The variability of stratospheric ozone in a 29 year assimilated data set and sensitivity calculations

Dissertation zur Erlangung des akademischen Grades

Dr. rer. nat.

der Universität Bremen
Fachbereich 1 – Physik und Elektrotechnik

Vorgelegt von

Mag. rer. nat. Gregor Kiesewetter

Gutachter:

Prof. Dr. John P. Burrows

Prof. Dr. Justus Notholt

Eingereicht am 30. 5. 2011

Kolloquium am 12. 10. 2011

Abstract

Consistent observation-based data sets of stratospheric ozone are needed in order to resolve many of the pending questions regarding stratospheric ozone. Satellite observations are available since the late 1970s; however, as most observational methods rely on backscattered sunlight, these do not provide complete long-term coverage of the stratosphere, in particular during polar night. In this PhD thesis, a 29 year data set of stratospheric ozone is introduced that has been generated from sequential assimilation of satellite observations into the Bremen 3D Chemistry Transport Model (CTM). In the method of data assimilation, a three-dimensional physical computer model is used to close the gaps between single measurements. Observations constrain the CTM where available, and at the same time the information is propagated into areas where no observations are available.
Here, profile ozone observations from the Solar Backscatter UV (SBUV and SBUV/2) instruments are used, which have been in orbit continuously since 1978. The resulting assimilated data set is validated against independent observations from other satellite platforms and in-situ observations with sondes. Agreement to independent observations is excellent throughout most of the stratosphere, and the assimilated data set can thus be used as a consistent extension of the satellite record beyond the limits of data coverage.
The assimilated data set, in conjunction with sensitivity calculations with the unconstrained CTM, is used to analyse the variability of stratospheric ozone during the last three decades on two distinctly different temporal and spatial domains.
The first research question deals with the short-term variability of polar ozone during winter. The Arctic ozone layer is subject to large inter-annual variations during spring; although statistical connections between dynamical quantities in winter and springtime total ozone abundance are known, little is known about how ozone anomalies develop and evolve in winter. With its coverage of polar latitudes during winter, the assimilated data set is ideally suited to address this issue. It is shown that ozone anomalies usually originate in the mid- to upper stratosphere and subsequently descend to the lower stratosphere, displaying a long lifetime of around six months. Ozone anomalies are strongly interrelated to anomalies in the stratospheric circulation, expressed here by the Northern Hemisphere Annular Mode (NAM). Extreme phases of the NAM, so-called strong and weak vortex events, lead to the formation of large and distinctively shaped ozone anomalies that traverse most of the stratosphere within days to weeks, and subsequently remain significant for five months in the lowermost stratosphere. A deeper analysis reveals that different mechanisms of interaction between chemistry and dynamics lead to the observed ozone anomaly pattern. Another source of mid-stratospheric Arctic ozone anomalies is the precipitation of energetic particles from the sun. Solar proton events lead to the formation of nitrogen oxides in the upper stratosphere and mesosphere that subsequently propagate down into the polar vortex and cause significant negative ozone anomalies lasting for up to six months in the assimilated data set.
In a second research question, the long-term evolution of ozone is analysed on a global basis. Stratospheric ozone showed large declines during the 1980s and 1990s in both hemispheres.

As a consequence of the regulation of ozone depleting substances (ODSs) by the Montreal Protocol in 1987, stratospheric chlorine loadings have peaked in the late 1990s and since begun to slowly decrease. A levelling off of negative ozone trends has been detected, raising the question whether this is already an onset of chemical recovery. The long-term evolution of column ozone is captured very well in the assimilated data set and the unconstrained CTM. Sensitivity calculations with the unconstrained CTM allow for a diligent attribution of observed ozone trends to their processes of origin. In particular, the relative contributions of anthropogenic (emissions of ODSs) and natural (changes in stratospheric circulation and temperature) factors are quantified. While a large part of ozone decreases in the 1980s and 1990s is attributable to ODS increases, only very small effects of chemical recovery are seen after the ODS turnaround. A significant trend change is observed between the phases of increasing and decreasing ODS loadings; however, this trend change is partly related to dynamics, and hence cannot yet be taken as significant evidence for an onset of ozone recovery in most regions.

Contents

Part I.

Introduction

1. Introduction

During recent decades, stratospheric ozone has been one of the most important research topics in atmospheric and environmental science, in particular since the discovery of the Antarctic ozone hole by Farman et al. (1985). Due to the strong absorption of solar ultraviolet (UV) radiation, the ozone layer acts as a protective shield against harmful UV radiation at the earth surface and is thus essential to life on Earth in its current form.
Ozone research dates back to 1839, when the German chemist C.F. Schönbein discovered a gas with a distinct smell in the laboratory and named it *ozone*, from the Greek word $o\zeta\epsilon\iota\nu$ (ozein, to smell). Soon it was suggested that ozone was formed from three oxygen atoms. Measurements of ozone in the atmosphere began shortly after its discovery, first by chemical analysis (starting in 1853 by Schönbein in Vienna, Austria), and later by spectroscopic techniques.Postulated by W. N. Hartley in 1880, the ozone layer was discovered in 1913 by the French physicists C. Fabry and H. Buisson in the stratosphere, the atmospheric region extending from roughly 10 to 45 km altitude. A first chemical theory of the formation and destruction of stratospheric ozone from oxygen was developed by Chapman (1930), which was accepted to be a complete description of ozone chemistry for about two decades. As laboratory measurements of reaction kinetics improved, it became obvious that Chapman's theory over-predicted stratospheric ozone concentrations by almost a factor of two. Consequently, additional mechanisms for ozone destruction were searched and found in the form of catalytic reaction cycles involving hydrogen compounds (Bates and Nicolet, 1950; Hunt, 1966), nitrogen oxides (Crutzen, 1970; Johnston, 1971), and chlorine (Stolarski and Cicerone, 1974; Cicerone et al., 1974; Molina and Rowland, 1974) as well as bromine (Wofsy et al., 1975) containing compounds.
Concerns first rose in the 1970s about different possibilities in which anthropogenic emissions could destroy the ozone layer, and the consequences for human health (Cutchis, 1974). While at first catalytic destruction of ozone through nitric oxides was considered (Crutzen, 1970; Johnston, 1971), it soon became obvious that the larger threat to the ozone layer came from chlorofluorocarbons (CFCs), which were widely used as chemically inert refrigerants (Molina and Rowland, 1974; Cicerone et al., 1974). Changes, however, were predicted to happen on a slow timescale – until the completely unexpected discovery of the Antarctic ozone hole by Farman et al. (1985). In view of recurring dramatic ozone losses in Antarctic spring, the international community acted quickly and banned the production and release of CFCs in the Montreal Protocol in 1987 and subsequent amendments. The discovery of the ozone hole triggered a huge increase in scientific efforts, and soon the chemical processes were understood which lead to the release of active chlorine from otherwise inert reservoir species during polar night and consequently to the rapid destruction of ozone in Antarctic spring (Solomon et al., 1986; Molina et al., 1987; Solomon, 1990). For their discoveries regarding ozone chemistry, Paul Crutzen, Mario Molina, and Sherwood Rowland were awarded the Nobel Prize in Chemistry in 1995.
As a consequence of the Montreal Protocol and its amendments, CFC emissions were phased out, and stratospheric concentrations of ozone depleting substances peaked in the late

1990s (Waugh et al., 2001). As a result, stratospheric ozone is expected to undergo gradual recovery in the course of the 21st century (Newman et al., 2006). Consequently, intense research has focused on a search for significant trend changes (Newchurch et al., 2003; Reinsel et al., 2005; Yang et al., 2008), as well as a prediction of the future evolution of the ozone layer (e.g., Eyring et al., 2007).

Stratospheric ozone continues to reside high on the international research agenda (documented by the quadrennial Scientific Assessment of Ozone Depletion, e.g., World Meteorological Organization, 2011). During the 1990s, depletion of Arctic ozone came into focus, as ozone destruction on scales similar to the Antarctic was detected in cold Arctic winters (Newman et al., 1997; Fioletov et al., 1997; Sinnhuber et al., 2000; Rex et al., 2002), whereas warm winters exhibit only very little heterogeneous ozone destruction. Incidentally, just at the time of writing of this thesis the world is witnessing what is probably the largest springtime ozone depletion in the Arctic ever detected. Several aspects regarding the large inter-annual variability of polar ozone are still not fully understood. While statistical relations between total ozone in spring and meteorological conditions during the preceding winter are known (Weber et al., 2003), little is understood about the development of ozone anomalies in the polar middle stratosphere during winter from a low-variability state in summer and autumn (Sinnhuber et al., 2006).

During recent years, the role of ozone as an important coupling element between chemistry and dynamics has been increasingly recognised, opening the wide field of research on chemistry-dynamics interactions in the stratosphere. Influences of dynamics on ozone are numerous – e.g., temperatures in the polar lower stratosphere determine the amount of ozone depletion during spring (Chipperfield and Jones, 1999; Newman et al., 2001; Weber et al., 2003). On longer timescales, stratospheric temperatures are expected to decrease significantly due to anthropogenic inputs of greenhouse gases (contrary to the troposphere where these gases induce a global warming effect) (Haigh and Pyle, 1979), an effect which may initially delay the recovery of polar ozone (Newman et al., 2006), but in the end may well lead to an over-recovery of global ozone (Eyring et al., 2007; Shepherd and Jonsson, 2008; Li et al., 2009; Waugh et al., 2009). Furthermore, greenhouse gas increases are projected to result in changed transport patterns such as an acceleration of the Brewer-Dobson circulation (Butchart et al., 2006; Li et al., 2008; Oman et al., 2009), modifying ozone distribution and recovery. On the other hand, ozone concentrations feed back on stratospheric dynamics through radiative heating; e.g., Antarctic ozone depletion has resulted in a colder and more stable Antarctic polar vortex which in turn lead to increased ozone depletion in a positive feedback cycle (Jones and Shanklin, 1995). Anomalies in stratospheric dynamics influence dynamical patterns near the surface (Baldwin and Dunkerton, 2001), and in particular the circulation anomalies induced by the ozone hole have been shown to dominate Antarctic surface climate trends (Thompson and Solomon, 2002; Gillett and Thompson, 2003).

The large advances in ozone research have been made possible by immense progress in the development of both observational systems such as satellite instruments, as well as increasingly complex and realistic computer models of the atmosphere, in particular chemistry transport models (CTMs) which use reanalysed observed meteorological fields, and chemistry climate models (CCMs) which generate meteorological fields internally from integration of the basic physical equations.

Some of the remaining uncertainties concerning stratospheric ozone are related to the lack of consistent long-time observation based data sets with global coverage, in particular also during polar night. Due to technical limitations native to each observational method, no

single observational data set is able to provide complete daily coverage of the stratosphere. Satellite instruments which have produced long data sets with near-global coverage rely on backscattered sunlight and are thus not available during polar winter, while satellite instruments measuring emitted radiation (which are thus able to operate in darkness) have not yet produced long continuous records of data. In-situ balloon borne sonde observations, on the other hand, are limited to single sites and altitudes below $\sim$ 30 km. Homogenised data sets have been published relying either on a series of identical satellite instruments (Stolarski and Frith, 2006), or by merging data from different satellite instruments with sonde observations (Randel and Wu, 2007), but these are restricted to zonal monthly means and either do not provide high-latitude information during winter at all (Stolarski and Frith, 2006) or rely on a single sonde site to represent ozone variability polewards of 55° latitude in each hemisphere (Randel and Wu, 2007). Hence, while these data sets are well suited for studying long-term ozone trends, they cannot be used to study ozone variability in three dimensions on a timescale of days to weeks.

A different approach is given by the method of data assimilation, which is essentially a combination of observational data with a computer model. Ideally, data assimilation merges the strengths of a 3D computer model (daily global coverage with selectable resolution and direct access to all model diagnostic quantities) with the realism of satellite observations. Information from satellite observations is used to constrain the CTM, while on the other hand, the CTM transport and chemistry scheme provide for the propagation of satellite information into areas where no observations are available.

In this thesis, a consistent long-term data set of stratospheric ozone is presented that has been generated by the assimilation of satellite data into a 3D chemistry transport model (CTM). The longest available satellite time series of global profile observations is used, which is provided by the Solar Backscatter UV (SBUV) instruments flown continuously since October 1978 on a series of satellites operated by the U.S. National Aeronautics and Space Administration (NASA) and the U.S. National Oceanic and Atmospheric Administration (NOAA).

Given that assimilation algorithms have been well established as a tool in atmospheric science (Khattatov et al., 2000; Ménard et al., 2000; Chipperfield et al., 2002), it is surprising that so far no long-term 3D data sets of assimilated profile ozone have been published (Lahoz et al., 2007), while several long-term assimilations of total column ozone have been produced (e.g., McPeters et al., 1998; Miller et al., 2002). Data sets of assimilated profile ozone have mostly been restricted to shorter time periods (e.g., Khattatov et al., 2000; Chipperfield et al., 2002). One attempt has been made as a part of the ERA-40 reanalysis conducted by the European Centre for Medium Range Weather Forecasts (ECMWF) (Uppala et al., 2005), but this data set has some known difficulties related to the synchronous assimilation of several parameters (Dethof and Hólm, 2004) and has a very limited representation of high-latitude ozone variability, as discussed in Chapter 6 of this thesis.

The research work presented in this thesis is focused on an analysis of ozone variability in the assimilated data set on different timescales. Two research questions are addressed, exploiting the possibilities of a consistent data set covering three decades with daily resolution.

Firstly, I investigate ozone variability in the polar middle stratosphere, addressing the questions: How do ozone anomalies develop during winter in the polar stratosphere under the influence of chemistry and dynamics? What are the processes involved in their creation and propagation, and how long do they survive before subsiding? Due to its coverage of polar latitudes during winter, the assimilated data set is ideally suited to address this issue.

Secondly, I investigate global long-term ozone trends and their attribution: Is the observed evolution of stratospheric ozone during the past decades consistent with our understanding of stratospheric dynamics and chemistry? What are the contributions of anthropogenic versus natural factors to the observed ozone changes, in particular with respect to the period of decreasing stratospheric chlorine loading since the turn of the century? Is an onset of recovery of the ozone layer already detectable?
The outline of this thesis is as follows. After a short review of stratospheric dynamics and ozone chemistry in Chapter 2, the CTM is described in Chapter 3. A theoretical introduction to the synthesis of model and observations, namely the assimilation algorithm, and its implementation in the CTM, are presented in Chapter 4. The SBUV satellite data used for the assimilation are introduced in Chapter 5.
Scientific results comprise mainly three chapters: The 29 year assimilated data set is introduced and validated against independent observations in Chapter 6, along with the CTM sensitivity runs without assimilation that are later used for attributing ozone changes to their origins. It is demonstrated that the assimilated data set is in excellent agreement to independent observations throughout most of the stratosphere, and can be seen as an extension of the satellite record into areas where satellite data are not available. The variability of polar ozone and the processes that contribute to it are analysed in Chapter 7, with a particular focus on the development of polar ozone anomalies during winter. Long-term trends, their relation to chemistry and dynamics, and the possible causes for trend changes are analysed in Chapter 8.
Chapter 9 presents conclusions and a short outlook on follow-up projects that are either planned or underway to address research questions arising from the current thesis.

2. The middle atmosphere

This chapter gives a short review of the middle atmosphere, focusing on stratospheric dynamics and chemistry. While a thorough introduction to this topic could (and does) fill several books, only a very basic approach is followed here owing to the limited space available, and mainly quantities of explicit relevance to the thesis are introduced. For a more exact treatment, the reader is referred to the cited literature, in particular Wallace and Hobbs (2006) for a general introduction, Brasseur et al. (1999) for a treatment of both dynamics and chemistry of the whole atmosphere, Brasseur and Solomon (2005) for one that is focused on the middle atmosphere, Holton (2004) for a thorough introduction to dynamic meteorology, and Andrews et al. (1987) for an in-depth treatment of stratospheric dynamics.

2.1. The Earth's atmosphere: An overview

The Earth's atmosphere is – compared to the size of the planet – a shallow layer of gas around the planet, that extends from the surface to the interplanetary space and is kept in place by Earth's gravitational attraction. Its total mass of $\sim 5 \cdot 10^{18}$ kg (Trenberth and Guillemot, 1994) is a tiny fraction – about a millionth – of the planetary mass of $\sim 6 \cdot 10^{24}$ kg. The atmosphere is usually divided in layers according to vertical temperature gradients. Zones of approximately homogeneous temperature gradient are referred to as "*spheres*", while their transition zones where the temperature gradient changes sign are termed "*pauses*".

An overview of the vertical temperature profile in the lower and middle atmosphere is shown in Figure 2.1. The lowest layer is the troposphere, supporting human life as well as weather and climate. It extends from the surface to the tropopause, located at an altitude of ~ 18 km in the tropics, ~ 12 km at mid-latitudes, and ~ 7 km in polar regions. The troposphere contains about $85 - 90\%$ of the total atmospheric mass. In the troposphere, the temperature decreases with height from the well-known average surface temperatures of ~ 290 K to around 210 K, with temperatures as low as ~ 180 K reached at the tropical tropopause. As the negative temperature gradient is close to the adiabatic lapse rate, the troposphere is often dynamically unstable, enabling fast vertical exchange of air parcels through convective activity. Globally, mixing timescales are of the order of weeks (Brasseur et al., 1999), and thus the troposphere is generally considered as "well-mixed". Above the tropopause, the temperature increases with altitude mainly due to the absorption of solar ultraviolet (UV) radiation by stratospheric ozone. 90% of atmospheric ozone are located in the stratosphere, which has led to the term *ozone layer* being used almost synonymously with the stratosphere. Contrary to the troposphere, due to the positive vertical temperature gradient the stratosphere is vertically stable, and the mean *age of air*, i.e. the time since air has been in contact with the tropopause, can be up to about 7 years in polar regions (Waugh and Hall, 2002). At around 50 km, temperatures reach a local maximum at the stratopause, and the gradient reverses again to become negative throughout the mesosphere, which extends

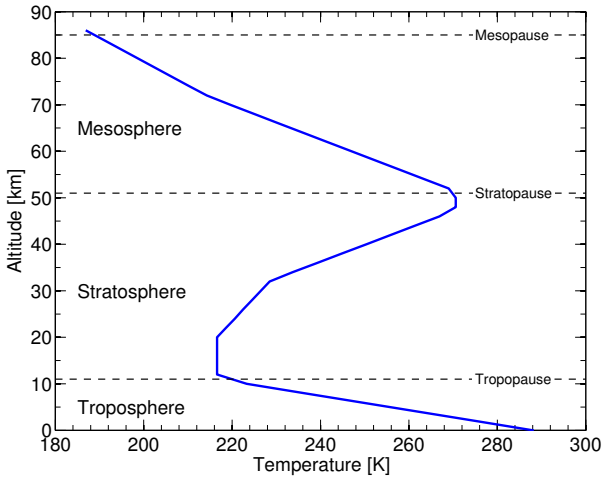

Figure 2.1.: *Average atmospheric temperature profile in the lower and middle atmosphere as defined in the U.S. Standard Atmosphere (NOAA, 1976)*

up to the mesopause around 90 km. Here, again, dynamical instability and hence fast vertical mixing is possible. The stratosphere and mesosphere together are usually referred to as "middle atmosphere", as opposed to the "lower atmosphere" (troposphere) and the "upper atmosphere" (thermosphere, exosphere). Above the mesopause, the temperature increases again in the thermosphere, mainly due to the absorption of very short-wave UV radiation by molecular oxygen. The transition from the mesosphere to the thermosphere also approximately marks the transition from the *homosphere* to the *heterosphere*. While long-lived air constituents are well mixed in the homosphere, and hence the molecular weight of air varies only little with altitude, in the heterosphere mixing is no longer ensured owing to the very low densities of air. Temperatures increase to reach maximum levels of several 100 K, which are, however, strongly dependent on solar irradiation and cannot be interpreted a thermodynamic equilibrium temperature. Due to the large mean free path, the different degrees of freedom decouple, and local thermodynamic equilibrium is no longer reached. Above the thermosphere, which extends up to about 700 km altitude, the exosphere marks the transition layer to interplanetary space.

