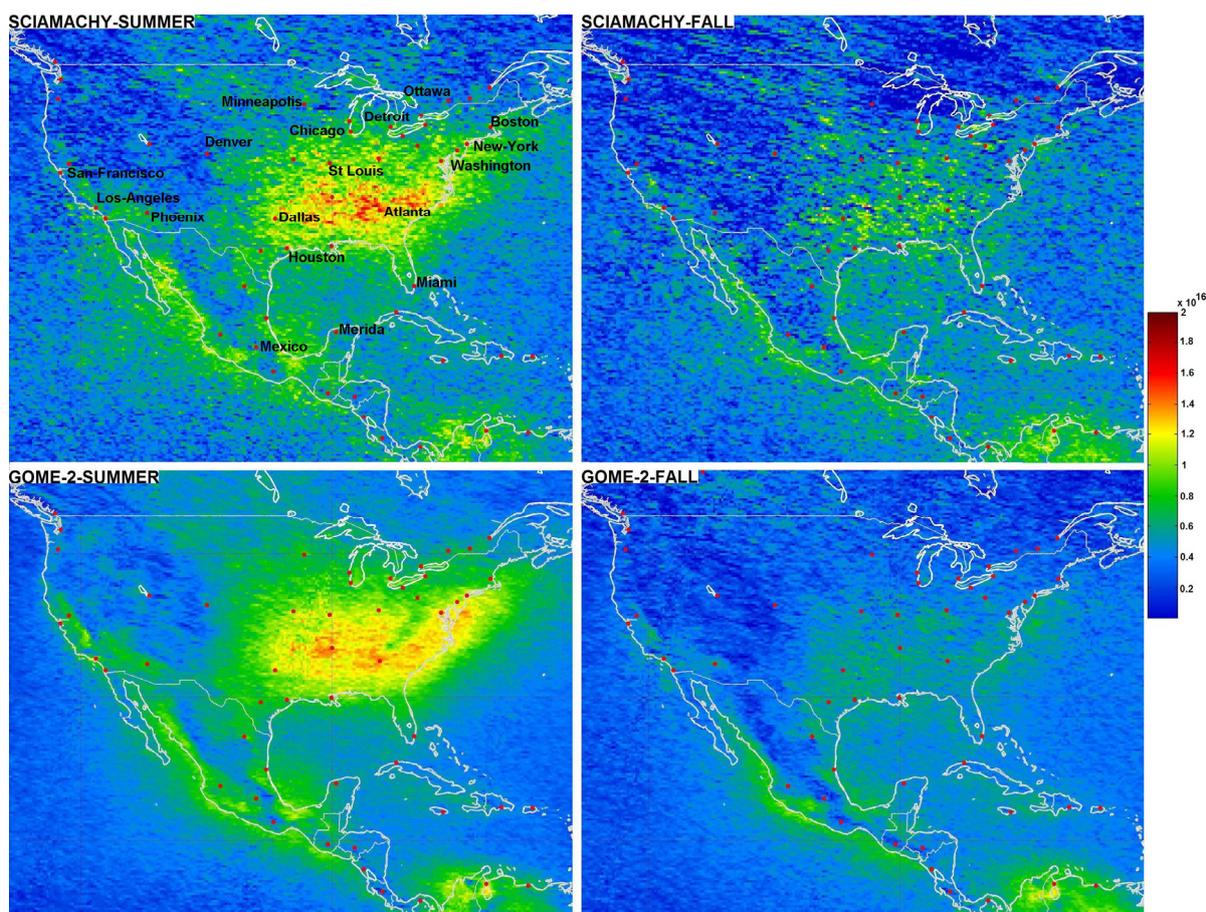
Figure 7-16: Comparison between H$_2$CO columns in central America and the United States retrieved from SCIAMACHY (upper panel) and GOME-2 (lower pane) during summer (left) and fall (right) seasons over the 2007-2010 period.

7.3.3 Russian wildfires in 2010

Temperature and solar insulation are the main drivers of isoprene emissions. It has been shown that climate change can potentially induce large long-term term variations in global biogenic NMVOC emissions [*Guenther et al.*, 2006] and that interannual variability of global isoprene emissions are induced by meteorological variability (in particular El Niño events) [*Müller et al.*, 2008]. On the local scale, measured isoprene and formaldehyde columns are correlated with air surface temperatures [*Palmer et al.*, 2006]. As another illustration of the interactions between climate and NMVOC emissions, we present here the Russian wildfire event which occurred during summer 2010, interesting as both biogenic and biomass burning sources contributed to the observed elevated H$_2$CO columns in the region.

During 2010, Russia experienced abnormal dry, hot weather starting early June. In late June, the heat wave slowly moved westward to the Ural Mountains and by July settled in European Russia. The persistent heat wave built in intensity as summer progressed and by late July and early August, temperatures were more than +10ºC warmer than normal at this time of the year. During the same period, many other abnormally high land temperatures were recorded, like in Eastern North America, Europe, and China. This can be seen in Figure 7-17 that shows the temperature anomalies in July 2010 calculated relative to the period 1970-2000. The heat wave continued up to mid-August, stopping around 19 August [http://www.ncdc.noaa.gov/sotc/global/2010/7]. Because of the record temperatures and drought, several hundred wildfires broke out across central and western

Russia, starting in late July (after the 20th). On 12 August 2010, a total of 26.739 natural fires had appeared on the Russian territory since the beginning of the fire period, burning a total area of 816.515 ha, both over grasslands and forests. The fires released enormous amounts of smoke and aerosols, carbon monoxide, and other pollutants severely degrading the air quality [http://www.gmes-atmosphere.eu/news/russian_fires].
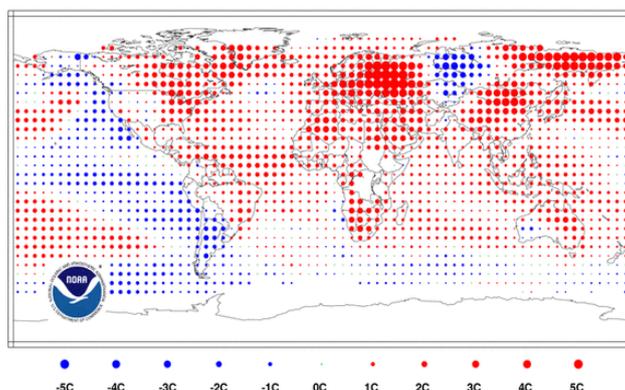
Figure 7-17: Map of temperature anomalies for July 2010. Anomalies are calculated relative to the period 1971-2000 [*National Climatic Data Center/NESDIS/NOAA*].
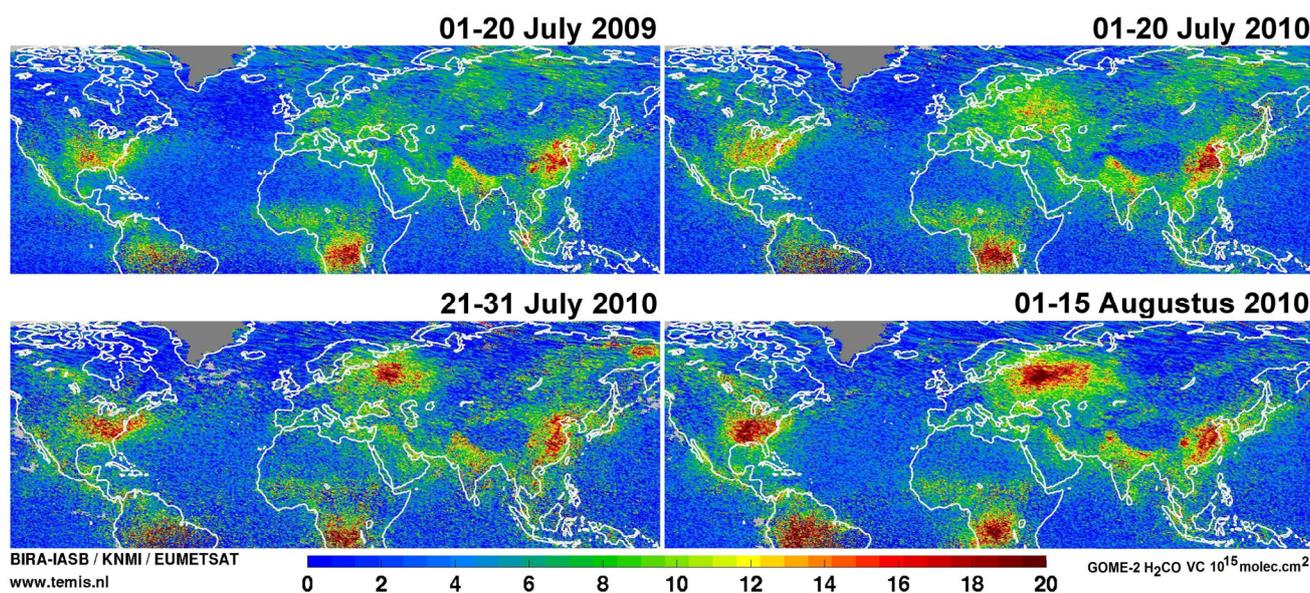
Figure 7-18: H₂CO columns retrieved from GOME-2 measurements between the first and 20th July 2009 and 2010, before the beginning of the Russian wildfires (upper panels), and between the 21st July and 15th August 2010, during the fires (lower panels).

Figure 7-18 shows formaldehyde columns retrieved from GOME-2 during this period. The upper panels compare H₂CO columns in 2009 and 2010, between the first and 20th July, before the beginning of the Russian fires. The H₂CO columns were noticeably higher in 2010 compared to 2009 over a large part of Europe and western Russia, most probably as a result of an increase of the biogenic emissions. The lower panels show the H₂CO columns during the fire period, with a first fire plume surrounding Moscow and its Southeastern region, extending later to Eastern Russia. Local fire effects were also detectable in some regions of Siberia and North of China. Note that the H₂CO satellite observations are not explicitly corrected for aerosol effects and the columns should

only be regarded qualitatively in this case of strong fires. In the future, we will consider applying an aerosol correction *a posteriori* for such special events, by performing off-line radiative transfer calculations using recorded or modelled aerosols properties.

7.4 Comparisons with MAX-DOAS ground-based observations

The validation of satellite H_2CO retrievals with data obtained by ground based remote sensing instruments, aircraft, *in situ* and/or other space based instruments is still sparse [*Vrekoussis et al.*, 2010]. Formaldehyde measurements by ground-based Fourier transform infrared spectrometers (FTIR) have been reported at the mountain station of Jungfraujoch (Switzerland) [*Demoulin et al.*, 1999], at polar stations [*Notholt et al.*, 1997], and during a ship cruise in the central Atlantic Ocean [*Notholt et al.*, 2000]. Multi-Axis-DOAS (MAX-DOAS) formaldehyde campaign measurements have been performed in the Italian Po-Valley in summer 2002 [*Heckel et al.*, 2005]. Comparisons between GOME and FTIR formaldehyde total columns have been presented in *Jones et al.* [2009] for Lauder (New Zealand), while comparisons between SCIAMACHY and MAX-DOAS H_2CO columns have been presented in *Wittrock et al.* [2006] in Nairobi (Kenya) and Cabauw (Netherlands). Comparisons between OMI H_2CO columns and aircraft measurements have been presented in *Millet et al.* [2008] and *Boeke et al.* [2011] for the North American region. Although these comparisons generally show satisfactory level of agreement, they rarely take into account the differences in vertical sensitivities of the observing systems.

7.4.1 Vertical sensitivities of ground-based measurements

An important issue when comparing ground-based and satellite data arise from the different vertical sensitivities of the measurement techniques. A rigorous formalism taking into account the different averaging kernels [*Rodgers and Conner;* 2003] was applied for the first time in *Vigouroux et al.* [2009] to the comparison of satellite and ground-based H_2CO total columns. MAX-DOAS and FTIR H_2CO measurements performed during several campaigns between 2004 and 2007 at the NDACC site of Reunion Island (21°S, 55°E) were intercompared in their common measurement period, from August to October 2004. Ground-based daily mean H_2CO total columns were also compared to correlative SCIAMACHY observations, accounting for the vertical sensitivity differences of the datasets.

The FTIR solar absorption measurements are performed in a wide spectral range (around 600–4500 cm^{-1}), which allow the retrieval of many species. The volume mixing ratio profiles of target gases are retrieved from the shapes of their absorption lines, which are pressure and temperature dependent. More detailed specifications of the FTIR experiments and of the H_2CO retrievals are respectively provided in *Senten et al.* [2008] and in *Vigouroux et al.* [2009]. MAX-DOAS instruments are designed to allow for quasi-simultaneous observations of scattered sunlight in a range of different line-of-sight (LOS) directions from the horizon to the zenith. As the light path through the troposphere increases with low elevation angles, off-axis measurements are very sensitive to tropospheric trace gas absorption. However, the length of the light path of the off-axis observations is very sensitive to aerosols and clouds and the retrievals generally require full radiative transfer model calculations [*Wittrock et al.*, 2004; *Heckel et al.*, 2005].

The left panel of Figure 7-19 shows an illustration of the SCIAMACHY, ground-based FTIR and MAX-DOAS averaging kernels, while the right panel shows the daily averaged observations made by each instrument from August to October 2004. In this Tropical site, away from the emission regions, formaldehyde observations are dominated by methane oxidation, with a possible

contribution from long-range transport of NMVOCs from Southeastern Africa. Both FTIR and MAX-DOAS H$_2$CO total columns were found to be in good agreement with SCIAMACHY values. The three instruments showed the same seasonal cycle with a minimum in austral winter, due to the lower OH concentrations during this period. However, the MAX-DOAS measurements showed reduced seasonal amplitude due to their lack of sensitivity in the free troposphere [*Vigouroux et al.* 2009].
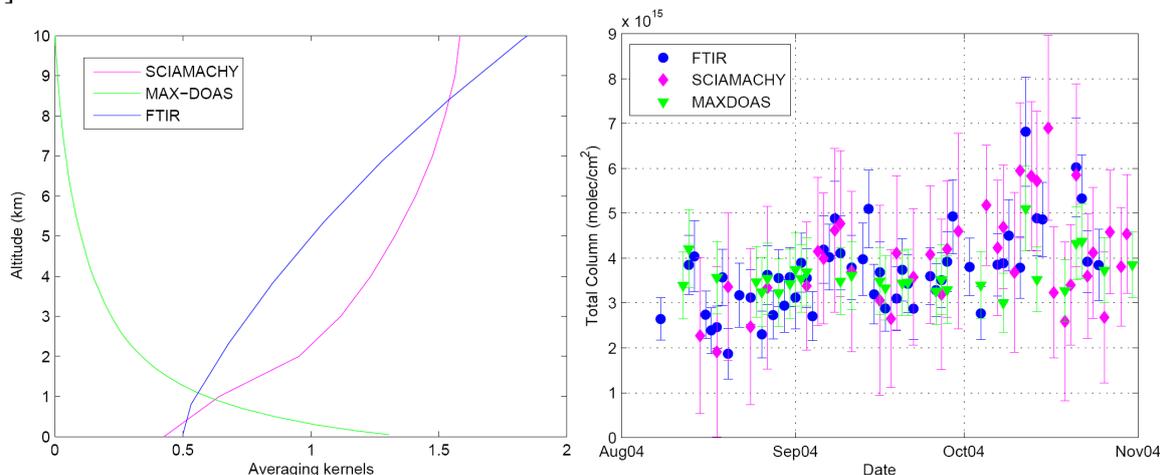
Figure 7-19: Typical FTIR (in blue), MAX-DOAS (in green) and satellite (in pink) averaging kernels and daily means of formaldehyde total columns above Saint-Denis during the 2004 campaign [*Vigouroux et al.*, 2009].

Based on this precursor study, we have applied the same formalism for comparison between GOME-2 H$_2$CO columns and MAX-DOAS measurements performed in Cabauw and in Beijing between 2008 and 2009. Measurements used here were made using a MAX-DOAS instrument designed and assembled at the Belgian Institute for Space Aeronomy. A retrieval algorithm based on the radiative transfer code LIDORT and the optimal estimation technique [*Rodgers*, 2000] has been developed by *Clémer et al.* [2010a and b] in order to provide simultaneous information on aerosol extinction and trace gas profiles. In a first step, the aerosol extinction vertical profiles are retrieved from measurements of the O$_4$ absorptions for different line-of-sight directions. Then, in a second step, these aerosol profiles are used as input for the retrieval of the vertical profiles of a number of tropospheric trace gases, such as NO$_2$, H$_2$CO, SO$_2$, O$_3$, and CHOCHO [*Clémer et al.*, 2010a and b].

The total column averaging kernels of the GOME-2 and MAX-DOAS instruments (**AK_1** and **AK_2**.) have a maximum sensitivity to formaldehyde in two different parts of the atmosphere (see the left panel of Figure 7-19). Consequently the retrieved total columns cannot be directly compared. If one assumes that the MAX-DOAS retrievals represent the best estimate of the true atmospheric state, the MAX-DOAS H$_2$CO profiles (**x_2^{pc}**) must be smoothed with the vertical sensitivity of the satellite to allow for geophysically consistent comparisons [*Rodgers and Conner;* 2003; *Boersma et al.*, 2004; *Vigouroux et al.*, 2009] (see also chapter 6). The smoothed differences between the total columns are then defined as:

$$\partial^{smooth} = VC_1 - AK_1 x_2^{pc} \tag{7-1}$$

7.4.2 Comparisons with MAX-DOAS observations in Cabauw

In summer 2009, the Cabauw Intercomparison Campaign of Nitrogen Dioxide measuring Instruments (CINDI) took place in Cabauw (52°N, 5°E), The Netherlands. The main objective of

this campaign was to intercompare NO_2 measuring instruments that can be used for validation of satellite tropospheric NO_2. However, other parameters were measured as well, among them O_3, aerosol, H_2CO, and CHOCHO. Here, we use MAX-DOAS measurements of H_2CO profiles performed by the BIRA-IASB team. Figure 7-20 shows the resulting daily mean MAX-DOAS total columns (in red) compared to the daily GOME-2 columns retrieved above Cabauw (in green). The MAX-DOAS and GOME-2 values generally coincide within their errors bars and the day-to-day variability is relatively well reproduced by both instruments. While the direct comparison result in satellite columns lower by 30% in average compared to the MAX-DOAS measurements, the mean smoothed difference is lower than 10%, reflecting a very good agreement of the observations.
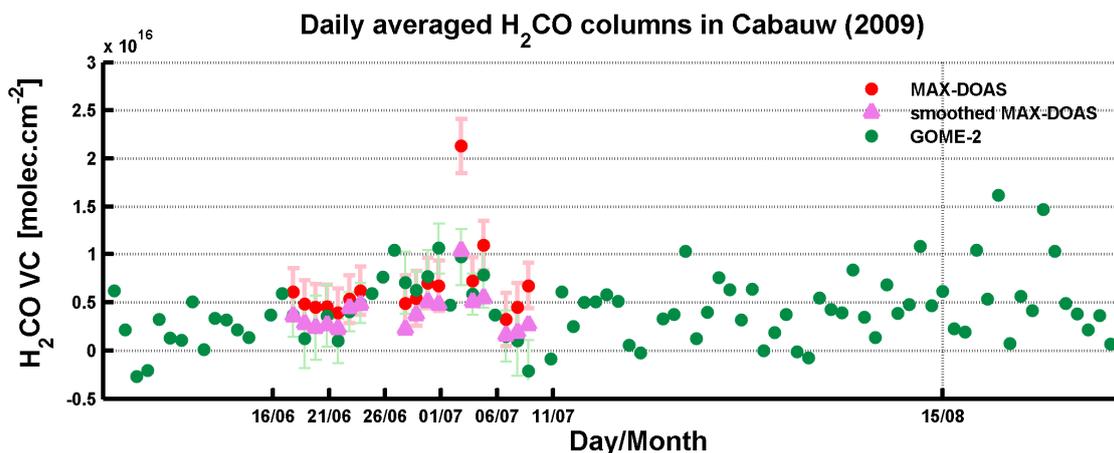
Figure 7-20: Daily averaged H₂CO vertical columns retrieved from MAX-DOAS (in red) and GOME-2 measurements (in green) in Cabauw during summer 2009. The plot shows also the MAX-DOAS columns smoothed with the satellite averaging kernels (in purple).
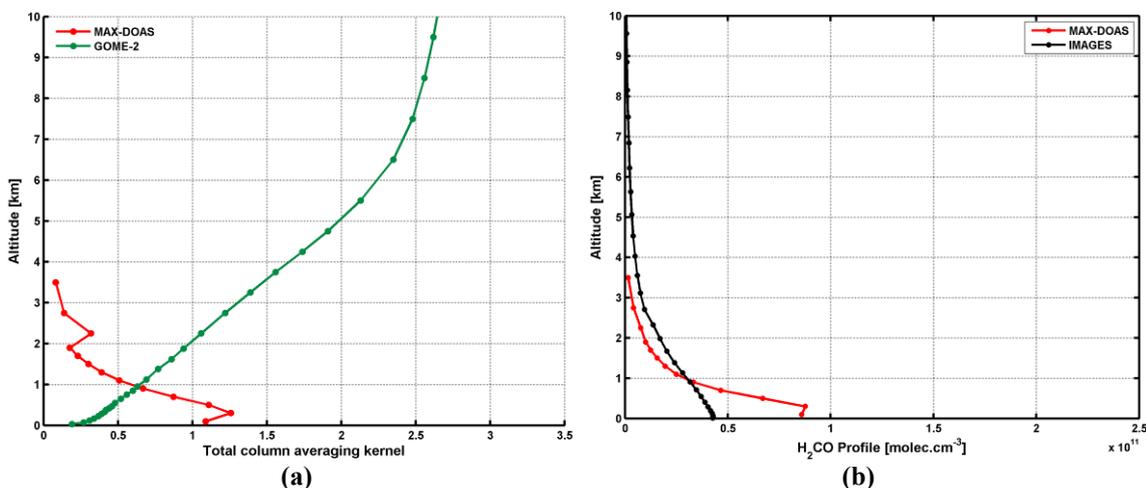
Figure 7-21: (a) Total column averaging kernels of the MAX-DOAS and GOME-2 H₂CO retrievals and (b) H₂CO vertical profiles simulated by IMAGES and retrieved from the MAX-DOAS measurements, in Cabauw in July 2009.

Figure 7-21 (a) shows the averaging kernels of the MAX-DOAS and GOME-2 H_2CO retrievals in Cabauw for July 2009. The corresponding IMAGES profile, used as *a priori* in the GOME-2 retrieval, and the retrieved MAX-DOAS H_2CO profile are represented in Figure 7-21 (b). As can be seen, the GOME-2 and MAX-DOAS instruments have a maximum of sensitivity to formaldehyde respectively above and below 1 km. Despite being not optimal, the comparison of both datasets is still relevant since the GOME-2 vertical sensitivity does not vanish at the surface. Also, the MAX-

DOAS averaging kernel is higher than 0.2 up to 3 km. Furthermore, 30% of the MAX-DOAS profile is above 1 km.

7.4.3 Comparisons with MAX-DOAS observations in Beijing

As part of the AMFIC project [www.amfic.eu], BIRA-IASB operated a MAX-DOAS instrument in Beijing (39.98°N, 116.38°E), in collaboration with the Institute of Atmospheric Physics (IAP) of the Chinese Academy of Sciences (CAS). The system was located a few hundred meters from the Olympic stadium, and was continuously operated from July 2008 to April 2009.
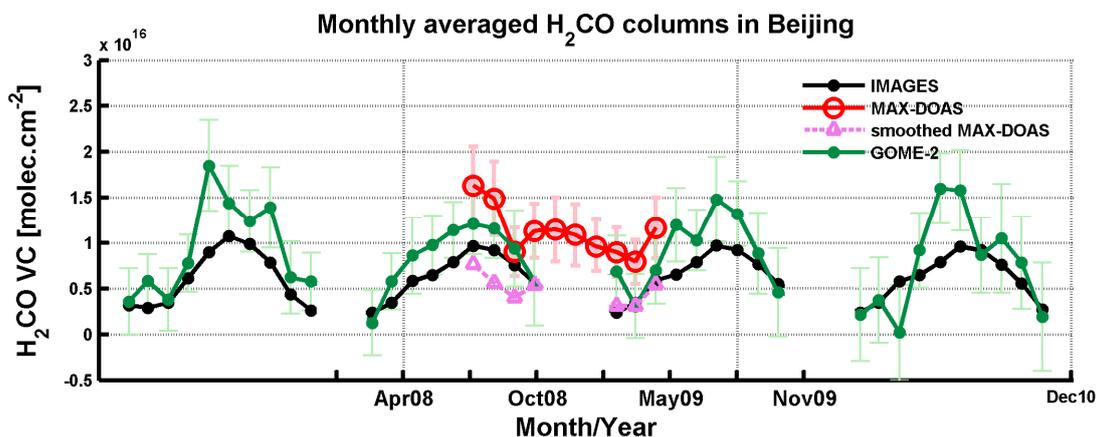
Figure 7-22: Monthly averaged H$_2$CO vertical columns retrieved from MAX-DOAS and GOME-2 measurements in Beijing. The plot also shows the MAX-DOAS columns smoothed with the satellite averaging kernels, and the H$_2$CO columns simulated by the IMAGES model.

(a)   **(b)**
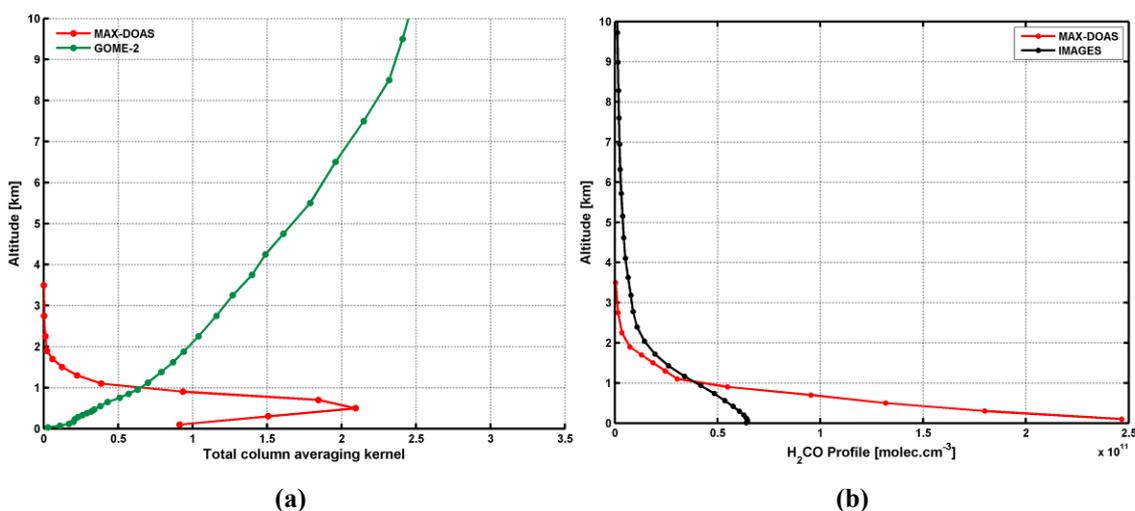
Figure 7-23: (a) Total column averaging kernels of the MAX-DOAS and GOME-2 H$_2$CO retrievals and (b) H$_2$CO vertical profiles simulated by IMAGES and retrieved from the MAX-DOAS measurements, in Beijing in July 2008.

Figure 7-22 shows the GOME-2 monthly averaged vertical columns extracted in a radius of 50 km around Beijing. Because of the limit on the solar zenith angle of 60°, there is no satellite result between November and January each year. The satellite columns are compared to monthly averaged MAX-DOAS H$_2$CO columns retrieved at the satellite overpass time (9:30 am). Both measurements capture the same seasonal variation. However the satellite columns are in average 30% lower than the ground-based measurements. In contrast to Cabauw, the smoothing of the MAX-DOAS profiles with the satellite averaging kernels is not found to significantly improve to comparison, the mean

difference after smoothing being 25%. This might be explained by the larger differences in the vertical sensitivity of the measurements. As an illustration, Figure 7-23 presents (a) the MAX-DOAS and GOME-2 averaging kernels and (b) the MAX-DOAS and IMAGES H₂CO profiles for July 2008. In this case, 90% of the retrieved MAX-DOAS profile lies below 1 km, where the satellite sensitivity is the lowest, while the sensitivity of the MAX-DOAS retrieval is zero above 2 km. The *a priori* profile shape provided by IMAGES is much less peaked near the surface than the MAX-DOAS profile. Although the agreement is within the error bars of both measurements, the differences in sensitivities are difficult to reconcile.