The atmosphere is a mixture of many chemical constituents, the most important being nitrogen (N_2, 79%) and oxygen (O_2, 21%), followed by the noble gases Argon (Ar, 0.9%), Neon (Ne; 18 ppmv, or parts per million by volume), and Helium (He, 5 ppmv), as well as the well-known greenhouse gases carbon dioxide (CO_2, 360 ppmv) and methane (CH_4, 1.7 ppmv). Ozone (O_3) is mainly present in the stratosphere with mixing ratios of up to 10 ppmv, while mixing ratios in the troposphere are around 50 ppbv (parts per billion by volume).

Due to their absorption properties, some of the minor species – in particular CO_2, CH_4, and O_3 – actually play dominant roles in the radiation budget. For the middle atmosphere, ozone is the single most important radiatively active component, as it determines to a large degree the stratospheric temperature distribution by its strong absorption of UV radiation.

The following sections are dedicated to providing an introduction to atmospheric dynamics

and chemistry, focusing on the middle atmosphere. A schematic overview of the entangled chemical, radiative, and dynamical processes determining the physical and chemical state of the stratosphere is given in Figure 2.2 (adapted from Brasseur et al., 1999). Physical components of the system are shown on the right-hand side of the figure, while chemical agents are situated on the left (corresponding to its fundamental role in stratospheric chemistry, ozone is displayed in its own box). The stratosphere is subject to external inputs of chemical (emissions) and physical nature (solar irradiation, input of energy and momentum through upward propagating planetary waves from the troposphere). Interactions between different components are indicated as arrows in Figure 2.2. From a single glance it is obvious that physical and chemical components can never be treated entirely separately, as they together form the stratospheric system. Details on the processes shown in Figure 2.2 will be given in the remaining part of this chapter: Basic stratospheric dynamics are discussed in Section 2.2, and an introduction to stratospheric chemistry is provided in Section 2.3.

2.2. Basic dynamics

The atmosphere is – at least in the homosphere – well described as a continuous compressible fluid, whose dynamics are governed by the Newtonian laws of motion in a rotating coordinate system. A rigid derivation of all governing equations is beyond the scope of this introductory chapter – see e.g. Holton (2004) for a detailed treatment. Instead, after introducing the fundamental quantities and relations in Section 2.2.1 and the assumption of hydrostatic equilibrium in Section 2.2.2, we shall state the set of governing equations in Section 2.2.3, and explain their application to the stratospheric circulation in Section 2.2.4.

2.2.1. Prerequisites

Before setting up the governing equations of motion for the atmosphere, a suitable coordinate system needs to be introduced. Throughout this study, a *Eulerian* framework is used, that is, a coordinate system fixed at with respect to the Earth, as opposed to the *Lagrangian* framework, in which physical processes are described in a coordinate system moving with the flow of the fluid. The coordinate system of choice for the description of atmospheric motion is naturally a spherical one. We shall denote $\mathbf{x}$ as the position vector,

$$\mathbf{x} = \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} a\lambda\cos\phi \\ a\phi \\ z \end{pmatrix}, \tag{2.1}$$

with a the radius of the Earth and λ and ϕ longitude and latitude, respectively. Instead of the geometric height, other vertical coordinates are often used in the atmospheric sciences, such as pressure or potential temperature (see below).
The velocity of a fluid particle is expressed by the time derivative of its position vector, $\mathbf{v} \equiv d\mathbf{x}/d\mathbf{t}$. Newton's second law relates the acceleration of a fluid particle to the forces acting on it, and is given in this coordinate system by

$$\frac{\partial \mathbf{v}}{\partial t} + \mathbf{v} \cdot \nabla \mathbf{v} = \mathbf{F}(\mathbf{x}, t) \tag{2.2}$$

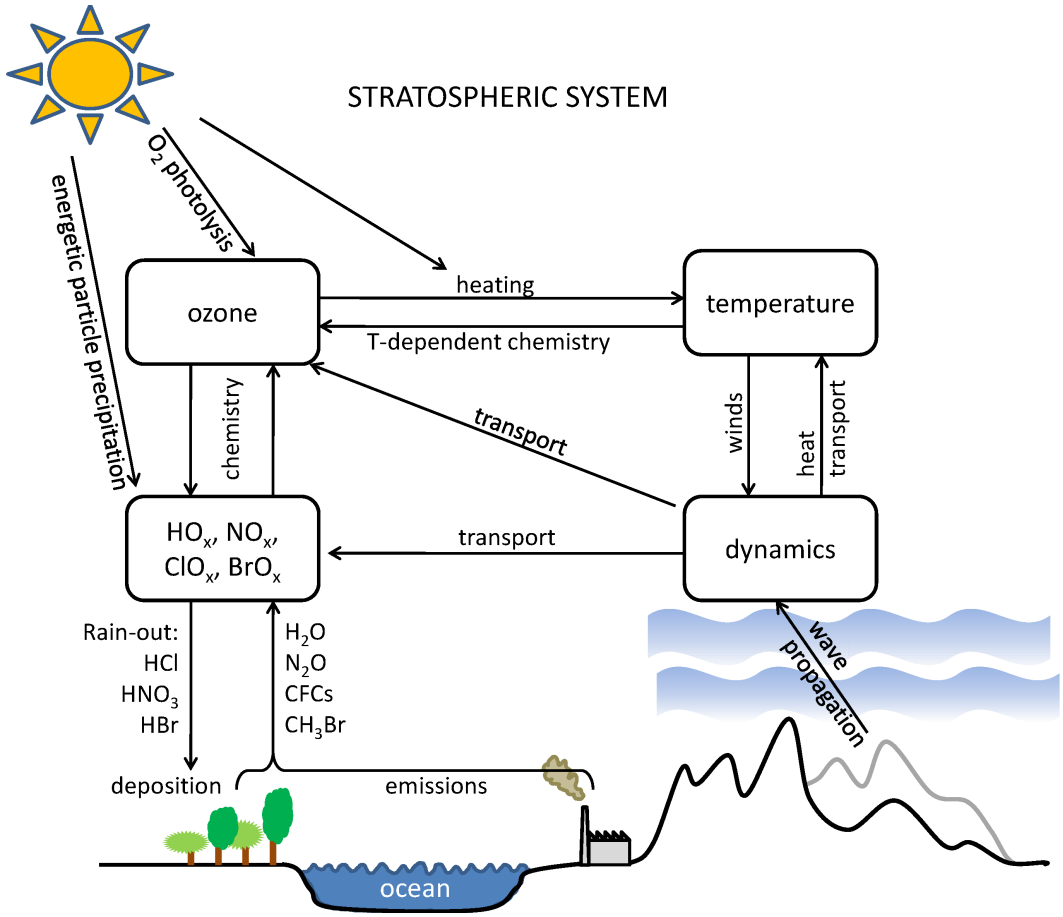

Figure 2.2.: *Schematic overview of the interactions between agents determining the physical and chemical state of the stratosphere. Solar UV radiation directly influences both ozone chemistry as well as temperature through radiative heating; in addition, also precipitation of high energetic particles from the sun leads to chemical changes. Influences from the troposphere include emissions (natural and anthropogenic) as well as input of energy through upward propagating Rossby waves. Chemical and physical processes are closely interrelated; e.g., ozone determines temperature through its properties as a UV absorber, while chemical ozone production and destruction rates are dependent on temperature. Plot adapted from Brasseur et al. (1999).*

where $\mathbf{F}$ is the sum of the forces acting on the fluid particle. The left side of (2.2) is obtained by considering that the total, or material derivative following the fluid is

$$\frac{d\mathbf{v}(\mathbf{x},t)}{dt} = \frac{\partial \mathbf{v}}{\partial t}\Big|_{\mathbf{x}} + \frac{\partial \mathbf{v}}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial \mathbf{v}}{\partial y}\frac{\partial y}{\partial t} + \frac{\partial \mathbf{v}}{\partial z}\frac{\partial z}{\partial t} = \frac{\partial \mathbf{v}}{\partial t} + \mathbf{v}\cdot\nabla\mathbf{v}. \tag{2.3}$$

2.2.2. Barometric height and hydrostatic equilibrium

Some simplifications are usually made when Newton's law (2.2) is applied to the atmosphere. One is the assumption of hydrostatic equilibrium, stating that vertical acceleration can be neglected. This is reasonably fulfilled in the absence of convection, which is usually the case

throughout the stratosphere, and generally for air motions on synoptic scales greater than about 1000 km (Holton, 2004). The vertical component of Equation (2.2) then reduces to a balance of downward acting gravity against upward acting pressure gradient, so that

$$\frac{1}{\rho}\frac{\partial p}{\partial z} = -g \tag{2.4}$$

with g the local gravitational acceleration, z the geometric height and ρ the air density. Integrating Equation (2.4), we find that

$$p(z) = \int_{\infty}^{z} g\rho(x, y, z')dz'. \tag{2.5}$$

Evaluation of the integral in (2.5) requires knowledge of the vertical distribution of air density, which is not known a priori. However, since the atmosphere obeys the ideal gas law reasonably well, we can express pressure as

$$p = \rho RT \tag{2.6}$$

where R is the gas constant of air, related to the universal gas constant R_0 by the molar mass of air M_{air} by $R = R_0/M_{air} = 8.31\,\mathrm{J\,mol^{-1}\,K^{-1}} / 0.028\,\mathrm{kg\,mol^{-1}} = 287\,\mathrm{J\,kg^{-1}\,K^{-1}}$. A reasonable simplification is then to assume a vertically isothermal atmosphere, i.e. $T = T_0 = const.$ (which we know is not exactly the case, see Figure 2.1). Under this assumption, and that of constant gravitational acceleration g (which is reasonably fulfilled for the lower and middle atmosphere) we can directly integrate (2.4) to yield

$$p(x, y, z) = p_0(x, y)\, e^{-z/H_0} \tag{2.7}$$

where p_0 is the surface pressure and H_0 is the atmospheric scale height,

$$H_0 = RT_0/g. \tag{2.8}$$

The e-folding height H_0 typically lies between 7 and 8 km for the lower and middle stratosphere. Consequently, inserting (2.6), we find that also density obeys the same vertical dependence,

$$\rho(x, y, z) = \rho_0(x, y)\, e^{-z/H_0}. \tag{2.9}$$

In spite of temperature variations in the real atmosphere, the barometric height formula is fulfilled quite well throughout the lower and middle atmosphere, and the real atmospheric scale height H is close to H_0. In particular, due to the monotonic dependence of pressure on height, equation (2.7) implies that pressure can be used as a vertical coordinate, which is often done in the form of *pressure height*, $z_p = H\ln(p_0/p)$.
Another important alternative vertical coordinate is *potential temperature*. Since this is employed as vertical coordinate in the chemistry transport model used in this thesis, it is introduced here briefly. The potential temperature θ of an air parcel is defined as the temperature this air parcel would reach, were it brought adiabatically from its pressure p to surface pressure p_0 (usually taken to be 1000 hPa). From the first law of thermodynamics, θ is given by Poisson's equation,

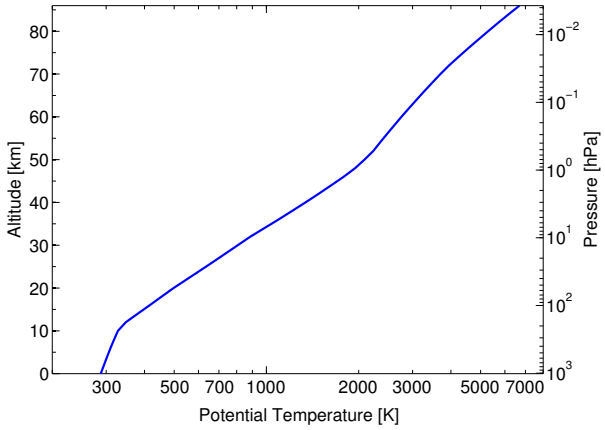

Figure 2.3.: *Vertical profile of potential temperature as calculated from the U.S. Standard Atmosphere (NOAA, 1976). Note the double vertical axis in terms of altitude (left) and pressure (right).*

$$\theta = T \left(\frac{p_0}{p} \right)^{R/c_p} , \tag{2.10}$$

where c_p represents the specific heat of air at constant pressure. Due to the monotonous increase of temperature with altitude, potential temperature is a well-defined vertical coordinate throughout the stratosphere. Its significant advantage as compared to geometrical height or log-pressure height lies in the fact that potential temperature is conserved following adiabatic motion, and the vertical velocity in this reference frame is directly related to the diabatic heating rate J,

$$\frac{d\theta}{dt} = \left(\frac{p_0}{p} \right)^{R/c_p} \frac{dT}{dt} = \frac{\theta}{T} J \ . \tag{2.11}$$

The primary diabatic heating sources are absorption and emission of radiation, and heating due to condensation (or evaporation) of water vapour (latent heating). Since adiabatic processes by definition conserve entropy, levels of constant potential temperature are levels of constant entropy and referred to as *isentropes*. The vertical distribution of potential temperature is displayed in Figure 2.3. The vertical axis is denoted in terms of geometrical height (left) as well as pressure as obtained from the barometric formula (right). Note the small variation of potential temperature in the troposphere, as opposed to the large increases in the stratosphere.
Having introduced the basic quantities and assumptions, we can now proceed to denote the complete set of equations governing atmospheric dynamics, also called the *primitive equations of motion*.

2.2.3. The primitive equations of motion

The governing set of equations for atmospheric motion in log-pressure coordinates is given by (Andrews et al., 1987)

$$\frac{\partial \Phi}{\partial z} = \frac{RT}{H} \tag{2.12}$$

$$\frac{du}{dt} - \left(f + \frac{u \tan\phi}{a}\right) v = -\frac{\partial \Phi}{\partial x} + X \tag{2.13}$$

$$\frac{dv}{dt} + \left(f + \frac{u\, tan\phi}{a}\right) u = -\frac{\partial \Phi}{\partial y} + Y \tag{2.14}$$

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \rho^{-1} \frac{\partial(\rho w)}{\partial z} = 0 \tag{2.15}$$

$$\frac{dT}{dt} + \left(\frac{\kappa T}{H}\right) w = \frac{J}{c_p} \tag{2.16}$$

where

$$\Phi = \int_0^z g(z')\, dz' \tag{2.17}$$

is the geopotential, and

$$f \equiv 2\Omega \sin\phi \tag{2.18}$$

is the *Coriolis parameter* or twice the vertical component of Earth's angular velocity Ω, that occurs in Equations (2.13) and (2.14) due to the Earth being a non-inertial rotating reference system. X and Y in Equations (2.13) and (2.14) represent longitudinal and latitudinal nonconservative forces, such as friction. Equation (2.12) is equivalent to (2.4), the hydrostatic equilibrium equation, Equations (2.13) and (2.14) are the horizontal momentum equations, (2.15) is the continuity equation stating the conservation of mass, and (2.16) is the thermodynamic energy equation, a reformulation of the first law of thermodynamics.
A general analytic solution to the set of Equations (2.12) – (2.16) is not available, and usually some further simplifications are used when applying them to study specific phenomena in the atmosphere.
A consequence of the non-inertial character of the reference frame is the formation of cyclones and anticyclones as *geostrophic winds*. If an air parcel is subject to a horizontal pressure gradient – which may arise for example by differential solar heating of distant regions – it will initially start to move in the direction of the pressure gradient, but then be deflected by the Coriolis force. In the geostrophic equilibrium situation, the pressure gradient force and the Coriolis force are exactly balanced, and the air parcel moves in a direction orthogonal to the pressure gradient. As a consequence, low-pressure systems are encircled counterclockwise (clockwise) in the Northern (Southern) hemisphere.
An important concept in atmospheric dynamics is that of *potential vorticity*. Potential vorticity $\hat{\zeta}$ is defined as the sum of planetary vorticity (i.e., the Coriolis parameter f) and relative vorticity ζ,

$$\zeta = \frac{\partial v}{\partial x} - \frac{1}{\cos\phi} \frac{\partial(u \cos\phi)}{\partial y}, \tag{2.19}$$

divided by the height h of the fluid parcel under consideration:

$$\hat{\zeta} = \left(\frac{\zeta + f}{h}\right). \tag{2.20}$$

It can be shown (Brasseur et al., 1999) that potential vorticity is a conserved quantity following the motion of an air parcel, if the nonconservative forces X, Y vanish. A frequently used form of potential vorticity is Ertel's potential vorticity (Ertel, 1942b), which is essentially equal to (2.20), but uses potential temperature as vertical coordinate,

$$\hat{P} = \frac{\zeta_\theta + f}{\hat{\sigma}} \tag{2.21}$$

with

$$\hat{\sigma} = -\frac{1}{g}\frac{\partial p}{\partial \theta} \tag{2.22}$$

the isentropic pseudo-density. $\hat{P}$ is conserved in the absence of diabatic heating and nonconservative forces (Ertel, 1942a). In many cases, the timescale of advection is much faster than the timescale of heating and friction, and hence $\hat{P}$ is usually approximately conserved following the motion of the air parcels. Thus, it may be used as a tracer of fluid motion, e.g. in order to distinguish polar vortex air masses from other high-latitude air, as it is done in Chapter 7.

Conservation of potential vorticity leads to quite strong constraints on the meridional motion of air parcels. Displacement in meridional direction leads to a restoring force, which results in an oscillatory movement of the particle about its initial latitude. Such oscillations lead to the formation of planetary scale *Rossby waves* (Rossby, 1939), which play a crucial role for stratospheric circulation due to their ability to deposit energy and angular momentum in wave breaking processes.

2.2.4. The general circulation of the stratosphere

The average motion of air in the atmosphere follows distinct, large-scale patterns, usually termed the *general circulation*. These patterns of rising, horizontal, and descending movement are governed by an interplay of radiative heating and dynamical forcing through the deposition of momentum and energy by atmospheric waves.

Figure 2.4 gives an overview of the general circulation patterns in the lower and middle atmosphere. The tropospheric circulation patterns, depicted in blue, consist of tropical Hadley cells, mid-latitude Ferrel cells, and high-latitude polar cells. In the tropics, large-scale ascent of air is driven by strong solar heating, which is compensated by descent in the extra-tropics to guarantee mass conservation. While the resulting tropical Hadley cells can be understood as a direct consequence of solar heating, the extra-tropical Ferrel cells are caused by eddy heat and momentum fluxes through baroclinic Rossby waves. Upwelling motion in the tropics can be so strong that some deep convective clouds are able to overshoot the tropopause and inject air parcels into the lower stratosphere. Brewer (1949) was the first to realise that convective updraught through the tropical tropopause constitutes the most important source of tropospheric air input into the stratosphere.

In the stratosphere, a comparably simple general circulation is established, as shown in red in 2.4. Strong upwelling motion is observed in the tropics, and from there the air is slowly transported poleward and downward in one single cell in each hemisphere, which is much stronger in the winter hemisphere. Owing to pioneering studies by Dobson et al. (1930) and Brewer (1949), who were the first to infer the stratospheric large-scale circulation from

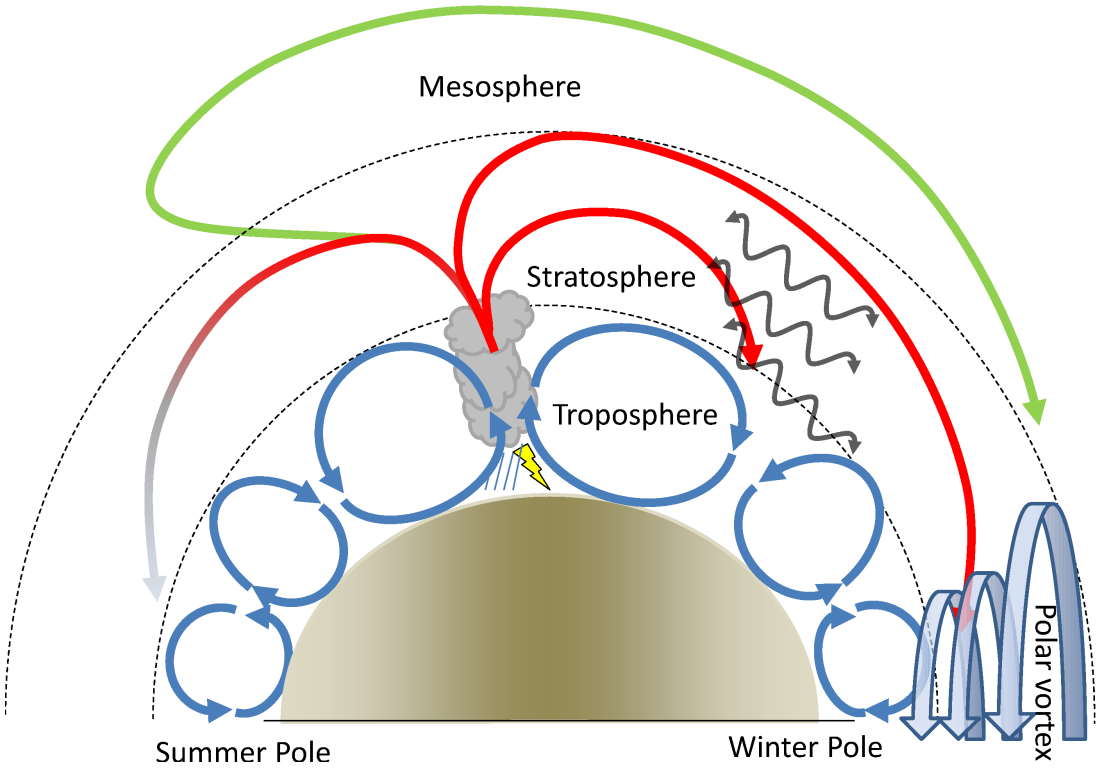

Figure 2.4.: *Schematic overview of the zonal mean general circulation of the atmosphere. Blue: tropospheric circulation, red: stratospheric circulation (Brewer-Dobson circulation), green: mesospheric circulation (meridional overturning circulation). Tropospheric air is injected into the stratosphere through deep convective clouds in the tropics and subsequently transported poleward under the influence of planetary waves (black arrows). Over the winter pole, a strong westerly wind regime is established, which is called the polar vortex. The figure is not to scale. Dashed lines symbolise tropopause and stratopause.*

the distribution of trace species, this stratospheric circulation pattern has been termed the *Brewer-Dobson circulation.*

Further up in the mesosphere, a meridional overturning circulation is found (green arrow in Figure 2.4), which transports air parcels from the summer pole to the winter pole as a result of breaking gravity waves (i.e., wave motions where the restoring force is buoyancy). As this thesis is concerned with the stratosphere, we shall limit a deeper discussion to the stratospheric circulation here.

While the shape of the Brewer-Dobson circulation might suggest at first sight a dominant role of diabatic heating in a similar mechanism as that leading to the tropospheric Hadley cell – rising air in the tropics through strong insolation and descent in the mid- and polar latitudes due to diabatic cooling and mass conservation –, the stratospheric circulation is in fact mainly driven by wave forcing. Absorption of solar UV radiation by ozone does, of course, play an important role for the temperature distribution in the stratosphere, but the true temperature distribution differs from a purely radiatively determined state by tens of Kelvins, particularly at high latitudes. Heat fluxes due to atmospheric waves are needed in

order to explain the differences.
Planetary waves are usually generated close to Earth's surface due to orographic effects and subsequently propagate upwards into the stratosphere, where they dissipate. As shown by Charney and Drazin (1961), Rossby waves can propagate vertically only if the mean zonal winds are weak and westerly, while no propagation is possible for easterly winds or westerly winds exceeding the critical *Rossby velocity*. The thermal gradient between the summer and the winter pole leads to zonal winds. During winter, these winds are strong and westerly, then reverse shortly after solstice, and in summer weaker easterlies prevail. Due to the westerly winds in the winter hemisphere, Rossby waves are able to propagate from the troposphere into the stratosphere in the mid-latitudes (symbolised as black arrows in Figure 2.4), thereby transporting energy and angular momentum. Since the critical velocity is dependent on the (horizontal) wave number, only the largest Rossby waves of planetary scale (typically wave numbers 1 – 3) are observed in the stratosphere. These planetary waves are able to overcome the constraints on meridional motion and lead to a poleward mass flux.
From the requirement of conservation of energy in the absence of dissipative processes, it follows as a consequence that the amplitude of upward-propagating atmospheric waves must increase exponentially with altitude (Brasseur et al., 1999). However, wave amplitudes cannot increase indefinitely, as beyond a certain point of perturbation from the undisturbed state the atmospheric configuration becomes unstable. In case of Rossby waves, this instability is of *barotropical* nature, that is, the local potential vorticity gradient is reversed by the wave (Charney and Stern, 1962). Wave breaking is induced, which leads to the deposition of energy and angular momentum, and to irreversible mixing of air parcels.
As a result of the meridional temperature gradient, a strong westerly circumpolar geostrophic circulation is observed in polar latitudes throughout the stratosphere during winter, which is known as the *polar vortex* (Holton, 2004). Due to the strong westerly winds exceeding the critical velocity, planetary waves are largely prevented from propagating into the polar stratosphere during the winter months. The strong meridional potential vorticity gradient along the edge of the polar vortex provides for an effective meridional transport barrier, and hence the polar vortex constitutes a zone of cold (due to lacking solar irradiation) and secluded air. Unlike the Arctic which is surrounded by alternating continental and oceanic patterns, favouring the development of planetary wave disturbances in zonal flow, Antarctica is situated in a near-perfect geographically symmetric environment, and hence the Antarctic polar vortex is much stabler, more secluded and colder than its Arctic counterpart. In these remote areas, unique chemical reactions can take place which lead to the formation of the ozone hole in austral spring. These are discussed in greater detail in the next section.

2.3. Stratospheric ozone

Due to the importance of stratospheric ozone both for terrestrial life and atmospheric dynamics and chemistry, it has long been a subject of intense research. While ozone itself was discovered by C. F. Schönbein in 1839, it took four decades until the existence of a stratospheric ozone layer was postulated by W. N. Hartley in 1880 and finally proved by C. Fabry and H. Buisson in 1913. Stratospheric ozone chemistry is now believed to be well understood, and there are good reasons to assume that this is more the case than the last time when this statement was generally accepted within the scientific community – in the early 1980s just before the completely unexpected detection of the Antarctic ozone hole.

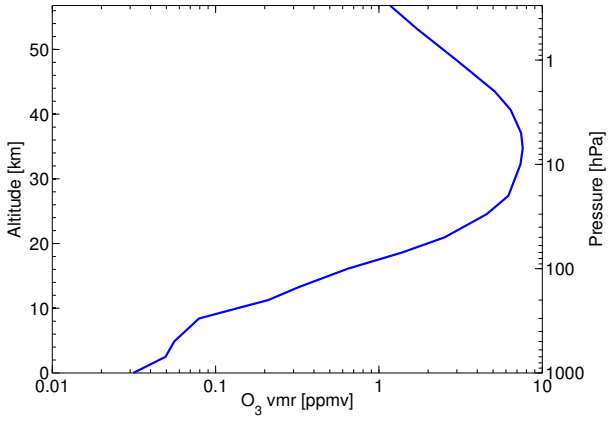

Figure 2.5.: *Climatological ozone profile in the NH mid-latitudes (30-60°), taken from the climatology of Fortuin and Kelder (1998). Note the double vertical axis displaying altitude and pressure (the native vertical coordinate of the climatology is pressure, altitude is calculated from the barometric height formula).*

Around 90% of the total ozone amount in the atmosphere are located in the stratosphere. A climatological average ozone profile for Northern Hemisphere mid-latitudes is shown in Figure 2.5 (taken from the climatology of Fortuin and Kelder, 1998). Mixing ratios are below 100 ppmv in the troposphere, then increase sharply to reach a maximum of $\sim$ 10 ppmv at 35 km, and decrease to $\sim$ 2 ppmv at the stratopause. The principal shape of the ozone layer is mainly explained by the photochemical balance of ozone production versus destruction, as described in the following sections. Ozone production requires short-wave UV radiation; consequently, the strongest ozone production and also highest *mixing ratios* are found in the tropical middle stratosphere. From there, ozone is transported poleward and downward by the Brewer-Dobson circulation, so that the largest *total column* amounts are actually found in the high latitudes. The total column abundance represents the integrated amount of ozone molecules over a unit surface area, and is thus the essential quantity determining harmful UV radiation at the Earth surface. Ozone columns are often measured in terms of *Dobson units* (DU), with 1 DU equivalent to a total column height of 0.001 cm at sea level pressure, or $2.69 \cdot 10^{16}$ ozone molecules per square centimetre. Typical ozone columns amount to 300 DU in the mid-latitudes, corresponding to a layer of merely 3 mm thickness at sea level pressure. Due to the asymmetric structure of the Brewer-Dobson circulation being significantly stronger in the winter hemisphere, total ozone outside the tropics follows a distinct annual cycle. In a very simplified picture, this cycle consists of an ozone increase during winter and a decline during summer, when advection of tropical ozone-rich air is not efficient.

Stratospheric ozone amounts are determined by chemistry and transport. In the upper stratosphere, where strong solar UV radiation is available, the photochemical lifetime of ozone is short, resulting in ozone concentrations always being close to the local photochemical equilibrium. As solar radiation penetrates deeper into the atmosphere, the short-wave UV part of the spectrum is gradually diminished due to absorption by ozone, leaving less UV photons for lower atmospheric layers. Chemically active reaction partners become scarce, and

ozone lifetimes increase, reaching up to several months or even years in the lower stratosphere. Thus, lower stratospheric ozone is dynamically controlled outside polar latitudes. At polar latitudes, heterogeneous reactions taking place on aerosol surfaces can transform chlorine compounds from their inactive form to reactive species, leading to massive and widespread ozone depletion in the Antarctic (and to a lesser extent in the Arctic) lower stratosphere during spring commonly known as the *ozone hole*.
The following sections provide an introduction to ozone chemistry, detailed into pure oxygen chemistry (Section 2.3.1), gas-phase catalytic cycles (Section 2.3.2), and polar heterogeneous chemistry or "ozone hole chemistry" (Section 2.3.3).

2.3.1. The Chapman cycle

Historically, Chapman (1930) was the first to propose a mechanism for ozone generation and destruction. According to the classical Chapman theory, ozone is produced in the stratosphere from the photolysis of oxygen by solar UV radiation ($\lambda < 242\,\mathrm{nm}$),

$$\mathrm{O_2} + h\nu \xrightarrow{j_a} 2\mathrm{O} \tag{2.23}$$

which can then recombine with molecular oxygen to form ozone,

$$\mathrm{O} + \mathrm{O_2} + M \xrightarrow{k_b} \mathrm{O_3} + M \tag{2.24}$$

if a third collision partner (M, typically N_2 or O_2) is available to take up the recoil, so that energy and momentum are conserved.
Conversely, ozone may be destroyed by photolysis ($\lambda < 366\,\mathrm{nm}$),

$$\mathrm{O_3} + h\nu \xrightarrow{j_c} \mathrm{O_2} + \mathrm{O}\ , \tag{2.25}$$

or by collisions with atomic oxygen,

$$\mathrm{O} + \mathrm{O_3} \xrightarrow{k_d} 2\mathrm{O_2}\ . \tag{2.26}$$

Reaction coefficients $j_{a,c}$, $k_{b,d}$ are temperature dependent. In terms of chemical families, reactions (2.23) to (2.26) describe the the *odd-oxygen* family, which is defined as $\mathrm{O_x} = \{\mathrm{O}, \mathrm{O_3}\}$. In (2.23), $\mathrm{O_x}$ is produced from the reservoir O_2, in (2.26) it is destroyed, and reactions (2.24) and (2.25) determine the partitioning between O and O_3. An additional reaction pathway for the destruction of $\mathrm{O_x}$ is the reaction

$$\mathrm{O} + \mathrm{O} + M \xrightarrow{k_e} \mathrm{O_2} + M, \tag{2.27}$$

but this reaction is too slow to be of importance in the stratosphere, while it does play a role in the mesosphere. Atomic oxygen can be produced in two states, the atomic ground state $\mathrm{O}(^3\mathrm{P})$, and the first excited state $\mathrm{O}(^1\mathrm{D})$. For the photolysis of ozone to produce $\mathrm{O}(^1\mathrm{D})$ by reaction (2.25), the wavelength of the absorbed photon must be below 320 nm, while the photolysis of molecular oxygen in (2.23) may produce $\mathrm{O}(^1\mathrm{D})$,

$$\mathrm{O_2} + h\nu \rightarrow \mathrm{O}(^1\mathrm{D}) + \mathrm{O}(^3\mathrm{P}) \tag{2.28}$$

if the wavelength of the photon is below 175.9 nm. Although the electronic state of atomic oxygen is not directly of importance within the Chapman cycle, the existence of $\mathrm{O}(^1\mathrm{D})$ is decisive for the generation of OH and NO involved in the catalytic cycles described in Section 2.3.2.

2.3.2. Gas phase catalytic cycles

For several decades, reactions (2.23) to (2.26) were believed to be a complete description of ozone chemistry, as they were able to reproduce the basic characteristic shape of the ozone layer. In the 1960s, however, more accurate measurements revealed that the Chapman theory severely over-predicts ozone concentrations. Hence, additional mechanisms for ozone destruction were needed, which were found by Hunt (1966) (odd-hydrogen), Crutzen (1970) (odd-nitrogen), and Molina and Rowland (1974) (odd-chlorine cycles). The basic idea common to these concepts is that the direct loss of odd-oxygen, described by (2.26), should be accelerated by catalytic mechanisms, in the form

$$\begin{aligned} X + \mathrm{O_3} &\rightarrow X\mathrm{O} + \mathrm{O_2} && (2.29)\\ X\mathrm{O} + \mathrm{O} &\rightarrow X + \mathrm{O_2} && (2.30)\\ \hline \text{Net} : \mathrm{O_3} + \mathrm{O} &\rightarrow 2\mathrm{O_2} && (2.31) \end{aligned}$$

whereby X represents the catalyst and $X\mathrm{O}$ the intermediate product. With suitable catalysts X, reactions (2.29) and (2.30) can proceed much faster than (2.26), even if concentrations of X are several orders of magnitude below those of ozone (Brasseur et al., 1999).