Besides the different vertical sensitivities of the measurements, another important issue is the smoothing of the satellite columns in the horizontal space, and the representativeness of a site for satellite validation. The spatial variability of the H₂CO concentrations between the centre of Beijing and its surroundings is difficult to capture with the GOME-2 resolution of 80×40 km². Since June 2010, the BIRA-IASB system has been operated in Xianghe, in the suburban area of Beijing, 50 km to the East. This slightly less polluted site may be more representative of the satellite observations. This will be investigated in future studies.

An additional issue comes from the diurnal variation of the H₂CO columns as shown in Figure 7-24, for MAX-DOAS observations in July 2008 and March 2009. It is interesting to note that the summertime H₂CO column increases by almost 1×10^{16} molec.cm^{-2} between 9:30 am and 13:30 pm, which should be taken into consideration for future studies involving the comparison of *e.g.* GOME-2 and OMI sensors.
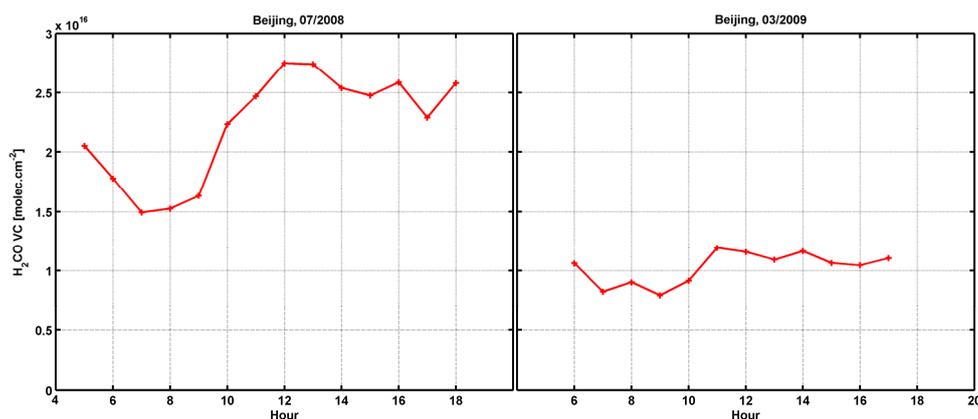
Figure 7-24: Averaged diurnal variation of the H₂CO column measured by the MAX-DOAS instrument in Beijing in July 2008 and in March 2009.

7.5 Conclusions

H₂CO columns have been retrieved from GOME-2 measurements using an optimized retrieval scheme, based on the experience gained from the GOME and SCIAMACHY instruments. The new improved settings consist in (1) fitting separately the BrO slant columns in a larger spectral window, thereby allowing for a decorrelation of the BrO and H₂CO spectral features and for a reduction of the scatter in the BrO slant columns, and (2) implementing a modified DOAS approach, using O₃ absorption cross-sections corrected with wavelength-dependent air mass factors, which minimizes residuals at moderate and large solar zenith angles. These improvements allow for a 30% reduction of the scatter in the H₂CO slant columns and lead to a highly satisfactory level of agreement between SCIAMACHY and GOME-2 H₂CO values, better than 7% everywhere, expect over Australia.

The higher sampling of GOME-2 observations is shown to improve the signal to noise ratio of the monthly averaged satellite H_2CO observations, allowing for a better definition of the H_2CO spatial distribution in comparison to GOME and SCIAMACHY. The higher sampling of GOME-2 also allows working with weekly H_2CO observations. However, we conclude from our analysis that instruments combining both high spatial resolution and daily Earth coverage, such as NASA OMI on AURA and the future TROPOMI on the GMES Sentinel-5 precursor mission, are best suited to monitor air quality at the urban level, even for weak absorbers such as formaldehyde whose retrievals are characterised by large individual random errors.

In an attempt to validate our GOME-2 product, daily averaged GOME-2 H_2CO columns have been compared with ground-based MAX-DOAS measurements performed in Cabauw during the CINDI campaign in June and July 2009. When taking into account the differences in vertical measurement sensitivities, the comparison is found to be better than 10%. Conversely, similar comparison using MAX-DOAS measurements in Beijing between July 2008 and April 2009 show larger discrepancies even when smoothing the retrieved MAX-DOAS profiles with the satellite averaging kernels. This might be attributed to the larger difference of sensitivities of the two instruments making the comparison of the two measurement datasets very difficult in the highly polluted megacity of Beijing. For further validation of the H_2CO satellite observations, more correlative sites are needed, in particular in tropical regions, characterized by strong NMVOC emissions, and at the suburban level in mid-latitudes. Furthermore, ground-based FTIR and MAX-DOAS measurements need to be combined because their vertical sensitivities are complementary.

The GOME, SCIAMACHY and GOME-2 H_2CO products developed at BIRA-IASB are available on the TEMIS web site [www.temis.nl]. Furthermore, an adapted version of this scientific algorithm has been implemented in the GOME-2 trace gas operational data processor (GDP v4.4), operated at DLR on behalf of EUMETSAT [http://atmos.caf.dlr.de/gome/product_hcho.html].

8   Trend Detection in Satellite Observations of Formaldehyde Tropospheric Columns

We present the first trend study performed on H_2CO satellite columns, retrieved from the GOME, SCIAMACHY and GOME-2 measurements between 1997 and 2010. A linear model with a seasonal component is used to fit the time series of monthly averaged H_2CO columns. The error and statistical significance of the inferred trends are estimated, taking into account the uncertainties on the satellite observations. The study focuses on Asia, but results are also provided for large cities worldwide. Statistically significant positive trends of formaldehyde columns are observed over Northeastern China (4% yr^{-1}) and India (1.7% yr^{-1}), and are related to strong increases in anthropogenic NMVOC emissions, whereas negative trends of about -2.5% yr^{-1} are observed over Tokyo as well as over cities of the Northeastern United States urban corridor as a result of effective pollution regulation measures. These results have been published in *Geophys. Res. Lett.* [*De Smedt et al.*, 2010][1].

8.1 Introduction

As further developed in chapter 2, non-methane volatile organic compounds (NMVOCs) play an important role in global tropospheric chemistry and air quality because they participate to the photochemical production of tropospheric ozone (in the presence of NO_x) and of secondary organic aerosols [*Houweling et al.,* 1998; *Kanakidou et al.,* 2005]. Albeit NO_x and NMVOC concentrations are expected to decrease in many industrialized countries, the rapid industrial development and urbanization in parts of Asia during the last decades resulted in significant increase of anthropogenic emissions, primarily from coal-fired power plants and automobile exhausts. In both cases, long-term measurements providing information on tropospheric trace gases are indispensable to assess and quantify these tendencies. Such time series can be used to monitor pollution trends, evaluate emission inventories, and the effect of measures taken to improve air quality [*Fishman et al.*, 2008].

Satellite observations of tropospheric nitrogen dioxide columns retrieved from GOME and SCIAMACHY measurements have been used in several studies to quantify trends in nitrogen oxide emissions [*Richter et al.*, 2005; *van der A et al.*, 2006 and 2008; *Stavrakou et al.*, 2008]. Substantial reductions in NO_x emissions have been found over some areas of the Eastern United States (-4.3% yr^{-1} in the Ohio River Valley) and to a lesser extent over Europe (-1.4% yr^{-1} in Germany, -1.0% yr^{-1} in the Po Basin). In contrast, a strong increase is observed in Asia, higher than suggested by bottom-up inventories. The largest emission increases (growth rate of 9.6% yr^{-1}) are found in industrial areas of Eastern China with a fast economical growth, in particular in the Beijing area [*Stavrakou et al.*, 2008].

---

[1] De Smedt, I., T. Stavrakou, J.-F. Müller, R. J. van Der A, and M. Van Roozendael (2010), Trend detection in satellite observations of formaldehyde tropospheric columns, *Geophys. Res. Lett.*, *37*(18), L18808, doi:10.1029/2010GL044245.

In this work, we take advantage of the 14-year long GOME-SCIAMACHY-GOME-2 record (chapters 6 and 7) to estimate trends and relate them to changes in anthropogenic NMVOC emissions using the IMAGESv2 global chemistry transport model [*Stavrakou et al.,* 2009a]. Our approach follows the method of *van der A et al* [2006 and 2008], used for the determination of trends in satellite NO_2 observations. However, trend detection of anthropogenic emissions in satellite formaldehyde observations suffers from several limitations. First, the uncertainties of H_2CO columns are larger than those of NO_2 because the H_2CO optical density is two orders of magnitude weaker than NO_2. Moreover, due to Rayleigh scattering, the sensitivity of satellite measurements near the surface is two to three times lower in the UV than in the visible wavelength range. Finally, whereas changes in NO_2 columns and changes in NO_x emissions are expected to be of the same order of magnitude [*Richter et al.,* 2005], anthropogenic emissions of NMVOCs are more difficult to identify because continental H_2CO columns are usually dominated by biogenic sources. To illustrate this last point, a simulation performed with the IMAGESv2 model for the year 2007 is shown in Figure 8-1. It indicates that a 100% increase of the anthropogenic NMVOC emissions in the model results in a variable increase of the H_2CO columns, comprised between 20% and 40% in the Northern hemisphere. Our main focus is on Asia, where the increases of anthropogenic emissions are the largest, but results for world megacities are also discussed.
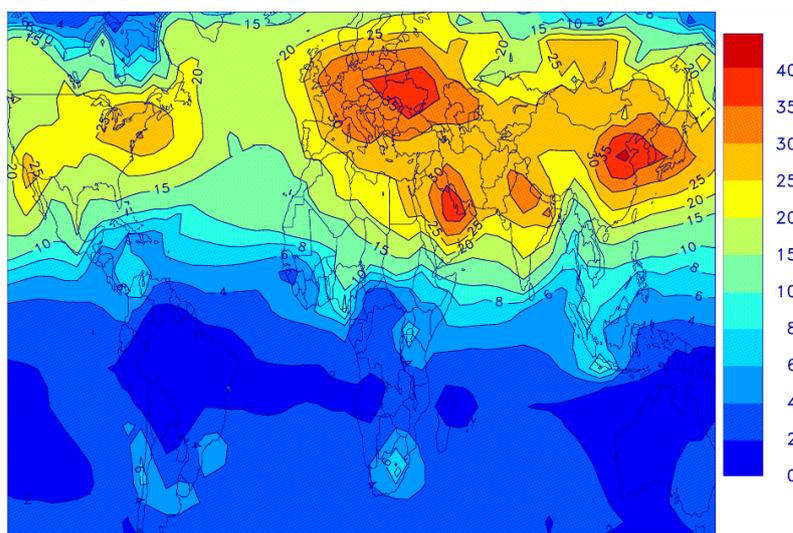
Figure 8-1: Modelled annual change in the H₂CO column for a 100% increase of the anthropogenic NMVOC emissions. Simulations were performed for the year 2007 with the IMAGESv2 model [*Stavrakou et al.,* 2009a].

8.2 Satellite Observations

A combination of H_2CO columns retrieved from GOME/ERS-2 (1997-2002) and from SCIAMACHY/ENVISAT (2003-2010) is used in this study, as originally done in *De Smedt et al.* [2010]. Here, in order to support the results reported from GOME and SCIAMACHY, the GOME-2/METOP-A observations between 2007 and 2010 have been included in a subsequent separated analysis. A complete description of the retrieval, based on the DOAS method, is provided in chapters 6 and 7. The retrieval settings have been homogenized as far as possible to avoid any inconstancies between the datasets.

Figure 8-2 shows the mean H_2CO vertical columns (*VC*) in Asia as retrieved from GOME between 1997 and 2002 and from SCIAMACHY between 2003 and 2010. Due to the decreasing sensitivity of tropospheric retrievals with increasing solar zenith angle, only columns for SZA lower than 60°

are accounted for. Observations with cloud fractions larger than 40% are also filtered out. No explicit correction is applied for aerosols but the cloud correction scheme accounts for a large part of the aerosol scattering effect. The total error on the monthly and regionally averaged columns is comprised between 20 and 40%. The detection limit (defined as three times the random error) is estimated at about 2×10^{15} molec.cm^{-2} [*Lerot et al.*, 2010b].
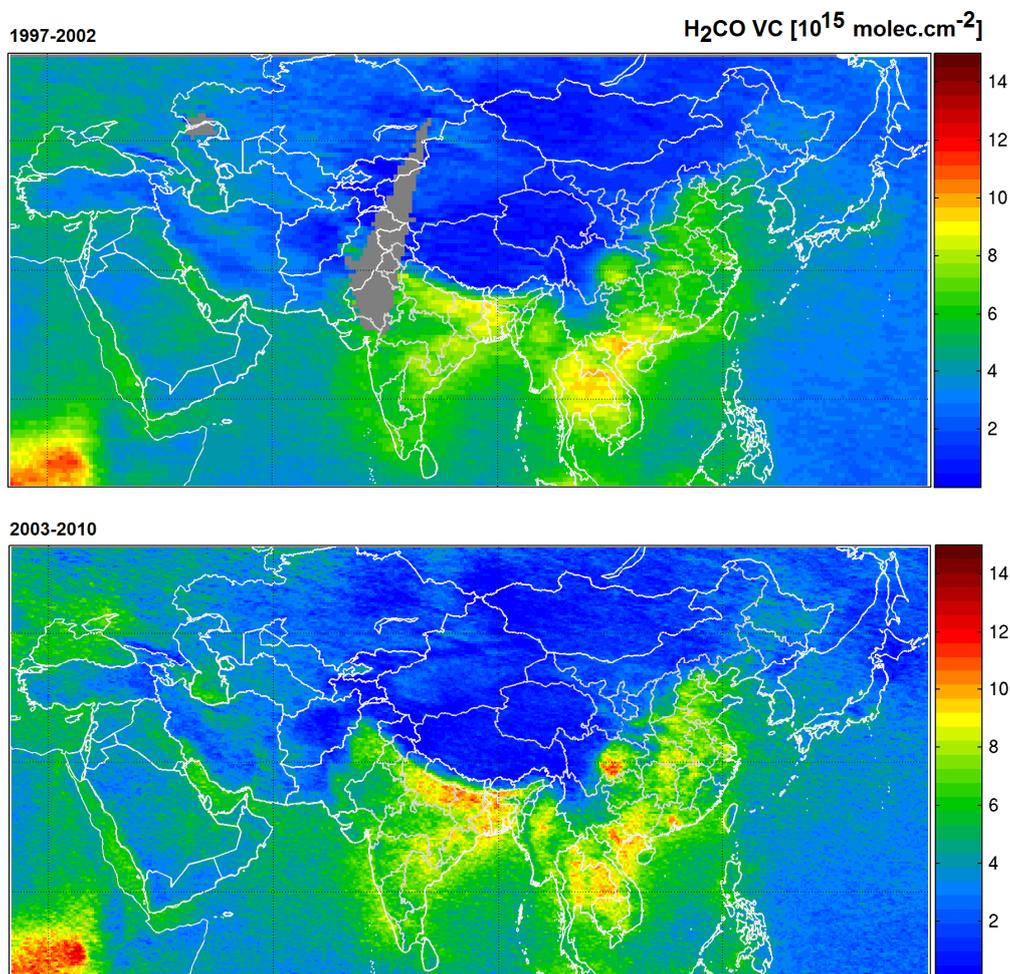
Figure 8-2: Averaged H₂CO vertical columns in Asia retrieved from GOME (upper panel) and SCIAMACHY (lower panel).

8.3 Trend Analysis Method

Formaldehyde columns generally show a marked seasonal variation dominated by biogenic emissions. To fit the time-series of monthly observations, a model with a linear trend and an intra-annual function, modelling the seasonal variations, has been used [*van der A et al.*, 2006]:

$$m(t) = A + Bt + \sum_{n=1}^{4} C_n \cos(n2\pi t) + \sum_{n=1}^{4} D_n \sin(n2\pi t) + U\delta \qquad \text{(8-1)}$$

where *m(t)* is the observed monthly H₂CO *VC* for month *t* (expressed in fractional year), and *A*, *B*, *Cₙ*, *Dₙ* and *U* are the fitted parameters. *A* is the annual mean of the first year, *B* is the annual trend, expressed in molec.cm² yr^{-1}. *Cₙ* and *Dₙ* are the Fourier terms used to model the seasonal variations. The number of terms has to be chosen in order to minimize the residuals without over-fitting the

data. In the case of formaldehyde, a 4th order series provided the best overall representation. U accounts for a possible bias between the satellite retrievals (δ switches from 0 to 1 in Jan. 2003, when the time series switches from GOME to SCIAMACHY). The local overpass time of GOME (10.30 am) and SCIAMACHY (10.00 am) is close enough to avoid systematic differences in their time series.

A linear least-squares method is used to fit Equation (8-1) with the observations, taking their errors into account. In order to assess the statistical significance of the derived annual trends, the uncertainties on the fitted parameters are calculated using the bootstrap resampling method [*Gardiner et al.*, 2008]. This technique enables to treat non-normally distributed data and is designed to accommodate for outliers, while providing uncertainties comparable with least-squares estimates when a Gaussian model is appropriate. Once the initial fit has been determined, the residuals are used randomly to create perturbations to the initial dataset, and the model is re-fitted. This operation of bootstrap resampling is repeated a large number of time (5000) to generate a distribution of trend parameters. The difference between the lower and the upper confidence limits of the inferred trends (the 2.5% and 97.5% quantiles) is equivalent to the $2\text{-}\sigma_B$ standard deviation for normal distributions [*Gardiner et al.*, 2008]. Uncertainties on the other fitted parameters are calculated in the same way. The statistical significance level of the trend is better than 95% for

$$t_B = \left| \frac{B}{\sigma_B} \right|$$ larger than 2 [*Weatherhead et al.*, 1998; *van der A et al.*, 2006].

To estimate the annual growth rate ($R=B/A$, expressed in % yr⁻¹), the fitted annual H₂CO concentration in 1997 ($A\pm\sigma_A$) is taken as reference. The growth rate uncertainty (σ_R) is therefore calculated as:

$$\left(\frac{\sigma_R}{R} \right)^2 = \left(\frac{\sigma_A}{A} \right)^2 + \left(\frac{\sigma_B}{B} \right)^2 - 2 \frac{\sigma_A \sigma_B}{AB} \rho_{AB} \qquad (8\text{-}2)$$

where ρ_{AB} is the correlation coefficient between the distributions of A and B [*Taylor*, 1982].
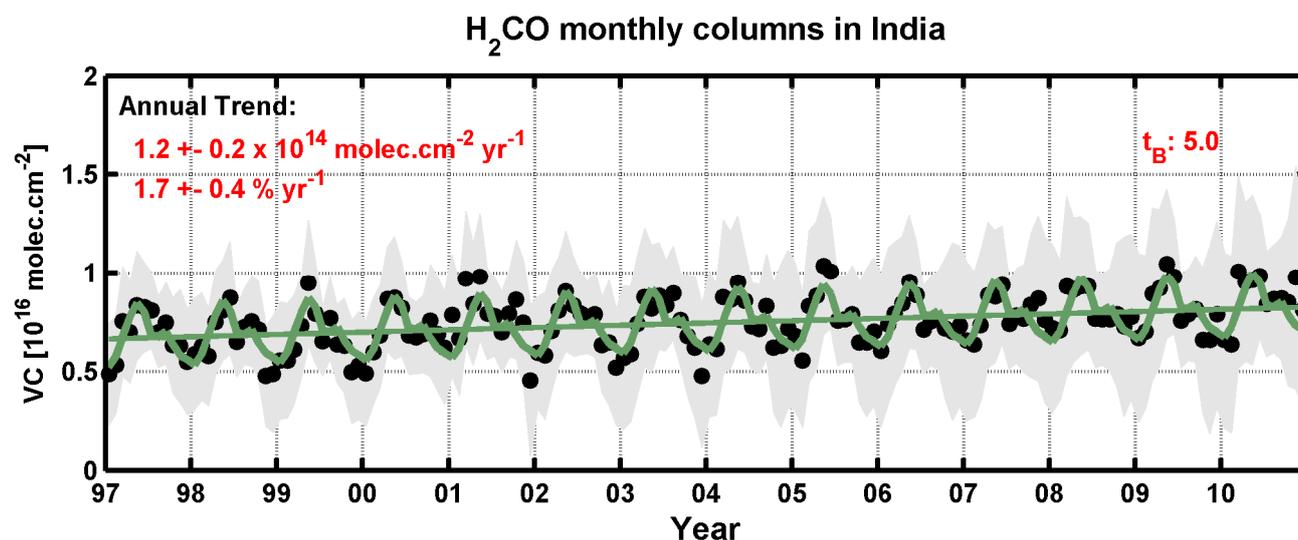
Figure 8-3: Example of measured (black dots) and fitted (green line) time series of H₂CO columns over India, including the underlying trend (green line).

The precision of the inferred trends is limited by the observation errors and by the length of the time series. To obtain significant results, the satellite monthly H₂CO columns need to be spatially

averaged over relatively large areas, limiting the spatial resolution of the trend study. Figure 8-3 presents an example of monthly averaged observations of H$_2$CO over the whole India and the fitted time series using Equation (8-1). In this particular case, the value of the annual trend **B** is 1.2±0.2x10^{14} molec.cm^{-2} yr^{-1} (**t$_B$** = 5) and the annual growth rate is 1.7±0.4% yr^{-1}. This represents a total increase of 24% between 1997 and 2010. In India, the fitted offset **U** between GOME and SCIAMACHY is 0.5x10^{14} molec.cm^{-2}. More generally, the magnitude of the offset is found to be always lower than 1.5x10^{15} molec.cm^{-2}, which is smaller than the minimum error on the observations.

8.4 Results

The model of Equation (8-1) has been applied over the large regions shown in chapters 6 and 7 (Southeastern US, Guatemala, Amazonia, Northern, Equatorial and Southern Africa, India, Southern China, Thailand, Indonesia, Europe and Northern Australia), and over the Pacific Ocean. Over these regions, significant trends were detected only in China and India, allowing us to exclude any systematic artificial trend in the H$_2$CO dataset, or in the background correction taken from the IMAGESv2 model over the Pacific Ocean. The trend analysis was also performed globally, using monthly H$_2$CO data averaged on a 5°x5° grid in latitude and longitude. The same features were observed, with more detailed patterns in Asia and around large urban areas. To draw a more precise picture of the inferred trends in Asia, monthly averaged time series of H$_2$CO observations have been calculated by country, and by province or state in China and India. The coordinates boundaries were obtained from to the Environmental Systems Research Institute (ESRI) which provides world administrative units in a geographic information system format (GIS) [http://www.esri.com/, http://www.cdc.gov/epiinfo/]. Additionally, a trend study was performed using monthly averaged H$_2$CO data around largest urban areas worldwide. In the following, we present the trend analysis results for China, India, other Asian countries, as well as for some world megacities, where the trends were found to be statistically significant.

8.4.1 Asia

The absolute trends derived from the GOME and SCIAMACHY observations between 1997 and 2010 and the corresponding growth rates relative to 1997 are shown on Figure 8-4. The values of the statistically significant trends are given in Table 8-1. As a support to the Tables and Figures, Figure 8-5 shows a map of Asian countries, while Figure 8-8 and Figure 8-12 present respectively the Chinese provinces and Indian states. China and India are the countries presenting the most significant trends, of respectively 3.6±0.9% yr^{-1} and 1.7±0.4% yr^{-1}. Results in China and India are detailed in the following sections. In other Asian countries, positive trends are found over South Korea (4.1% yr^{-1}), Turkey (3.5% yr^{-1}), Israel (3.4% yr^{-1}) and Jordan (1.8% yr^{-1}). The largest annual growth rate is found in Mongolia (22.5±15.9% yr^{-1}) but the related uncertainty is large because of the very low H$_2$CO columns in 1997. Afghanistan presents also a significant growth rate (3.8±1.6% yr^{-1}). A negative trend is found in Myanmar (Burma) (-1.3% yr^{-1}). Thailand exhibits no trend while it presents the highest annual H$_2$CO columns, dominated by biogenic sources and biomass burning (see Figure 8-2). In a separated analysis, the GOME-2 observations between 2007 and 2010 have been added to the time series. In this case, both SCIAMACHY and GOME-2 columns are used between 2007 and 2010, and two offset parameters are fitted in Equation (8-1). The corresponding results are provided in Table 8-2. The values on the inferred trends are equivalent to Table 8-1

(within their error bars), demonstrating the robustness of the method and the consistency of the satellite observations.
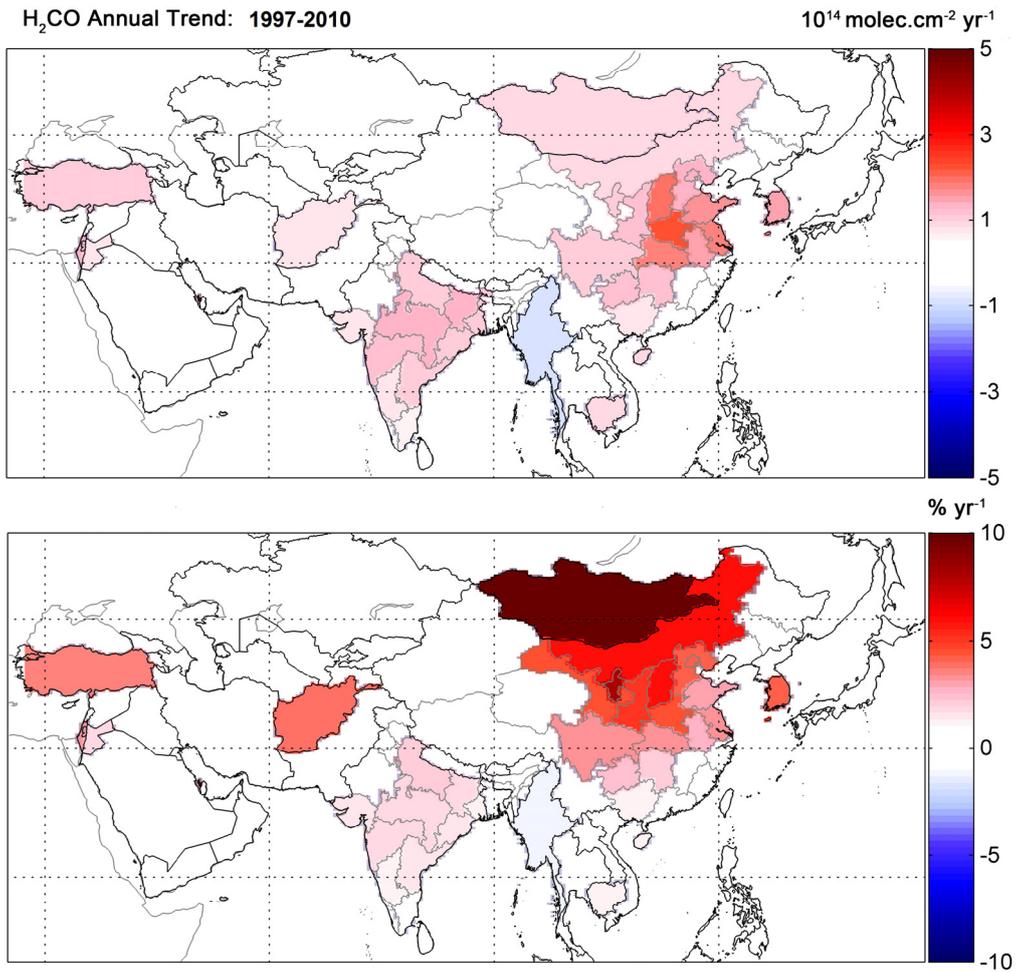
Figure 8-4: Absolute and relative annual trends over 1997-2010. The relative trends are calculated with 1997 as reference year. Blank areas represent not statistically significant results.

Figure 8-5: Map of Asian countries.

Table 8-1: Most statistically significant annual trends in Asian countries based on GOME and SCIAMACHY observations between 1997 and 2010.