$\mathrm{HO_x}$ chemistry

The first candidate to be identified as a possible catalyst was the hydroxyl radical OH (Hunt, 1966), which has several sources in the upper stratosphere:

$$\begin{aligned} \mathrm{H_2O} + \mathrm{O}(^1\mathrm{D}) &\rightarrow 2\,\mathrm{OH} && (2.32)\\ \mathrm{CH_4} + \mathrm{O}(^1\mathrm{D}) &\rightarrow \mathrm{CH_3} + \mathrm{OH} && (2.33)\\ \mathrm{H_2} + \mathrm{O}(^1\mathrm{D}) &\rightarrow \mathrm{H} + \mathrm{OH} && (2.34) \end{aligned}$$

At high altitudes ($\gtrsim 40\,\mathrm{km}$), where atomic oxygen is abundant, the major catalytic cycle for the destruction of ozone based on odd-hydrogen ($\mathrm{HO_x} = \{\mathrm{H}, \mathrm{OH}, \mathrm{HO_2}\}$) is

$$\begin{aligned} \mathrm{OH} + \mathrm{O} &\rightarrow \mathrm{H} + \mathrm{O_2} && (2.35)\\ \mathrm{H} + \mathrm{O_3} &\rightarrow \mathrm{OH} + \mathrm{O_2} && (2.36)\\ \hline \text{Net} : \mathrm{O} + \mathrm{O_3} &\rightarrow 2\,\mathrm{O_2} && (2.37) \end{aligned}$$

As atomic oxygen concentrations decrease below 40 km, the following cycle becomes more important:

$$\begin{aligned} \mathrm{OH} + \mathrm{O_3} &\rightarrow \mathrm{HO_2} + \mathrm{O_2} && (2.38)\\ \mathrm{HO_2} + \mathrm{O} &\rightarrow \mathrm{OH} + \mathrm{O_2} && (2.39)\\ \hline \text{Net} : \mathrm{O} + \mathrm{O_3} &\rightarrow 2\,\mathrm{O_2} && (2.40) \end{aligned}$$

Below $\sim 30\,\mathrm{km}$, atomic oxygen is almost absent, and the dominant odd-hydrogen mechanism is

$$\begin{aligned} \mathrm{OH} + \mathrm{O_3} &\rightarrow \mathrm{HO_2} + \mathrm{O_2} && (2.41)\\ \mathrm{HO_2} + \mathrm{O_3} &\rightarrow \mathrm{OH} + 2\,\mathrm{O_2} && (2.42)\\ \hline \text{Net} : 2\,\mathrm{O_3} &\rightarrow 3\,\mathrm{O_2} && (2.43) \end{aligned}$$

NO_x chemistry

Another series of catalytic cycles, which were found by Crutzen (1970) and Johnston (1971), involve nitrogen oxides, usually summarised in the NO_x family ($NO_x = \{N,\ NO,\ NO_2\}$). Nitric oxide (NO) is mainly produced in the stratosphere by the oxidation of nitrous oxide (N_2O), which has natural as well as anthropogenic sources in the troposphere:

$$\mathrm{N_2O} + \mathrm{O}(^1\mathrm{D}) \rightarrow 2\,\mathrm{NO} \qquad (2.44)$$

The conversion of odd-oxygen to O_2 is then possible by the two reaction cycles

$$\begin{aligned} \mathrm{NO} + \mathrm{O_3} &\rightarrow \mathrm{NO_2} + \mathrm{O_2} && (2.45)\\ \mathrm{NO_2} + \mathrm{O} &\rightarrow \mathrm{NO} + \mathrm{O_2} && (2.46)\\ \hline \text{Net} : \mathrm{O} + \mathrm{O_3} &\rightarrow 2\,\mathrm{O_2} && (2.47) \end{aligned}$$

and

$$\begin{aligned} \mathrm{NO} + \mathrm{O_3} &\rightarrow \mathrm{NO_2} + \mathrm{O_2} && (2.48)\\ \mathrm{NO_2} + \mathrm{O_3} &\rightarrow \mathrm{NO_3} + \mathrm{O_2} && (2.49)\\ \mathrm{NO_3} + h\nu &\rightarrow \mathrm{NO} + \mathrm{O_2} && (2.50)\\ \hline \text{Net} : 2\,\mathrm{O_3} &\rightarrow 3\,\mathrm{O_2} && (2.51) \end{aligned}$$

Reactions (2.45) – (2.47) represent the dominant ozone reaction cycle throughout the middle stratosphere.

ClO_x chemistry

Chronologically, the next species that were found to be important catalysts for ozone destruction were reactive chlorine compounds (Molina and Rowland, 1974; Cicerone et al., 1974). While there are natural sources for reactive chlorine from the photolysis or oxidation of methyl chloride, a more important source are anthropogenic emissions of chlorofluorocarbons (CFCs), in particular $CFCl_3$ and CF_2Cl_2. These industrially manufactured compounds, first synthesised in 1928, are nontoxic and nonflammable, and soon became were widely used as refrigerants, propellants, solvents, and cleansing agents. Due to their chemical inertness, these gases have practically no sinks in the troposphere, and can be injected into the stratosphere through the tropical tropopause layer. In the stratosphere, however, CFCs can be photolysed by short-wave UV radiation ($\lambda < 220\,\mathrm{nm}$),

$$\begin{aligned} \mathrm{CF_2Cl_2} + h\nu &\rightarrow \mathrm{CF_2Cl} + \mathrm{Cl} && (2.52)\\ \mathrm{CFCl_3} + h\nu &\rightarrow \mathrm{CFCl_2} + \mathrm{Cl} && (2.53) \end{aligned}$$

The chlorine released in (2.52) and (2.53) may then destroy ozone by the catalytic cycle

$$\begin{aligned} \mathrm{Cl} + \mathrm{O_3} &\rightarrow \mathrm{ClO} + \mathrm{O_2} && (2.54)\\ \mathrm{ClO} + \mathrm{O} &\rightarrow \mathrm{Cl} + \mathrm{O_2} && (2.55)\\ \hline \mathrm{Net}: \mathrm{O_3} + \mathrm{O} &\rightarrow 2\,\mathrm{O_2} && (2.56) \end{aligned}$$

which, as it involves atomic oxygen, is most important in the upper stratosphere. Furthermore, a reaction cycle involving the ClO dimer is important for the formation of the ozone hole in the polar lower stratosphere in spring, and is described in Section 2.3.3 below (Equations 2.66–2.70). Other gas phase reactive cycles with chlorine involve members of the other reactive species families – HO_x, NO_x, and, as described below, BrO_x.

BrO_x chemistry

Bromine can act as a catalyst for ozone destruction in a similar way as chlorine (Wofsy et al., 1975). Its most important source is methyl bromine (CH_3Br) released by biological activity in the ocean, by agriculture, and biomass burning (Brasseur et al., 1999). Anthropogenic sources are mainly halons (CF_3Br, CF_2ClBr), which have been regulated under amendments of the Montreal protocol. These source gases are destroyed in the stratosphere to generate active bromine (Br, BrO).
While stratospheric bromine concentrations are more than 100 times lower than those of chlorine, bromine is about 60 times more efficient than chlorine in destroying ozone (Sinnhuber et al., 2009), since the chemically inactive reservoir species for bromine, $BrONO_2$ and HBr, are less stable than their chlorine counterparts.
Owing to this large ozone destruction potential of bromine, a potential influence of climate change on bromine concentrations and hence ozone depletion has recently come into focus of research, as warming sea temperatures may lead to enhanced biological activity in the oceans and increased natural bromine emissions (Butler et al., 2007), which may then be transported into the stratosphere more efficiently through enhanced convective activity (Ramanathan and Collins, 1991; Dessens et al., 2009; Aschmann et al., 2011).
Besides the possibility to insert Br directly as catalytic agent X in (2.29) – (2.30), an important cycle coupling Cl_x and Br_x families is (McElroy et al., 1986)

$$\begin{aligned} \mathrm{BrO} + \mathrm{ClO} &\rightarrow \mathrm{Br} + \mathrm{Cl} + \mathrm{O_2} && (2.57)\\ \mathrm{Br} + \mathrm{O_3} &\rightarrow \mathrm{BrO} + \mathrm{O_2} && (2.58)\\ \mathrm{Cl} + \mathrm{O_3} &\rightarrow \mathrm{ClO} + \mathrm{O_2} && (2.59)\\ \hline \mathrm{Net}: 2\,\mathrm{O_3} &\rightarrow 3\,\mathrm{O_2} && (2.60) \end{aligned}$$

This cycle is of special importance for ozone hole chemistry taking place in the polar lower stratosphere, which is described further in Section 2.3.3.
In order to describe the combined effects of a variety of chlorine and bromine containing compounds which possess the potential to destroy stratospheric ozone – also referred to as "ozone depleting substances (ODSs)" – the measures of *equivalent effective chlorine (EECl)* and *equivalent effective stratospheric chlorine (EESC)* are often used. EECl is the sum of tropospheric concentrations of ozone depleting substances, weighted by the number of

chlorine or bromine atoms they contain, and an empirically determined fractional release factor that describes how efficiently the compound is photolysed in the stratosphere (World Meteorological Organization, 2007, chapter 8). Additionally, each Br atom is weighted by a bromine efficiency factor $\alpha \approx 60$ to express the higher ozone depleting potential of bromine as compared to chlorine. EESC is essentially the same as EECl, but for the stratosphere; here, tropospheric concentrations are delayed by an age of air spectrum to account for the different transport pathways and residence times in the stratosphere. Consequently, while EECl is uniquely defined by observations, EESC curves differ according to the assumptions taken. The peak in EESC is usually seen around 1997 for mid-latitudes and 2001 for polar latitudes (Newman et al., 2007).

Limits to catalytic cycle efficiency

Due to interaction between different catalytic families, the effects of different cycles are not additive. The efficiency of the individual catalytic ozone destruction cycles is mainly limited by conversion of chemically active species (HO_x, NO_x, Cl_x, Br_x) into inert reservoir species, e.g. by the following reactions:

$$\begin{aligned}
HO_2 + HO_2 &\rightarrow H_2O_2 + O_2 && (2.61)\\
NO_2 + OH + M &\rightarrow HNO_3 + M && (2.62)\\
Cl + CH_4 &\rightarrow HCl + CH_3 && (2.63)\\
ClO + NO_2 + M &\rightarrow ClONO_2 + M && (2.64)\\
BrO + NO_2 + M &\rightarrow BrONO_2 + M && (2.65)
\end{aligned}$$

The inactive reservoir species on the right hand side of equations (2.61) – (2.65) are then slowly transported from the lower stratosphere to the troposphere and may eventually be washed out, unless they are re-converted into chemically active radicals by photolysis or oxidation beforehand. All reactions discussed so far take place in the gas phase. There are, however, additional pathways for activating chlorine and deactivating nitric oxides on aerosol surfaces, which play a crucial role for the formation of the ozone hole. These are discussed in Section 2.3.3.

2.3.3. Polar heterogeneous chemistry

At high latitudes of both hemispheres, atmospheric conditions are reached in the lower stratosphere during winter which allow for unique chemistry taking place, leading to catastrophic ozone destruction in spring. As discussed in Section 2.2.4, a strong circumpolar westerly wind regime known as the polar vortex develops at high latitudes each fall, providing for secluded and cold air masses encircling the pole during polar night. In the lower stratosphere, temperatures become so low ($\lesssim 195\,\mathrm{K}$) that polar stratospheric clouds (PSCs; McCormick et al., 1982) can form. These PSCs are formed by aerosol particles consisting essentially of nitric acid (HNO_3) and water, forming a solid nitric acid trihydrate (NAT) cluster (type Ia), a supercooled ternary solution of HNO_3, H_2SO_4, and H_2O (STS, type Ib), or pure water (type II). Critical temperatures for the formation of type I and type II PSCs are around 195 K and 185 K, respectively. These temperatures are reached in the Antarctic

polar vortex regularly during winter, whereas in the Arctic PSC occurrence is more episodic and variable.
On the surfaces of PSC particles, heterogeneous reactions can take place which have two important effects: The activation of chlorine from inactive reservoir species (see Section 2.3.2), and the removal of NO_x from the system by the production of HNO_3. These reactions are (Solomon, 1999; Brasseur et al., 1999)

$$\begin{aligned}
ClONO_{2_{(g)}} + HCl_{(s)} &\rightarrow HNO_{3_{(s)}} + Cl_{2_{(g)}} && (2.66)\\
ClONO_{2_{(g)}} + H_2O_{(s)} &\rightarrow HNO_{3_{(s)}} + HOCl_{(g)} && (2.67)\\
HOCl_{(g)} + HCl_{(s)} &\rightarrow Cl_{2_{(g)}} + H_2O_{(s)} && (2.68)\\
N_2O_{5_{(g)}} + H_2O_{(s)} &\rightarrow 2\,HNO_{3_{(s)}} && (2.69)\\
N_2O_{5_{(g)}} + HCl_{(s)} &\rightarrow ClONO_{2_{(g)}} + HNO_{3_{(s)}} && (2.70)
\end{aligned}$$

where subscripts indicate a reactant in solid (s) or gas (g) phase. When HNO_3 is trapped in PSC particles and eventually removed by sedimentation, nitric oxides are no longer available to convert active chlorine and bromine into reservoir species, a situation which is known as *denitrification* (Waibel et al., 1999).
During polar night, Cl_2 and HOCl produced by reactions (2.66) – (2.68) accumulate in the polar vortex. With the return of sunlight, these gases are quickly photolysed, releasing Cl,

$$\begin{aligned}
Cl_2 + h\nu &\rightarrow 2\,Cl && (2.71)\\
HOCl + h\nu &\rightarrow OH + Cl && (2.72)
\end{aligned}$$

Chlorine then destroys ozone in the chlorine-bromine cycle (2.57) – (2.59), or by

$$\begin{aligned}
2\ (Cl + O_3 &\rightarrow ClO + O_2) && (2.73)\\
ClO + ClO + M &\rightarrow ClOOCl + M && (2.74)\\
ClOOCl + h\nu &\rightarrow Cl + ClOO && (2.75)\\
ClOO + M &\rightarrow Cl + O_2 + M && (2.76)\\
\hline
\text{Net}: 2\,O_3 &\rightarrow 3\,O_2 && (2.77)
\end{aligned}$$

These two cycles are efficient enough to almost completely destroy Antarctic ozone in the lower stratosphere within a few weeks, leading to a reduction of the ozone column by more than 50% (Solomon, 1999). Typically, ozone depletion sets in around the end of August, and the ozone hole (defined as area with ozone column below 220 DU) reaches its maximum extent in October, with a size of up to $2.7 \cdot 10^7\,km^2$ and local daily minimum values as low as ~ 100 DU. After disappearance of PSCs, the conversion of active chlorine to inactive reservoirs becomes effective again, unless the vortex is severely denitrified. The breakdown of the Antarctic polar vortex around November leads to mixing in of ozone rich air that has accumulated around the vortex during winter, and a spread of the ozone depleted air masses into mid-latitudes (*dilution*). Besides in-situ gas phase chemistry in the middle and upper stratosphere, export of ozone depleted air masses from polar latitudes is thus a major driver of mid-latitude ozone declines during the 1980s and 1990s (see, e.g., Chapter 8 of this thesis).

As the Arctic polar vortex is warmer and less stable, PSC formation and chlorine activation is less widespread than in the Antarctic. Thus, ozone depletion by heterogeneously activated chlorine is less severe than in the Antarctic, but underlies a large year-to-year variability. Owing to asymmetric transport by the Brewer-Dobson circulation, the Arctic ozone column is in general higher than the Antarctic column (500 DU vs. 300 DU, in pre-ozone-hole conditions), and even in years with strong springtime ozone depletion the remaining Arctic average ozone column is well above 220 DU; this limit is touched only locally.

Part II.

Setup

3. The Bremen Chemistry Transport Model

In this chapter, the stratospheric chemistry transport model (CTM) used throughout this thesis is described. It has been introduced by Sinnhuber et al. (2003) and has been used in several studies (Sinnhuber et al., 2006, 2009; Kiesewetter et al., 2010a,b; Aschmann et al., 2009; Aschmann et al., 2011). Although no unique name has been fixed so far, it has been termed "Bremen 3D CTM (B3DCTM)" in some studies, and this convention shall be continued in the current thesis wherever a distinction from other CTMs is necessary.

3.1. Overview

The CTM is a Eulerian model run at a horizontal resolution of $3.75^\circ \times 2.5^\circ$ (longitudinal $\times$ latitudinal). Potential temperature is used as vertical coordinate, with 24 isentropic levels ranging from 330 K to 3000 K (roughly $10 - 55$ km). Levels are listed in Table 3.1, together with their approximate global average pressure and altitude. An illustration of the vertical domain of the model is provided in Figure 3.1. The vertical level spacing amounts to roughly 1 km in the lower stratosphere and then gradually increases to about 5 km in the upper stratosphere. Horizontal advection is driven by analysed wind fields and temperatures from the European Centre for Medium-Range Weather Forecasts (ECMWF). A single re-analysed meteorological data set covering the period of model runs in this study does not exist; hence meteorological fields from the ERA-40 reanalysis (Uppala et al., 2005) are used for the period 1979–1999 and from the ERA-Interim reanalysis (Dee et al., 2011) 1989–2009. The horizontal advection scheme is based on conservation of second order moments (Prather, 1986), a method which is characterised by low diffusion. Vertical transport is derived directly from interactively calculated diabatic heating rates using the MIDRAD scheme (Shine, 1987).

3.2. Gas phase chemistry

In this study, a time-dependent version of the linearised ozone chemistry ("Linoz v2") described by Hsu and Prather (2009) is used. This is an update to the Linoz scheme introduced by McLinden et al. (2000) which has been used in the Bremen CTM before (Kiesewetter et al., 2010a). Although this chemistry scheme is simple, using only one ozone tracer and net production rates parametrised according to ozone volume mixing ratio (O3), temperature (T), and ozone column above the respective grid cell ($CO3$), it has been shown to generate realistic ozone fields (Hsu and Prather, 2009).
In the Linoz scheme, it is assumed that the key tracer families (NO_y, Cl_y) and long-lived source gases (N_2O, CH_4, H_2O) experience low variability and that their small-scale fluctuations do not produce ozone tendencies significantly different from the mean distribution. Under these assumptions, it is possible to expand the ozone tendency to the first order

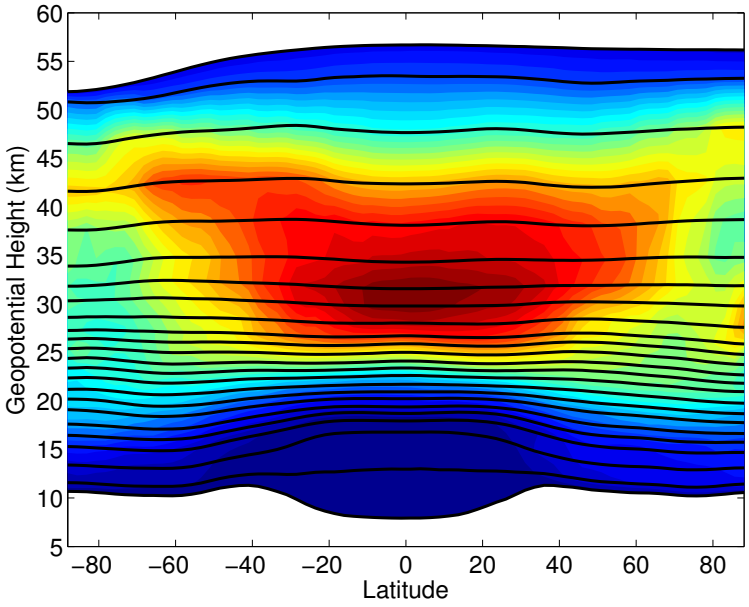

Figure 3.1.: *Vertical domain of the B3DCTM as used in this study. The 24 isentropic levels are shown as black lines. For illustrative purposes, a typical configuration of the ozone layer (ozone mixing ratio increasing from blue to red) is shown as contour plot in the background.*

about the local ozone mixing ratio O3, temperature T, and overhead ozone column $CO3$. For this purpose, the ozone tendency $P - L \equiv dO3/dt$ (production minus loss) is written as a function of these three parameters,

$$\frac{dO3}{dt} = (P - L)[O3, T, CO3] \tag{3.1}$$

and expanded to the first order,

$$\frac{dO3}{dt} = (P-L)_0 + \left.\frac{\partial(P-L)}{\partial O3}\right|_0 \cdot (O3 - O3_0) + \left.\frac{\partial(P-L)}{\partial T}\right|_0 \cdot (T - T_0) + \left.\frac{\partial(P-L)}{\partial CO3}\right|_0 \cdot (CO3 - CO3_0). \tag{3.2}$$

$(P-L)_0$, $O3_0$, T_0, and $CO3_0$ denote climatological values of $P-L$, $O3$, T, and $CO3$, which are tabulated according to month, latitude and log-pressure height. Partial derivatives are evaluated at the climatological values. From $\mathrm{O3} - \mathrm{O3}_0$, $T - T_0$, $CO3 - CO3_0$, a photochemical steady state ozone volume mixing ratio (vmr) in the absence of dynamics can be calculated by demanding

$$\left.\frac{dO3}{dt}\right|_{ss} = 0, \tag{3.3}$$

Level Nr.	Potential Temperature [K]	Approx. pressure [hPa]	Approx. altitude [km]
1	2726	0.2	57
2	2250	0.5	52
3	1857	1	47
4	1533	2	42
5	1265	3	38
6	1044	6	35
7	904	8	32
8	822	11	30
9	746	15	28
10	694	18	27
11	662	21	26
12	631	24	25
13	600	28	24
14	575	33	23
15	550	37	22
16	521	44	21
17	496	52	20
18	475	59	19
19	450	70	18
20	430	81	17
21	410	95	16
22	380	125	15
23	350	188	12
24	335	271	10

Table 3.1.: *Potential temperature levels in the B3DCTM configuration as used in this study, and their approximate pressure and altitude (global averages for the period 1989–2007).*

which is solved by

$$O3_{ss} = O3_0 + \left[(P-L)_0 + \left.\frac{\partial(P-L)}{\partial T}\right|_0 (T-T_0) + \left.\frac{\partial(P-L)}{\partial CO3}\right|_0 (CO3 - CO3_0) \right] \cdot \tau, \quad (3.4)$$

where

$$\tau = -\left[\left.\frac{\partial(P-L)}{\partial O3}\right|_0 \right]^{-1} \quad (3.5)$$

is a time constant. With equations 3.4 and 3.5, equation 3.2 can be rewritten as

$$\frac{dO3}{dt} = \frac{O3_{ss} - O3}{\tau}. \quad (3.6)$$

Consequently, the ozone mixing ratio O3 relaxes against the photochemical equilibrium state $O3_{ss}$ with a time constant τ (the photochemical lifetime of ozone). As an example, Figure 3.2 shows the lifetime of ozone in the Arctic (North of 70°N), demonstrating the large variation

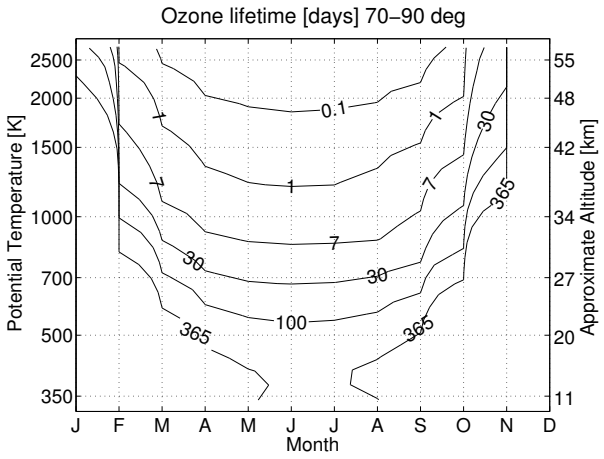

Figure 3.2.: *Ozone lifetime in the Arctic, as calculated from the Linoz scheme (Hsu and Prather, 2009).*

of ozone lifetimes in the middle stratosphere, which range from days (or even hours) in summer to almost infinite during polar night.

Linoz coefficients are based on calculations with a complete photochemical box model (Prather, 1992) that has been updated using state of the art rate coefficients (Sander et al., 2006, also known as JPL 2006 recommendations) and cross sections (Atkinson et al., 2004). Stratospheric tracer concentrations are derived by scaling climatological profiles to the tropospheric abundance of long-lived source gases (N_2O, CH_4, and halocarbons), see Hsu and Prather (2009) for details. Tropospheric source gas mixing ratios are taken from the A1 scenario of the CCMVal 2 Report (Eyring et al., 2010). A detailed validation of the Linoz scheme is provided by Hsu and Prather (2009).

Model simulations presented in this thesis cover three decades (1978–2009). During this period, the chemical composition of the stratosphere experienced significant changes, most importantly concerning the stratospheric halogen loading. As a result of the widespread use of chlorofluorocarbons and halons, concentrations of these substances increased approximately linearly until the early 1990s, when the phase-out of production and emissions required by the Montreal Protocol caused a turnaround which occurred in the troposphere around 1995, and correspondingly a few years delayed in the stratosphere (the delay is latitude-dependent, corresponding to the stratospheric age of air of up to 7 years).

In order to account for the rising and falling stratospheric chlorine loadings, three different Linoz tables are used. They were generated for stratospheric trace gas concentrations corresponding to the years 1978, 2000, and 2010, (more specifically, for stratospheric tracer profiles scaled to tropospheric concentrations of three years earlier, in order to account for the mean stratospheric age of air). These years are chosen as the first and last years of the model integrations, and a "peak-EESC" year in between. The tables are then interpolated linearly in time, leading to a linear increase of EESC by a factor of 2 between 1978 and 1999, and a linear decrease by $\sim 10\%$ until 2009. Equivalent effective chlorine (EECl) calculated from the halocarbon concentrations used for scaling the Linoz coefficients is shown in Figure 3.3. The selection of 1999/2000 as the transition from rising to falling EESC is

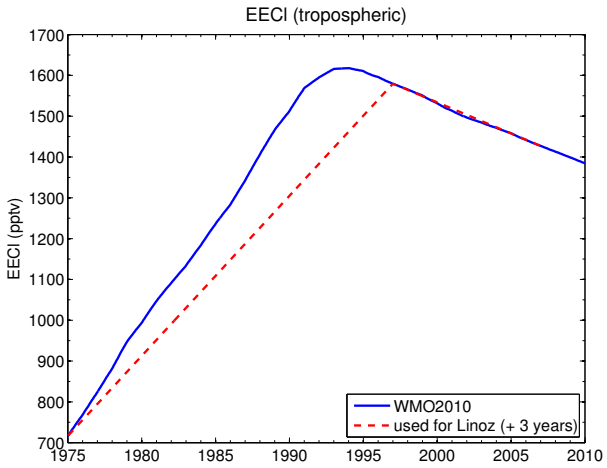

Figure 3.3.: *Tropospheric Effective Equivalent Chlorine (EECl) used to scale the linearised ozone chemistry scheme (Linoz). The red dashed curve is used in Linoz with a time delay of 3 years.*

a compromise between high and low latitudes, rather inclined towards high latitudes. The practical advantage of this selection is that in the assimilation run, the chemistry scheme is most appropriate at high latitudes where satellite data are scarce, while at mid-latitudes and tropical latitudes, small imperfections in the chemistry scheme are cancelled by a constant input of satellite data.

Although Linoz does include mid-latitude heterogeneous ozone chemistry on aerosols, aerosol levels are kept at background levels in this study (year 1990 of the SAGE II climatology, Thomason et al., 1997). Volcanic eruptions, such as El Chichon 1981 and Pinatubo 1991, lead to large enhancements of stratospheric sulphur aerosol via the injection of SO_2 into the stratosphere. This enables heterogeneous chlorine activation on aerosol surfaces similar to processes encountered on polar stratospheric clouds during winter and spring, and thus has been responsible for considerable ozone declines in the mid-latitudes in the years following the eruptions, i.e. early 1980s and 1990s. Ozone changes caused by chemical effects of volcanic aerosols are missed by the chemistry scheme. However, they are in parts present in the external meteorological fields used to drive the model, and are inferred by satellite data in the assimilation run.

Linoz does not account for heterogeneous activation of chlorine or bromine on polar stratospheric clouds. These effects are included in a separate polar chemistry scheme described in Section 3.3.

Consequently, in the sensitivity runs without assimilation, ozone changes induced by Linoz through changing chemical composition are referred to as “changes in gas-phase chemistry” throughout this study.

The underlying idea of Linoz is the “outsourcing” of the computationally expensive chemistry calculations into the generation of the coefficient tables. This concept allows for the inclusion of a reasonable ozone chemistry also in model environments where employing a complex chemistry scheme would not be feasible. Naturally, a parametrised chemistry scheme like Linoz has limitations, e.g. as it cannot account for short-term variations in concentrations of ozone reaction partners. The linearisation around a climatological temperature

profile implies that the coefficients may not be ideal if the temperature profile used in the model – which is taken directly from meteorological reanalyses – differs significantly from the Linoz temperature climatology. Due to its design, the degree of realism achieved with the Linoz scheme is limited in the uppermost stratosphere and above, where ozone follows a daily cycle that is not accounted for in the Linoz scheme. Furthermore, the quality of the meteorological reanalyses is limited in the upper stratosphere, and unrealistic temperatures lead to very unrealistic ozone concentrations in the uppermost CTM levels that may disrupt the whole CTM at single instances. Therefore, temperature dependency of the Linoz scheme is switched off in the uppermost two levels (above 50 km) at all times. Effectively, this is equivalent to setting $T = T_0$ in Equation 3.4.
A full chemistry scheme including 50+ tracers is available for the B3DCTM that has for example been used in recent studies by Aschmann et al. (2009); Aschmann et al. (2011). It is not used in this study for several reasons. In the assimilation run, ozone is altered while all other tracers (if there are any) are left unchanged. However, for a realistic atmospheric tracer configuration, different tracers are not independent, and the internal balances between reactants have to be taken into account. These have been shown to be disrupted due to alteration of only one tracer, and would require additional correction (Chipperfield et al., 2002). Secondly, although computational facilities have been strongly improved during recent years, computing time is still an issue. In particular, for the several sensitivity runs without assimilation, use of the full chemistry scheme was beyond computational resources available; moreover, the simplified chemistry scheme is an advantage in the sense that it allows for a simpler attribution of ozone changes to their processes of origin, while in the complete chemistry scheme different processes are more difficult to separate.
Nonetheless, use of the full chemistry scheme for assimilation may become an interesting option in the near-term future in order to include other tracers in the assimilated data set.

3.3. Polar heterogeneous chemistry

The effects of heterogeneous chlorine and bromine activation, which are not contained in the Linoz scheme, are included in the CTM in the form of a simple parametrised polar chemistry. As discussed in Chapter 2, heterogeneous activation of chlorine takes place in polar winter and spring, when temperatures in the lower stratospheric polar vortex get cold enough to allow the formation of polar stratospheric clouds. As mentioned in Section 2.3.3, critical temperatures for the formation of type I and type II PSCs are around 195 K and 185 K, respectively. Since the onset of chlorine activation is determined by particles which form first in a cooling stratosphere, it is reasonable to assume only one threshold temperature, T_{NAT}, for the onset of chlorine activation. This is supported by observations investigating the process of chlorine activation (Santee et al., 1997; Kühl et al., 2004). T_{NAT} is the highest PSC formation temperature that is thermodynamically possible. In the polar chemistry module, T_{NAT} is calculated from climatological HNO_3 and H_2O concentrations.
When conditions for PSC formation are met in a CTM grid box, i.e. when $T < T_{NAT}$, and enough sunlight is present, ozone is destroyed exponentially at an empirical rate. Hence, the overall effects of heterogeneous activation of halogens and subsequent catalytic ozone destruction given in Section 2.3.3 are expressed by a single (pseudo first order) reaction rate r_{pc}. The ozone lifetime $\tau_{pc} \equiv 1/r_{pc}$ and critical solar zenith angle θ_{crit} are tuned to achieve best agreement of lower stratospheric ozone to sonde observations at NyÅlesund in the cold

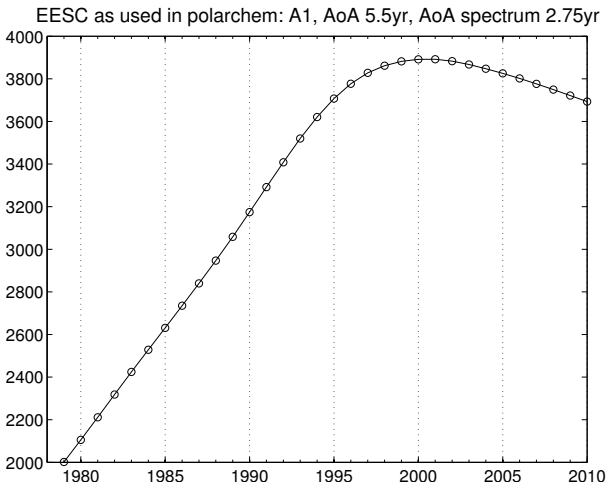

Figure 3.4.: *Equivalent Effective Stratospheric Chlorine (EESC) used to scale the empirical ozone decay rate in the polar heterogeneous chemistry scheme.*

Arctic winter of 1999/2000. Optimal parameters were chosen as $\tau \equiv 1/r_{pc} = 10\,\text{days}$ (for 2000) and critical solar zenith angle $\theta_{crit} = 92.5°$ (note that due to Earth's curvature, light is present at solar zenith angles as high as $\sim 94°$ in the lower stratosphere).
In order to account for the changing stratospheric halogen loading in the stratosphere during the period of model integrations, the ozone decay rate r_{pc} is scaled with EESC, calculated as described by Newman et al. (2006) (Age of air = 5.5yr, Age spectrum width = 2.75 yr, Bromine scaling factor $\alpha = 50$). This EESC curve, which is shown in Figure 3.4, almost doubles between 1978 and 2000, peaks in 2001, and then decreases by $\sim 5\%$ until 2009.
The dependence of the ozone decay rate on EESC is usually estimated to be close to linear (Harris et al., 2010); however, there are also arguments in favour of a stronger dependence (Searle et al., 1998; Hsu and Prather, 2009). Reasonable values for the exponent n ($r_{pc} = 1/\tau \propto \text{EESC}^n$) range between 1 and 2 (Searle et al., 1998). In the assimilation run, n is chosen as 1. In the sensitivity runs, either $1/\tau \propto \text{EESC}$ or constant year 2000 conditions are used. In order to obtain an upper estimate for trends related to polar chemistry, one additional model run uses a higher scaling of $1/\tau \propto \text{EESC}^2$.

4. Data Assimilation

Originally developed for use in numerical weather prediction, data assimilation has become a mighty tool in atmospheric and oceanographic sciences during recent decades. The general idea is both straightforward and intriguing: Combine two entirely different forms of describing natural processes, namely computer models and observations. While computer models can only provide a simplified representation of our best understanding of how nature works, they have the big advantage of global and temporal coverage. Observations, on the other hand, are the only way of directly gaining knowledge about nature; however, there are practical and fundamental limitations in the coverage observed data sets may provide. A measurement will always be a snapshot taken at a certain place and time, and no observer will ever be able to observe every parameter of interest at all possible locations at once. Satellite observations, for instance, have greatly improved our knowledge about a vast variety of atmospheric parameters during the last ~ 40 years. However, most observational concepts rely on backscattered sunlight (such as the SBUV instruments which are used in this study), which is not available for several months at high latitudes every winter. Other observational methods have been developed which avoid this problem, but then face other difficulties – e.g., sondes provide highly accurate in-situ measurements of ozone, but only at distinct locations; satellite remote sensors which measure emitted rather than backscattered radiation (and are thus able to operate during polar night) have not been able to deliver continuous long time series of observations so far. Hence, data gaps are a fundamental issue in the atmospheric sciences, and data assimilation provides one way of dealing with it.
Let us assume that we have at hand two things – a set of observations of a physical quantity (e.g., satellite observations of ozone mixing ratio), and a computer model that is able to predict how this physical quantity will evolve in space and time, to the best of our understanding (e.g., the chemical transport model used in this study). Then we can use the information contained in the observations to pull the model towards reality, and at the same time the model to extend the information from the observations into areas where no observations are available. This integration of external observational information into the model is the process of data assimilation. In the end, we arrive at a global "assimilated" state that has the coverage of the computer model but contains all the information present in the observations. This assimilated state, also called the *analysis*, represents the *optimal estimate* for the state of the observed quantity, in our case the global distribution of ozone.
It should be emphasised that assimilation does more than a simple interpolation of observed data. Assimilation takes into account the temporal evolution of the modelled quantity between inputs of data, and provides for a physically adequate propagation of information beyond the limits of data coverage. In the case of satellite observations, hours may lie between single measurements at different locations. The ozone field may look entirely different at position A when the satellite has moved from A to B, and thus closing the gaps by interpolation – e.g. to provide a daily ozone field – can in fact never represent the state of the atmosphere correctly at a single time. Moreover, interpolation cannot solve the problem of data gaps during darkness: Extrapolating satellite data into polar night, although math-

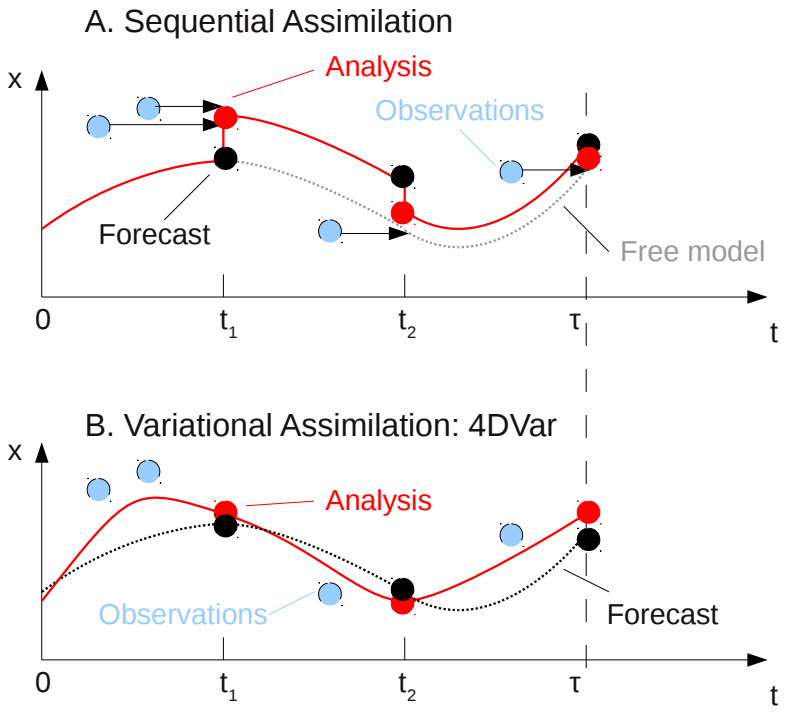

Figure 4.1.: *Assimilation schemes: Sequential versus continuous variational (4DVAR). The sequential scheme produces jumps at the analysis times t_k, while the variational analysis is smooth and consistent with model evolution. Realisations of forecasts and analysis at the sequential assimilation time steps t_k are marked with black and red dots, respectively. Note that these sequential assimilation intervals t_k do not necessarily encompass every model time step (the model output may be finer in time).*

ematically possible, does not make sense, as an adequate representation of chemistry and transport is needed.