Country	1997 H$_2$CO VC $A \pm \sigma_A$ [10^{14} molec.cm^{-2}]	Linear Trend $B \pm \sigma_B$ [10^{14} molec.cm^{-2} yr^{-1}]	Growth Rate $R \pm \sigma_R$ [% yr^{-1}]	Statistical Significance $t_B=|B/\sigma_B|$
India	66.3 ± 1.1	1.2 ± 0.2	1.7 ± 0.4	5.0
China	26.0 ± 0.8	0.9 ± 0.2	3.6 ± 0.9	4.6
Turkey	30.8 ± 1.5	1.1 ± 0.3	3.5 ± 1.1	3.6
Mongolia	04.1 ± 2.1	0.9 ± 0.3	22.5 ± 15.9	3.4
Israel	40.3 ± 2.4	1.4 ± 0.4	3.4 ± 1.2	3.4
South Korea	36.2 ± 2.7	1.5 ± 0.5	4.1 ± 1.5	3.3
Myanmar (Burma)	68.8 ± 1.5	-0.9 ± 0.3	-1.3 ± 0.4	3.0
Afghanistan	20.2 ± 1.4	0.8 ± 0.3	3.8 ± 1.6	2.9
Jordan	36.2 ± 1.4	0.7 ± 0.3	1.8 ± 0.8	2.4

Table 8-2: same as Table 8-1 but including GOME-2 observations between 2007 and 2010.

Country	1997 H$_2$CO VC $A \pm \sigma_A$ [10^{14} molec.cm^{-2}]	Linear Trend $B \pm \sigma_B$ [10^{14} molec.cm^{-2} yr^{-1}]	Growth Rate $R \pm \sigma_R$ [% yr^{-1}]	Statistical Significance $t_B=|B/\sigma_B|$
India	65.9 ± 1.2	1.4 ± 0.2	2.1 ± 0.4	6.2
China	26.8 ± 1.3	0.8 ± 0.2	3.0 ± 0.9	4.0
Turkey	30.1 ± 1.7	0.7 ± 0.3	2.5 ± 1.1	2.6
Mongolia	03.0 ± 2.3	0.7 ± 0.3	21.8 ± 23.2	2.3
Israel	39.0 ± 2.1	1.5 ± 0.3	3.9 ± 1.0	4.7
South Korea	35.5 ± 2.6	1.6 ± 0.4	4.5 ± 1.4	4.0
Myanmar (Burma)	70.2 ± 1.5	-1.0 ± 0.3	-1.5 ± 0.4	3.7
Afghanistan	20.2 ± 1.3	0.7 ± 0.2	3.2 ± 1.4	2.7
Jordan	36.1 ± 1.4	0.5 ± 0.3	1.5 ± 0.8	2.1

To make sure that the observed trends are not due to an indirect effect of the aerosols on the retrieved H$_2$CO vertical columns, we performed similar trend analysis on the cloud fractions and air mass factors used in the retrieval. Indeed, as a matter of fact, any negative trend in ***AMF*** would results in positive trend in the vertical columns (***VC=SC/AMF***). Results of this additional trend study are shown in Figure 8-6. Although a decrease of the cloud fractions is detected over Central China and Mongolia, this has no impact on the ***AMFs*** which present no statistically significant trends, confirming the actual increase of the tropospheric H$_2$CO columns [*van der A et al.*, 2006].
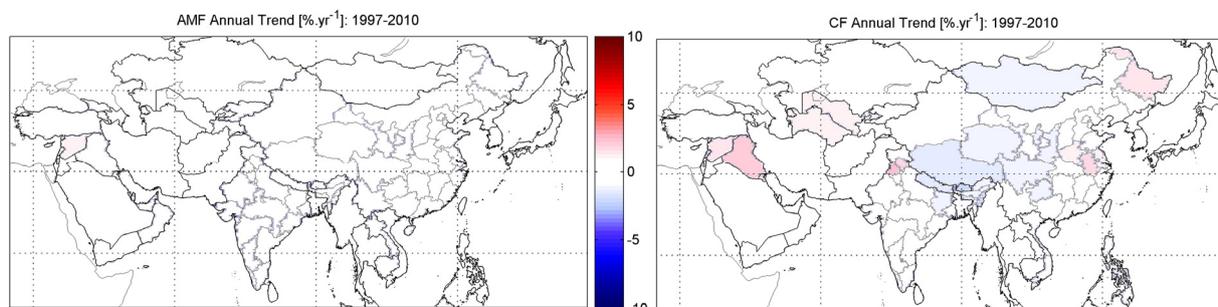
Figure 8-6: Annual trends in air mass factors and cloud fractions over 1997-2010. Blank areas represent not statistically significant results.

8.4.2 China

Over China, the inferred annual trend is $0.9\pm0.2\times10^{14}$ molec.cm^2 yr^{-1} ($3.6\pm0.9\%$ yr^{-1} between 1997 and 2010). As seen in Figure 8-4 and in Table 8-3, the largest absolute trends are found in the Northeastern part of the country (2×10^{14} molec.cm^2 yr^{-1}), *i.e.* in the region with the highest anthropogenic emissions [*Ohara et al.,* 2007]. Figure 8-7 shows the H₂CO trend in the Beijing - Tientsin - Hebei area, where the growth rate is $4.1\pm1.3\%$ yr^{-1}. Note that the H₂CO columns in summer 2008 are 20% lower than in summer 2007, reducing the importance of the growth rate (amounting to 5.6%.yr^{-1} until 2007). This might reflect the effect of the restrictions applied during the Olympics games, as shown with NO₂ satellite observations [*Mijling et al.*, 2009] and with ground-based measurements of secondary pollutants, including anthropogenic NMVOCs [*Wang et al.*, 2010]. In Northern China, the sparsely populated province of Inner Mongolia shows a strong trend of $6.2\pm2.5\%$ yr^{-1} between 1997 and 2010, whereas the Ningxia province, bordering Inner Mongolia, has the highest relative trend of $8\pm3.5\%$ yr^{-1}, indicative of a very fast industrial development most likely related to the construction of new power plants [*Zhang et al.,* 2009].

Table 8-3: Inferred statistically significant annual trends in Chinese provinces.

Province in China	1997 H₂CO VC $A \pm \sigma_A$ [10^{14} molec.cm^{-2}]	Linear Trend $B \pm \sigma_B$ [10^{14} molec.cm^{-2} yr^{-1}]	Growth Rate $R \pm \sigma_R$ [% yr^{-1}]	Statistical Significance $t_B=|B/\sigma_B|$
Henan	48.9 ± 1.9	2.2 ± 0.3	4.5 ± 0.9	6.4
Shanxi	32.4 ± 2.1	2.0 ± 0.4	6.1 ± 1.5	5.6
Hubei	52.6 ± 2.0	1.8 ± 0.3	3.4 ± 0.8	5.1
Jiangsu	54.1 ± 2.1	1.8 ± 0.4	3.3 ± 0.8	5.0
Anhui	58.1 ± 1.9	1.5 ± 0.3	2.6 ± 0.7	4.4
Sichuan	30.8 ± 1.3	1.1 ± 0.3	3.4 ± 1.0	4.1
Shandong	50.9 ± 2.4	1.6 ± 0.4	3.1 ± 0.9	4.0
Shaanxi	24.1 ± 1.8	1.2 ± 0.3	5.2 ± 1.7	3.8
Hebei	35.1 ± 2.3	1.4 ± 0.4	4.1 ± 1.3	3.7
Guizhou	51.5 ± 1.8	1.2 ± 0.3	2.3 ± 0.7	3.6
Hunan	57.8 ± 2.1	1.2 ± 0.4	2.2 ± 0.7	3.4
Ningxia	13.7 ± 1.9	1.1 ± 0.3	8.0 ± 3.5	3.2
Inner Mongolia	15.3 ± 1.7	0.9 ± 0.3	6.1 ± 2.5	3.1
Hainan	65.0 ± 1.9	0.9 ± 0.3	1.3 ± 0.6	2.5
Gansu	11.9 ± 1.0	0.5 ± 0.2	4.6 ± 2.2	2.5
Guangxi	69.0 ± 1.9	0.8 ± 0.3	1.1 ± 0.5	2.4

H₂CO monthly columns in Hebei/China
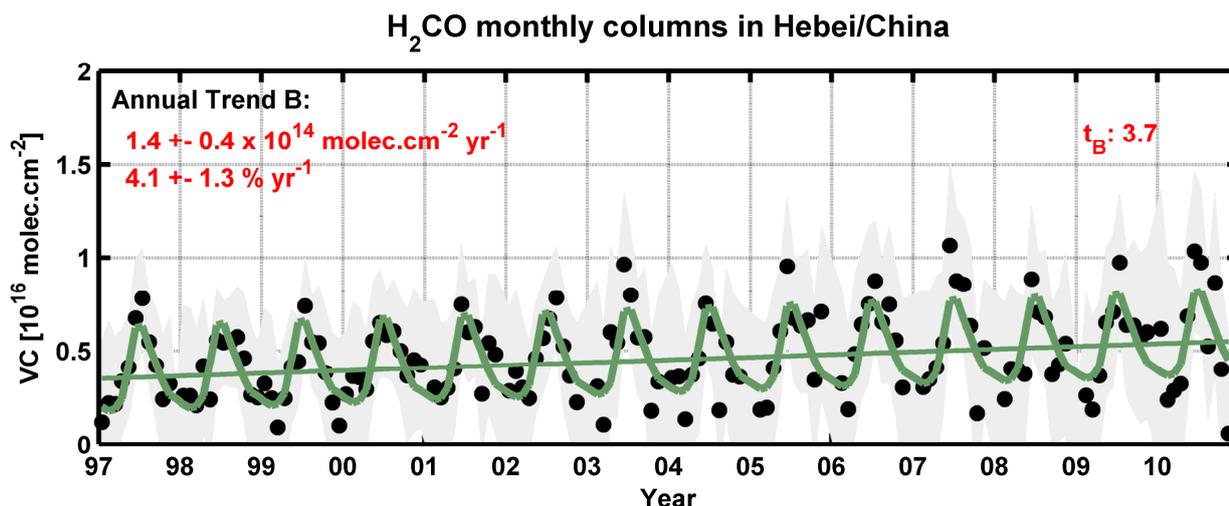
Figure 8-7: Measured (black dots) and fitted (green line) time series of H₂CO columns over the Beijing-Tientsin-Hebei area in China.

Based on geographic and economic criteria, the 31 Chinese provinces can be grouped into six distinct regions (see Figure 8-8) [*Cai et al.,* 2009]. The time series of both the annual and winter means of the regionally averaged columns are shown in Figure 8-9 and the corresponding trend values are provided in Table 8-4. The increasing trend of the annual means is significant in almost all regions. The highest growth rate (5.2±2% yr⁻¹) is found in the Northwestern region (mainly driven by the Shaanxi and Ningxia provinces), followed by the Northern and Central regions. When considering only the October to March period, when the contribution of natural sources of H₂CO precursors is lowest, the trend is much stronger, particularly in Northern China. For example, although the Northeastern region presents the weakest annual trend, it exhibits a strong trend of 4.6±2.0% yr⁻¹ in the Oct.-Mar. period. Similarly, the trend in the Northern region increases from 2.9 to 6.5% yr⁻¹, pointing to changes in anthropogenic NMVOC emissions. Again, including GOME-2 measurements between 2007 and 2010 leads to very consistent results, as shown in the two last columns of Table 8-4.

Figure 8-8: Maps of Chinese provinces divided in 6 distinct regions.
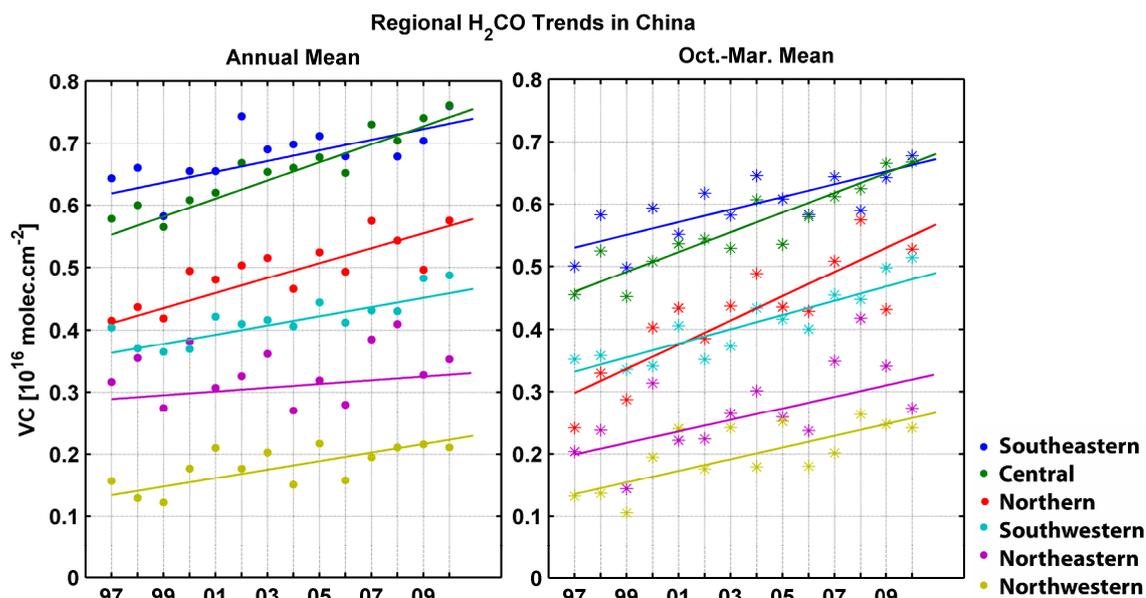
Figure 8-9: Annual (left) and winter (right) averaged H$_2$CO columns in 6 regions of China between 1997 and 2010 (dots and stars), and the corresponding annual trends (lines).

Table 8-4: H$_2$CO Trends between 1997 and 2010 for the regions shown in Figure 8-8. Absolute trends are in 10^{14} molec.cm^2 yr^{-1} and relative trends (in parentheses) are in % yr^{-1}. The two first columns present the results obtained with GOME and SCIAMACHY data only, while the two last columns present the results obtained including the GOME-2 data between 2007 and 2010.

	GOME, SCIAMACHY		GOME, SCIAMACHY, and GOME-2	
Region	**Annual Trend**	**Oct.-Mar. Trend**	**Annual Trend**	**Oct.-Mar. Trend**
Southeastern	0.9 ± 0.3 (1.4 ± 0.5)	1.0 ± 0.3 (1.9 ± 0.6)	0.7 ± 0.3 (1.2 ± 0.4)	1.0 ± 0.3 (1.8 ± 0.6)
Central	1.5 ± 0.3 (2.6 ± 0.6)	1.6 ± 0.3 (3.5 ± 0.8)	1.5 ± 0.3 (2.7 ± 0.6)	1.7 ± 0.3 (3.6 ± 0.7)
Northern	1.2 ± 0.3 (2.9 ± 0.9)	1.9 ± 0.4 (6.5 ± 1.4)	1.3 ± 0.3 (3.1 ± 0.9)	1.8 ± 0.4 (5.9 ± 1.4)
Southwestern	0.7 ± 0.2 (2.0 ± 0.7)	1.1 ± 0.3 (3.4 ± 0.8)	0.9 ± 0.2 (2.4 ± 0.7)	1.2 ± 0.3 (3.6 ± 0.8)
Northeastern	0.3 ± 0.3 (1.0 ± 1.2)	0.9 ± 0.4 (4.6 ± 2.0)	0.5 ± 0.3 (1.8 ± 1.3)	0.6 ± 0.4 (3.2 ± 2.0)
Northwestern	0.7 ± 0.2 (5.2 ± 2.0)	0.9 ± 0.3 (6.9 ± 2.1)	0.9 ± 0.2 (7.5 ± 2.4)	1.0 ± 0.3 (7.3 ± 2.2)

A trend of 5.1% yr^{-1} in the Chinese anthropogenic NMVOC emissions between 1997 and 2009 is reported in REAS (Regional Emission inventory in ASia), representing a total increase of 61% over this period [*Ohara et al.,* 2007]. The spatial distribution of these emission trends are illustrated in Figure 8-10. As illustrated in Figure 8-11, simulations performed with the IMAGESv2 model [*Stavrakou et al.,* 2009a] using the REAS NMVOC emission inventory for 1997 and 2009 show that the REAS emission trend leads to a total increase of 17-25% on the modeled H$_2$CO columns in Eastern China (25-34% in winter), which is much lower than the observed increase of 38±12% derived from the satellite H$_2$CO columns in the Northern region (Figure 8-9) between 1997 and 2009. This result is most likely pointing to an underestimation of the reported REAS NMVOC emission trend. A similar underestimation of the NO$_x$ emission in REAS (by a factor of 2-4) has also been reported by *Uno et al.* [2007], based on GOME NO$_2$ observations. A stronger increase in Chinese anthropogenic NMVOC emissions of 108% between 1997 and 2009 is reported in *Bo et al.* [2008]. It is mostly attributed to an increase of the emissions due to traffic (particularly motorcycles), fossil fuel combustion and industrial processes. Further, the heavy industrialization of the coastal areas tends to expand towards Central and Northwestern inland regions, due to the

transfer of emission-intensive plants from cities to rural areas with much less stringent pollution controls, coupled to the fact that fuel emission regulations are mainly implemented in the cities, which causes the transfer of the cheaper high emission vehicles to underdeveloped regions [*Bo et al., 2008; Cai et al., 2009*].
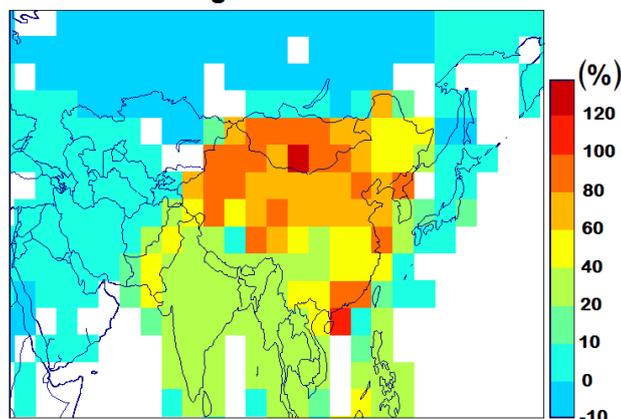
Figure 8-10: Increase in anthropogenic NMVOC emissions in the Regional Emission inventory in ASia inventory (REAS) between 1997 and 2009 (61% in China), at the IMAGES model resolution [*Ohara et al. 2007*].
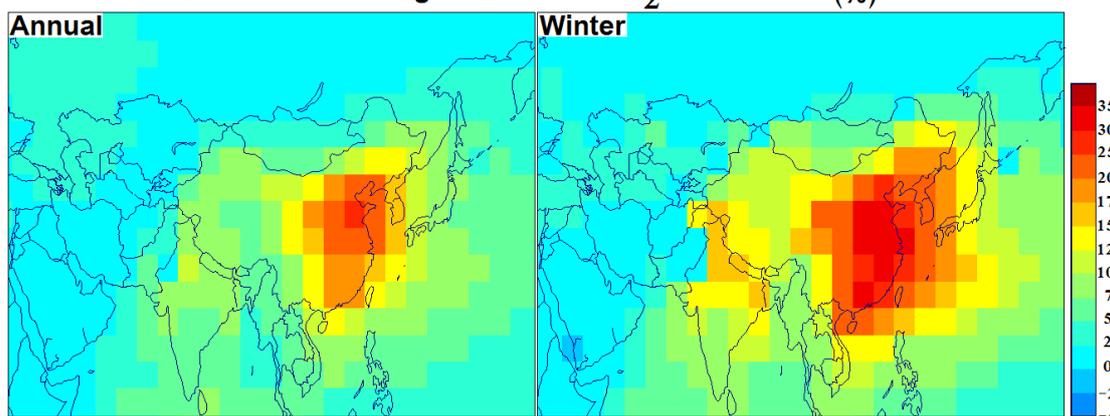
Figure 8-11: Increase in H₂CO columns calculated by the IMAGESv2 model, using the REAS NMVOC emissions in 1997 and in 2009. The left panel shows the annual increase while the left panel shows the winter increase [© *T. Stavrakou*].

8.4.3 India

Over India, we find an annual trend of $1.3\pm0.2\times10^{14}$ molec.cm^2 yr^{-1}, which represents a growth rate of $1.7\pm0.4\%$ yr^{-1} (see also Figure 8-3). The country presents spatially homogeneous trends, with increasing H₂CO columns over almost all states, except for the Northwestern states where trend analysis is not conclusive due to the systematic lack of GOME observations (Figure 8-2). The highest annual trend is observed over Uttar-Pradesh and New Delhi ($2.1\pm0.6\%$ yr^{-1}), whereas slightly lower trends are found in the Central states of Madhya Pradesh, Andhra Pradesh and Maharashtra, including Bombay (between 1.5 and 1.9% yr^{-1}) and in the Northeastern states of Bihar, Orissa and West Bengal, including Calcutta (1.8, 1.6 and 1.3% yr^{-1}) (see Table 8-5).

Table 8-5: Statistically Significant Annual Trends for Indian States.

Province in India	1997 H$_2$CO VC $A \pm \sigma_A$ [10^{14} molec.cm^{-2}]	Linear Trend $B \pm \sigma_B$ [10^{14} molec.cm^{-2} yr^{-1}]	Growth Rate $R \pm \sigma_R$ [% yr^{-1}]	Statistical Significance $t_B = |B/\sigma_B|$
Madhya-Pradesh	69.0 ± 1.3	1.3 ± 0.3	1.9 ± 0.4	5.1
Bihar	74.6 ± 1.6	1.3 ± 0.3	1.8 ± 0.4	3.8
Orissa	71.6 ± 1.6	1.1 ± 0.3	1.6 ± 0.4	3.8
Maharashtra	64.3 ± 1.4	1.2 ± 0.3	1.8 ± 0.5	4.3
Andhra-Pradesh	62.6 ± 1.3	1.0 ± 0.3	1.5 ± 0.4	3.7
Tamil-Nadu	60.0 ± 1.1	0.6 ± 0.2	1.0 ± 0.4	2.6
Uttar-Pradesh	49.6 ± 1.4	1.1 ± 0.3	2.1 ± 0.6	3.9
West-Bengal	77.6 ± 1.8	1.0 ± 0.3	1.3 ± 0.4	3.2
Gujarat	50.9 ± 1.5	0.8 ± 0.3	1.6 ± 0.6	2.9
Karnataka	60.9 ± 1.4	0.7 ± 0.3	1.2 ± 0.5	2.7
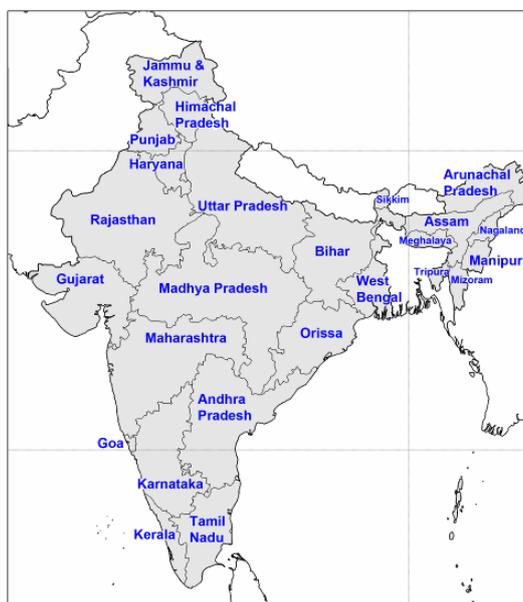
Figure 8-12: Maps of Indian states.

8.4.4 Principal Agglomerations of the World

We calculated trends in the H$_2$CO vertical columns over the largest agglomerations of the world, based on monthly averages of satellite observations within radius of 100, 200 and 500 km around city centers. The optimal radius is a trade off between the spatial resolution and the quality of the H$_2$CO time series. The results are summarized in Table 8-6 and represented in Figure 8-13 for the most important agglomerations.

In a radius of 100 km around Tokyo, we find a significant negative trend of -2.6% yr^{-1}, while the trend is zero for the whole Japan and slightly positive around Osaka and Nagoya. Large positive trends are found in Chinese and Indian megacities, as well as in the Asian megacities of Seoul, Jakarta, Cairo, Istanbul and Tehran. In the Eastern United States, we find significant negative trends of about -2.5% yr^{-1} over New York, Washington, Dallas, Philadelphia and Miami. More surprisingly, Los Angeles and San Francisco present positive annual growth rates of 2 and 3% yr^{-1}.

According to the US Environmental Protection Agency National Emission Inventory (EPA-NEI), the US anthropogenic NMVOC emissions present a negative trend of about -2% yr^{-1} for the whole country [http://www.epa.gov/ttnchie1/trends]. In Europe, while the European Monitoring and Evaluation Program centre on emission inventories (EMEP-CEI) reports negative trends in the anthropogenic NMVOC emissions of -3% yr^{-1} [http://www.ceip.at], the H₂CO column trends in the megacities of London and Paris are slightly negative but not statistically significant. This might be due to the poor quality of the satellite H₂CO columns in Europe resulting from the low sensitivity of the UV instruments in the troposphere at these latitudes, coupled to the rather weak local H₂CO concentrations compared to other regions at the same latitudes [*De Smedt et al.,* 2008].

H₂CO Annual Trend [% yr^{-1}]: 1997-2010
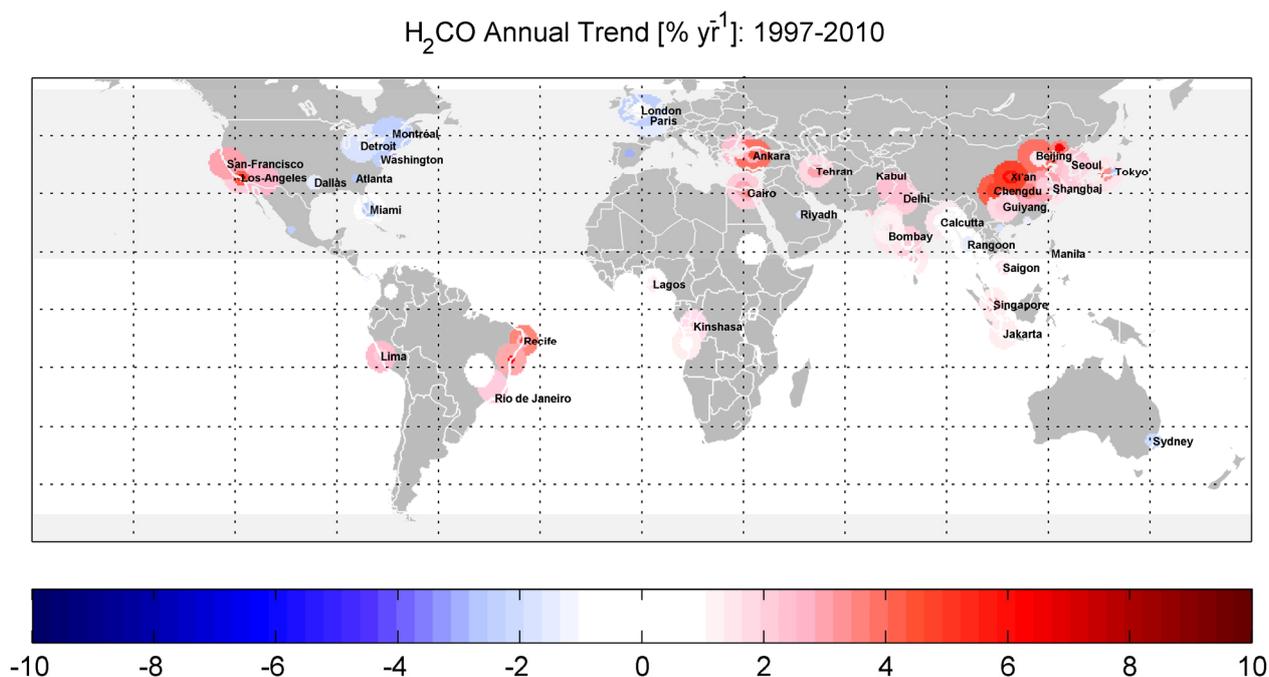
Figure 8-13: Significant annual growth rates of the H₂CO columns between 1997 and 2010 over the most important agglomerations worldwide. The different circle sizes correspond to radius of 100, 200 or 500 km.