Different assimilation methods have been developed, which can be basically distinguished into sequential and variational schemes. A schematic comparison is given in Figure 4.1. In a sequential scheme, as it is used in this study, the model computes a forecast of the quantity of interest, which is then "updated" periodically by a weighted comparison against observational data that have been gathered during the period of model integration (the "assimilation window"). The updated values, called "analysis", are then passed back to the model as initial conditions for the next iteration. Hence, the model output will be subject to jumps at the end of every assimilation window that are not consistent with the physical equations contained in the model. Conversely, a continuous assimilation scheme such as the 4D variational scheme (4DVAR, Fisher and Lary, 1995), which has been frequently used (e.g. Khattatov et al., 1999; Lahoz et al., 2007; Errera et al., 2008), requires compliance of the assimilated state with the model evolution. Here the analysis state is obtained by varying

the complete model trajectory in the considered assimilation interval (extending from $t = 0$ to $t = \tau$ in Figure 4.1). Also observations from the future are taken into account in order to determine the analysis at a certain time t, thus ensuring a smooth evolution of the analysis, which may be desirable. The downside of the 4DVAR method is that it requires the model to be perfect, whereas the sequential scheme explicitly accounts for uncertainties in the model. Khattatov et al. (1999) explicitly compared sequential and 4DVAR assimilation schemes and concluded that the sequential scheme produced an analysis that was in better agreement with the observations where available, due to the missing perfect model assumption. Furthermore, as noted by Khattatov et al. (1999), the inclusion of an internal error tracer is a strong point of the sequential assimilation method.
With the relatively simple chemistry scheme used in our CTM in this study, we cannot expect it to provide a perfect description of nature. Thus, the sequential assimilation method is better suited to address our objective of generating a data set that is strongly constrained by observations but extends their information into temporal and spatial data gaps.
The sequential assimilation scheme is introduced in Section 4.1, and its implementation in the B3DCTM is described in 4.2. A stringent derivation of the mathematics in Section 4.1 is given in Section 4.3.

4.1. Sequential data assimilation

The method of assimilation chosen for this study is sequential assimilation, which is equivalent to a simplified Kalman filtering scheme (Kalman, 1960). Essentially, this scheme consists of two steps which are repeated continuously: forecast step and analysis step. Each forecast step (which is in our case identical to the model time step, 30 min), the CTM produces a forecast of the state of the atmosphere, denoted as a vector of ozone mixing ratios $\mathbf{x^f}(t_k) \equiv \mathbf{x^f}_k$ at time t_k. Along with the forecast of the ozone tracer itself, the variance of the tracer field as caused by errors in transport, chemistry and discretisation is transported in the model to provide an error covariance forecast matrix $\mathbf{B^f}_t$. At the same time, observations of the "true" (unknown) state of the atmosphere $\mathbf{x^t}_k$ are made, denoted by the vector $\mathbf{y}_k$. These observations are then assimilated in the analysis step, resulting in an updated ozone field, the "analysis" $\mathbf{x^a}_k$. At the same time, an updated covariance $\mathbf{B^a}_k$ is computed. Both $\mathbf{x^a}_k$ and $\mathbf{B^a}_k$ are then passed back to the model and used as initial conditions for computing the next forecast at time t_{k+1}, $\mathbf{x^f}_{k+1}$ and $\mathbf{B^f}_{k+1}$. This repetitive scheme is illustrated in Figure 4.2. During the forecast period, the increase of uncertainty in the state vector $\mathbf{x}$ due to model imperfections is accounted for by an error growth term which is added to $\mathbf{B^f}$ after model integration. Due to the sheer size of $\mathbf{B}$, evolution of the full covariance matrix is not feasible in the model, and the diagonal elements or variances are evolved instead.
The following sections provide a mathematical formulation of the assimilation scheme, divided into forecast step (Section 4.1.1) and analysis step (Section 4.1.2). The essential quantities involved in the assimilation scheme (at a given time t_k) are

- $\mathbf{x^f} \in \mathbb{R}^{n_{mod}}$ ... the model forecast, which is of dimension $n_{mod} = n_{lon} \times n_{lat} \times n_{lev} = 96 \times 72 \times 24 = 165888$ = number of model grid cells ($n_{lon}, n_{lat}, n_{lev}$ ... numbers of longitudes, latitudes and vertical levels in the model)
- $\mathbf{x^t} \in \mathbb{R}^{n_{mod}}$ ... the unknown true state of the atmosphere, discretised to the model grid
- $\mathbf{x^a}$ ... the analysed state or best estimate for $\mathbf{x^t}$

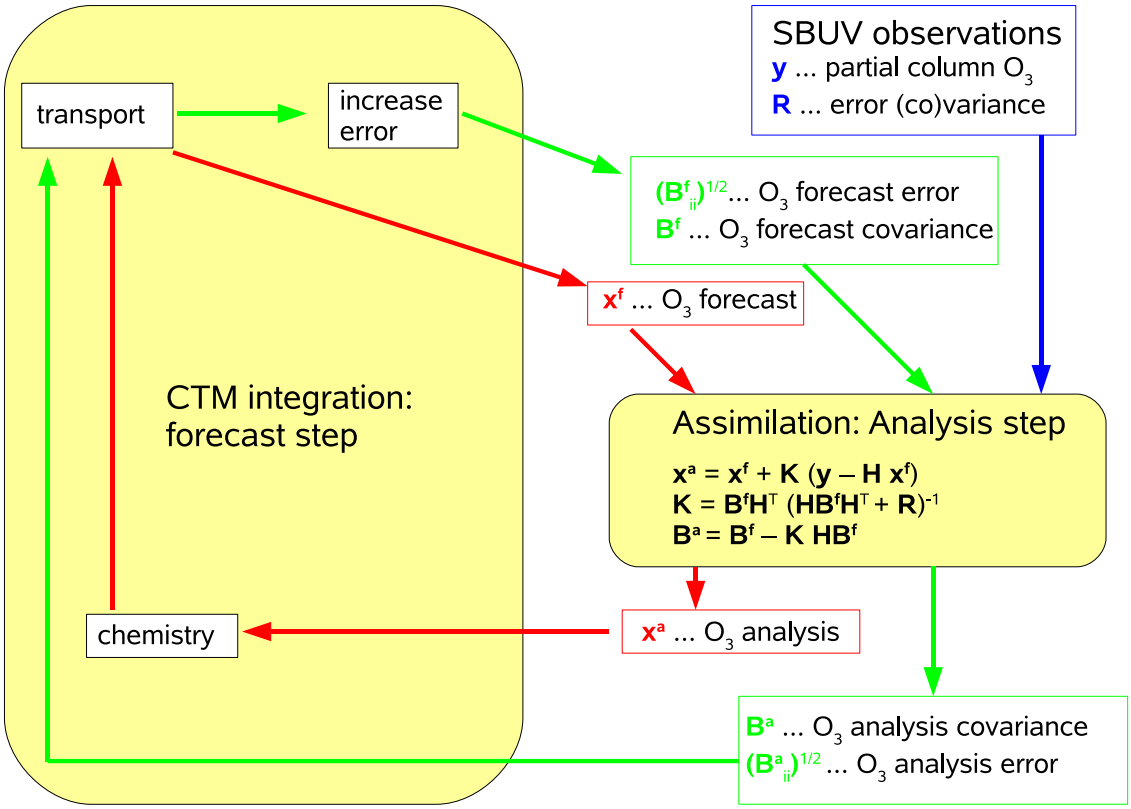

Figure 4.2.: *The sequential assimilation scheme (Kalman filter) used in this study. The Kalman filter is divided into two steps which are repeated periodically: the forecast step (CTM integration) and the analysis step, in which all satellite observations made during the forecast interval are assimilated. Red and green arrows refer to the evolution of the ozone tracer and its covariance, respectively. For technical reasons, the square root of the variance, or "error", is transported in the forecast step.*

- $\mathbf{B^f} \in \mathbb{R}^{n_{mod} \times n_{mod}}$ ... the forecast covariance matrix
- $\mathbf{B^a} \in \mathbb{R}^{n_{mod} \times n_{mod}}$ ... the analysis covariance matrix
- $\mathbf{y} \in \mathbb{R}^{n_{obs}}$ ... the vector of observations, which is of dimension $n_{obs} = n_{obsloc} \cdot n_{obslev} =$ total number of observation points (n_{obsloc} ... number of observation locations, n_{obslev} ... number of vertical observation levels. For SBUV observations $n_{obs} < 1000$ in the 30 min assimilation window)
- $\mathbf{R} \in \mathbb{R}^{n_{obs} \times n_{obs}}$ ... the observational error covariance matrix

Exact definitions of the covariance matrices require a background of conditional probabilities and are given in Section 4.3. Note that different states (forecast, analysis, true) are indicated

as upper indices, while a lower index will be used to indicate time dependency. Since the assimilation process takes place at discrete points in time, this is done by a discrete index k corresponding to times t_k ($k = 1 \dots m$, with m total number of model/assimilation time steps).

4.1.1. Forecast step

As described above, the essence of the Kalman filter is that the analysis state at a certain time t_k, $\mathbf{x^a}_k$, is used as initial condition for the calculation of the forecast state at the next model time step, $\mathbf{x^f}_{k+1}$. Evolution of the state vector from time t_k to t_{k+1} is given by the (abstract) model integration operator $\mathbf{m}$, so that

$$\mathbf{x^f}_{k+1} = \mathbf{m}(\mathbf{x^a}_k) = \mathbf{M}\mathbf{x^a}_k. \tag{4.1}$$

Together with the state vector, its error covariance matrix $\mathbf{B}$ is evolved from the analysis covariance at time t_k, $\mathbf{B^a}_t$, to the forecast covariance at time t_{k+1}, $\mathbf{B}_{k+1}$. In a strict mathematical sense, the evolution of the error covariance matrix involves the tangent linear model

$$\mathbf{M} = \frac{d\mathbf{x^f}_{k+1}}{d\mathbf{x^a}_k} \tag{4.2}$$

and is computed by

$$\mathbf{B^f}_{k+1} = \mathbf{M}\mathbf{B^a}_k\mathbf{M}^{\mathrm{T}} + \mathbf{Q}, \tag{4.3}$$

$\mathbf{Q}$ denoting some kind of error increase due to imperfections in the model.
However, since the dimension of the covariance matrix $\mathbf{B}$ is $n_{mod} \times n_{mod} = 165888 \times 165888$, the evolution of the covariance cannot be calculated as in (4.3), but has to be simplified. Following the treatment of Khattatov et al. (2000), only the diagonal elements b^a_{ii} of $\mathbf{B^a}$ are transported in the model integration (Equation 4.3) to obtain the diagonal elements b^f_{ii} of the forecast variance $\mathbf{B^f}$. For the analysis step (Equation 4.10), the off-diagonal elements (covariances) are parametrised as

$$b_{ij} = \sqrt{b_{ii}b_{jj}}\ exp\left(-\frac{\Delta r^2_{xy}}{2L^2_{xy}}\right) \cdot exp\left(-\frac{\Delta r^2_z}{2L^2_z}\right), \tag{4.4}$$

with Δr_{xy} and Δr_z distances in horizontal and vertical direction. L_{xy} and L_z are tunable coherence lengths. During model integration, the analysis error variance $b^a_{ii,t}$ is first transported as a tracer,

$$\tilde{b}^f_{ii,k+1} = \mathbf{m}b^a_{ii,k} \tag{4.5}$$

and then increased by the error growth term q^t_{ii} in order to account for the growing uncertainty in the modelled tracer in the absence of data,

$$b^f_{ii,k+1} = \tilde{b}^f_{ii,k+1} + q_{ii,k} \tag{4.6}$$

$$q_{ii,k} = \left[\epsilon \cdot x^f_{i,k+1} \cdot \Delta t\right]^2. \tag{4.7}$$

4.1.2. Analysis step

The relation between the true state of the atmosphere at time t_k, $\mathbf{x^t}_k$, and the observations $\mathbf{y}_k$ can be expressed by the observation operator $\mathbf{H}$,

$$\mathbf{y}_k = \mathbf{H}(\mathbf{x^t}_k) + \varepsilon_k^m \tag{4.8}$$

where ε_k^m is the overall error encountered during the measurement process, composed of a measurement error ε_k^o and a representativeness error ε_k^r,

$$\varepsilon_k^m = \sqrt{(\varepsilon_k^o)^2 + (\varepsilon_k^r)^2}\ , \tag{4.9}$$

with covariance matrices $\mathbf{O}$ and $\mathbf{R}$, respectively (in practise, both are assumed diagonal).
Note that the "true" state $\mathbf{x^t}_k$ is actually a discretisation of the true state of the continuous atmospheric parameter $\mathbf{x}_k$, mapped onto the spatial domain of the model. For simplicity, the discussion here is limited to the discrete spaces of observations and model. For a more thorough discussion of this issue, the reader is referred to Cohn (1997). Essentially, the assimilation process performs an inversion of Equation (4.8) to solve for $\mathbf{x^t}_t$ – which is not possible in the strict mathematical sense since Equation (4.8) is, in general, both under- and overdetermined: The size of the data vector $\mathbf{y}_k$ is usually much smaller than the size of the model state $\mathbf{x^t}_t$ to be determined; however, in data-dense areas, observations may be located closer than the model grid and hence Equation 4.8 may be locally over-determined. Inversion of Equation 4.8 is performed by the process of optimal estimation (Rodgers, 2000) to provide the analysed (or "optimal") state $\mathbf{x^a}_k$ (called "analysis" in the following), which is a best linear unbiased estimator to $\mathbf{x^t}_k$. From optimal estimation theory, the analysis is given by (Chipperfield et al., 2002)

$$\mathbf{x^a}_k = \mathbf{x^f}_k + \mathbf{K}_k \left(\mathbf{y}_k - \mathbf{H}\mathbf{x^f}_k\right) \tag{4.10}$$

where

$$\mathbf{K}_k = \mathbf{B^f}_k\mathbf{H}^{\mathrm{T}} \left(\mathbf{H}\mathbf{B^f}_k\mathbf{H}^{\mathrm{T}} + \mathbf{O} + \mathbf{R}\right)^{-1}\ . \tag{4.11}$$

$\mathbf{K}_k$ is called the Kalman gain. The analysis error covariance $\mathbf{B^a}_k$ is given by

$$\mathbf{B^a}_k = \mathbf{B^f}_k - \mathbf{B^f}_k\mathbf{H}^{\mathrm{T}} \left(\mathbf{H}\mathbf{B^f}_k\mathbf{H}^{\mathrm{T}} + \mathbf{O} + \mathbf{R}\right)^{-1} \mathbf{H}\mathbf{B^f}_k\ . \tag{4.12}$$

An explicit derivation of the Kalman filter forecast and analysis steps is provided in Section 4.3.
A goodness-of-fit diagnostic χ^2 is calculated along with the analysis,

$$\chi_k^2 = (\mathbf{y}_k - \mathbf{H}_k\mathbf{x^a}_k)^{\mathrm{T}}(\mathbf{H}_k\mathbf{B^f}_k\mathbf{H}_k^{\mathrm{T}} + \mathbf{O} + \mathbf{R})^{-1}(\mathbf{y}_k - \mathbf{H}_k\mathbf{x^a}_k)\ . \tag{4.13}$$

As shown in Section 4.3 below, this diagnostic effectively compares the Kalman filter covariance to the innovation, that is, the part of the current observation not already explained by past observations. In case of Gaussian error statistics, it is χ^2 distributed with expectation value $n_{obs,k}$. This diagnostic is used to tune the empirical assimilation parameters ϵ (error growth rate), L_{xy} and L_z (forecast error correlation lengths).

4.2. Implementation of the assimilation scheme

The sequential assimilation scheme used here is basically an update from Khattatov et al. (2000) and Chipperfield et al. (2002). A version of the code has been provided by Prof. Martyn Chipperfield (personal communication) and was modified by the author to be implemented in the Bremen 3D CTM.

Besides a shared-memory parallelisation of the code that led to significant speedup, essential modifications concern the construction of a proper observation operator **H** which maps the model space to data space. In the case of direct assimilation of satellite observations, this is basically a horizontal and vertical interpolation operator. Since SBUV are nadir-viewing instruments, their vertical resolution is rather coarse, amounting to roughly 7 km in the middle stratosphere (Bhartia et al., 1996). Thus, a vmr measurement may not be treated as a point-like observation at a certain pressure level (as it may be done in the case of, e.g., solar occultation measurements which offer a much higher vertical resolution). In the retrieval process this limited resolution is expressed by averaging kernels, which are specific to each observed profile and would also be needed for constructing the correct observation operator between the SBUV and CTM domain, along with the a priori profile used in the retrieval process. Since SBUV averaging kernels and a priori profiles are not publicly available, in this study an alternative method is employed by assimilating not vmrs but partial columns, which have already been integrated vertically to a coarser resolution, and by introducing the representativeness error ε^r described in Section 4.1.2. In this case, the observation operator consists of a horizontal interpolation of model ozone mixing ratios to the observation location, and a vertical integration to obtain partial columns between the same pressure levels the SBUV retrievals use. Since the lowest SBUV partial column includes the whole troposphere which is not included in our model, and also contains little information due to the instrument configuration (McPeters et al., 1984), only 12 out of the 13 partial columns provided are assimilated.