Table 8-6: Inferred Trends in Selected Large Cities Worldwide

City (Country)	Annual Trend [10^{14} molec.cm⁻² yr^{-1}]	Growth Rate [% yr^{-1}]
Tokyo (Japan)	-1.2 ± 0.5	-2.6 ± 1.1
Manila (Philippines)	-0.3 ± 0.3	-0.7 ± 0.5
Beijing (China)	1.3 ± 0.3	4.0 ± 1.2
Shanghai (China)	1.0 ± 0.4	1.9 ± 0.9
Delhi (India)	1.5 ± 0.4	2.4 ± 0.7
Bombay (India)	0.8 ± 0.3	1.4 ± 0.5
Seoul (South Korea)	0.9 ± 0.3	2.5 ± 1.0
Jakarta (Indonesia)	0.9 ± 0.3	1.6 ± 0.5
Cairo (Egypt)	1.0 ± 0.3	2.6 ± 0.8
Istanbul (Turkey)	1.0 ± 0.3	2.4 ± 0.9
Tehran (Iran)	1.1 ± 0.3	3.0 ± 1.1
Lagos (Nigeria)	1.8 ± 0.4	1.6 ± 0.4

City (Country)	Annual Trend [10^{14} molec.cm^{-2} yr^{-1}]	Growth Rate [% yr^{-1}]
Kinshasa (Congo)	1.9 ± 0.5	1.5 ± 0.4
Los Angeles (USA)	0.7 ± 0.3	2.1 ± 0.9
San Francisco (USA)	0.7 ± 0.3	3.0 ± 1.4
New York (USA)	-1.4 ± 0.7	-3.2 ± 1.1
Washington (USA)	-1.5 ± 0.6	-2.5 ± 0.9
Dallas (USA)	-1.0 ± 0.4	-1.6 ± 0.6
Philadelphia (USA)	-1.2 ± 0.6	-2.3 ± 1.0
Miami (USA)	-1.3 ± 0.4	-2.4 ± 0.6
London (UK)	-0.8 ± 0.5	-2.2 ± 2.0
Paris (France)	-1.3 ± 0.7	-2.4 ± 2.0

8.5 Conclusions

A consolidated time series of tropospheric formaldehyde columns retrieved from GOME and SCIAMACHY observations between 1997 and 2010 has been used to determine trends over Asia and over large agglomerations of the world. Observations of GOME-2 between 2007 and 2010 were included in a separated analysis, leading to a consistent picture of the trends. Over China, the trend is larger in the Northeastern part and reaches 4% yr^{-1} in the Beijing area, whereas in the Southeastern region the trend is lower (1.4% yr^{-1}). India exhibits spatially homogeneous trends ranging between 1 and 2% yr^{-1}. The detected trends in H₂CO columns are explained by changes in the corresponding anthropogenic NMVOC emissions. Simulations performed with the IMAGESv2 model suggest that the reported trend of the REAS NMVOC emission inventory in Asia might be largely underestimated, particularly in Central-East china. Negative trends (-2.5% yr^{-1}) are found around Tokyo and several cities of Eastern U.S., illustrating the effective policies of emission reduction in these countries.

9   Constraining Non-Methane Volatile Organic Compound Emissions with Satellite Formaldehyde Observations

Abstract

In this chapter, we illustrate the utilisation of the satellite formaldehyde dataset built throughout this thesis for modelling applications. In particular, we focus on two key studies performed by the tropospheric chemistry modelling team of BIRA-IASB (J.-F. Müller and co-workers) and reported in *Stavrakou et al.* [2009a[1] and 2009b[2]].

These two studies have been conducted using an updated version of the global chemistry transport model IMAGES, which includes an optimized chemical scheme with respect to H$_2$CO production from biogenic and pyrogenic sources. The focus was on Tropical ecosystems and Southeast Asia, which experience strong biogenic and biomass burning NMVOC emissions, and on North America, where isoprene emissions are dominant in summertime. The biogenic NMVOC emissions were provided either by the GEIA inventory, or by a newly developed isoprene inventory (called MEGAN-ECMWF) based on the MEGAN model and driven by meteorological fields from the ECMWF. Biomass burning NMVOC emissions were provided by the GFED inventory, version 1 and 2.

The first objective was to test the ability of the IMAGES model to reproduce the observed features of the global H$_2$CO distribution retrieved from GOME and SCIAMACHY measurements, and to provide an assessment on whether the newly developed MEGAN-ECMWF and GFEDv2 inventories improve the agreement with the observations. The time series of monthly averaged satellite H$_2$CO columns were compared to three IMAGES simulations, obtained using different combinations of the *a priori* emission inventories [*Stavrakou et al.*, 2009a]. The purpose of the second study was to derive updated biogenic and biomass burning NMVOC flux estimates using the IMAGES model in its inverse mode and the SCIAMACHY observations as top-down constraints [*Stavrakou et al.*, 2009b].

The inversion of reactive species emissions at the global scale requires wide coverage satellite data of high quality, available over long time periods. The detailed error budget of our dataset allows for a precise selection of the best observations. The inversion results obtained by *Stavrakou et al.* [2009b] show the added value of the satellite H$_2$CO observations to improve NVMOC emissions inventories and the optimization brings the model closer to the measurements at all latitudes and all seasons. However, the lack of an aerosol correction over fire scenes in the satellite H$_2$CO retrieval is a potential source of error for the inferred biomass burning NMVOC fluxes. Moreover, random and systematic errors on the satellite columns should be further reduced in order to best take advantage of the measurements.

---

[1] Stavrakou, T., J.-F. Müller, I. De Smedt, M. Van Roozendael, *et al.* (2009a), Evaluating the performance of pyrogenic and biogenic emission inventories against one decade of space-based formaldehyde columns, *Atmos. Chem. Phys., 9*, 1037–1060.

[2] Stavrakou, T., J.-F. Müller, I. De Smedt, M. Van Roozendael, *et al.* (2009b), Global emissions of non-methane hydrocarbons deduced from SCIAMACHY formaldehyde columns through 2003–2006, *Atmos. Chem. Phys., 9*, 3663–3679.

9.1 Introduction

Current bottom-up models predict global isoprene emission of about 400±150 Tg yr^{-1} [*Ehhalt et al., IPCC*, 2001; *Guenther et al.*, 2006; *Müller et al.*, 2008], 70% being emitted in the Tropics, but the magnitude and distribution of these emissions are the subject of much speculation, particularly in tropical ecosystems because of the scarcity of field measurements [*Karl et al.*, 2007; *Cohen et al.*, 2008; *Barkley et al.*, 2008; *Müller et al.*, 2008]. The emission inventories of other NMVOCs from biogenic, biomass burning and anthropogenic sources are still more uncertain due to their large diversity and rapid spatiotemporal variability. The top-down or inverse modeling approach consists in adjusting the surface emissions used as input in a chemistry transport model (CTM), in such a way that the discrepancy between atmospheric observations and the model predictions becomes minimal. This is achieved by defining a set of emission parameters to be optimized and then minimizing a function of these parameters, which quantifies the overall discrepancy between the model and the data. The simplest inversion method to relate the observed H₂CO columns to the emitted NMVOCs is the use of empirical linear relationships between H₂CO and NMVOCs, derived from simulations performed with CTMs. This method is only valid in first approximation and when the local H₂CO production can be attributed to a single NMVOC oxidation (typically isoprene emissions over Northeastern US during summertime). In other cases, the use of the full CTM in its inverse mode is needed, in order to account for non-linearity (*e.g.* the feedback of NMVOC emission updates on OH concentrations), simultaneous production of H₂CO (from biogenic, biomass burning and anthropogenic sources), and for model and observations errors [*Stavrakou et al.*, 2009b].

Several studies have demonstrated the usefulness of H₂CO columns measurements from satellites as top-down constraints on NMVOC emissions. First studies based on H₂CO observations from the GOME instrument (from 1996 to 2001) focused on the isoprene emission estimates over the North American continent [*Palmer et al.*, 2003; 2006; *Abbot et al.*, 2003], in East and South Asia [*Fu et al.*, 2007], or on the global scale [*Shim et al.*, 2005]. More recently, North American isoprene emissions have been derived from H₂CO columns measured from the OMI instrument for the summer 2006 [*Millet et al.*, 2008]. These modelling studies have been carried out using the GEOS-CHEM chemistry transport model and the Harvard retrieval of satellite H₂CO columns [*Chance et al.*, 2000]. The IUP-Bremen SCIAMACHY and the Harvard OMI H₂CO products have also been used together with the regional CTM CHIMERE to constrain isoprene emissions over Europe [*Dufour et al.*, 2009b; *Curci et al.*, 2010]. In regions where isoprene is the dominant H₂CO precursor, the uncertainty on these emissions is mainly related to uncertainties in the isoprene chemical mechanism. In Tropical regions, the derivation of emissions is more difficult, to a large extent because of the strong contribution of biomass burning to the H₂CO column, difficult to separate from the biogenic contribution, and because the degradation mechanisms of many NMVOCs are still very uncertain [*Barkley et al., 2008; Müller et al.*, 2008].

Our dataset of satellite H₂CO observations has been used in several NMVOC studies. In *Barkley et al.* [2009] the most significant spatial and temporal H₂CO variations were determined on the global scale. In particular, unusually low H₂CO columns were consistently observed over the Amazon rainforest during the transition from the wet-to-dry season, suggesting a large-scale annual shutdown of Amazonian isoprene emissions prior to the dry season. In *Barkley et al.* [2011], simulations performed with a high resolution version of the GEOS-CHEM model over Tropical South America are compared with SCIAMACHY and OMI H₂CO observations, in order to evaluate the chemistry of the GEOS-CHEM model and the MEGAN isoprene emission inventory in Amazonia. In *Kim et al.* [2011], statistical analyses are performed using global GOME,

SCIAMACHY and OMI H_2CO observations, together with those of MOPITT CO, to study the seasonal variability of biogenic and pyrogenic emissions on the global scale and over long measurement periods. Here we focus on the main results obtained by *Stavrakou et al.* [2009a and b] using our GOME and SCIAMACHY H_2CO observations and an updated version of the IMAGES model. In particular, the study by *Stavrakou et al.* [2009b] is the first to report estimates of pyrogenic and biogenic NMVOC emission strengths on the global scale constrained by spatial observations of H_2CO columns. A brief description of the IMAGESv2 model which contains an optimized chemical scheme with respect to H_2CO production and uses state-of-the-art *a priori* emission inventories is given in 9.2, while the results of the two studies are summarized respectively in sections 9.3 and 9.4.

9.2 The IMAGES global chemistry transport model

9.2.1 General description

A comprehensive description of the IMAGES global CTM is provided in *Müller and Stavrakou* [2005] and in *Stavrakou et al.* [2009a and b]. The model provides the global distribution of about 20 short-lived and 48 long-lived chemical compounds between the Earth surface and the pressure level of 44 hPa, through a chemical mechanism comprising 164 gas-phase reactions, 39 photodissociations and 3 heterogeneous reactions on the surface of sulphate aerosols. The model is run at a horizontal resolution of 5°x5° degrees (or alternatively 2°x2.5°), with 40 pressure levels in the vertical. Monthly mean ECMWF/ERA40 wind fields drive the advection, whereas daily ECMWF fields are used for surface pressure, temperature and humidity. The model time step is equal to three hours and the model calculates daily averaged concentrations. It accounts for the impact of diurnal variations of the chemical compounds through correction factors computed off-line via a diurnal cycle simulation with a 20-min time step. Diurnal profiles are also used to estimate the H_2CO concentration at the satellite overpass time.

9.2.2 Optimization of the chemical scheme with respect to H₂CO production

In view of the large number and diversity of NMVOCs involved in the production of formaldehyde, an extensive preliminary study has been conducted using the Master Chemical Mechanism V3.1 (MCM) [*Saunders et al.*, 2003] in order to determine the H_2CO yields in the oxidation of the most prominent NMVOCs. These results have been used to optimize the chemical mechanisms in IMAGES with respect to H_2CO production from isoprene and pyrogenic NMVOCs [*Stavrakou et al.*, 2009a]. The current version of the IMAGESv2 model comprises emissions from 18 explicit NMVOCs. H_2CO yields from the oxidation of isoprene and different pyrogenic NMVOCs are given in Table 9-1. Short-term and final H_2CO yields are calculated after one day or two months. The final yield determines the global budget of formaldehyde, while the short-term yield provides a better indication of the H_2CO production that may be detected by satellite instruments. The emissions of the most reactive H_2CO precursors can be optimized using the satellite H_2CO data, whereas the emissions of slowly reacting species (*e.g.* methanol) should be optimized using direct observations of these compounds.

Biogenic isoprene is the most emitted NMVOC and presents also the highest H_2CO yield, ranging between 2 and 2.6 depending on the NO_x level, resulting in a final H_2CO production of about 455 Tg yr^{-1}. Note that the yield of H_2CO from isoprene in IMAGESv2 and MCM are about 20% higher than the yield calculated with the GEOS-CHEM mechanism [*Evans et al.*, 2003] under high NO_x

conditions [*Palmer et al.*, 2006]. The yield dependence on the abundance of NO_x is still more uncertain.

The total H_2CO production from biomass burning is estimated at 66 Tg yr^{-1}, half of which is produced after one day. From Table 9-1, it can be seen that only three NMVOCs, namely acetic acid, methanol and ethene, account for one third of the total short-term H_2CO production and 43% of the final production. Among them, the oxidation of ethene contributes the most, 27% in the short-term and 18% in the long-term H_2CO production, whereas the slowly reacting methanol and acetic acid, contribute mainly to the final production term. Glyoxal, although abundantly emitted during fire events, leads to a negligible H_2CO formation. The compounds explicitly included in the IMAGESv2 chemical scheme represent 87% of the total short- and long-term H_2CO production from biomass burning NMVOC emissions. The remainder is accounted for by the surrogate species C_4H_{10} [*Müller and Stavrakou*, 2005].

Anthropogenic NMVOCs emissions are much more difficult to model with a limited number of NMVOCs. There are hundreds of reported anthropogenic VOCs with large uncertainties in the speciation. Furthermore, the degradation mechanisms for most of these compounds are unavailable which makes the H_2CO production difficult to estimate. For these reasons, anthropogenic emissions are not optimized in the studies presented here [*Stavrakou et al.*, 2009a and b].

Table 9-1: Production of H₂CO from biomass burning NMVOCs and from biogenic isoprene. The emissions are 10-year averages based on GFEDv2 and the MEGAN-ECMWF inventories. [a] This total includes 10 Tg yr^{-1} of compounds with negligible H₂CO yields, among which formic acid and acetylene [*Stavrakou et al.*, 2009a].

Pyrogenic NMVOC	*Molar Weight*	*Global Emission Tg yr^{-1}*	*H₂CO Yields (mol/mol)*		*H₂CO Production Tg yr^{-1}*	
			Short-term	**Final**	**Short-term**	**Final**
Acetic acid (CH_3COOH)	60	17.1	0.12	1	1.03	8.55
Methanol (CH_3OH)	32	08.8	0.14	1	1.16	8.25
Ethene (C_2H_4)	28	06.3	1.38	1.76	9.32	11.88
Glyoxal (CHOCHO)	58	06.0	0.06	0.06	0.19	0.19
Formaldehyde (H_2CO)	30	05.0	1	1	5.00	5.00
Glycolaldehyde ($C_2H_4O_2$)	60	04.8	0.80	0.87	1.92	2.09
Methylglyoxal ($C_3H_4O_2$)	72	03.9	1	1	1.63	1.63
Biacetyl ($C_4H_6O_2$)	86	03.7	2	2	2.58	2.58
Acetaldehyde (CH_3CHO)	44	03.6	0.94	1	2.31	2.45
Propene (C_3H_6)	42	03.5	1.78	1.92	4.45	4.80
Ethane (C_2H_6)	30	03.2	0.03	0.99	0.10	3.17
Acetone (($CH_3)_2CO$)	58	03.0	0.11	2	0.17	3.10
Propane (C_3H_8)	44	01.9	0.03	1.65	0.04	2.14
2-butanone (Mek) (C_4H_8O)	72	01.7	0.53	2.15	0.38	1.52
Benzene (C_6H_6)	78	01.7	0.05	0.31	0.03	0.20
Toluene ($C_6H_5CH_3$)	92	01.0	0.62	1.34	0.20	0.44
Others (C_4H_{10})	58	20.8[a]			5.04	7.60
Biogenic Isoprene	*MW*	*Emission Tg yr^{-1}*	*H₂CO Yields (mol/mol)*		*H₂CO Production Tg yr^{-1}*	
			Short-term	**Final**	**Short-term**	**Final**
Isoprene (C_5H_8) - high NO_x	68	410	2.26	2.58	409	467
low NO_x			2.05	2.46	371	445

9.2.3 A priori emissions

Two different biogenic emissions of isoprene are used to drive the model: the Global Emissions Inventory Activity (GEIA) database [*Guenther et al.*, 1995] and the newly developed MEGAN-ECMWF bottom-up inventory [*Müller et al.*, 2008]. The latter is based on the Model of Emissions of Gases and Aerosols from Nature (MEGAN) model [*Guenther et al.*, 2006] and coupled with a

detailed canopy environment model including the treatment of radiative transfer in the canopy and the calculation of leaf temperature [*Wallens*, 2004]. Operational ECMWF analyses are used to drive the model with the solar flux, the cloud cover, the temperature, and the windspeed. The averaged annual global isoprene emission between 1996 and 2005 retrieved from this new inventory is about 410 Tg yr^{-1}, *i.e.* about 30% less than the standard MEGAN and GEIA estimates [*Guenther et al.*, 1995 and 2006]. In *Müller et al.* [2008], the MEGAN-ECMWF isoprene emission inventory has been evaluated against flux measurement campaigns in Eastern US and in Amazonia, and against our satellite H₂CO columns, using the IMAGES model.

The distribution of vegetation fires is provided by the Global Fire Emissions Database (GFED) version 1 or version 2 [*van der Werf et al.*, 2004 and 2006]. In both inventories, carbon fire emissions are estimated by using satellite fire activity data converted into burned area. Trace gas emissions are derived from the carbon emissions using emission factors from *Andreae and Merlet* [2001] with updates from *Andreae* [personal communication, 2007]. A diurnal profile of biomass burning emissions, peaking in the afternoon, is applied in the CTM calculations [*Giglio*, 2007].

Anthropogenic CO and total NMVOC emissions are taken from the Emission Database for Global Atmospheric Research (EDGAR) v3.3 inventory for 1997 [*Olivier et al.*, 2001]. Speciation of NMVOC emissions is obtained from the POET database [*Olivier et al.*, 2003], except for biofuel emissions where the updated speciation by *Andreae* is used [*Andreae and Merlet*, 2001 and 2007].

9.2.4 Grid-based inversion analysis

The NMVOC emission estimates that best reproduce the H₂CO observations within their uncertainties, and account for the errors on the *a priori* emissions, are quantified through inverse modelling performed with the IMAGESv2 model on the global scale. The inversion is based on the grid-based approach, previously applied to the optimization of CO fluxes in the global 3-D CTM IMAGES [*Stavrakou and Muller*, 2006], and thoroughly presented *Müller and Stavrakou* [2005]. This approach allows for the optimization of the emission strengths at the spatial resolution of the model, providing a differentiation among the emission sources. Here we explain in short the main steps of the inversion.

Let $\phi_j^a(x,t)$ be the *a priori* emission distribution, where j denote the different emission categories (here biogenic or biomass burning), x the space variables (latitude, longitude and altitude), and t the time (month). The *a priori* $\phi_j^a(x,t)$ and optimized $\phi^{opt}(x,t,f_j)$ emission distributions are linked through the expression:

$$\phi^{opt}(x,t,f_j) = \sum_{j=1}^{m} e^{f_j} \phi_j^a(x,t) \qquad\qquad\qquad (9\text{-}1)$$

Where $f = (f_j)$ is the vector of control parameters to be determined in order to minimize the cost function J, which quantifies the bias between the modelled columns $H(f)$ and the observed columns y:

$$J(f) = \frac{1}{2}\left[(H(f) - y)^T E^{-1}(H(f) - y) + f^T B^{-1} f\right] \qquad\qquad (9\text{-}2)$$

In this equation, E and B are respectively the error matrices of the observations and the emission parameters. By using J and its derivatives with respect to each component of f in an iterative minimization algorithm, an updated estimate for f is obtained until J has reached its minimum. The observation error matrix E, comprises the error on the vertical columns (estimated in chapter 6) and

a model error of 20%, based on sensitivity calculations [*Stavrakou et al.*, 2009a]. The emission fluxes are estimated at the model resolution for each emitting process and each month, leading to a huge number of unknowns to be determined for each source category (larger than 120000). To reduce the number of unknowns, grid cells with very low *a priori* emissions are excluded from the analysis. Moreover, extra constraints are brought in the inversion in the form of correlations among the *a priori* errors on the emission parameters. These correlations are implemented through the off-diagonal elements of the matrix **B** [*Stavrakou and Müller*, 2006]. The spatiotemporal coexistence of the different H₂CO sources is a recognized inherent difficulty in the derivation of the underlying emissions of H₂CO precursors. This challenging issue is tackled in IMAGESv2 by using a dedicated spatiotemporal correlation setup, based on physical assumptions on how the emission from a source type for a given month and location is expected to correlate with the emissions estimated at neighbouring months and locations. Through this framework, the inversion is expected to find a compromising state in between the emission pattern borrowed from the *a priori* inventories and the H₂CO observations within their assigned and measured uncertainties, while the seasonality and geographical distribution of the emissions are also optimized [*Stavrakou et al.*, 2009b].

This inversion technique has two important advantages: it is able to address any non-linear problem, and it can handle any large number of control parameters. Its main limitations are that it requires important programming efforts and large computing resources, and that it does not provide an exact evaluation of the errors on the optimized emissions [*Stavrakou and Müller*, 2006].

9.2.5 Satellite data selection

The simulated columns are evaluated against monthly averaged H₂CO observations, binned at the model horizontal grid (5°x5°) between 1997 and 2006. The monthly mean H₂CO columns used in these studies are calculated using observations with a cloud fraction below 0.4. The spatial resolution of the model being rather coarse, the total errors on the monthly averaged observations are mostly due to systematic errors, which are on average comprised between 20 and 40% for regions with a significant H₂CO signal (see chapter 6). Satellite data are excluded when the error on a model grid cell exceeds 70%, which occur principally at mid and high latitudes and for H₂CO columns close to the detection limit. The data in the region influenced by the South Atlantic Anomaly are also filtered out. Furthermore, oceanic data are excluded from the analysis as H₂CO production from the methane oxidation is not optimized.

9.3 *Evaluating the performance of pyrogenic and biogenic emission inventories against one decade of space-based formaldehyde columns [Stavrakou et al., 2009a]*

In this study, the dataset of formaldehyde columns retrieved from GOME and SCIAMACHY between 1997 and 2006 is compared with H₂CO columns simulated by the IMAGESv2 global CTM. Table 9-2 summarizes the three different model simulations that have been conducted, using different combinations of the biogenic and pyrogenic emission inventories.

Table 9-2: Simulations conducted with the IMAGES global CTM [*Stavrakou et al.*, 2009a]

	Biomass burning emission inventory	Biogenic emission inventory
S1	GFEDv1 [*van der Werf et al.*, 2004]	GEIA [*Guenther et al.*, 1995]
S2	GFEDv1 [*van der Werf et al.*, 2004]	MEGAN-ECMWF [*Müller et al.*, 2008]
S3	GFEDv2 [*van der Werf et al.*, 2006]	MEGAN-ECMWF [*Müller et al.*, 2008]

Figure 9-1, Figure 9-2, and Figure 9-3 show comparisons between monthly averaged satellite observations and the results of the three IMAGES simulations (S1, S2 and S3) over the Eastern US and a few Tropical regions. The aim is to test the ability of the model to reproduce the observed features of the H_2CO distribution and to provide a first assessment about whether the improved methodologies and input data implemented in GFEDv2 and MEGAN-ECMWF lead to better results in the comparisons of modelled with observed H_2CO column measurements.
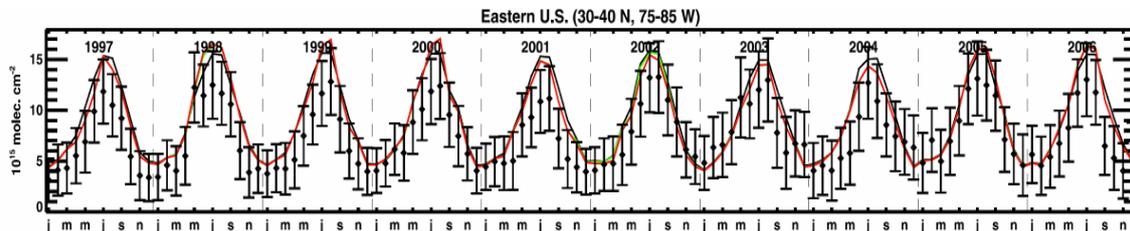
Figure 9-1: Monthly H₂CO columns in Eastern US. Observations are shown in black diamonds with their error bars. *A priori* modelled columns are shown in solid lines: S1 (GFEDv1/GEIA) in black, S2 (GFEDv1/MEGAN-ECMWF) in green, S3 (GFEDv2/MEGAN-ECMWF) in red.
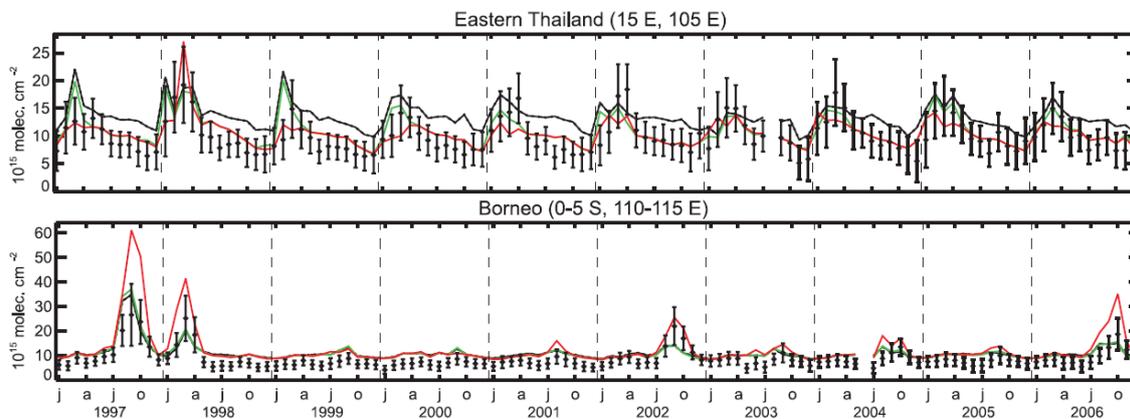
Figure 9-2: Monthly H₂CO columns in Southeastern Asia.
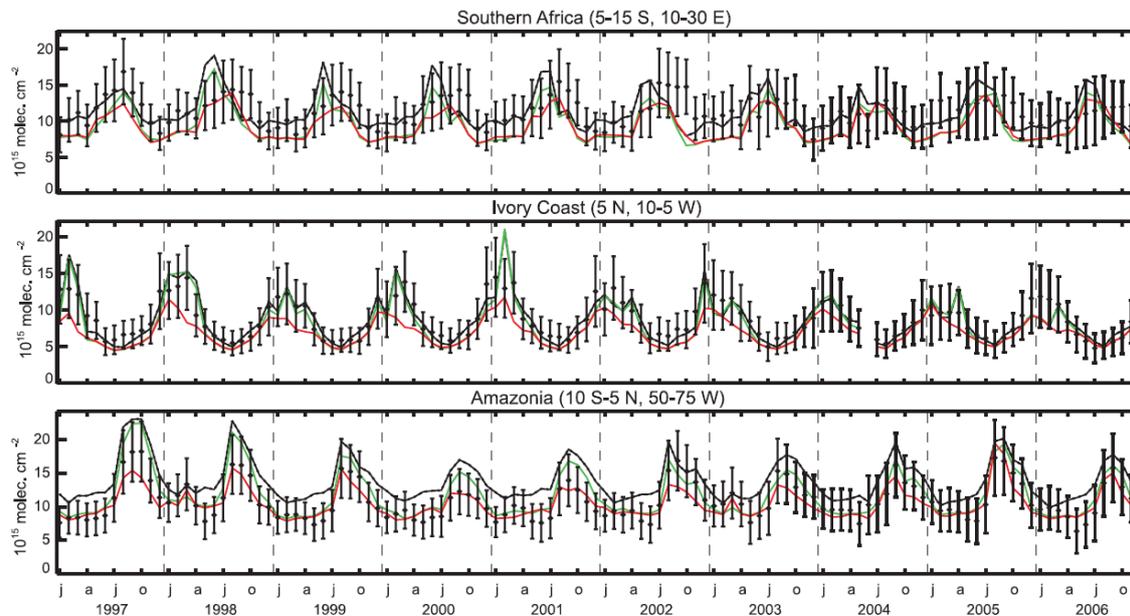
Figure 9-3: Monthly H₂CO columns in Africa and Amazonia.