The observational error ε^o includes accuracy and precision of the instruments. In this study, the values reported by Bhartia et al. (1996) are used, which are around 5% in the middle stratosphere (see Chapter 5). The representativeness error is more difficult to estimate, since it must include not only the effects of different vertical grids (profile structures smaller than the observational grid are smeared out and thus lost by the observation operator) but also the vertical over-sampling of the satellite data since the model forecast profile is not convolved with averaging kernels before comparing it to the satellite data. An estimate of the magnitude of the observational error was derived by comparing CTM ozone profiles directly integrated to the SBUV partial column layers to the same CTM profiles, integrated by convoluting them with an idealised Gaussian-shaped averaging kernel of 8 km full width of half maximum. The resulting absolute value of mean deviations was then assumed as representativeness error and added to the observational error. Values of observational and representativeness errors are given in Table 5.2.

The tunable error correlation lengths were chosen in accordance with Chipperfield et al. (2002), who used a very similar CTM to assimilate ozone observations. Parameter values are given in table 4.1. The numerical value for the error growth factor $\epsilon = 0.01/\text{hour}$ is an update from Chipperfield et al. (2002) to ensure a value of $\tilde{\chi}^2 \equiv \chi^2/n_{obs}$ close to 1. As the long-term mean of the χ^2 diagnostic is close to 1 (see Figure 6.10), the treatment of errors and error growth is reasonable in this assimilation scheme.

Parameter	Symbol	Value
Horizontal error correlation	L_{xy}	1000 km
Vertical error correlation	L_z	2.8 km
Error growth	ϵ	0.01/hour

Table 4.1.: *Overview of the most important assimilation parameters.*

4.3. Derivation from estimation theory

While Section 4.1 provided an overview of the mathematics used in data assimilation, and Section 4.2 described the practical implementation in our model, this section gives a more exact and rigorous derivation of the mathematics in Section 4.1 from the principles of estimation theory.

4.3.1. Principals of estimation theory

Basic model equations revisited

First, let us re-state the basic equations governing the assimilation process. Following Cohn (1997), we can describe the evolution of the unknown true state of the atmosphere, $\mathbf{x^t}$, from time t_{k-1} to t_k by a propagator $\mathbf{m}$,

$$\mathbf{x^t}_k = \mathbf{m}(\mathbf{x^t}_{k-1}) + \varepsilon^{\mathbf{t}}{}_{k-1}, \tag{4.14}$$

where $\varepsilon^{\mathbf{t}}{}_{k-1}$ denotes the model error from time t_{k-1} to t_k, caused by errors in discretisation, physics and chemistry. In general, $\varepsilon^{\mathbf{t}}$ is dependent on the true state, and is thus not only unknown but also unknowable in a deterministic sense. However, for the model $\mathbf{m}$ to be reasonable, we shall expect that $\varepsilon^{\mathbf{t}}{}_{k-1}$ is small in some sense, i.e. that $\mathbf{m}$ approximates the real evolution of the (continuum) true state well. In this case, $\varepsilon^{\mathbf{t}}{}_{k-1}$ can be treated as a small perturbation to Equation 4.14. For convenience, we shall assume here that $\varepsilon^{\mathbf{t}}{}_{k-1}$ follows a Gaussian distribution with zero mean and covariance matrix $\mathbf{Q}_{k-1}$, and that it is white in time,

$$\varepsilon^{\mathbf{t}}{}_k \propto N(\mathbf{0}, \mathbf{Q}_k), \left\langle \varepsilon^{\mathbf{t}}{}_k (\varepsilon^{\mathbf{t}}{}_\ell)^T \right\rangle = \mathbf{0} \quad \forall k \neq \ell, \tag{4.15}$$

where N represents the normal distribution and $\langle . \rangle$ denotes the expectation value. In practise, this assumption is problematic, but since little is known about the model error statistics, it is still the most reasonable one can make.
The second basic equation is the discrete stochastic observation model, basically Equation 4.8, which describes what happens in case of a measurement. The vector of observations $\mathbf{y}_k$ is related to the true state of the atmosphere $\mathbf{x^t}_k$ via the observation operator $\mathbf{H}$,

$$\mathbf{y}_k = \mathbf{H}(\mathbf{x^t}_k) + \varepsilon^m_k. \tag{4.16}$$

We shall assume here that the observation operator $\mathbf{H}$ is linear, which is well justified for the assimilation of a directly measured quantity such as ozone mixing ratio. In the assimilation

scheme used here, $\mathbf{H}$ is actually a linear interpolation in horizontal and vertical direction. Again, as a simplification, we assume the measurement error ε_k^m to be white in time,

$$\left\langle \varepsilon_k^m (\varepsilon_\ell^m)^T \right\rangle = \mathbf{0} \quad \forall k \neq \ell \ , \tag{4.17}$$

and Gaussian distributed with zero mean,

$$\varepsilon_k^m \propto N(\mathbf{0}, \mathbf{R}_k), \tag{4.18}$$

i.e. we assume the observations to be *unbiased* – which we know will probably not be fulfilled in practise, but is accepted here in order to keep things simple. Although theoretically biases might be corrected before assimilation, in practise little information is available about their magnitudes, and the assumptions of both unbiased observations as well as unbiased model is often the only reasonable approach. $\mathbf{R}_k$ in Equation 4.18 represents the total measurement error covariance matrix. In the discussion in Section 4.1, the sum $\mathbf{R} + \mathbf{O}$ was used instead of $\mathbf{R}$ alone to express the fact that the total measurement error can be split up into a representativeness error and an observational error. For notational convenience, only the total measurement error covariance will be used in this section to derive the Kalman filter equations. The equations in Section 4.1 are then obtained by replacing $\mathbf{R}$ with $\mathbf{R} + \mathbf{O}$.

Conditional probabilities

Since the true state of the atmosphere which we would like to estimate, $\mathbf{x^t}_k$, is given by a stochastic-dynamic model (Equation 4.14), it follows a probability distribution function $p(\mathbf{x^t}_k)$. Basically, the aim of the estimation process is to determine this probability distribution function $p(\mathbf{x^t}_k)$ – or, more precisely, its maximum, as we would like to know the best, or optimal, estimate $\tilde{\mathbf{x}}_k^{\mathbf{t}} \equiv \mathbf{x^a}_k$ for $\mathbf{x^t}_k$. In the absence of observations, the complete solution of the prediction problem is given by the unconditional density $p(\mathbf{x^t}_k)$. However, if observations are available, the desired density of $\mathbf{x^t}_k$ becomes a conditional one, since we have to take into account all observations made up to some time t_ℓ (usually $\ell = k$ for the analysis step and $\ell = k - 1$ for the forecast step, see below):

$$p(\mathbf{x^t}_k) = p(\mathbf{x^t}_k | \mathcal{Y}_\ell), \tag{4.19}$$

where $\mathcal{Y}_\ell$ denotes the set of all realisations of observations made up to time t_ℓ,

$$\mathcal{Y}_\ell := \{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_\ell\} \, . \tag{4.20}$$

The conditional probability density of $\mathbf{a}$ given $\mathbf{b}$ is defined as (Jazwinski, 1970; Cohn, 1997)

$$p(\mathbf{a}|\mathbf{b}) \equiv p_{\mathbf{a}|\mathbf{b}}(\mathbf{a}, \mathbf{b}) \equiv \frac{p_{\mathbf{a},\mathbf{b}}(\mathbf{a}, \mathbf{b})}{p_{\mathbf{b}}(\mathbf{b})} \tag{4.21}$$

On the right side of Equation 4.21, the probabilities and their dependencies are mentioned explicitly (e.g. $p_{\mathbf{a},\mathbf{b}}(\mathbf{a}, \mathbf{b})$ is the probability of $\mathbf{a}$ *and* $\mathbf{b}$, depending on $\mathbf{a}$ and $\mathbf{b}$). For convenience, usually only the short (left) notation is used further on.
The expected value of $\mathbf{a}$ given $\mathbf{b}$ is then defined as

$$\langle \mathbf{a}|\mathbf{b} \rangle \equiv \int \alpha \; p_{a|b}(\alpha|\mathbf{b}) \; d\alpha. \tag{4.22}$$

4.3.2. Solution: The Kalman filter

The estimation problem may now be re-stated using the definitions in Section 4.3.1. The forecast and forecast covariance at time t_k are the mean and covariance matrix, respectively, of the conditional probability distribution $p(\mathbf{x^t}_k|\mathcal{Y}_{k-1})$ of $\mathbf{x^t}_k$ taking into account all observations up to time t_{k-1},

$$\mathbf{x^f}_k \equiv \left\langle \mathbf{x^t}_k \middle| \mathcal{Y}_{k-1} \right\rangle \ , \tag{4.23}$$

$$\mathbf{B^f}_k \equiv \left\langle \left(\mathbf{x^t}_k - \mathbf{x^f}_k\right) \left(\mathbf{x^t}_k - \mathbf{x^f}_k\right)^T \middle| \mathcal{Y}_{k-1} \right\rangle \ . \tag{4.24}$$

Conversely, the analysis and analysis covariance at time t_k are nothing else but the mean and covariance matrix of the probability density of $\mathbf{x^t}_k$ given all observations up to time t_k,

$$\mathbf{x^a}_k \equiv \left\langle \mathbf{x^t}_k \middle| \mathcal{Y}_k \right\rangle \ , \tag{4.25}$$

$$\mathbf{B^a}_k \equiv \left\langle \left(\mathbf{x^t}_k - \mathbf{x^a}_k\right) \left(\mathbf{x^t}_k - \mathbf{x^a}_k\right)^T \middle| \mathcal{Y}_k \right\rangle \ . \tag{4.26}$$

Now the the Kalman filter may be derived as a minimum variance (conditional mean) solution of the estimation problem. A few assumptions have to be made about the quantities involved, which have been mentioned before and are summarised here: Both model error ε^t_k and measurement error ε^m_k are assumed Gaussian with zero mean and white in time. In addition, they are taken to be uncorrelated and independent of $\mathbf{x^t}_k$,

$$\left\langle \varepsilon^t_k (\varepsilon^m_k)^T \right\rangle = \left\langle \varepsilon^t_k (\mathbf{x^t}_k)^T \right\rangle = \left\langle \varepsilon^m_k (\mathbf{x^t}_k)^T \right\rangle = \mathbf{0}. \tag{4.27}$$

Further, we assume that the *initial* probability distribution of $\mathbf{x^t}_0$ is Gaussian with mean $\hat{\mathbf{x}}^\mathbf{t}_0$ and covariance matrix $\mathbf{B}_0$,

$$\mathbf{x^t}_0 \propto N(\hat{\mathbf{x}}^\mathbf{t}_0, \mathbf{B}_0). \tag{4.28}$$

From induction, it will follow that also $p(\mathbf{x^t}_k|\mathcal{Y}_k)$ is Gaussian for all k. Also, we require both the model propagator $\mathbf{m}$ and the observation operator $\mathbf{H}$ to be linear in their arguments, so that equations 4.14 and 4.16 can be rewritten as

$$\mathbf{x^t}_k = \mathbf{M}_{k-1}\mathbf{x^t}_{k-1} + \varepsilon^t_{k-1} \tag{4.29}$$

and

$$\mathbf{y}_k = \mathbf{H}_k \mathbf{x^t}_k + \varepsilon^m_k \ . \tag{4.30}$$

Forecast step

The aim of the forecast step is to determine the expected value of the true state of the atmosphere and its error covariance matrix at time t_k, conditioned on all observations up to the earlier time t_{k-1} (Cohn, 1997),

$$\mathbf{x^f}_k = \left\langle \mathbf{x^t}_k | \mathcal{Y}_{k-1} \right\rangle \ , \tag{4.31}$$

$$\mathbf{B^f}_k = \left\langle \left(\mathbf{x^t}_k - \mathbf{x^f}_k\right) \left(\mathbf{x^t}_k - \mathbf{x^f}_k\right)^T \middle| \mathcal{Y}_{k-1} \right\rangle \ . \tag{4.32}$$

Substituting $\mathbf{x^t}_k$ from Equation 4.29 into 4.31, we get

$$\begin{aligned} \mathbf{x^f}_k &= \left\langle \mathbf{M}_{k-1}\mathbf{x^t}_{k-1} + \varepsilon^t_{k-1} | \mathcal{Y}_{k-1} \right\rangle && (4.33) \\ &= \mathbf{M}_{k-1} \langle \mathbf{x^t}_{k-1} | \mathcal{Y}_{k-1} \rangle + \langle \varepsilon^t_{k-1} | \mathcal{Y}_{k-1} \rangle && (4.34) \end{aligned}$$

since $\mathbf{M}_{k-1}$ is a constant matrix. According to definition 4.25, the first term is the analysis at time t_{k-1}, $\mathbf{x^a}_{k-1}$. The second term is the unconditional expectation $\langle \varepsilon^t_{k-1} \rangle$ – according to assumption 4.15, ε^t does not depend on observations, and $\langle \varepsilon^t \rangle = 0$ under the same assumption. Thus we get

$$\mathbf{x^f}_k = \mathbf{M}_{k-1}\mathbf{x^a}_{k-1} \ , \tag{4.35}$$

identical to Equation 4.1 (here an explicit time dependence of the model $\mathbf{M}$ has been included). The best forecast at time t_k is – not surprisingly – indeed obtained by applying the propagator $\mathbf{M}_{k-1}$ to the analysis state at time t_{k-1}.
In order to obtain the forecast covariance, we insert (4.29) and (4.35) in (4.32), yielding

$$\begin{aligned} \mathbf{B^f}_k &= \left\langle \left[\mathbf{M}_{k-1}(\mathbf{x^t}_{k-1} - \mathbf{x^a}_{k-1}) + \varepsilon^t_{k-1}\right] \left[\mathbf{M}_{k-1}(\mathbf{x^t}_{k-1} - \mathbf{x^a}_{k-1}) + \varepsilon^t_{k-1}\right]^T \Big| \mathcal{Y}_{k-1} \right\rangle \\ &= \mathbf{M}_{k-1} \left\langle (\mathbf{x^t}_{k-1} - \mathbf{x^a}_{k-1})(\mathbf{x^t}_{k-1} - \mathbf{x^a}_{k-1})^T | \mathcal{Y}_{k-1} \right\rangle \mathbf{M}^T_{k-1} \\ &\quad + \langle \varepsilon^t_{k-1} {\varepsilon^t_{k-1}}^T | \mathcal{Y}_{k-1} \rangle \\ &= \mathbf{M}_{k-1}\mathbf{B^a}_{k-1}\mathbf{M}^T_{k-1} + \mathbf{Q}_{k-1} \ , \end{aligned} \tag{4.36}$$

as the cross-terms vanish due to (4.27). Hence we have derived Equation 4.3.

Analysis step

In developing the analysis update equations, we will need the Gaussian assumptions more directly than in the forecast step. First we note that the probability density of the analysis state $p(\mathbf{x^t}_k | \mathcal{Y}_k)$ can be rewritten in terms of earlier observations by repeatedly applying the definition of conditional probabilities, (4.21):

$$\begin{aligned} p(\mathbf{x^t}_k | \mathcal{Y}_k) &= p(\mathbf{x^t}_k | \mathbf{y}_k, \mathcal{Y}_{k-1}) \\ &= \frac{p(\mathbf{x^t}_k, \mathbf{y}_k, \mathcal{Y}_{k-1})}{p(\mathbf{y}_k, \mathcal{Y}_{k-1})} \\ &= \frac{p(\mathbf{y}_k | \mathbf{x^t}_k, \mathcal{Y}_{k-1}) p(\mathbf{x^t}_k, \mathcal{Y}_{k-1})}{p(\mathbf{y}_k, \mathcal{Y}_{k-1})} \\ &= \frac{p(\mathbf{y}_k | \mathbf{x^t}_k, \mathcal{Y}_{k-1}) p(\mathbf{x^t}_k | \mathcal{Y}_{k-1}) p(\mathcal{Y}_{k-1})}{p(\mathbf{y}_k | \mathcal{Y}_{k-1} p(\mathcal{Y}_{k-1})} \\ &= \frac{p(\mathbf{y}_k | \mathbf{x^t}_k, \mathcal{Y}_{k-1}) p(\mathbf{x^t}_k | \mathcal{Y}_{k-1})}{p(\mathbf{y}_k | \mathcal{Y}_{k-1})} \ . \end{aligned} \tag{4.37}$$

This expression can be simplified by noting that

$$p(\mathbf{y}_k | \mathbf{x^t}_k, \mathcal{Y}_{k-1}) = p(\mathbf{y}_k | \mathbf{x^t}_k) \tag{4.38}$$

since given $\mathbf{x^t}_k$, $\mathbf{y}_k$ depends only on ε_k^m, which is independent of $\mathcal{Y}_{k-1} = \{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_{k-1}\}$ according to (4.17). Hence (4.37) becomes

$$p(\mathbf{x^t}_k|\mathcal{Y}_k) = \frac{p(\mathbf{y}_k|\mathbf{x^t}_k)p(\mathbf{x^t}_k|\mathcal{Y}_{k-1})}{p(\mathbf{y}_k|\mathcal{Y}_{k-1})} \ . \tag{4.39}$$