Mean biases between the *a priori* modelled H$_2$CO columns and the observations as well as the spatiotemporal correlation coefficients are provided separately for the non-burning and burning seasons in Table 9-3 (for the two biogenic emission inventories GEIA and MEGAN-ECMWF) and Table 9-4 (for the two pyrogenic emission inventories GFEDv1 and v2). As can be seen, the high correlation coefficients between the observed and simulated columns over most regions indicate a good consistency between the model, the implemented inventories and the H$_2$CO dataset. Over Eastern US (Figure 9-1), both the GEIA and MEGAN-ECMWF inventories are in good consistency and capture the seasonal variations very well, even though the model is biased high by 16–19% on average in summertime, the overestimation being more significant for Southeastern US (about 35%). The use of the MEGAN-ECMWF inventory improves the model/data agreement in almost all regions, suggesting that this new isoprene inventory provides a better representation of the temporal variability in the emissions, but biases persist over parts of Africa and Australia. The situation is more complex regarding the biomass burning inventories as none GFED version appears to be consistent with the observation over all regions. Yet, the differences between the two inventories can be evaluated based on the model/data comparisons. A better agreement is achieved over Indonesia (*e.g.* Borneo in Figure 9-2) and Southern Africa when GFEDv2 is used, but GFEDv1 succeeds better in getting the correct seasonal patterns and intensities of the fire episodes over the Amazon basin, as reflected in the significantly higher correlations calculated in this region (see also Figure 9-3).

Table 9-3: Non-burning season mean biases (model-observations) and correlation coefficients calculated for the model simulations S1 and S2 over 1997–2006.

Regions	GEIA (S1)		MEGAN-ECMWF (S2)	
	Bias (%)	Correl	Bias (%)	Correl
Indochina	34.6	0.66	16.2	0.77
Indonesia	38.4	0.54	33.2	0.58
Northern Africa	13.1	0.64	00.6	0.76
Southern Africa	-00.6	0.44	-21.3	0.58
Amazonia	31.3	0.79	05.7	0.83
North America	19.1	0.77	15.9	0.80

Table 9-4: Burning season mean biases (model-observations) and correlation coefficients calculated for the S2 and S3 simulations over 1997–2006.

Regions	GFEDv1 (S2)		GFEDv2 (S3)	
	Bias (%)	Correl	Bias (%)	Correl
Indochina	19.1	0.78	02.8	0.71
Indonesia	36.3	0.71	59.0	0.82
Northern Africa	14.7	0.58	16.3	0.56
Southern Africa	-06.1	0.67	-10.1	0.78
Amazonia	20.7	0.85	-06.5	0.79

9.4 *Global emissions of non-methane hydrocarbons deduced from SCIAMACHY formaldehyde columns through 2003–2006 [Stavrakou et al., 2009b[2]]*

In this second study, formaldehyde columns retrieved from SCIAMACHY through 2003 to 2006 have been used as top-down constraints to derive updated global biogenic and biomass burning flux estimates for the NMVOC precursors of H$_2$CO. The study focuses on regions experiencing strong emissions, and hence exhibiting a high signal to noise ratio and lower measurement uncertainties. The inversion is implemented in the grid-based framework described in section 9.2.4, within which emission fluxes are derived at the model resolution, together with a differentiation of the sources in a grid cell. Two inversion studies are conducted using either the GFED v1 or v2 as *a priori* for the pyrogenic fluxes, while the *a priori* biogenic emissions are taken from the MEGAN-ECMWF isoprene emission database (simulations S2 and S3 in Table 9-2). Seasonal averages of H$_2$CO satellite data in 2006 are displayed in the left column of Figure 9-4, the middle column shows the *a priori* modelled H$_2$CO columns calculated using the pyrogenic GFEDv2 and the isoprene MEGAN-ECMWF emission inventories (S3), while the right column presents the updated H$_2$CO columns

after optimization. This figure serves as an illustration of the inversion results that were obtained between 2003 and 2006 and using the two biomass burning inventories as input. As already shown in the first study, the *a priori* columns capture quite well the patterns and seasonality observed from space, as well as the observed column enhancements in most of the regions of interest (continental Tropics, US and Asia), although noticeable exceptions are detected (*e.g.* over Southern Africa in Sept.-Oct.-Nov., or over Indonesia and Australia throughout the year). The optimization brings the model closer to the measurements at all latitudes and all seasons compared to the *a priori* simulated columns. This is reflected in a cost function reduction by about 25% in both inversions (S2 and S3).
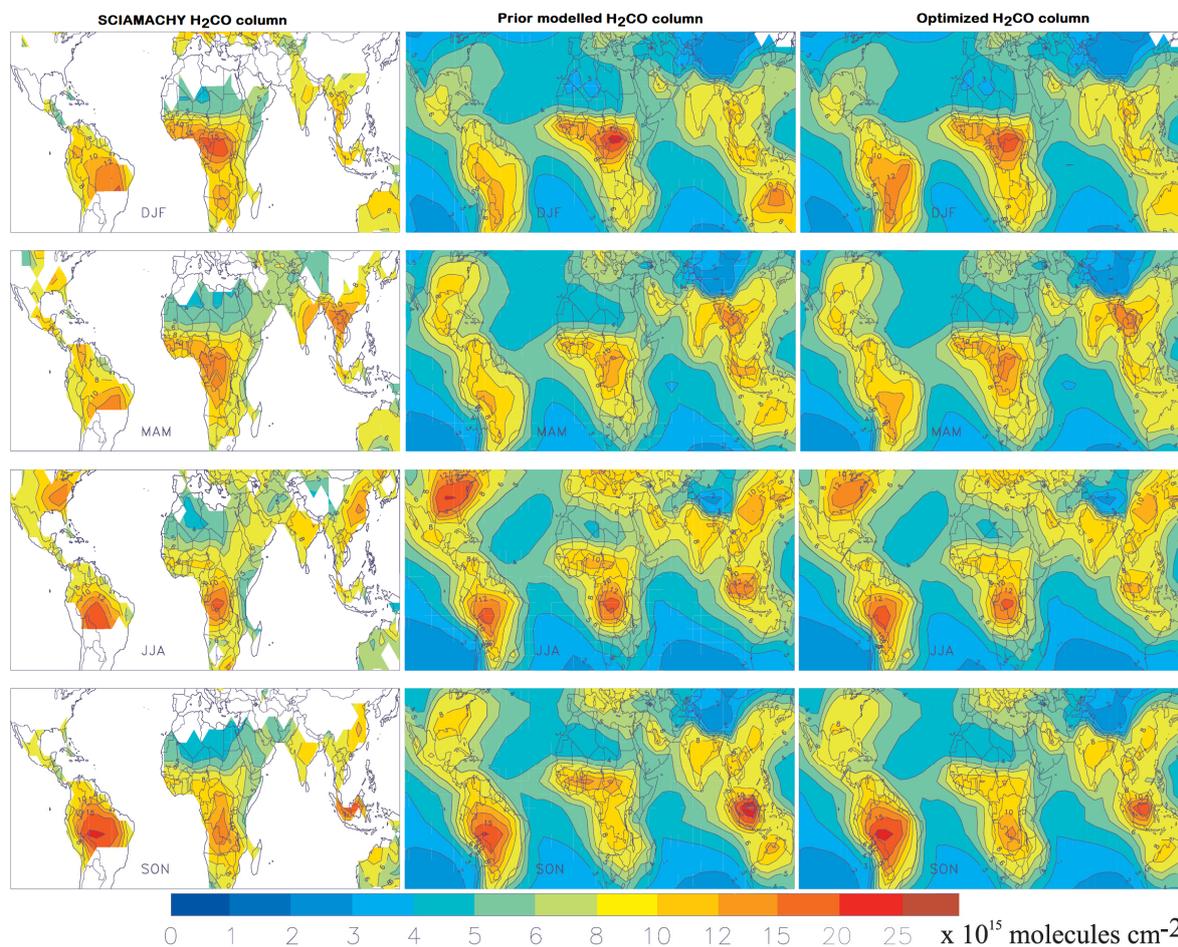
Figure 9-4: Left panels: Averaged SCIAMACHY H₂CO columns for winter (DJF), spring (MAM), summer (JJA), and fall (SON) 2006. Center panels: H₂CO columns simulated using IMAGESv2 driven by the GFEDv2 and MEGAN-ECMWF inventories. Right panels: IMAGESv2 columns derived from the inversion study [*Stavrakou et al.*, 2009b].

The modelled H₂CO column updates imply changes in the underlying biogenic and pyrogenic NMVOC emissions, as illustrated by the ratio of the optimized to the *a priori* annual emissions for 2006 shown in Figure 9-5, respectively for biogenic emissions (left) and pyrogenic emissions (right) in 2006. The strongest isoprene source change is encountered in Southern Africa, where the inferred average increase amounts to 55%. The optimized isoprene emission estimated at 65 Tg yr^{-1} agree quite well with literature isoprene flux estimates in this region [*Otter et al.*, 2003]. Over Indonesia the biogenic emissions appear to be overestimated by 20–30%, whereas over the Indochina peninsula the *a priori* inventory captures both the seasonality and the magnitude of the observed columns (see also Figure 9-2). The *a priori* simulations reproduce very well the observations over

the Amazon basin, and thus no significant emission update is inferred over this region, in contrast to previous work suggesting a large increase of the isoprene source. Over North America, the posterior isoprene emissions, amounting to about 38.5 Tg yr^{-1}, are estimated to be on average by 25% lower than the *a priori* over 2003–2006, the reduction being more important in the Eastern US. This suggested decrease, not reported in previous inversion studies based on GOME observations, is due to the differences in H₂CO yields from isoprene oxidation between IMAGESv2 and GEOS-CHEM, coupled to the fact that our H₂CO columns are lower over this region compared to the Harvard GOME dataset. However, the OMI satellite measurements imply also a decrease of up to 23% in the MEGAN estimate, corroborating the derived fluxes.
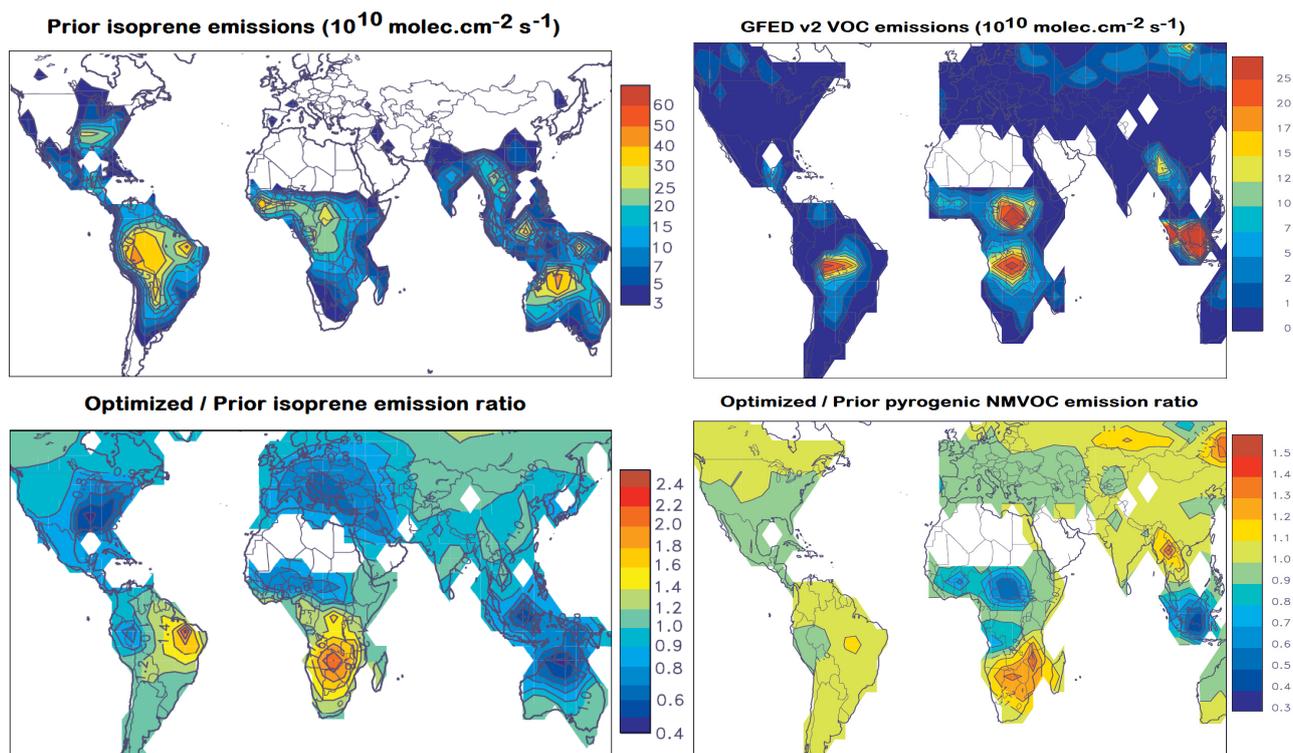
Figure 9-5: Left: *a priori* **isoprene emissions taken from the MEGAN-ECMWF inventory (top) and inferred ratios of the posterior to the prior biogenic emissions from the optimization (bottom). Right:** *a priori* **pyrogenic GFEDv2 NMVOC emissions (top) and ratios of the posterior to the prior pyrogenic NMVOC emissions deduced from the inversion (bottom). Maps are shown for 2006 but results are given for the 2003 -2006 period in** *Stavrakou et al.* **[2009b].**

Despite their *a priori* deviations, pyrogenic estimates inferred from the two inversions (using either GFED v1 or v2) converge on reasonably similar values. In Africa North of the equator, the inversions clearly suggest a decrease in the biomass burning source, by 16% and 35% compared to GFEDv1 and GFEDv2 respectively. Nevertheless, the lack of aerosol correction over fire scenes in the satellite H₂CO retrieval might introduce a bias in the H₂CO columns during important fire events. Note, however, that an independent study using the same emission database pointed also to a large model overestimation when compared to observed NO₂ columns over this region [*Stavrakou et al.*, 2008]. Whereas both GFEDv1 and v2 are consistent over Indochina and result in a very good agreement with the observed H₂CO columns, a serious overestimation of the *a priori* columns is found over Indonesia when GFEDv2 is used (see Figure 9-2). Over the Central-Eastern Amazon basin, the use of GFEDv2 leads to strongly underestimated columns. The inversion decreases the model/data biases over this region by strongly increasing the biogenic source. This result highlights the fact that, due to the overlap of the different emission source types, the quantification of the

NMVOC emissions using inverse modelling of H$_2$CO columns, should be conducted with particular care in order to avoid misinterpretations of the results.

A number of sensitivity experiments have been conducted in order to assess the impact of uncertainties related to the inversion setup and the chemical mechanism. Whereas changes in the background error covariance matrix have only a limited impact on the posterior fluxes, showing the robustness of the inversion, the use of an alternative isoprene mechanism characterized by lower H$_2$CO yields (the GEOS-CHEM mechanism) increases the posterior isoprene source estimate by 11% over Northern America and by up to 40% in Tropical regions. Such discrepancies reflect important differences between the GEOS-CHEM and MCM mechanisms, in particular regarding the dependence of the H$_2$CO yield from isoprene on the NO$_x$ abundances.

9.5 Conclusions

The H$_2$CO columns derived from GOME and SCIAMACHY during this thesis have been used in several studies to improve our knowledge of NMVOC emissions [*Müller et al.*, 2008; *Stavrakou et al.*, 2009a and 2009b; *Barkley et al.* 2009 and 2011, *Kim et al.*, 2011]. In particular, they have served as an input for extensive comparisons performed with a recently updated version of the IMAGES model, that includes optimized chemical mechanisms for isoprene and the most prominent pyrogenic NMVOCs. Particular emphasis has been placed on the continental Tropics, where H$_2$CO abundances are generally dominated by biogenic and pyrogenic NMVOC emissions, and on North America, experiencing strong biogenic emissions during the growing season. Comparisons between observed and simulated monthly averaged H$_2$CO columns between 1997 and 2006 have allowed assessing the performance of two biomass burning inventories (GFEDv1 and GFEDv2) and two biogenic inventories (GEIA and MEGAN-ECMWF). The satellite H$_2$CO observations show that the new MEGAN-ECMWF inventory describes better isoprene emissions on the global scale, particularly over the Amazon forest, and point out which new features in GFEDv2 improve the model/data comparisons, depending on the region.

A second key study made used of the SCIAMACHY observations between 2003 and 2006 as input for inverse modelling of biogenic and pyrogenic NMVOC emissions, at the model spatial resolution and on a monthly basis, using the IMAGESv2 model in its inverse mode. The model/data biases are found to be reduced in all regions and seasons, showing the success of the inversion approach and the information brought by the satellite observations. The largest isoprene increase is found in South Africa, while significant reductions are deduced in Northeastern US, Indonesia and Australia. Satellite observations suggest an increase of the GFEDv2 pyrogenic NMVOC emissions in South Africa and the Indochina peninsula, and significant flux decreases in North Africa and Indonesia. However, the omission of aerosol correction in the satellite observations is a recognized important source of error for the inferred biomass burning NMVOC fluxes. The effect of the inclusion of aerosols in the H$_2$CO retrievals is under investigation and will be quantified in the future.

Although these two studies put emphasis on tropical regions, where the detected H$_2$CO signal is significant, anthropogenic NMVOC emissions also produce a detectable signal in industrialized areas. In order to quantify this contribution based on satellite retrievals, the chemical mechanism of the model has to be extended in order to include all anthropogenic precursors of formaldehyde. In parallel, the signal to noise ratio of the satellite H$_2$CO observations should be improved, particularly over mid-latitude regions. In this regard, the GOME-2 H$_2$CO columns since 2007 are providing additional insight into the observations of tropospheric formaldehyde from space, thanks to a much better Earth sampling.

10 Conclusions and perspectives

During this work, global distributions of formaldehyde columns have been derived from the spaceborne nadir spectrometers GOME, SCIAMACHY and GOME-2, over the period 1996 to 2010, using the differential optical absorption spectroscopy (DOAS) technique. The ultimate goal was to provide a consistent long-term multi-sensor dataset of H_2CO observations and thereby to contribute to a better knowledge of NMVOCs emissions and to support scientific studies related to air quality and climate change issues.

Formaldehyde DOAS retrieval

Retrieval settings have been carefully optimized for H_2CO and consistently applied to all instruments. The key difference of our retrieval compared to previous studies [*e.g. Chance et al.*, 2000; *Wittrock et al.*, 2006] is the selected fitting window (328.5-346 nm), which has been shifted to shorter wavelengths. This allowed minimizing two major sources of bias in the H_2CO retrieval, namely (1) the spectral interference with the strong O_4 absorption band peaking at 360 nm, (2) the imperfectly corrected polarisation anomaly affecting SCIAMACHY around 350 nm. The improved quality of our H_2CO retrieval is demonstrated by a reduction of the fitting errors in the Tropics, a global decrease of the noise on the columns, and geophysically more consistent H_2CO abundances over arid and semi arid regions as well as over Oceans. For GOME-2 measurements, a further improved version of the algorithm has been developed that includes a two-step fitting procedure to minimize interferences between H_2CO and BrO spectral structures, and a modified DOAS approach to better handle strong O_3 absorption effects. For the three satellite instruments, a common set of high–resolution cross-sections, degraded to the resolution of each sensor, is used. The air mass factors are also consistently determined from radiative transfer calculations taking into account the geometry of the observation, the cloud fraction, the effective cloud top height, the surface albedo and the surface altitude. *A priori* vertical H_2CO profile shapes are taken from the IMAGES model.

The resulting time series of GOME (1996-2002) and SCIAMACHY (2003-2010) monthly averaged H_2CO columns are found to be very consistent over the main region of NMVOC emissions. Likewise, SCIAMACHY and GOME-2 observations between 2007 and 2010 are in excellent agreement, displaying correlation coefficients higher than 0.8 and mean column differences lower than 7% almost everywhere.

Our H_2CO retrieval algorithms have been developed in the framework of the TEMIS and PROMOTE projects and the whole dataset is available on the corresponding websites [http://www.temis.nl and http://www.gse-promote.org]. Moreover, our retrieval settings have been adapted and implemented in the operational processors of SCIAMACHY and GOME-2, operated by the German Aerospace Centre DLR, respectively on behalf of ESA and EUMETSAT.

Error and information content

A substantial effort has been devoted to characterize the retrieval and to provide a detailed error analysis of the H_2CO columns, together with their vertical averaging kernels. For individual measurements, the random error on the slant column is the largest source of uncertainty. The differences between the random errors obtained from GOME, SCIAMACHY or GOME-2 retrievals were found to be consistent with the different ground pixel sizes of each instrument. However, the higher sampling rate of GOME-2 in comparison to GOME and SCIAMACHY leads to a significant reduction of the scatter on the monthly averaged H_2CO observations, allowing for a better

resolution of the H$_2$CO spatial distribution. By monitoring the long-term evolution of the measurement errors, the effects of the degradation of the GOME, SCIAMACHY and GOME-2 instruments could also be assessed.

Systematic errors are found to be due to uncertainties on the cross-section datasets, in the *a priori* H$_2$CO vertical distribution, the cloud altitude and the surface albedo, with relative contributions depending on the observation conditions. This comprehensive error budget is provided for every satellite pixel, allowing the selection of the best observations. Consequently, pixels with cloud fractions larger than 40% are excluded for quantitative analysis because they are characterized by large air mass factor errors and low information content below the cloud. Observations made at solar zenith angles larger than 60° have been excluded because the slant column retrievals present large systematic errors due to strong ozone absorption and because of the low vertical sensitivity of these measurements in the boundary layer. In this regard, it should be noticed that the morning overpass time of the European platforms is not the most favourable to monitor air quality at mid-latitudes since it results in very low sun observations in winter. A satellite overpass time around noon allows for observations at minimum solar zenith angles, increasing the sensitivity to the boundary layer.

We performed two exploratory studies to validate GOME-2 observations using MAX-DOAS measurements in Cabauw (The Netherlands) and Beijing (China). When taking into account the differences in vertical measurement sensitivities, differences of less than 10% are found in Cabauw, while similar comparisons in Beijing show larger discrepancies. This might be attributed to the high horizontal variability of the H$_2$CO columns around Beijing, which makes the comparison difficult given the coarse spatial resolution of GOME-2.

Implications for NMVOC emissions

On the global scale, the largest annual H$_2$CO columns are consistently retrieved over Tropical ecosystems. Seasonal variations are principally related to biogenic emission changes with temperature and solar insulation, while the interannual variability is linked to exceptional events, such as strong fire episodes (*e.g.* 1997 in Indonesia, 2004 in Alaska, 2010 in Russia) or to a lesser extent, particularly warm summer seasons (*e.g.* 2003 in Europe, or 2007 in US, or 2010 in Russia). Although more difficult to detect, anthropogenic signals have been identified unambiguously around industrial areas in Asia and in America.

Owing to the efforts made during this thesis to consolidate H$_2$CO datasets from three sensors covering almost 15 years of continuous global observations, it has been possible for the first time to quantify long-term changes in H$_2$CO columns and therefore in NMVOC emissions. Statistically significant positive trends, in the range of 1.5 to 4% yr^{-1}, were found in Asia, more particularly in Eastern China and India, and around some of the largest agglomerations worldwide. In contrast, negative trends (-2.5% yr^{-1}) were detected around Tokyo and several cities of Eastern US. The inferred trends in H$_2$CO columns are explained by changes in the corresponding anthropogenic NMVOC emissions.

A second essential outcome of our study has been to provide data suitable to perform inverse modelling of NMVOC emissions with the IMAGES model. The results demonstrate the strength of satellite data as top-down constrain for biogenic and biomass burning NMVOC emission inventories. On the global scale the inferred emissions from the two categories exhibit only weak deviations from the corresponding *a priori* estimates. However, the regional updates often present large departures from their prior values. For example, the posterior isoprene emissions over North America are estimated to be about 25% lower than the MEGAN-ECMWF estimate, whereas an

increase of more than 50% is deduced over the South African continent. Despite important *a priori* deviations between the biomass burning inventories (GFED version 1 or version 2), pyrogenic estimates inferred from the inversions converge on reasonably similar values. Our dataset has also been used by other modelling teams to perform the same kind of studies, in combination with OMI observations [*Barkley et al.*, 2009 and 2011, *Kim et al.*, 2011].

Perspectives

Retrieval improvements

The main limitation of current formaldehyde observations is the random error on the retrieved slant columns which effectively limits the resolution of the measurements in both space and time. The fitting procedure might possibly be improved by extending the fitting interval towards shorter wavelengths in order to capture an additional strong H_2CO absorption band. However, to do so, the DOAS approach needs to be modified to handle the AMF variations within the wavelength interval, caused by the strong ozone absorption at these shorter wavelengths. This can be done by developing the idea of the wavelength and solar zenith angle dependant ozone cross-section correction applied for the GOME-2 retrievals. This approach is already used to improve the retrieval of ozone [*Coldewey-Egbers et al.*, 2005; *Lerot et al.,* 2010a], but has to be extended to weak absorbers.

One of the main improvements required in the AMF calculation is the implementation of a correction to take into account the absorbing aerosol effects which are especially important, *e.g.* in biomass burning conditions. This could be achieved using modelled or measured aerosol optical properties, which should be included in the radiative transfer calculations. Other additional refinements include improvements to the accuracy and the spatial resolution of the input parameters (the albedo database, the *a priori* profile shapes, or the cloud retrieval algorithm). For example, an updated albedo database with a finer spatial resolution is needed and the effects of changes in land use on surface reflectivity might also be taken into account. These improvements are expected to increase the spatial and temporal accuracy of the H_2CO columns.

Validation

The validation of satellite retrievals in the lower troposphere is challenging. Ground-based measurements, where available, are often sampling the atmosphere at different spatial and temporal scales than the satellite measurements, which leads to ambiguous comparisons. Additional measurements are definitely needed over a variety of regions, in particular in the Tropics and at the suburban level in mid-latitudes. Suitable ground-based measurements obtained with FTIR and MAX-DOAS instruments are progressively becoming available [*Jones et al.*, 2009; *Vigouroux et al.*, 2009; *Clemer et al.*, 2010] and will be used in future studies to validate satellite results. Also, aircraft measurements and *in situ* data can be used to validate, directly or indirectly, the H_2CO total columns.

Existing concurrent satellite H_2CO datasets should also be intercompared. Such comparisons should be performed taking into account relevant parameters such as the averaging kernels, the spatial resolution and the time of observations. For example, a comparison study could be performed between the GOME-2 and OMI observations over Beijing and the suburban area of Xianghe, taking into account their different spatial resolution (80x40 versus 24x13 km^2) and their different overpass times (9:30 am versus 01:45 pm), based on the aerosol and H_2CO profiles diurnally retrieved from MAX-DOAS measurements.

Future applications for tropospheric chemistry studies

In parallel, the development of inversion capabilities to relate satellite observations to emissions and surface concentrations will be continued. In particular, the representation of anthropogenic NMVOC in the IMAGES model needs to be more detailed, by including a larger number of NMVOC species, in order to better represent the H_2CO production. The uncertainties in the chemical degradation schemes need also to be reduced for the numerous anthropogenic NMVOCs, but also for the isoprene oxidation at low NO_x levels [*Stavrakou et al.,* 2010; *Barkley et al.*, 2011]. The spatial and temporal resolution of the global IMAGES model (currently of 2°x2.5°, 1 day), could be refined but at important computational costs. Besides, satellite H_2CO and NO_2 observations are increasingly used in regional chemistry transport model such as CHIMERE (spatial resolution of 0.5°x0.5°) [*Bessagnet et al.*, 2004] for air quality studies and inverse modelling of emissions [*Dufour et al.*, 2009b; *Mijling et al.*, 2009; *Curci et al.*, 2010]. For such applications, a good spatial resolution of the satellite measurements is an important requirement and OMI is currently the best suited instrument to monitor air quality at the urban level.

Finally, formaldehyde observations could also be used synergistically with other satellite trace gases observations in inverse modelling applications. For example tropospheric glyoxal and formaldehyde columns retrieved from SCIAMACHY have been used with the IMAGESv2 CTM in a two-compound inversion scheme designed to estimate the continental source of glyoxal [*Stavrakou et al.*, 2009c]. Over North America, the ratio of tropospheric columns of H_2CO to NO_2 from OMI observations have been used to investigate the variations in the relative sensitivity of surface ozone formation in summer to precursor species concentrations of NMVOCs and NO_x [*Duncan et al.*, 2010]. To perform this kind of multi-compounds studies, it will be essential to homogenize as far as possible the settings and the external parameters used for the satellite retrievals, in order to reduce systematic errors.

Future satellite instruments

The time series of the global morning observations of formaldehyde columns created during this work will be extended with measurements from the next two GOME-2 instruments, respectively on METOP-B and METOP-C, which are planned for launch in mid 2012 and in the 2016/2017 timeframe. Future instruments of higher spatial resolution will allow improving the identification of the sources at the urban scale. TROPOMI, to be launched in 2014 on the GMES Sentinel 5 precursor mission on a sun-synchronous early afternoon orbit and latter on Sentinel 5 in 2020, will provide a spatial resolution of ~7x7 km^2 with a daily global Earth coverage, and should therefore improve the monitoring of air quality. With the GMES Sentinel 4 mission, to be launched on the EUMETSAT geostationary platform MTG in 2018, atmospheric composition measurements will for the first time be available at high time resolution (30 min). Although the sentinel 4/UV-N instrument will only cover the European region, its measurements will provide a new insight into the evolution and extent of air pollution with the temporal resolution necessary to address air quality on a diurnal basis.

Publications

Here follows a list of the publications and conference proceedings published in the framework of this thesis.

Peer-reviewed papers

De Smedt, I., J.-F. Müller, T. Stavrakou, R. van der A, H. Eskes, and M. Van Roozendael (2008), Twelve years of global observation of formaldehyde in the troposphere using GOME and SCIAMACHY sensors, *Atmos. Chem. Phys.*, *8*(16), 4947-4963.

De Smedt, I., T. Stavrakou, J.-F. Müller, R. J. van Der A, and M. Van Roozendael (2010), Trend detection in satellite observations of formaldehyde tropospheric columns, *Geophys. Res. Lett.*, *37*(18), L18808, doi:10.1029/2010GL044245.

Barkley, M., P. Palmer, I. De Smedt, T. Karl, A. Guenther, and M. Van Roozendael (2009), Regulated large-scale annual shutdown of Amazonian isoprene emissions?, *Geophys. Res. Lett.*, *36*(L04803), 1-5, doi:10.1029/2008GL036843.

Barkley, M., P. Palmer, L. Ganzeveld, A. Almut, D. Hagberg, T. Karl, A. Guenther, F. Paulot, P. O. Wennberg, J. Mao, T. Kurosu, K. Chance, J.-F. Müller, I. De Smedt, M. Van Roozendael, D. Chen, Y. Wang, R. M. Yantosca (2011), Can a 'state of the art' chemistry transport model really simulate Amazonian tropospheric chemistry?, submitted to *J. of Geophys. Res.*.

Boersma, K. F., D. J. Jacob, M. Trainic, Y. Rudich, I. De Smedt, R. Dirksen, and H. J. Eskes (2009), Validation of urban NO_2 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*(12), 3867-3879.

Blond, N., K. F. Boersma, H. J. Eskes, M. Van Roozendael, and I. De Smedt (2007), Intercomparison of SCIAMACHY nitrogen dioxide observations, in situ measurements and air quality modeling results over Western Europe, *J. of Geophys. Res.*, *112*(D10), doi:10.1029/2006JD007277.

Kim, J. H., S. M. Kim, K. H. Baek, L. Wang, T. Kurosu, I. De Smedt, K. Chance, and M. J. Newchurch (2011), Evaluation of satellite-derived HCHO using statistical methods, *Atmos. Chem. Phys. Discuss.*, 11, 8003-8025, doi:10.5194/acpd-11-8003-2011.

Lerot, C., T. Stavrakou, I. De Smedt, J.-F. Müller, and M. Van Roozendael (2010), Glyoxal vertical columns from GOME-2 backscattered light measurements and comparisons with a global model, *Atmos. Chem. Phys.*, *10*(24), 12059-12072.

Mijling, B., R. J. van der A, K. F. Boersma, M. Van Roozendael, I. De Smedt, and H. M. Kelder (2009), Reductions of NO_2 detected from space during the 2008 Beijing Olympic Games, *Geophys. Res. Lett.*, *36*(L13801), 1-6, doi:10.1029/2009GL038943.

Müller, J.-F., T. Stavrakou, S. Wallens, I. De Smedt, M. Van Roozendael, M. J. Potosnak, J. Rinne, B. Munger, A. Goldstein, and A. Guenther (2008), Global isoprene emissions estimated using MEGAN, ECMWF analyses and a detailed canopy environment model, *Atmos. Chem. Phys.*, *8*(5), 1329-1341.

Stavrakou, T., J.-F. Müller, K. F. Boersma, I. De Smedt, and R. van der A (2008), Assessing the distribution and growth rates of NOx emission sources by inverting a 10-year record of NO_2 satellite columns, *Geophys. Res. Lett.*, *35*(10), 2-6, doi:10.1029/2008GL033521.

Stavrakou, T., J.-F. Müller, I. De Smedt, M. Van Roozendael, G. R. van Der Werf, L. Giglio, and A. Guenther (2009a), Evaluating the performance of pyrogenic and biogenic emission inventories against one decade of space-based formaldehyde columns, *Atmos. Chem. Phys.*, *9*(3), 1037-1060.

Stavrakou, T., J.-F. Müller, I. De Smedt, M. Van Roozendael, G. R. van Der Werf, L. Giglio, and A. Guenther (2009b), Global emissions of non-methane hydrocarbons deduced from SCIAMACHY formaldehyde columns through 2003–2006, *Atmos. Chem. Phys.*, *9*, 3663-3679.

Stavrakou, T., J.-F. Müller, I. De Smedt, M. Van Roozendael, M. Kanakidou, M. Vrekoussis, F. Wittrock, A. Richter, and J. P. Burrows (2009c), The continental source of glyoxal estimated by the synergistic use of spaceborne measurements and inverse modelling, *Atmos. Chem. Phys.*, *9*(21), 8431-8446.

Theys, N., M. Roozendael, F. Hendrick, X. Yang, I. De Smedt, A. Richter, M. Begoin, Q. Errera, P. V. Johnston, K. Kreher, and M. De Mazière (2011), Global observations of tropospheric BrO columns using GOME-2 satellite data, *Atmos. Chem. Phys.*, *11*(4), 1791-1811.

van der A, R. J., H. J. Eskes, K. F. Boersma, T. P. van Noije, M. Van Roozendael, I. De Smedt, D. H. Peters, and E. W. Meijer (2008), Trends, seasonal variability and dominant NOx source derived from a ten year record of NO2 measured from space, *J. Geophys. Res.*, *113*(D04302), 6-12, doi:10.1029/2007JD009021.

van der A, R. J., D. H. Peters, H. Eskes, K. F. Boersma, M. Van Roozendael, I. De Smedt, and H. M. Kelder (2006), Detection of the trend and seasonal variation in tropospheric NO_2 over China, *J. Geophys. Res.*, *111*(D12317), 1-10, doi:10.1029/2005JD006594.

Vigouroux, C., F. Hendrick, T. Stavrakou, B. Dils, I. De Smedt, C. Hermans, et al. (2009), Ground-based FTIR and MAX-DOAS observations of formaldehyde at Réunion Island and comparisons with satellite and model data, *Atmos. Chem. Phys.*, *9*(24), 9523-9544.

Proceedings

De Smedt, I., T. Stavrakou, J.-F. Müller, N. Hao, *et al.* (2009), H_2CO columns retrieved from GOME-2: first scientific results and progress towards the development of an operational product, *Proceedings of the 2009 EUMETSAT Meteorological Satellite Conference*, Bath, U.K.

De Smedt, I., M. Van Roozendael, T. Stavrakou, J.-F. Müller, R. Van der A, H. Eskes (2007), Global Observation of Formaldehyde in the Troposphere by Satellites: Gome and Sciamachy Results, *Proceedings of the 2007 ESA Envisat Symposium*, Montreux, Switzerland.

De Smedt, I., M. Van Roozendael, T. Stavrakou, J.-F. Müller, R. Van der A, Tropospheric CH_2O Observations from Satellites: Error Budget Analysis of 12 Years of Consistent Retrieval from GOME and SCIAMACHY Measurements (2007), *Proceedings of the Second ACCENT Symposium*, Urbino, Italy.

De Smedt, I., M. Van Roozendael, R. Van der A, H. Eskes, and J.-F. Müller (2006), Retrieval of formaldehyde columns from GOME as part of the GSE PROMOTE and comparison with 3D-CTM calculations, *Proceedings of the Atmospheric Science Conference*, ESA ESRIN, Frascati, Italy.

De Smedt, I., M. Van Roozendael and J.-F. Müller (2005), Formaldehyde Columns from GOME: Improved Retrievals and their Comparison with the IMAGES Model, *Proceedings of the First ACCENT Symposium,* Urbino, Italy.

References

Abbot, D. S., P. I. Palmer, R. V. Martin, K. V. Chance, D. J. Jacob, and A. Guenther (2003), Seasonal and interannual variability of North American isoprene emissions as determined by formaldehyde column measurements from space, *Geophys. Res. Lett.*, *30*(17), 1886.

Afe, O. T., A. Richter, B. Sierk, F. Wittrock, and J. P. Burrows (2004), BrO emission from volcanoes: A survey using GOME and SCIAMACHY measurements, *Geophys. Res. Lett.*, *31*, L24113, doi:10.1029/2004GL020994.

Aliwell, S. R., M. Van Roozendael, P.V.Johnston, A. Richter, T. Wagner, et al. (2002), Analysis for BrO in zenith-sky spectra: An intercomparison exercise for analysis improvement, *J. Geophys. Res.*, 107, D140, doi: 10.1029/2001JD000329.

Andreae, M. O. and P. Merlet (2001), Emission of trace gases and aerosols from biomass burning, *Global Biogeochem. Cycles*, 15, 955-966.

Andres, R. J, G. Marland, Fung I., Matthews E. (1996), A one degree by one degree distribution of carbon dioxide emissions from fossil fuel consumption and cement manufacture, 1950-1990, *Global Biogeochem. Cycles*, 10:419-429.

Barkley, M. P., P. I. Palmer, U. Kuhn, J. Kesselmeier, K. V. Chance, T. P. Kurosu, R. V. Martin, D. Helmig, and A. Guenther (2008), Net ecosystem fluxes of isoprene over tropical South America inferred from Global Ozone Monitoring Experiment (GOME) observations of HCHO columns, *J. Geophys. Res.*, *113*(D20), D20304.

Barkley, M., P.I. Palmer, I. De Smedt, T. Karl, *et al.* (2009), Regulated large-scale annual shutdown of Amazonian isoprene emissions?, *Geophys. Res. Lett., 36*, L04803, doi: 10.1029/2008GL036843.

Barkley, M., P. Palmer, L. Ganzeveld, A. Almut, D. Hagberg, T. Karl, A. Guenther, F. Paulot, P. O. Wennberg, J. Mao, T. Kurosu, K. Chance, J.-F. Müller, I. De Smedt, M. Van Roozendael, D. Chen, Y. Wang, R. M. Yantosca (2011), Can a 'state of the art' chemistry transport model really simulate Amazonian tropospheric chemistry?, submitted to *J. of Geophys. Res.*.

Barret, B., S. Turquety, D. Hurtmans, C. Clerbaux, J. Hadji-Lazaro, I. Bey, M. Auvray, and P.-F. Coheur (2005), Global carbon monoxide vertical distributions from spaceborne high-resolution FTIR nadir measurements, *Atmos. Chem. Phys.*, 5, 2901-2914, doi:10.5194/acp-5-2901-2005.

Bates, D. R. and M. Nicolet (1950), Atmospheric hydrogen, *Pup. Astron. Soc. Pacific*, 62.

Beer, R., *et al.* (2006), TES on the Aura Mission: Scientific Objectives, Measurements, and Analysis Overview, *IEEE Trans. Geosci. Remote Sens.* 44, 1102-1105.

Beirle, S., S. Kühl, J. Puķīte, and T. Wagner (2010), Retrieval of tropospheric column densities of NO_2 from combined SCIAMACHY nadir/limb measurements, *Atmosp. Meas. Tech.*, *3*(1), 283-299, doi:10.5194/amt-3-283-2010.

Bessagnet, B., A. Hodzic, R. Vautard, M. Beekmann, S. Cheinet, C. Honoré, C. Liousse and L. Rouil (2004), Aerosol modeling with CHIMERE: preliminary evaluation at the continental scale, *Atmos. Env.*, 38, 2803-2817.

Bergamaschi, P. (2005), Inverse modelling of national and European CH4 emissions using the atmospheric zoom model TM5, *Atmos. Chem. Phys.*, *5*, 2431-2460.

Bey, I., D. J. Jacob, R. M. Yantosca, J. A. Logan, B. D. Field, A. M. Fiore, Q. Li, H. Y. Liu, L. J. Mickley, and M. G. Schultz (2001), Global modeling of tropospheric chemistry with assimilated meteorology: Model description and evaluation, J. Geophys. Res., 106(D19), 23073-23096, doi:029/2001JD000807.

Bo, Y., H. Cai, and S.D. Xie (2008), Spatial and temporal variation of historical anthropogenic NMVOCs emission inventories in China, *Atmos. Chem. Phys., 8*, 7297-7316.

Boeke, N. L., J. D. Marshall, S. Alvarez, K. V. Chance, A. Fried, T. P. Kurosu, B. Rappenglück, *et al.* (2011), Formaldehyde columns from the Ozone Monitoring Instrument: Urban versus background levels and evaluation using aircraft data and a global model, *J. Geophys. Res., 116*(D5), D05303.

Boersma, K. F., H. J. Eskes, and E. J. Brinksma (2004), Error analysis for tropospheric NO_2 retrieval from space, *J. Geophys. Res.*, 109(D04311), doi:10.1029/2003JD003962.

Bogumil, K. , J. Orphal, S. Voigt, H. Bovensmann, O. C. Fleischmann, M. Hartmann, T. Homann, P. Spietz, and J. P. Burrows (1999), Reference spectra of atmospheric trace gases measured with the SCIAMACHY PFM satellite spectrometer, *in ESAMS '99 - European Symposium on atmospheric Measurements from Space*, volume 2, pages 443–447.

Bogumil, K., *et al.* (2003), Measurements of molecular absorption spectra with the SCIAMACHY pre-flight model: Instrument characterization and reference data for atmospheric remote sensing in the 230-2380 nm region, *J. Photochem. Photobiol. A.: Photochem.*, 157, 167-184.

Bousquet, P. et al. (2006), Contribution of anthropogenic and natural sources to atmospheric methane variability, *Nature, 443*(7110), 439-43, doi:10.1038/nature05132.

Bovensmann, H., J. P. Burrows, M. Buchwitz, J. Frerick, S. Noël, V. V. Rozanov, K. V. Chance, and A. H. Goede (1999), SCIAMACHY - Mission objectives and measurement modes, *J. Atmos. Sci.*, 56, 2, 127-150.

Boynard, A., C. Clerbaux, P.-F. Coheur, D. Hurtmans, S. Turquety, M. George, J. Hadji-Lazaro, C. Keim, and J. Meyer-Arnek (2009), Measurements of total and tropospheric ozone from IASI: comparison with correlative satellite, ground-based and ozonesonde observations, *Atmos. Chem. Phys., 9*(16), 6255-6271.

Brasseur, G. P., J. J. Orlando, and G. S. Tyndall (1999), Atmospheric Chemistry and Global Change, *Oxford University Press*.

Buchwitz, M., R. de Beek, J. P. Burrows, H. Bovensmann, T. Warneke, J. Notholt, J. F. Meirink, A. P. H. Goede, P. Bergamaschi, S. Körner, M. Heimann, J.-F. Müller, and A. Schulz (2005), Atmospheric methane and carbon dioxide from SCIAMACHY satellite data: Initial comparison with chemistry and transport models, *Atmos. Chem. Phys.*, 5, 941-962.

Buchwitz, M., I. Khlystova, H. Bovensmann, and J. P. Burrows (2007), Three years of global carbon monoxide from SCIAMACHY: comparison with MOPITT and first results related to the detection of enhanced CO over cities, *Atmos. Chem. Phys.*, 7, 2399–2411.

Bucsela, E. J., E. A. Celarier, M. O. Wenig, J. F. Gleason, J. P. Veefkind, K. F. Boersma, and E. J. Brinksma (2006), Algorithm for NO_2 vertical column retrieval from the ozone monitoring instrument, *IEEE Trans. Geosci. Remote Sens.*, 44, 1245–1258.

Burrows, J. P., A. Dehn, B. Deters, S. Himmelmann, A. Richter, S. Voigt and J. Orphal (1998), Atmospheric remote-sensing reference data from GOME: 1. Temperature-dependent absorption cross sections of NO_2 in the 231-794 nm range , *J. Quant. Spectrosc. Rad. Transfer*, 60, 1025.

Burrows, J. P., A. Richter, A. Dehn, B. Deters, S. Himmelmann, S. Voigt, and J. Orphal (1999a), Atmospheric remote-sensing reference data from GOME: 2. Temperature-dependent absorption cross sections of O_3 in the 231-794 nm range, *J. Quant. Spectrosc. Rad. Transfer*, 61, 509-517.

Burrows, J. P., M. Weber, M. Buchwitz, V. V. Rozanov, A. Ladstädter-Weissenmayer, A. Richter, R. de Beek, R. Hoogen, K. Bramstedt, K.-U. Eichmann, M. Eisinger and D. Perner (1999b), The Global Ozone Monitoring Experiment (GOME): mission concept and first scientific results, *J. Atmos. Sci.*, 56, 151-175.

Brasseur, G., R. G. Prinn, A. P. Pszenny (2003), Atmospheric chemistry in a changing world, The IGBP series, *Springer-Verlag, Berlin-Heidelberg*.

Brauers, T., J. Bossmeyer, H.-P. Dorn, E. Schlosser, R. Tillmann, R. Wegener, and A. Wahner (2007), Investigation of the formaldehyde differential absorption cross section at high and low spectral resolution in the simulation chamber SAPHIR, *Atmos. Chem. Phys.*, 7(1), 3579-3586, doi:10.5194/acp-7-3579-2007.

Brion, J., *et al.* (1998), J., Absorption spectra measurements for the ozone molecule in the 350-830 nm region, *J. Atmos. Chem.*, 30, 291-299.

Brion, J., *et al.* (1993), J., High-resolution laboratory absorption cross section of O3. Temperature effect, *Chem. Phys. Lett.*, 213 (5-6), 610-512.

Cai, H. and S.D. Xie (2009), Tempo-spatial variation of emission inventories of speciated volatile organic compounds from on-road vehicles in China, *Atmos. Chem. Phys., 9*, 6983-7002.

Callies, J., Corpaccioli, E., Eisinger, M., Hahne, A., and Lefebvre, A. (2000), GOME-2- Metop's second-generation sensor for operational ozone monitoring, *ESA Bull.*, 102, 28–36.

Cantrell, C. A., J. A. Davidson, A. H. McDaniel, R. E. Shetter, and J. G. Calvert (1990), Temperature-dependent formaldehyde cross sections in the near-ultraviolet spectral region, *J. Phys. Chem.*, 94, 3902-3908.

Chahine, M., C. Barnet, E. T. Olsen, L. Chen, E. Maddy (2005), On the determination of atmospheric minor gases by the method of vanishing partial derivatives with application to CO_2, *Geophys. Res. Lett.*, 32, L22803, doi:10.1029/2005GL024165.

Chance, K. and R. J. Spurr (1997), Ring effect studies: Rayleigh scattering including molecular parameters for rotational Raman scattering, and the Fraunhofer spectrum, *Applied Optics*, 36, 5224-5230.

Chance, K., P. I. Palmer, R. J. Spurr, R. V. Martin, T. P. Kurosu, and D. J. Jacob (2000), Satellite observations of formaldehyde over North America from GOME, *Geophys. Res. Lett.*, 27, 21, 3461-3464.

Chandrasekhar, S. (1960): Radiative transfer, *Dover Publications, Inc., New York*.

Chen, Y., and R. G. Prinn (2006), Estimation of atmospheric methane emissions between 1996 and 2001 using a three-dimensional global chemical transport model, *J. Geophys. Res.*, *111*(D10), D10307.

Clémer, K. et al. (2009), Simultaneous retrieval of tropospheric aerosol extinction and NO_2 vertical profiles from MAXDOAS measurements in Beijing, *Proceedings of the 8th International Symposium on Tropospheric Profiling, ISBN 978-90-6960-233-2, Delft, The Netherlands*.

Clémer, K., M. Van Roozendael, C. Fayt, F. Hendrick, C. Hermans, G. Pinardi, R. Spurr, P. Wang, and M. De Mazière (2010a), Multiple wavelength retrieval of tropospheric aerosol optical properties from MAXDOAS measurements in Beijing, *Atmos. Meas. Tech.*, *3*(4), 863-878.

Clémer, K., I. De Smedt, C. Fayt, F. Hendrick, C. Hermans, C. Lerot, G. Pinardi, R. Spurr, P. Wang, M. Van Roozendael (2010b), Tropospheric pollutant retrievals from MAXDOAS measurements in Beijing, *Extendend abstract of the 8th International Symposium on Advanced Environmental Monitoring*, Hokkadio, Japan.

Clerbaux, C., J. Hadji-Lazaro, S. Turquety, G. Mégie, and P.-F. Coheur (2003), Trace gas measurements from infrared satellite for chemistry and climate applications, *Atmos. Chem. Phys.*, *3*(5), 1495-1508.

Clerbaux, C., J. C. Gille, and D. P. Edwards (2004), New Directions: Infrared measurements of atmospheric pollution from space, *Atmos. Environ.*, *38*(27), 4599-4601, doi:10.1016/j.atmosenv.2004.05.005.

Clerbaux, C. et al. (2009), Monitoring of atmospheric composition using the thermal infrared IASI/MetOp sounder, *Atmos. Chem. Phys.*, *9*(16), 6041-6054.

Cohen, R. C., D. Helmig, and J. Lelieveld, J. (2008), The "Guyanas Atmosphere-Biosphere exchange and Radicals Intensive Experiment with a Learjet" (GABRIEL) measurement campaign in October 2005, *Atmos. Chem. Phys.*, Special Issue, available at: http://www.atmos-chem-phys.net/special issue88.html.

Coheur, P.-F., H. Herbin, C. Clerbaux, D. Hurtmans, C. Wespes, M. Carleer, S. Turquety, et al. (2007), ACE-FTS observation of a young biomass burning plume: first reported measurements of C2H4, C3H6O, H2CO and PAN by infrared occultation from space, *Atmos. Chem. Phys.*, *7*(20), 5437-5446.

Coheur, P.-F., L. Clarisse, S. Turquety, D. Hurtmans, and C. Clerbaux (2009), IASI measurements of reactive trace species in biomass burning plumes, *Atmos. Chem. Phys.*, *9*(15), 5655-5667.

Coldewey-Egbers, M., M. Weber, L. N. Lamsal, R. de Beek, M. Buchwitz, and J. P. Burrows (2005), Total ozone retrieval from GOME UV spectral data using the weighting function DOAS approach, *Atmos. Chem. Phys.*, *5*(4), 1015-1025.

Crevoisier, C., A. Chédin, H. Matsueda, T. Machida, R. Armante, and N. A. Scott (2009), First year of upper tropospheric integrated content of CO_2 from IASI hyperspectral infrared observations, *Atmos. Chem. Phys.*, *9*(14), 4797-4810.

Curci, G., P. I. Palmer, T. P. Kurosu, K. Chance, and G. Visconti (2010), Estimating European volatile organic compound emissions using satellite observations of formaldehyde from the Ozone Monitoring Instrument, *Atmos. Chem. Phys.*, *10*(23), 11501-11517.

Deeter, M. N., et al. (2003), Operational carbon monoxide retrieval algorithm and selected results for the MOPITT instrument, *J. Geophys. Res.,* 108(D14), 4399, doi:10.1029/2002JD003186.

Delmas, R., G. Mégie, V.-H. Peuch (2005), Physique et chimie de l'atmosphère, *Editions Belin, Paris*.

Demoulin, P., R. Zander, F. Mélen, E. Mahieu, and C. Servais (1999), Column abundance measurements of formaldehyde above the Jungfraujoch, *in Proceedings of Atmospheric Spectroscopy Applications 1999*, Reims, France, 1–3 September, 59–62.

Denman, K. L., G. Brasseur, A. Chidthaisong, P. Ciais, P. M. Cox, R. E. Dickinson, D. Hauglustaine, C. Heinze, E. Holland, D. Jacob, U. Lohmann, S Ramachandran, P.L. da Silva Dias, S.C. Wofsy and X. Zhang (2007), Couplings between changes in the climate system and biogeochemistry, in: Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change [Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K.B. Averyt, M.Tignor and H.L. Miller (eds.)], *Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA*.

De Smedt, I., M. Van Roozendael, R. Van der A, H. Eskes, and J.-F. Müller (2006), Retrieval of formaldehyde columns from GOME as part of the GSE PROMOTE and comparison with 3D-CTM calculations, *Proceedings of the Atmospheric Science Conference*, ESA.

De Smedt, I., J.-F. Müller, T. Stavrakou, R. van der A, *et al.* (2008), Twelve years of global observation of formaldehyde in the troposphere using GOME and SCIAMACHY sensors, *Atmos. Chem. Phys., 8*, 4947-4963.

De Smedt, I., T. Stavrakou, J.-F. Müller, N. Hao, P. Valks, D. Loyola, M. Van Roozendael (2009), H_2CO columns retrieved from GOME-2: first scientific results and progress towards the development of an operational product, *Proceedings of the 2009 EUMETSAT Meteorological Satellite Conference*, Bath, U.K..

De Smedt, I., T. Stavrakou, J.-F. Müller, R. J. van Der A, and M. Van Roozendael (2010), Trend detection in satellite observations of formaldehyde tropospheric columns, *Geophys. Res. Lett.*, *37*(18), L18808, doi:10.1029/2010GL044245.

Dlugokencky, E. J., S. Houweling, L. M. Bruhwiler, K. A. Masarie, P. M. Lang, J. B. Miller, and P. P. Tans (2003), Atmospheric methane levels off: Temporary pause or a new steady-state?, *Geophys. Res. Lett.*, *30*(19), 1992, doi:10.1029/2003GL018126.

Dufour, G., S. Szopa, M. P. Barkley, C. D. Boone, A. Perrin, P. I. Palmer, and P. F. Bernath (2009a), Global upper-tropospheric formaldehyde: seasonal cycles observed by the ACE-FTS satellite instrument, *Atmos. Chem. Phys.*, *9*(12), 3893-3910.

Dufour, G., F. Wittrock, M. Camredon, M. Beekmann, A. Richter, B. Aumont, and J. P. Burrows (2009b), SCIAMACHY formaldehyde observations: constraint for isoprene emission estimates over Europe?, *Atmos. Chem. Phys.*, *9*(5), 1647-1664.

Dufour, G., M. Eremenko, J. Orphal, and J.-M. Flaud (2010), IASI observations of seasonal and day-to-day variations of tropospheric ozone over three highly populated areas of China: Beijing, Shanghai, and Hong Kong, *Atmos. Chem. Phys.*, *10*(8), 3787-3801.

Duncan, B. N., Y. Yoshida, J. R. Olson, S. Sillman, R. V. Martin, L. Lamsal, Y. Hu, K. E. Pickering, C. Retscher, and D. J. Allen (2010), Application of OMI observations to a space-based indicator of NOx and VOC controls on surface ozone formation, *Atmospheric Environment*, *44*(18), 2213-2223, doi:10.1016/j.atmosenv.2010.03.010.

Edwards, D. P. (2006), Air Quality Remote Sensing From Space, *Eos Trans. AGU*, *87*(33), doi:10.1029/2006EO330005.

Ehhalt, D., M. Prather, F. Dentener, R. Derwent, E. Dlugokencky, E. Holland, I. Isaksen, J. Katima, V. Kirchhoff, P. Matson, P. Midgley, M. Wang (2001), Atmospheric Chemistry and Greenhouse Gases, in Climate Change 2001: The Scientific Basis. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change [Houghton, J.T.,Y. Ding, D.J. Griggs, M. Noguer, P.J. van der Linden, X. Dai, K. Maskell, and C.A. Johnson (eds.)], *Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA*.

Emmons, L.K., D.A. Hauglustaine, J.-F. Muller, M.A. Carroll, G.P. Brasseur, D. Brunner, J. Staehelin, V. Thouret, A. Marenco (2000), Data composites of airborne observations of tropospheric ozone and its precursors, *J. Geophys. Res.*, 105, 16, 20497-20538.

Eisinger, M. and J. P. Burrows (1998), Tropospheric sulfur dioxide observed by the ERS-2 GOME instrument, *Geophys. Res. Lett.*, 25, 22, 4177-4180.

Eskes, H. J. et al. (2003a), GOME assimilated and validated ozone and NO2 fields for scientific users and model validation, *Final Report (http://www.knmi.nl/goa), European Commission, Fifth Framework Programme, Environment and Sustainable Development*, 1998-2002.

Eskes, H. J. and K. F. Boersma (2003b), Averaging kernels for DOAS total-column satellite retrievals, *Atmos. Chem. Phys.*, 3, 1285-1291.

ESPERE - Environmental Science Published for Everybody Round the Earth (2003-2010), Climate encyclopedia, *http://www.atmosphere.mpg.de/enid/2cc9087c8fc3f933bfd33eba65c5050a,0 /English/CLIMATE_ENCYCLOPAEDIA_144.html*.

ETOPO5 (1988), Data Announcement 88-MGG-02, Digital relief of the Surface of the Earth. NOAA, *National Geophysical Data Center, Boulder, Colorado*.

EUMETSAT (2009), GOME-2 Products Guide, *Ref.: EUM/OPS-EPS/MAN/07/0445,* Issue: v2D, Date: 6 Mar 2009.

Evans M. J., A. Fiore, and D. J. Jacob (2003), The GEOS-Chem chemical mechanism version 5-07-8, *Harvard University, Cambridge, MA, USA*, http://homepages.see.leeds.ac.uk/_lecmje/GEOS-CHEM/ geoschem mech.pdf.

Farman, J. C., B. G. Gardiner and J. D Shanklin (1985), Large losses of total ozone in Antarctica reveal seasonal ClO$_x$/NO$_x$ interaction, *Nature*, 315, 207-210.

Fayt, C. and M. Van Roozendael (2001), Windoas 2.1, *Software User Manual*, *BIRA-IASB*.

Fiore, A. M., H. Levy II, and D. A. Jaffe (2011), North American isoprene influence on intercontinental ozone pollution, *Atmos. Chem. Phys., 11*(4), 1697-1710, doi:10.5194/acp-11-1697-2011.

Fishman, J., C. Watson, J. Larsen, and J. Logan (1990), Distribution of tropospheric ozone determined from satellite data, *J. Geophys. Res.*, 95(D4), 3599-3617.

Fishman, J., J. M. Jr. Hoell, R. J. Bendura, R. J. McNeal, and V. W. Kirchhoff (1996), NASA GTE TRACE-A experiment (September-October, 1992): Overview, *J. Geophys. Res.*, 101(D19), 23865-23880, doi:10.1029/96JD00123.

Fishman, J. et al. (2008), Remote Sensing of Tropospheric Pollution from Space, *Bulletin of the American Meteorological Society*, 89(6), 805-821.

Fleischmann, O. C., et al. (2004), New ultraviolet absorption cross-sections of BrO at atmospheric temperatures measured by time-windowing Fourier transform spectroscopy, *J. Photochem. Photobiol. A*, 168, 117–132.

Fu, T.-M., D. J. Jacob, P. I. Palmer, K. V. 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(D6), D06312.

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(D15), D15303.

Forster, P, V. Ramaswamy, P. Artaxo, T. Berntsen, R. Betts, D.W. Fahey, J. Haywood, J. Lean, D.C. Lowe, G. Myhre, J. Nganga, R. Prinn, G. Raga, M. Schulz and R. Van Dorland (2007), Changes in atmospheric constituents and in radiative forcing, in Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change [Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K. B. Averyt, M. Tignor and H. L. Miller (eds.)], *Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.*

Frankenberg, C., J. F. Meirink, M. vanWeele, U. Platt, T.Wagner (2005), Assessing methane emissions from global space-borne observations, *Science*, 308 (5724), 1010-1014, 10.1126/science.1106644.

Fried, A., J. G. Walega, J. R. Olson, J. H. Crawford, G. Chen, P. Weibring, D. Richter, C. Roller, F. Tittel, B. G. Heikes, *et al.* (2008), Formaldehyde over North America and the North Atlantic during the Summer 2004 INTEX Campaign: Methods, Observed Distributions, and Measurement-Comparisons, *J. Geophys. Res.*, 113, D10302, doi:10.1029/2007JD009185.

Kanakidou, M. et al. (2005), Organic aerosol and global climate modelling: a review, *Atmos. Chem. Phys.*, 5(4), 1053-1123.

Karl, T., A. Guenther, R. J. Yokelson, J. Greenberg, M. Potosnak, D. R. Blake and P. Artaxo (2007), The tropical forest and fire emissions experiment: Emission, chemistry, and transport of biogenic volatile organic compounds in the lower atmosphere over Amazonia, *J. Geophys. Res.*, 112, D18302, doi:10.1029/2007JD008539.

Kurucz, R. L., I. Furenlid, J. Brault, and L. Testerman (1984), Solar flux atlas from 296 to 1300 nm, *National Solar Observatory, Sunspot, New Mexico, U.S.A..*

Galbally, I. (1989), The International Global Atmospheric Chemistry Programme, *Commission on Atmospheric Chemistry and Global Pollution of the International Association of Meteorology and Atmospheric Physics.*

Gardiner, T., A. Forbes, M. de Mazière, C. Vigouroux, *et al.* (2008), Trend analysis of greenhouse gases over Europe measured by a network of ground-based remote FTIR instruments, *Atmos. Chem. Phys., 8*, 6719-6727.

Giglio, L. (2007), Characterization of the tropical diurnal fire cycle using VIRS and MODIS observations, *Remote Sensing of Environment*, 108, 407-421.

Gottwald, M., H. Bovensmann, et al. (2006), SCIAMACHY, Monitoring the Changing Earth's Atmosphere, *DLR, Institut für Methodik der Fernerkundung (IMF).*

Gottwald, M., H. Bovensmann, et al. (2011), SCIAMACHY, Exploring the Changing Earth's Atmosphere, *Springer Dordrecht Heidelberg London New-York*.

Grainger, J. F. and J. Ring (1962), Anomalous Fraunhofer line profiles, *Nature*, 193, 762.

Granier, C., J. F. Lamarque, A. Mieville, J.-F. Muller, J. Olivier, J. Orlando, J. Peters, G. Petron, G. Tyndall, S. Wallens (2005), POET, a database of surface emissions of ozone precursors, *available on internet at http://www.aero.jussieu.fr/projet/ACCENT/POET.php*.

Gratien, A., B. Picquet-Varrault, J. Orphal, E. Perraudin, J.-F. Doussin, J.-M. Flaud (2007), Laboratory intercomparison of the formaldehyde absorption cross sections in the infrared (1660–1820 cm^{-1}) and ultraviolet (300–360 nm) spectral regions, *J. Geophys. Res.*, 112(D5), D05305, doi:10.1029/2006JD007201.

Greenblatt, G. D., J.J. Orlando, J.B. Burkholder, and A.R. Ravishankara (1990), Absorption measurements of oxygen between 330 and 1140 nm, *J. Geophys. Res.*, 95(D11), 18577-18582, doi:10.1029/90JD01375.

George, M. et al. (2009), Carbon monoxide distributions from the IASI/METOP mission: evaluation with other space-borne remote sensors, *Atmos. Chem. Phys.*, *9*(21), 8317-8330.

Guenther, A., C. N. Hewitt, D. Erickson, R. Fall, C. Geron, T. Graedel, P. Harley, L. Klinger, M. Lerdau, W. A. McKay, T. Pierce, B. Scholes, R. Steinbrecher, R. Tallamraju, J. Taylor, and P. Zimmerman (1995), A global model of natural volatile organic compound emissions, *J. Geophys. Res.*, 100(D5), 8873-8892, doi:10.1029/94JD02950.

Guenther, A., T. Karl, P. Harley, C. Wiedinmyer, P.I. Palmer, and C. Geron (2006), Estimates of the global terrestrial isoprene emissions using MEGAN (Model of Emissions of Gases and Aerosols from Nature), *Atmos. Chem. Phys.*, 6, 3181-3210.

Hak, C. et al. (2005), Intercomparison of four different in-situ techniques for ambient formaldehyde measurements in urban air, *Atmos. Chem. Phys.*, *5*(3), 2881-2900, doi:10.5194/acp-5-2881-2005.

Hauglustaine, D., J. Jouzel and H. Le Treut (2004), Climat : chronique d'un bouleversement annoncé, *Editions Le Pommier – Presses de la Cité*.

Heckel, A. , A. Richter, T. Tarsu, F. Wittrock, C. Hak, I. Pundt, W. Junkermann, and J. P. Burrows (2005), MAX-DOAS measurements of formaldehyde in the Po-Valley, *Atmos. Chem. Phys.*, 5, 909-918.

Hourdin, F. et al. (2006), The LMDZ4 general circulation model: climate performance and sensitivity to parametrized physics with emphasis on tropical convection, *Clim. Dyn.*, *27*(7-8), 787-813, doi:10.1007/s00382-006-0158-0.

Houweling, S., F. Dentener, and J. Lelieveld (1998), The impact of nonmethane hydrocarbon compounds on tropospheric photochemistry, *J. Geophys. Res.*, *103*, doi:10.1029/97JD03582.

Houweling, S., F. Dentener, and J. Lelieveld (2000), Simulation of preindustrial atmospheric methane to constrain the global source strength of natural wetlands, *J. Geophys. Res.*, *105*(D13), 17243-17255.

IGAC (2007), Satellite observations of tropospheric trace gases and aerosols, in: *IGACtivities Newsletter*.

IPCC (2001): Climate Change 2001: The Scientific Basis. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change [Houghton, J.T.,Y. Ding, D.J. Griggs, M. Noguer, P.J. van der Linden, X. Dai, K. Maskell, and C.A. Johnson (eds.)]. *Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 881pp*.

IPCC (2007): Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change [Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K.B. Averyt, M. Tignor and H.L. Miller (eds.)]. *Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA*, 996 pp.

Jacob, D. J. (1999), Introduction to Atmospheric Chemistry, *Princeton University Press, Princeton, New-Jersey*.

Jones, N. B., K. Riedel, W. Allan, S. Wood, P. I. Palmer, K. Chance, and J. Notholt (2009), Long-term tropospheric formaldehyde concentrations deduced from ground-based fourier transform solar infrared measurements, *Atmos. Chem. Phys.*, *9*(18), 7131-7142.

Kanakidou, M., J.H. Seinfeld, S.N. Pandis, I. Barnes, *et al.* (2005), Organic aerosol and global climate modelling: a review, *Atmos. Chem. Phys.*, *5*, 1053-1123.

Khokhar, M. F., C. Frankenberg, S. Beirle, S. Kühl, M. Van Roozendael, A. Richter, U. Platt and T. Wagner (2005), Satellite observations of atmospheric SO2 from volcanic eruptions during the time period of 1996 to 2002, *Journal of Advances in Space Research*, 36(5), 879-887, 10.1016/j.asr.2005.04.114.

Kim, J. H., S. M. Kim, K. H. Baek, L. Wang, T. Kurosu, I. De Smedt, K. Chance, and M. J. Newchurch (2011), Evaluation of satellite-derived HCHO using statistical methods, *Atmos. Chem. Phys. Discuss.*, 11, 8003-8025, doi:10.5194/acpd-11-8003-2011.

Kinnison, D. E. et al. (2007), Sensitivity of chemical tracers to meteorological parameters in the MOZART-3 chemical transport model, *J. Geophys. Res.*, *112*(D20), D20302.

Koelemeijer, R. B., P. Stammes, J. W. Hovenier, and J. F. de Haan (2002), Global distributions of effective cloud fraction and cloud top pressure derived from oxygen A band spectra measured by the Global Ozone Monitoring Experiment: comparison to ISCCP data, *J. Geophys. Res.*, 107(D12), 4151, doi:10.1029/2001JD000840.

Koelemeijer, R. B. A., J. F. de Haan, and P. Stammes, A database of spectral surface reflectivity in the range 335-772nm derived from 5.5 years of GOME observations (2003), *J. Geophys. Res.*, 108(D2), 4070, doi:10.1029/2002JD002429.

Kopacz, M. et al. (2010), Global estimates of CO sources with high resolution by adjoint inversion of multiple satellite datasets (MOPITT, AIRS, SCIAMACHY, TES), *Atmos. Chem. Phys.*, *10*(3), 855-876.

Kromminga, H., J. Orphal, P. Spietz, S. Voigt, and J.P. Burrows (2003), New measurements of OClO absorption cross sections in the 325-435 nm and their temperature dependence between 213-293 K, *J. Photochem. Photobiol. A.: Chemistry*, 157, 149-160.

Kühl, S., W. Wilms-Grabe, S. Beirle, C. Frankenberg, M. Grzegorski, J. Hollwedel, F. Khokhar, S. Kraus, U. Platt, S. Sanghavi, C. von Friedeburg and T. Wagner (2004), Stratospheric chlorine activation in the arctic winters 1995/96 to 2001/02 derived from GOME OClO measurements, *Adv. Space Res.*, 34, 798-803.

Kylling, K., P. Stamnes, and S.-C. Tsay (1995), A reliable and efficient two-stream algorithm for spherical radiative transfer: Documentation of accuracy in realistic layered media, *J. Atmos. Phys.*, 21, 115-150.

Lee, C., A. Richter, M. Weber, and J. P. Burrows (2008), SO$_2$ Retrieval from SCIAMACHY using the Weighting Function DOAS (WFDOAS) Technique: comparison with Standard DOAS retrieval, *Atmos. Chem. Phys.*, 8, 6137-6145, doi:10.5194/acp-8-6137-2008.

Leitão, J., A. Richter, M. Vrekoussis, A. Kokhanovsky, Q. J. Zhang, M. Beekmann, and J. P. Burrows (2010), On the improvement of NO$_2$ satellite retrievals – aerosol impact on the airmass factors, *Atmos. Meas. Tech.*, 3, 475–493, doi:10.5194/amt-3-475-2010.

Le Treut, H., R. Somerville, U. Cubasch, Y. Ding, C. Mauritzen, A. Mokssit, T. Peterson and M. Prather (2007), Historical Overview of Climate Change, in: Climate Change 2007: The Physical Science Basis, Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change [Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K.B. Averyt, M. Tignor and H.L. Miller (eds.)], *Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.*

Lerot, C., M. Van Roozendael, J.-C. Lambert, J. Granville, J. van Gent, D. Loyola, and R. Spurr (2010a), The GODFIT algorithm: a direct fitting approach to improve the accuracy of total ozone measurements from GOME, *Int. J. Rem. Sens.*, 31:2, 543 – 550.

Lerot, C., T. Stavrakou, I. De Smedt, J.-F. Müller, and M. Van Roozendael (2010b), Glyoxal vertical columns from GOME-2 backscattered light measurements and comparisons with a global model, *Atmos. Chem. Phys.*, *10*(24), 12059-12072.

Levelt, P. F., van den Oord, G. H. J., Dobber, M. R., M̈alkki, A., Visser, H., de Vries, J., Stammes, Lundell J., and Saari, H. (2006), The Ozone Monitoring Instrument, *IEEE Trans. Geo. Rem. Sens.*, 44, 5, 1093–1101, doi:10.1109/TGRS.2006.872333.

Liou, K. N. (2002), An introduction to atmospheric radiation, second edition, *Academic Press, Elsevier Science, USA*.

Liu, G., D. W. Tarasick, V. E. Fioletov, C. E. Sioris, and Y. J. Rochon (2009), Ozone correlation lengths and measurement uncertainties from analysis of historical ozones onde data in North America and Europe, *J. Geophys. Res.*, 114, D04112, doi:10.1029/2008JD010576.

Lopez, J. P., M. Luo, L. E. Christensen, M. Loewenstein, H. Jost, C. R. Webster, and G. Osterman (2008), TES carbon monoxide validation during two AVE campaigns using the Argus and ALIAS instruments on NASA's WB-57F, *J. Geophys. Res.-Atmos.*, 113, D16S47, doi:10.1029/2007JD008811.

Marbach, T., S. Beirle, J. Hollwedel, U. Platt, and T. Wagner (2004), Identification of tropospheric emission sources from satellite observations: Synergistic use of trace gas measurements of formaldehyde (HCHO), and nitrogen dioxide (NO_2), Proceedings of the ENVISAT & ERS Symposium, ESA publication SP-572.

Martin, R. V. et al. (2002), An improved retrieval of tropospheric nitrogen dioxide from GOME, *J. Geophys. Res.*, *107*(D20), doi:10.1029/2001JD001027.

Martin, R. V., D. J. Jacob, K. V. Chance, T. P. Kurosu, P. I. Palmer, and M. J. Evans (2003), Global inventory of nitrogen oxide emissions constrained by space-based observations of NO_2 columns, *J. Geophys. Res.*, *108*(D17), 1-12, doi:10.1029/2003JD003453.

Martin, R. V., A. M. Fiore, and A. Van Donkelaar (2004a), Space-based diagnosis of surface ozone sensitivity to anthropogenic emissions, *Geophys. Res. Lett.*, *31*(6), 2-5, doi:10.1029/2004GL019416.

Martin, R. V., D. D. Parrish, T. B. Ryerson, D. K. Nicks, K. V. Chance, T. P. Kurosu, D. J. Jacob, E. D. Sturges, A. Fried, and B. P. Wert (2004b), Evaluation of GOME satellite measurements of tropospheric NO_2 and HCHO using regional data from aircraft campaigns in the southeastern United States, *J. Geophys. Res.*, *109*(D24), 1-11, doi:10.1029/2004JD004869.

Martin, R. V. (2008), Satellite remote sensing of surface air quality, *Atmospheric Environment*, *42*(34), 7823-7843, doi:10.1016/j.atmosenv.2008.07.018.

McMillan, W. W., et al. (2005), Daily global maps of carbon monoxide: First views from NASA's Atmospheric Infrared Sounder. *Geophys. Res. Lett.*, 32, L11801, doi:10.1029/2004GL021821.

Meller, R., and G. K. Moortgat (2000), Temperature dependence of the absorption cross section of HCHO between 223 and 323K in the wavelength range 225–375 nm, *J. Geophys. Res.*, 105(D6), 7089–7102, doi:10.1029/1999JD901074.

Michel, C., C. Liousse, J.-M. Grégoire, K. Tansey, G.R. Carmichael and J.-H. Woo (2005), Biomass burning emission inventory from burnt area data given by the SPOT-VEGETATION system in the frame of the TRACE-P and ACE-Asia campaigns, *J. Geophys. Res*, 110, D09304, doi:10.1029/2004JD005461.

Mieville, A., C. Granier, C. Liousse, B. Guillaume, F. Mouillot, J. Lamarque, J. Grégoire, and G. Pétron (2010), Emissions of gases and particles from biomass burning during the 20th century using satellite data and an historical reconstruction, *Atmospheric Environment*, *44*(11), 1469-1477, doi:10.1016/j.atmosenv.2010.01.011.

Mijling, B., R. J. van Der A, K. F. Boersma, M. Van Roozendael, et al. (2009), Reductions of NO_2 detected from space during the 2008 Beijing Olympic Games, *Geophys. Res. Lett.*, *36*, doi: 10.1029/2009GL038943.

Mikaloff Fletcher, S. E., P. P. Tans, L. M. Bruhwiler, J. B. Miller, and M. Heimann (2004), CH4 sources estimated from atmospheric observations of CH4 and its C-13/C-12 isotopic ratios: 1. Inverse modeling of source processes, *Global Biogeochem. Cycles*, *18*(4).

Millet, D, 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 emissions, *J. Geophys. Res.*, 111, D24S02, doi:10.1029/2005JD006853.

Millet, D. B., D. J. Jacob, K. F. Boersma, T.-M. Fu, T. P. Kurosu, K. V. Chance, C. L. Heald, and 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*(D2), 1-18, doi:10.1029/2007JD008950.

Monks, P. (2005), Gas-phase radical chemistry in the troposphere, *Chem. Soc. Rev.*, 34, 376-395.

Müller, J.-F., and T. Stavrakou (2005), Inversion of CO and NOx emissions using the adjoint of the IMAGES model, *Atmos. Chem. Phys.*, 5, 1157–1186.

Müller, J.-F., T. Stavrakou, S. Wallens, I. De Smedt, M. Van Roozendael, M. J. Potosnak, J. Rinne, B. Munger, A. Goldstein, and A. Guenther (2008), Global isoprene emissions estimated using MEGAN, ECMWF analyses and a detailed canopy environment model, *Atmos. Chem. Phys.*, *8*(5), 1329-1341.

Munro, R., R. Siddans, W. J. Reburn, and B. Kerridge (1998), Direct measurement of tropospheric ozone from space, *Nature,* 392, 168–171.

Munro, R., M. Eisinger, C. Anderson, J. Callies, E. Corpaccioli, R. Lang, A. Lefebvre, Y. Livschitz, and A. P. Albinana (2006), GOME-2 on MetOp, *Proc. of The 2006 EUMETSAT Meteorological Satellite Conference, Helsinki, Finland, 12–16 June 2006, EUMETSAT p.48.*

Notholt, J., G. C. Toon, R. Lehmann, B. Sen, and J.-F. Blavier (1997), Comparison of Arctic and Antarctic trace gas column abundances from groud-based Fourier transform infrared spectrometry, *J. Geophys. Res.*, 102, 12863–12869.

Notholt, J., G. C. Toon, C. P.Rinsland, N. S.Pougatchev, N. B.Jones, B. J.Connor, R. Weller, M. Gautrois, and O. Schrems (2000), Latitudinal variations of trace gas concentrations measured by solar absorption spectroscopy during a ship cruise, *J. Geophys. Res.*, 105, 1337–1349.

Ohara, T., H. Akimoto, J. Kurokawa, N. Horii, *et al.* (2007), An Asian emission inventory of anthropogenic emission sources for the period 1980–2020, *Atmos. Chem. Phys.*, *7*, 4419-4444, http://www.jamstec.go.jp/frcgc/research/p3/index.html.

Olivier, J., A. F. Bouwman, C. W. M. Van der Maas, J. J. M. Berdowski, C. Veldt, J.-P. J. Bloos, A. J. H. Visschedijk, P. Y. Zandveld, and J. L. Haverlag (1996), Description of EDGAR Version 2.0: A set of emission inventories of greenhouse gases and ozone depleting substances for all anthropogenic and most natural sources on a per country basis and on 1° x 1° grid, *RIVM Rep. 771060002, Bilthoven: National Institute of Public Health and the Environment.*

Olivier, J., J. J. Berdowski, J. A. Peters, J. Bakker, A. J. Visschedijk, and J.-P. J. Bloos (2001), Applications of EDGAR. Including a description of EDGAR 3.0: reference database with trend data for 1970-1995, *RIVM Rep. 773301 001/NOP report no. 410200 051, Bilthoven: National Institute of Public Health and the Environment.*

Olivier, J., J. Peters, C. Granier, G. Pétron, J.-F. Müller, and S. Wallens (2003), Present and future surface emissions of atmospheric compounds, *POET Report #2, EU project EVK2-1999-00011.*

Olivier, J., J. A. Van Aardenne, F. Dentener, L. Ganzeveld and J. A. Peters (2005), Recent trends in global greenhouse gas emissions: regional trends and spatial distribution of key sources, in: "Non-CO2 Greenhouse Gases (NCGG-4)", A. van Amstel (coord.), page 325-330. *Millpress, Rotterdam, ISBN 90 5966 043 9.*

Olivier, J., T. Pulles and J. A. van Aardenne (2008), EDGAR data on non-CO_2 greenhouse gas, in CO2 emissions from fuel combustion 1971-2006, *Ed. International Energy Agency (IEA), Paris*, ISBN 978-92-64-04238-4.

Orphal, J. and K. Chance (2003), Ultraviolet and visible absorption cross-sections for HITRAN, *J. of Quantitative Spectroscopy & Radiative Transfer*, 82, 491–504, doi:10.1016/S0022-4073(03)00173-0.

Otter, L., A. Guenther, C. Wiedinmyer, G. Fleming, P. Harley, and J. Greenberg (2003), Spatial and temporal variations in biogenic volatile organic compound emissions for Africa south of the equator, *J. Geophys. Res.*, 108(D13), 8505, doi:10.1029/2002JD002609.

Pacyna, E. G., J. M. Pacyna, F. Steenhuisen, and S. Wilson (2006), Global anthropogenic mercury emission inventory for 2000, *Atmospheric Environment*, 40(22), 15, doi:10.1016/j.atmosenv.2006.03.04.

Palmer, P. I., D. J. Jacob, K. Chance, R. V. Martin, R. J. D. Spurr, T. P. Kurosu, I. Bey, R. Yantosca, A. Fiore, and Q. Li (2001), Air-mass factor formulation for spectroscopic measurements from satellites: application to formaldehyde retrievals from GOME, *J. Geophys. Res.*, 106(D13), 14539-14550, 10.1029/2000JD900772.

Palmer, P. I., D. J. Jacob, A. M. 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. Kurosu, A. Guenther, C. Wiedinmyer, J. Stanton, M. Pilling, S. Pressley, B. Lamb, and A. L. Sumner (2006), Quantifying the seasonal and interannual variability of North American isoprene emissions using satellite observations of formaldehyde column, *J. Geophys. Res..*, 111, D12315, doi:10.1029/2005JD006689.

Perrin, a, D. Jacquemart, F. Kwabia Tchana, and N. Lacome (2009), Absolute line intensities measurements and calculations for the 5.7 and 3.6μm bands of formaldehyde, *Journal of Quantitative Spectroscopy and Radiative Transfer*, 110(9-10), 700-716, doi:10.1016/j.jqsrt.2008.11.005.

Petty, G. W. (2006), A first course in atmospheric radiation, second edition, *Sundog Publishing, Madison, Wisconsin*.

Pfister, G. G., L. K. Emmons, P. G. Hess, J. Lamarque, J. J. Orlando, S. Walters, A. Guenther, P. I. Palmer, and P. J. Lawrence (2008), Contribution of isoprene to chemical budgets: A model tracer study with the NCAR CTM MOZART-4, *J. Geophys. Res..*, 113(D5).

Platt, U. (1994), Differential optical absorption spectroscopy (DOAS), in Air Monitoring by Spectroscopic Techniques, *M.W. Sigrist ed., Chemical Analysis Series, Wiley, New York*, 127, 27-84.

Platt, U and J. Stutz (2008), Differential Optical Absorption Spectroscopy: Principles and Applications (Physics of Earth and Space Environments), *Springer-Verlag, Berlin, Heidelberg,* ISBN 978-3540211938.

Pöschl, U., R. von Kuhlmann, N. Poisson, and P. J. Crutzen, (2000), Development and intercomparison of condensed isoprene oxidation mechanisms for global atmospheric modeling, *J. Atmos. Chem.*, 37, 29–52.

Prentice, I. C., Farquhar, G.D., M. J. R. Fasham, M. L. Goulden, M. Heimann, V. J. Jaramillo, H. S. Kheshgi, C. Le Quéré, R.J. Scholes, D.W.R. Wallace (2001), The Carbon Cycle and Atmospheric Carbon Dioxide, in Climate Change 2001: The Scientific Basis. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change [Houghton, J.T.,Y. Ding, D.J. Griggs, M. Noguer, P.J. van der Linden, X. Dai, K. Maskell, and C.A. Johnson (eds.)], *Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA*.

Razavi, A., C. Clerbaux, C. Wespes, L. Clarisse, D. Hurtmans, S. Payan, C. Camy-Peyret, and P. F. Coheur (2009), Characterization of methane retrievals from the IASI space-borne sounder, *Atmos. Chem. Phys.*, 9(20), 7889-7899.

Razavi, A., F. Karagulian, L. Clarisse, D. Hurtmans, P. F. Coheur, C. Clerbaux, J.-F. Müller, and T. Stavrakou (2011), Global distributions of methanol and formic acid retrieved for the first time from the IASI/MetOp thermal infrared sounder, *Atmos. Chem. Phys.*, 11, 857-872, doi:10.5194/acp-11-857-2011.

Richter, A., F. Wittrock, M. Eisinger, and J. P. Burrows (1998), GOME observations of tropospheric BrO in northern hemispheric spring and summer 1997, *Geophys. Res. Lett.*, 25, 2683-2686.

Richter, A. and T. Wagner (2001), Diffuser plate spectral structures and their influence on GOME slant columns, *Tech. rep.*, *Inst. Env. Phys., Univ. Bremen and Inst. Env. Phys., Univ. Heidelberg*.

Richter, A. and J. P. Burrows (2002), Tropospheric NO2 from GOME measurements, *Adv. Space Res.*, 29, 1673-1683.

Richter, A., J.P. Burrows, H. Nüß, C. Granier, and U. Niemeier (2005), Increase in tropospheric nitrogen dioxide over China observed from space, *Nature*, 04092(437), doi:10.1038.

Rigby, M. et al. (2008), Renewed growth of atmospheric methane, *Geophysical Research Letters*, *35*(22), doi:10.1029/2008GL036037.

Richter, A., M. Begoin, A. Hilboll, and J. P. Burrows (2011), An improved NO_2 retrieval for the GOME-2 satellite instrument, *Atmos. Meas. Tech. Discuss.*, *4*(1), 213-246.

Rodgers, C. D. (2000), Inverse Methods for Atmospheric Sounding, Theory and Practice, *World Scientific Publishing, Singapore-New-Jersey-London-Hong Kong*.

Rodgers, C. D., and B. J. Connor (2003), Intercomparison of remote sounding instruments, *J. Geophys. Res.*, *108*, doi:10.1029/2002JD002299.

Rozanov, A., Rozanov, V., and Burrows, J. P. (2001), A numerical radiative transfer model for a spherical planetary atmosphere: Combined differential integral approach involving the Piccard iterative approximation, *J. Quant. Spectrosc. Radiat. Transfer*, 69, 491–512.

Sander, S. P. et al. (2006), Chemical Kinetics and Photochemical Data for Use in Atmospheric Studies. Evaluation Number 15, *National Aeronautics and Space Administration, Jet Propulsion Laboratory, California Institute of Technology*.

Sandu, A., and R. Sander (2006), Technical note: Simulating chemical systems in Fortran90 and Matlab with the Kinetic PreProcessor KPP-2.1, *Atmos. Chem. Phys.*, *6*(1), 187-195.

Saunders, S. M., M. E. Jenkin, R. G.Derwent, and M. J. Pilling (2003), Protocol for the development of the Master Chemical Mechanism, MCM v3 (Part A): tropospheric degradation of non-aromatic volatile organic compounds, *Atmos. Chem. Phys.*, 3, 161-180.

Schneising, O., M. Buchwitz, J. P. Burrows, H. Bovensmann, M. Reuter, J. Notholt, R. Macatangay, and T. Warneke (2008a), Three years of greenhouse gas column-averaged dry air mole fractions retrieved from satellite - Part 1: Carbon dioxide, *Atmos. Chem. Phys.*, 8, 3827-3853.

Schneising, O., M. Buchwitz, J. P. Burrows, H. Bovensmann, P. Bergamaschi, and W. Peters (2008b), Three years of greenhouse gas column-averaged dry air mole fractions retrieved from satellite - Part 2: Methane, *Atmos. Chem. Phys. Discuss.*, 8, 8273-8326.

Schultz, M., S. Rast, M. van het Bolscher, T. Pulles, J. Pereira, A. Spessa, S. Dalsøren, T. van Nojie, S. Szopa, S. (2007), REanalysis of the TROpospheric chemical composition over the past 40 years, a long-term global modeling study of tropospheric chemistry funded under the 5th EU framework programme, http://www.retro.enes.org/reports/D1-6 final.pdf, *Tech. rep., EU-Contract No. EVK2-CT-2002-00170*.

Seinfeld, J. H. and S. N. Pandis (2006), Atmospheric Chemistry and Physics: From air pollution to climate change, *second edition, John Wiley and Sons, New-York*.

Senten, C., M. De Mazière, B. Dils, C. Hermans, M. Kruglanski, E. Neefs, F. Scolas, A. C. Vandaele, G. Vanhaelewyn, C. Vigouroux, M. Carleer, P. F. Coheur, S. Fally, B. Barret, et al. (2008), Technical Note:

New ground-based FTIR measurements at Ile de La Réunion: observations, error analysis, and comparisons with independent data, *Atmos. Chem. Phys.*, 8, 3483–3508.

Shim, C., Y. Wang, Y. Choi, P. I. Palmer, D. S. Abbot, and K. Chance (2005), Constraining global isoprene emissions with Global Ozone Monitoring Experiment (GOME) formaldehyde column measurements, *J. Geophys. Res.*, 110, D24301, doi:10.1029/2004JD005629.

Shindell, D. T. et al. (2006), Multimodel simulations of carbon monoxide: Comparison with observations and projected near-future changes, *J. Geophys. Res.*, *111*(D19), D19306.

Siddans, R., B. G. Latter, B. J. Kerridge, J. Smeets, G. Otter, S. and Slijkhuis (2006), Analysis of GOME-2, Slit function Measurements, Final Report, *Eumetsat Contract No. EUM/CO/04/1298/RM*.

Singh, H. B., W. H. Brune, J. H. Crawford, D. J. Jacob, and P. B. Russell (2006), Overview of the summer 2004 Intercontinental Chemical Transport Experiment–North America (INTEX-A), *J. Geophys. Res.*, 111, D24S01, doi:10.1029/2006JD007905.

Slijkhuis, S. and G. Lichtenberg (2008), ENVISAT-1 SCIAMACHY ATBD, Doc.No.: ENV-ATB-DLR-SCIA-0041, Issue: 5, Date: 20.5.08.

Smith Sr., W. L., H. Revercomb, G. Bingham, A. Larar, H. Huang, D. Zhou, J. Li, X. Liu, and S. Kireev (2009), Technical Note: Evolution, current capabilities, and future advance in satellite nadir viewing ultra-spectral IR sounding of the lower atmosphere, *Atmos. Chem. Phys.*, *9*(15), 5563-5574.

Solomon, S., A. L. Schmeltekopf, and R. W. Sanders (1987), On the interpretation of zenith sky absorption measurements, *J. Geophys. Res.*, 92, 8311-8319.

Spurr, R. (2001), Linearized radiative transfer theory, A general discrete ordinate approach to the calculation of radiances and analytic weighting functions, with applications to atmospheric remote sensing, thesis manuscript, proefschrift aan de *Technische Universiteit Eindhoven*.

Spurr, R. (2008), LIDORT and VLIDORT: Linearized pseudo-spherical scalar and vector discrete ordinate radiative transfer models for use in remote sensing retrieval problems, *Light Scattering Reviews, 3*, 229–271, A. Kokhanovsky, Berlin, Springer.

Stavrakou, T., and 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*(D15), D15304.

Stavrakou, T., J.-F. Müller, I. De Smedt, M. Van Roozendael, G. van der Werf, L. Giglio (2007), One decade of pyrogenic NMVOC emissions deduced from HCHO satellite data, *Proceedings of the 2nd ACCENT Symposium*, Urbino, July 2007.

Stavrakou, T., J.-F. Müller, K. F. Boersma, I. De Smedt, and R. van der A (2008), Assessing the distribution and growth rates of NOx emission sources by inverting a 10-year record of NO_2 satellite columns, *Geophys. Res. Lett.*, *35*(10), 2-6, doi:10.1029/2008GL033521.

Stavrakou, T., J.-F. Müller, I. De Smedt, M. Van Roozendael, *et al.* (2009a), Evaluating the performance of pyrogenic and biogenic emission inventories against one decade of space-based formaldehyde columns, *Atmos. Chem. Phys., 9*, 1037–1060.

Stavrakou, T., J.-F. Müller, I. De Smedt, M. Van Roozendael, *et al.* (2009b), Global emissions of non-methane hydrocarbons deduced from SCIAMACHY formaldehyde columns through 2003–2006, *Atmos. Chem. Phys., 9*, 3663–3679.

Stavrakou, T., J.-F. Müller, I. De Smedt, M. Van Roozendael, M. Kanakidou, M. Vrekoussis, F. Wittrock, A. Richter, and J. P. Burrows (2009c), The continental source of glyoxal estimated by the synergistic use of spaceborne measurements and inverse modelling, *Atmos. Chem. Phys.*, *9*(21), 8431-8446.

Stavrakou, T., J. Peeters, and J.-F. Müller (2010), Improved global modelling of HOx recycling in isoprene oxidation: evaluation against the GABRIEL and INTEX-A aircraft campaign measurements, *Atmos. Chem. Phys.*, *10*(20), 9863-9878.

Stavrakou, T., A. Guenther, A. Razavi, L. Clarisse, C. Clerbaux, P.-F. Coheur, D. Hurtmans, F. Karagulian, M. De Mazière, C. Vigouroux, C. Amelynck, N. Schoon, Q. Laffineur, B. Heinesch, M. Aubinet, J.-F. Müller (2011), First space-based derivation of the global atmospheric methanol emission fluxes, *Atmos. Chem. Phys. Disc.*, *11*(2), 5217-5270, doi:10.5194/acpd-11-5217-2011.

Steck, T. *et al.* (2008), Retrieval of global upper tropospheric and stratospheric formaldehyde (H_2CO) distributions from high-resolution MIPAS-Envisat spectra, *Atmos. Chem. Phys.*, *8*(3), 463-470, doi:10.5194/acp-8-463-2008.

Stevenson, D. S. *et al.* (2006), Multimodel ensemble simulations of present-day and near-future tropospheric ozone, *J. Geophys. Res.*, *111*(D8), D08301.

Susskind, J., C. D. Barnet, and J. M. Blaisdell (2003), Retrieval of atmospheric and surface parameters from AIRS/AMSU/HSB data in the presence of clouds, *IEEE Trans. Geosci. Remote Sens.*, 41, 2, doi:10.1109/ TGRS.2002.808236.

Tilstra, L. G. and P. Stammes (2005), Alternative polarisation retrieval for SCIAMACHY in the ultraviolet, *Atmos. Chem. Phys.*, 5, 2099-2107, doi:10.5194/acp-5-2099-2005.

Taylor, J.R. (1982), An introduction to error analysis, *University Science Books, Oxford University Press*.

Theys, N, M. Van Roozendael, F. Hendrick, C. Fayt, C. Hermans, J.-L. Baray, F. Goutail, J.-P. Pommereau, and M. De Mazière (2007), Retrieval of stratospheric and tropospheric BrO columns from multi-axis DOAS measurements at Reunion Island (21°S, 56°E), *Atmos. Chem. Phys.*, 7, 4733–4749.

Theys, N. (2010), Atmospheric Bromine Monoxide: multi-platform observations and model calculations, *PhD Thesis, Université Libre de Bruxelles*.

Theys, N., M. Roozendael, F. Hendrick, X. Yang, I. De Smedt, A. Richter, M. Begoin, Q. Errera, P. V. Johnston, K. Kreher, and M. De Mazière (2011), Global observations of tropospheric BrO columns using GOME-2 satellite data, *Atmos. Chem. Phys.*, *11*(4), 1791-1811.

Thomas, W., E. Hegels, S. Slijkhuis, R. Spurr, and K. Chance (1998), Detection of biomass burning combustion products in Southeast Asia from backscatter data taken by the GOME spectrometer, *Geophys. Res. Lett.*, 25, 1317– 1320.

Turquety, S., D. Hurtmans, J. Hadji-Lazaro, P.-F. Coheur, C. Clerbaux, D. Josset, and C. Tsamalis (2009), Tracking the emission and transport of pollution from wildfires using the IASI CO retrievals: analysis of the summer 2007 Greek fires, *Atmos. Chem. Phys.*, *9*(14), 4897-4913.

Uno, I., Y. He, T. Ohara, K. Yamaji, *et al.* (2007), Systematic analysis of interannual and seasonal variations of model-simulated tropospheric NO_2 in Asia and comparison with GOME-satellite data, *Atmos. Chem. Phys., 7*, 1671-1681.

U.S. Standard Atmosphere (1976), *U.S. Government Printing Office, Washington, D.C..*

Van Aardenne, J. A., F. J. Dentener, J. Olivier, C. G. Klein Goldewijk and J. Lelieveld (2001), A 1x1 degree resolution dataset of historical anthropogenic trace gas emissions for the period 1890-1990, *Global Biogeochem. Cycles*,15(4), 909-928.

Vandaele, A.C., et al. (2002), High-resolution Fourier transform measurement of the NO_2 visible and near-infrared absorption cross-section: Temperature and pressure effects, *J. Geophys. Res.*, 107, D18, 4348, doi:10.1029/2001JD000971.

van der A, R. J., D. H. 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 2 over China, *J. Geophys. Res.*, 111, D12317, doi:10.1029/2005JD006594.

van der A, R.J., H.J. Eskes, K.F. Boersma, T.P. van Noije, *et al.* (2008), Trends, seasonal variability and dominant NOx source derived from a ten year record of NO$_2$ measured from space, *J. Geophys. Res., 113*, D04302, doi: 10.1029/2007JD009021.

van der Werf, G. R., J. T. Randerson, G. J. Collatz, L. Giglio, P. S. Kasibhatla, A. F. Arellano Jr., S. C. Olsen, and E. S. Kasischke (2004), Continental scale partitioning of fire emissions during the 1997 to 2001 El Nino/La Nina period, Science, 303, 73-76.

van der Werf, G.R., J. T. Randerson, L. Giglio, G. J. Collatz, P. S. Kasibhatla, A. F. Arellano Jr. (2006), Interannual variability in global biomass burning emissions from 1997 to 2004, *Atmos. Chem. Phys.*, 6, 3423-3441.

Van Roozendael, M., T. Wagner, A. Richter, *et al.* (2002a), Intercomparison of BrO measurements from ERS-2 GOME, ground-based and balloon platforms, *Adv. Space Res.*, 29, 1661-1666.

Van Roozendael, M., V. Soebijanta, C. Fayt, and J.-C. Lambert (2002b), Investigation of DOAS Issues Affecting the Accuracy of the GDP Version 3.0 Total Ozone Product, in ERS-2 GOME GDP 3.0 Implementation and Delta Validation, *ERSE-DTEX-EOAD-TN-02-0006, ESA/ESRIN, Frascati, Italy*, Chap.6, pp.97-129.

Van Roozendael, M. et al. (2006), Ten years of GOME/ERS-2 total ozone data - The new GOME data processor (GDP) version 4: 1. Algorithm description, *J. Geophys. Res., 111*(D14), D14311.

Vigouroux, C., F. Hendrick, T. Stavrakou, B. Dils, I. De Smedt, C. Hermans, A. Merlaud, F. Scolas, C. Senten, G. Vanhaelewyn, S. Fally, M. Carleer, J.-M. Metzger, J.-F. Müller, M. Van Roozendael, and M. De Mazière (2009), Ground-based FTIR and MAX-DOAS observations of formaldehyde at Réunion Island and comparisons with satellite and model data, *Atmos. Chem. Phys.*, 9, 9523-9544, doi:10.5194/acp-9-9523-2009.

Vountas, M., V. V. Rozanov, and J. P. Burrows (1998), Ring effect: Impact of rotational Raman scattering on radiative transfer in earth's atmosphere, *J. Quant. Spectrosc. Radiat. Transfer.*, 60, 943-961.

Vrekoussis, M., F. Wittrock, A. Richter, and J. P. Burrows (2010), GOME-2 observations of oxygenated VOCs: what can we learn from the ratio glyoxal to formaldehyde on a global scale?, *Atmos. Chem. Phys.*, *10*(21), 10145-10160.

Wahner, A., A.R. Ravishankara, S.P.Sander, and R.R.Friedl (1988), Absorption cross section of BrO between 312 and 385 nm at 298 and 223K, Chem. Phys. Lett., 152, 507-512.

Wagner, T. (1999), Satellite Observations of Atmospheric Halogen Oxides, *PhD Thesis, Heidelberg University*.

Wagner, T., C. von Friedeburg, M. Wenig, C. Otten, and U. Platt (2002), UV-visible observations of atmospheric O4 absorptions using direct moonlight and zenith-scattered sunlight for clear-sky and cloudy sky conditions, *J. Geophys. Res.*, *107*(D20), 4424.

Wallens, S. (2004), Modélisation des émissions de composés organiques volatils par la végétation, *PhD Thesis, Université Libre de Bruxelles.*

Wang, P., P. Stammes, R. van der A, G. Pinardi, and M. Van Roozendael (2008), FRESCO+: an improved O$_2$ A-band cloud retrieval algorithm for tropospheric trace gas retrievals, *Atmos. Chem. Phys., 8*, 6565-6576.

Wang, T., W. Nie, J. Gao, L. K. Xue, *et al.* (2010), Air quality during the 2008 Beijing Olympics: secondary pollutants and regional impact, *Atmos. Chem. Phys*, 10, 7603–7615.

Warneke, C., S. McKeen, and J. A. de Gouw et al. (2007), Determination of Biogenic Emissions from Aircraft Measurements During TEXAQS2006 and ICARTT2004 Campaigns and Comparison with Biogenic Emission Inventories, Eos Trans. AGU, 88(52), *Fall Meet. Suppl., Abstract A14D-05*.

Weatherhead, E. C., *et al.* (1998), Factors affecting the detection of trends: Statistical considerations and applications to environmental data, *J. Geophys. Res.*, *103*, 17.149–17.161.

Wittrock, F., A. Richter, A. Ladstätter-Weißenmayer, and J.P. Burrows (2000), Global observations of formaldehyde, Proceedings of the ERS-ENVISAT symposium, ESA publication SP-461.

Wittrock, F., H. Oetjen, A. Richter, S. Fietkau, T. Medeke, A. Rozanov, and J. P. Burrows (2004) MAX-DOAS measurements of atmospheric trace gases in Ny-Alesund – Radiative transfer studies and their application, *Atmos. Chem. Phys.*, 4, 955–966, doi:10.5194/acp-4-955-2004.

Wittrock, F. (2006), The Retrieval of oxygenated volatile organic compounds by remote sensing techniques, *PhD Thesis, Bremen University*.

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.

World Meteorological Organization (1998), Scientific assessment of ozone depletion: 1998, *WMO, Geneva, Switzerland*.

World Meteorological Organization (2002), Scientific assessment of ozone depletion: 2002, *WMO, Geneva, Switzerland*.

Zhang, Q., D.G. Streets, and K. He (2009), Satellite observations of recent power plant construction in Inner Mongolia, China, *Geophys. Res. Lett., 36*, L15809, doi: 10.1029/2009GL038984.

Zhou, Y., D. Brunner, K.F. Boersma, R. Dirksen, and P. Wang (2009), An improved tropospheric NO_2 retrieval for OMI observations in the vicinity of mountainous terrain, *Atmos. Meas. Tech., 2,* 401-416.

Zhou, Y., D. Brunner, R. J. Spurr, K. F. Boersma, M. Sneep, C. Popp, and B. Buchmann (2010), Accounting for surface reflectance anisotropy in satellite retrievals of tropospheric NO_2, *Atmos. Meas. Tech.*, *3*(5), 1185-1203.

Ziemke, J. R., S. Chandra, and P. K. Bhartia (2001), Cloud slicing: a new technique to derive upper tropospheric ozone from satellite measurements, *J. Geophys. Res.,* 106, 9853–9867.

List of acronyms

ADEOS: Advanced Earth Observing System

ACE: Atmospheric Chemistry Experiment

AIRS: Atmospheric IR Sounder

AMF: Air Mass Factor

AMFIC: Air Quality Monitoring and Forecasting in China

AOD: Aerosol optical depth

BIRA-IASB: Belgisch Instituut voor Ruimte-Aëronomie - Institut d'Aéronomie Spatiale de Belgique

BUV: Backscatter Ultraviolet

IUP: Institut für Umweltphysik (Institute of Environmental Physics), University of Bremen

CCD: charge coupled device

CEIP: Centre on Emission Inventories and Projections

CNES: Centre National d'Études Spatiales

CSA : Canadian space agency

CTM: Chemistry transport model

DLR: Deutsches zentrum für Luft und Raumfahrt (German Aerospace Center)

DOAS: Differential Optical Absorption Spectroscopy

ECMWF: European Centre for Medium-Range Weather Forecasts

EDGAR: Emission Database for Global Atmospheric Research

EMEP: European Monitoring and Evaluation Program

ENVISAT: Environmental Satellite

EPA: US Environmental Protection Agency

EPS: European Polar System

ERA-40: Forty-Year European Re-Analysis

ERS: European Research Satellite

ESA: European Space Agency

EOS: Earth Observing System

EUMETSAT: European Organisation for the Exploitation of Meteorological Satellites

FRESCO: Fast Retrieval Scheme for Cloud Observables

FTIR: Fourier Transform Spectroscopy

GEIA: Global Emissions Inventory Activity

GEMS: Global and regional Earth-system (Atmosphere) Monitoring using Satellite and in-situ data

GEO: Geostationary Orbit

GeoSCIA: Geostationary Scanning Imaging Absorption Spectrometer

GEOSS: Global Earth Observation System of Systems

GeoTROPE: Geostationary Tropospheric Explorer

GFED: Global Fire Emissions Database

GMES: Global Monitoring for Environment and Security

GOME: Global Ozone Monitoring Experiment

GOMOS: Global ozone monitoring by occultation of stars

GOSAT: Greenhouse gases Observing Satellite

HALOE: Halogen Occultation Experiment

IASI: Interféromètre atmosphérique de sondage dans l'infrarouge

IGAC: International Global Atmospheric Chemistry Project

IGACO: Integrated Global Atmospheric Chemistry Observation

IMAGES: Intermediate Model for the Annual and Global Evolution of Species

IMF: Institut für Methodik der Fernerkundung (Remote Sensing Technology Institute)

IMG: Interferometric monitor for greenhouse gases

IPCC: Intergovernmental Panel on Climate Change

IUPAC: International Union of Pure and Applied Chemistry

IRS: InfraRed Sounder

JAXA: Japan Aerospace Exploration Agency

JPL ATMOS: Jet Propulsion Laboratory Atmospheric Trace Molecule Spectrometry Experiment

KNMI: Royal Netherlands Meteorological Institute

LEO: Low Earth Orbit

MACC: Monitoring Atmospheric Composition and Climate

MAX-DOAS: Multi-AXis DOAS

MEGAN: Model of Emissions of Gases and Aerosols from Nature

METEOSAT: Meteorological Satellites

MIPAS: Michelson inferometer for passive atmospheric sounding

MLS: Microwave Limb Sounder

MOPITT: Measurements Of Pollution In The Troposphere

MTG: METEOSAT Third Generation

NASA: National Aeronautics and Space Administration

NASDA: National Space Development Agency of Japan

NDACC: Network for the Detection of Atmospheric Composition Change

NEI: National Emissions Inventories

NEC: National Emission Ceilings

NIR: Near infra-red

NMVOC: non methane volatile organic compound

NOAA: National Oceanic and Atmospheric Administration

NPP: NPOESS Preparatory Project

NPOESS: National Polar-orbiting Operational Environmental Satellite

OCO: Orbiting Carbon Observatory

OH: hydroxyl radical

OMI: Ozone Monitoring Instrument

OMPS: Ozone Monitoring and Profiling Suite

OSIRIS: Optical Spectrograph and Infra-Red Imager System

PMD: Polarization Measurement Device

PROMOTE: Protocol Monitoring for the GMES Service Element

RASA: Russian Aviation and Space Agency

REAS: Regional Emission Inventory in Asia

RMS: Root Mean Square

SAA: Southern Atlantic Anomaly

SAGE: Stratospheric Aerosol and Gas Experiment

SBUV: Solar Backscatter Ultraviolet

SC: Slant Column

SCIAMACHY: Scanning Imaging Absorption SpectroMeter for Atmospheric CHartographY

SFM: Spectrophotometer

SNSB: Swedish National Space Board

SNR: Signal-to-Noise Ratio

SME: Solar Mesospheric Explorer

SOA: secondary organic aerosols

SWIR: Short-wave infra-red

SZA: Solar Zenith Angle

TANSO-FTS: Thermal And Near infrared Sensor for carbon Observation

TEMIS: Tropospheric Emission Monitoring Internet Service

TES: Tropospheric Emission Spectrometer

TIR: Themal infra-red

TOMS: Total Ozone Mapping Spectrometer

TROPOMI: Tropospheric Ozone Monitoring Instrument

UARS: Upper Atmosphere Research Satellite

UVN: Ultraviolet Visible Near-infrared spectrometer

VC: Vertical Column

VMR: Volume Mixing Ratio

VOC: volatile organic compound

WMO: World Meteorological Organisation

Chemical species

CFC: chlorofluorocarbon

CH_3: methyl radical

CH_3O: methoxy radical

CH_3O_2: methyl peroxy radical

CH_3OH: methanol

CH_3OOH: methyl hydroperoxide

CH_4: methane

C_3H_6: propene

C_5H_8: isoprene

CO: carbon monoxide

CO_2: carbon dioxide

DMS: dimethyl sulfide

HCO: formyl radical

H_2: hydrogen

H_2CO: formaldehyde

H_2O_2: hydrogen peroxide

HCFC: hydrochlorofluorocarbons

HCOOH: formic acid

HNO_3: nitric acid

HO_2: hydroperoxyl radical

N: azote

N_2: diazote or nitrogen

NH_3: ammonia

NMVOC: non methane volatile organic compound

NO: nitrogen monoxide

NO_2: nitrogen dioxide

NO_3: nitrate radical

N_2: nitrogen

N_2O: dinitrogen monoxide or nitrous oxide

N_2O_3: dinitrogen trioxide or nitrous anhydride

N_2O_5: dinitrogen pentoxide or nitric anhydride

O_2: oxygen

O_3: ozone

OH: hydroxyl radical

PAN: peroxyacyl nitrate

RO_2: alkyl peroxy radical

SO_2: sulphur dioxide

VOC: volatile organic compound

Volatile organic species in the atmosphere [*Seinfeld and Pandis,* 2006]

Type of compound	General chemical formula	Examples
Alkanes	$R{-}H$	CH_4, methane CH_3CH_3, ethane
Alkenes	$R_1C{=}CR_2$	$CH_2{=}CH_2$, ethene $CH_3{-}CH{=}CH_2$, propene
Alkynes	$RC{\equiv}CR$	$HC{\equiv}CH$, acetylene
Aromatics	C_6R_6 (cyclic)	C_6H_6, benzene $C_6H_5(CH_3)$, toluene
Alcohols	$R{-}OH$	CH_3OH, methanol CH_3CH_2OH, ethanol
Aldehydes	$R{-}CHO$	HCHO or H_2CO, formaldehyde CHOCHO, glyoxal CH_3CHO, acetaldehyde
Ketones	RCOR	$CH_3C(O)CH_3$, acetone
Peroxides	$R{-}OOH$	CH_3OOH, methyl hydroperoxide
Organic acids	$R{-}COOH$	HCOOH, formic acid CH_3COOH, acetic acid
Organic nitrates	$R{-}ONO_2$	CH_3ONO_2, methyl nitrate $CH_3CH_2ONO_2$, ethyl nitrate
Alkyl peroxy nitrates	RO_2NO_2	$CH_3O_2NO_2$, methyl peroxynitrate
Acylperoxy nitrates	$R{-}C(O)OONO_2$	$CH_3COOONO_2$, peroxyacetyl nitrate (PAN)
Biogenic compounds	C_5H_8 $C_{10}H_{16}$	$CH_2{=}C(CH_3){-}CH{=}CH_2$, isoprene α-pinene, β-pinene (monoterpenes)

Notes