Now the three probability density functions in (4.39) can be evaluated separately. For $p(\mathbf{x^t}_k|\mathcal{Y}_{k-1})$, this is straightforward, as it is none else but the density of the forecast $\mathbf{x^f}_k$, which is Gaussian with mean $\mathbf{x^f}_k$ as given by (4.35) and covariance matrix $\mathbf{B^f}_k$, given by (4.36). The mean of the density $p(\mathbf{y}_k|\mathbf{x^t}_k)$ can be evaluated by inserting the observation equation (4.30):

$$\langle \mathbf{y}_k|\mathbf{x^t}_k\rangle = \langle \mathbf{H}_k\mathbf{x^t}_k + \varepsilon_k^m \big| \mathbf{x^t}_k\rangle = \mathbf{H}_k\mathbf{x^t}_k \ , \tag{4.40}$$

since ε_k^m is state-independent and has zero mean according to (4.18). The covariance matrix can be calculated as

$$\begin{aligned} & \left\langle \left(\mathbf{y}_k - \langle \mathbf{y}_k|\mathbf{x^t}_k\rangle\right)\left(\mathbf{y}_k - \langle \mathbf{y}_k|\mathbf{x^t}_k\rangle\right)^T \Big| \mathbf{x^t}_k \right\rangle \\ = \ & \left\langle (\mathbf{y}_k - \mathbf{H}_k\mathbf{x^t}_k)(\mathbf{y}_k - \mathbf{H}_k\mathbf{x^t}_k)^T \big| \mathbf{x^t}_k \right\rangle \\ = \ & \langle \varepsilon_k^m(\varepsilon_k^m)^T|\mathbf{x^t}_k\rangle \\ = \ & \mathbf{R}_k \ , \end{aligned} \tag{4.41}$$

where again (4.30) and (4.18) have been used. Since ε_k^m is Gaussian, $p(\mathbf{y}_k|\mathbf{x^t}_k)$ is Gaussian with mean $\mathbf{H}_k\mathbf{x^t}_k$ and covariance matrix $\mathbf{R}_k$.
The remaining density is $p(\mathbf{y}_k|\mathcal{Y}_{k-1})$. Its expectation value is again calculated by inserting (4.30),

$$\begin{aligned} \langle \mathbf{y}_k|\mathcal{Y}_{k-1}\rangle &= \langle \mathbf{H}_k\mathbf{x^t}_k + \varepsilon_k^m|\mathcal{Y}_{k-1}\rangle \\ &= \mathbf{H}_k\langle \mathbf{x^t}_k|\mathcal{Y}_{k-1}\rangle + \langle \varepsilon_k^m|\mathcal{Y}_{k-1}\rangle \\ &= \mathbf{H}_k\mathbf{x^f}_k \ , \end{aligned} \tag{4.42}$$

where we have used (4.18) and definition (4.23). Its covariance matrix is then

$$\begin{aligned} & \left\langle \left(\mathbf{y}_k - \langle \mathbf{y}_k|\mathcal{Y}_{k-1}\rangle\right)\left(\mathbf{y}_k - \langle \mathbf{y}_k|\mathcal{Y}_{k-1}\rangle\right)^T \Big| \mathcal{Y}_{k-1} \right\rangle \\ = \ & \left\langle \left[\mathbf{H}_k\left(\mathbf{x^t}_k - \mathbf{x^f}_k\right) + \varepsilon_k^m\right]\left[\mathbf{H}_k\left(\mathbf{x^t}_k - \mathbf{x^f}_k\right) + \varepsilon_k^m\right]^T \Big| \mathcal{Y}_{k-1} \right\rangle \\ = \ & \mathbf{H}_k\left\langle (\mathbf{x^t}_k - \mathbf{x^f}_k)(\mathbf{x^t}_k - \mathbf{x^f}_k)^T \Big| \mathcal{Y}_{k-1}\right\rangle \mathbf{H}_k^T + \left\langle \varepsilon_k^m(\varepsilon_k^m)^T|\mathcal{Y}_{k-1}\right\rangle \\ & + \mathbf{H}_k\left\langle (\mathbf{x^t}_k - \mathbf{x^f}_k)(\varepsilon_k^m)^T \Big| \mathcal{Y}_{k-1}\right\rangle + \left\langle \varepsilon_k^m(\mathbf{x^t}_k - \mathbf{x^f}_k)^T \Big| \mathcal{Y}_{k-1}\right\rangle \mathbf{H}_k^T \\ = \ & \mathbf{H}_k\mathbf{B^f}_k\mathbf{H}^T + \mathbf{R}_k \ . \end{aligned} \tag{4.43}$$

The cross-terms again vanish by (4.27). The distribution $p(\mathbf{y}_k|\mathcal{Y}_{k-1})$ is Gaussian since ε_k^m was assumed Gaussian (4.18), and the density of the forecast, $p(\mathbf{x^t}_k|\mathcal{Y}_{k-1})$, is Gaussian.

Hence, the density $p(\mathbf{x^t}_k|\mathcal{Y}_k)$ is calculated by inserting these results into (4.39). The multivariate Gaussian density of a random vector $\mathbf{a} \in \mathbb{R}^n$ with mean $\mu \in \mathbb{R}^n$ and covariance matrix $\mathbf{\Sigma} \in \mathbb{R}^{n \times n}$ is defined as (Jazwinski, 1970)

$$p(\mathbf{a}) = (2\pi)^{-\frac{n}{2}} |\mathbf{\Sigma}|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(\mathbf{x}-\mu)^T \Sigma^{-1}(\mathbf{x}-\mu)\right] , \tag{4.44}$$

($|\cdot|$ representing the determinant), and thus (4.39) can be written as

$$p(\mathbf{x^t}_k|\mathcal{Y}_k) = \frac{p_1 p_2}{p_3} \tag{4.45}$$

with

$$p_1 = (2\pi)^{-\frac{n_{obs}}{2}} |\mathbf{R}|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(\mathbf{y}-\mathbf{H}\mathbf{x^t})^T \mathbf{R}^{-1}(\mathbf{y}-\mathbf{H}\mathbf{x^t})\right] , \tag{4.46}$$

$$p_2 = (2\pi)^{-\frac{n_{mod}}{2}} |\mathbf{B^f}|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(\mathbf{x^t}-\mathbf{x^f})^T \mathbf{B^f}^{-1}(\mathbf{x^t}-\mathbf{x^f})\right] , \tag{4.47}$$

$$p_3 = (2\pi)^{-\frac{n_{obs}}{2}} |\mathbf{H}\mathbf{B^f}\mathbf{H}^T + \mathbf{R}|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(\mathbf{y}-\mathbf{H}\mathbf{x^f})^T (\mathbf{H}\mathbf{B^f}\mathbf{H}^T + \mathbf{R})^{-1}(\mathbf{y}-\mathbf{H}\mathbf{x^f})\right] . \tag{4.48}$$

For notational convenience, the time index k has been omitted in equations (4.46)–(4.48). Since (4.45) is a product of exponential functions, it can be written as

$$p(\mathbf{x^t}_k|\mathcal{Y}_k) = c \, \exp\left(-\frac{1}{2}J\right) , \tag{4.49}$$

with

$$c = (2\pi)^{-\frac{n_{mod}}{2}} |\mathbf{R}|^{-\frac{1}{2}} |\mathbf{B^f}|^{-\frac{1}{2}} |\mathbf{H}\mathbf{B^f}\mathbf{H}^T + \mathbf{R}|^{\frac{1}{2}} , \tag{4.50}$$

and

$$\begin{aligned} J &= (\mathbf{y}-\mathbf{H}\mathbf{x^t})^T \mathbf{R}^{-1}(\mathbf{y}-\mathbf{H}\mathbf{x^t}) + (\mathbf{x^t}-\mathbf{x^f})^T (\mathbf{B^f})^{-1}(\mathbf{x^t}-\mathbf{x^f}) \\ &\quad -(\mathbf{y}-\mathbf{H}\mathbf{x^t})^T (\mathbf{H}\mathbf{B^f}\mathbf{H}^T + \mathbf{R})^{-1}(\mathbf{y}-\mathbf{H}\mathbf{x^f}) . \end{aligned} \tag{4.51}$$

It is now possible to show that $p(\mathbf{x^t}_k|\mathcal{Y}_k)$ is Gaussian with mean $\mathbf{x^a}_k$ and covariance matrix $\mathbf{B^a}_k$, and to derive the correct expressions for $\mathbf{x^a}_k$ and $\mathbf{B^a}_k$ this way (cf, e.g., Cohn, 1997). However, we avoid this lengthy calculation here, since we are mainly interested in obtaining $\mathbf{x^a}_k$ and $\mathbf{B^a}_k$. Since $\mathbf{x^a}_k$ should be the best, or maximum likelihood, estimator for $\mathbf{x^t}_k$, we can derive it by maximising $p(\mathbf{x^t}_k|\mathcal{Y}_k)$. Indeed, following Jazwinski (1970) (Example 7.3), this is equivalent to minimising $-\log\, p(\mathbf{x^t}_k|\mathcal{Y}_k)$, or J in (4.51).
The gradient of J is

$$\begin{aligned} \nabla J(\mathbf{x^t}) &= -\mathbf{H}^T \mathbf{R}^{-1}(\mathbf{y}-\mathbf{H}\mathbf{x^t}) - (\mathbf{y}-\mathbf{H}\mathbf{x^t})^T \mathbf{R}^{-1}\mathbf{H} \\ &\quad +(\mathbf{B^f})^{-1}(\mathbf{x^t}-\mathbf{x^f}) + (\mathbf{x^t}-\mathbf{x^f})^T (\mathbf{B^f})^{-1} , \end{aligned} \tag{4.52}$$

and setting it equal to zero, we get

$$\mathbf{H}^T\mathbf{R}^{-1}(\mathbf{y}-\mathbf{H}\hat{\mathbf{x}}^\mathbf{t})+(\mathbf{B}^\mathbf{f})^{-1}(\hat{\mathbf{x}}^\mathbf{t}-\mathbf{x}^\mathbf{f})=0\ . \tag{4.53}$$

Solving for $\hat{\mathbf{x}}^\mathbf{t}\equiv\mathbf{x}^\mathbf{a}$, we get

$$\hat{\mathbf{x}}^\mathbf{t}=\left[\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}+(\mathbf{B}^\mathbf{f})^{-1}\right]^{-1}\ \left[\mathbf{H}^T\mathbf{R}^{-1}\mathbf{y}+(\mathbf{B}^\mathbf{f})^{-1}\mathbf{x}^\mathbf{f}\right]\ . \tag{4.54}$$

It remains to demonstrate that this expression is equal to (4.10). For this, we use the following matrix identities (Jazwinski, 1970)
Lemma 4.1:

$$\left[\mathbf{B}^{\mathbf{f}\,-1}+\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\right]^{-1}=\mathbf{B}^\mathbf{f}-\mathbf{B}^\mathbf{f}\mathbf{H}^T(\mathbf{H}\mathbf{B}^\mathbf{f}\mathbf{H}^T+\mathbf{R})^{-1}\mathbf{H}\mathbf{B}^\mathbf{f} \tag{4.55}$$

and
Lemma 4.2:

$$\left[\mathbf{B}^{\mathbf{f}\,-1}+\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\right]^{-1}\mathbf{H}^T\mathbf{R}^{-1}=\mathbf{B}^\mathbf{f}\mathbf{H}^T(\mathbf{H}\mathbf{B}^\mathbf{f}\mathbf{H}^T+\mathbf{R})^{-1}. \tag{4.56}$$

Proof of *Lemma* 4.1:
In order to show that the right side of (4.55) is the inverse of the term in brackets on the left, we multiply them, obtaining

$$\begin{aligned}
&\left[(\mathbf{B}^\mathbf{f})^{-1}+\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\right]\left[\mathbf{B}^\mathbf{f}-\mathbf{B}^\mathbf{f}\mathbf{H}^T(\mathbf{H}\mathbf{B}^\mathbf{f}\mathbf{H}^T+\mathbf{R})^{-1}\mathbf{H}\mathbf{B}^\mathbf{f}\right]\\
=\ &\mathbf{1}-\mathbf{H}^T(\mathbf{H}\mathbf{B}^\mathbf{f}\mathbf{H}^T+\mathbf{R})^{-1}\mathbf{H}\mathbf{B}^\mathbf{f}+\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\mathbf{B}^\mathbf{f}\\
&-\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\mathbf{B}^\mathbf{f}\mathbf{H}^T(\mathbf{H}\mathbf{B}^\mathbf{f}\mathbf{H}^T+\mathbf{R})^{-1}\mathbf{H}\mathbf{B}^\mathbf{f}\\
=\ &\mathbf{1}+\left[-\mathbf{H}^T+\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\mathbf{B}^\mathbf{f}(\mathbf{B}^\mathbf{f})^{-1}\mathbf{H}^{-1}(\mathbf{H}\mathbf{B}^\mathbf{f}\mathbf{H}^T+\mathbf{R})-\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\mathbf{B}^\mathbf{f}\mathbf{H}^T\right]\\
&\cdot\left[(\mathbf{H}\mathbf{B}^\mathbf{f}\mathbf{H}^T+\mathbf{R})^{-1}\mathbf{H}\mathbf{B}^\mathbf{f}\right]\\
=\ &\mathbf{1}+\left[-\mathbf{H}^T+\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\mathbf{B}^\mathbf{f}\mathbf{H}^T+\mathbf{H}^T-\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\mathbf{B}^\mathbf{f}\mathbf{H}^T\right]\left[(\mathbf{H}\mathbf{B}^\mathbf{f}\mathbf{H}^T+\mathbf{R})^{-1}\mathbf{H}\mathbf{B}^\mathbf{f}\right]\\
=\ &\mathbf{1}\square
\end{aligned} \tag{4.57}$$

which proves *Lemma* 4.1 since the inverse is unique and exists as all involved matrices are assumed non-singular.
The proof of *Lemma* 4.2 is now straightforward, using (4.57):

$$\begin{aligned}
&\left[(\mathbf{B}^\mathbf{f})^{-1}+\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\right]^{-1}\mathbf{H}^T\mathbf{R}^{-1}\\
=\ &\left[\mathbf{B}^\mathbf{f}-\mathbf{B}^\mathbf{f}\mathbf{H}^T(\mathbf{H}\mathbf{B}^\mathbf{f}\mathbf{H}^T+\mathbf{R})^{-1}\mathbf{H}\mathbf{B}^\mathbf{f}\right]\mathbf{H}^T\mathbf{R}^{-1}\\
=\ &\mathbf{B}^\mathbf{f}\mathbf{H}^T(\mathbf{H}\mathbf{B}^\mathbf{f}\mathbf{H}^T+\mathbf{R})^{-1}\left[(\mathbf{H}\mathbf{B}^\mathbf{f}\mathbf{H}^T+\mathbf{R})\mathbf{R}^{-1}-\mathbf{H}\mathbf{B}^\mathbf{f}\mathbf{H}^T\mathbf{R}^{-1}\right]\\
=\ &\mathbf{B}^\mathbf{f}\mathbf{H}^T(\mathbf{H}\mathbf{B}^\mathbf{f}\mathbf{H}^T+\mathbf{R})^{-1}\left[\mathbf{H}\mathbf{B}^\mathbf{f}\mathbf{H}^T\mathbf{R}^{-1}+\mathbf{1}-\mathbf{H}\mathbf{B}^\mathbf{f}\mathbf{H}^T\mathbf{R}^{-1}\right]\\
=\ &\mathbf{B}^\mathbf{f}\mathbf{H}^T(\mathbf{H}\mathbf{B}^\mathbf{f}\mathbf{H}^T+\mathbf{R})^{-1}\ \square
\end{aligned} \tag{4.58}$$

Now, using (4.55) and (4.56), (4.54) becomes

$$\begin{aligned}
\hat{\mathbf{x}}^{\mathbf{t}} &= \left[\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H} + (\mathbf{B}^{\mathbf{f}})^{-1}\right]^{-1}\left[\mathbf{H}^T\mathbf{R}^{-1}\mathbf{y} + (\mathbf{B}^{\mathbf{f}})^{-1}\mathbf{x}^{\mathbf{f}}\right] \\
&= \left[\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H} + (\mathbf{B}^{\mathbf{f}})^{-1}\right]^{-1}\mathbf{H}^T\mathbf{R}^{-1}\mathbf{y} + \left[\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H} + (\mathbf{B}^{\mathbf{f}})^{-1}\right]^{-1}(\mathbf{B}^{\mathbf{f}})^{-1}\mathbf{x}^{\mathbf{f}} \\
&= \mathbf{B}^{\mathbf{f}}\mathbf{H}^T(\mathbf{H}\mathbf{B}^{\mathbf{f}}\mathbf{H}^T + \mathbf{R})^{-1}\mathbf{y} + \left[\mathbf{B}^{\mathbf{f}} - \mathbf{B}^{\mathbf{f}}\mathbf{H}^T(\mathbf{H}\mathbf{B}^{\mathbf{f}}\mathbf{H}^T + \mathbf{R})^{-1}\mathbf{H}\mathbf{B}^{\mathbf{f}}\right](\mathbf{B}^{\mathbf{f}})^{-1}\mathbf{x}^{\mathbf{f}} \\
&= \mathbf{B}^{\mathbf{f}}\mathbf{H}^T(\mathbf{H}\mathbf{B}^{\mathbf{f}}\mathbf{H}^T + \mathbf{R})^{-1}\mathbf{y} + \mathbf{x}^{\mathbf{f}} - \mathbf{B}^{\mathbf{f}}\mathbf{H}^T(\mathbf{H}\mathbf{B}^{\mathbf{f}}\mathbf{H}^T + \mathbf{R})^{-1}\mathbf{H}\mathbf{x}^{\mathbf{f}} \\
&= \mathbf{x}^{\mathbf{f}} + \mathbf{B}^{\mathbf{f}}\mathbf{H}^T(\mathbf{H}\mathbf{B}^{\mathbf{f}}\mathbf{H}^T + \mathbf{R})^{-1}(\mathbf{y} - \mathbf{H}\mathbf{x}^{\mathbf{f}}) \\
&= \mathbf{x}^{\mathbf{f}}{}_k + \mathbf{K}_k(\mathbf{y} - \mathbf{H}\mathbf{x}^{\mathbf{f}}{}_k)
\end{aligned} \tag{4.59}$$

which is identical to (4.10), with the Kalman gain $\mathbf{K}_k$ defined as in (4.11),

$$\mathbf{K}_k = \mathbf{B}^{\mathbf{f}}{}_k\mathbf{H}_k^{\mathrm{T}}\left(\mathbf{H}\mathbf{B}^{\mathbf{f}}{}_k\mathbf{H}^T + \mathbf{R}\right)^{-1} . \tag{4.60}$$

It remains to demonstrate that the analysis error covariance matrix $\mathbf{B}^{\mathbf{a}}{}_k$ is given by (4.12). $\mathbf{B}^{\mathbf{a}}{}_k$ is defined as

$$\begin{aligned}
\mathbf{B}^{\mathbf{a}}{}_k &\equiv \left\langle(\mathbf{x}^{\mathbf{t}}{}_k - \langle\mathbf{x}^{\mathbf{t}}{}_k|\mathcal{Y}_k\rangle)(\mathbf{x}^{\mathbf{t}}{}_k - \langle\mathbf{x}^{\mathbf{t}}{}_k|\mathcal{Y}_k\rangle)^T\middle|\mathcal{Y}_k\right\rangle \\
&= \left\langle(\mathbf{x}^{\mathbf{t}}{}_k - \mathbf{x}^{\mathbf{a}}{}_k)(\mathbf{x}^{\mathbf{t}}{}_k - \mathbf{x}^{\mathbf{a}}{}_k)^T|\mathcal{Y}_k\right\rangle .
\end{aligned} \tag{4.61}$$

Using (4.59), the analysis error $\varepsilon_k^a \equiv \mathbf{x}^{\mathbf{t}} - \mathbf{x}^{\mathbf{a}}$ can be calculated as

$$\begin{aligned}
\varepsilon_k^a &\equiv \mathbf{x}^{\mathbf{t}} - \mathbf{x}^{\mathbf{a}} \\
&= \mathbf{x}^{\mathbf{t}}{}_k - \left[\mathbf{x}^{\mathbf{f}}{}_k + \mathbf{K}_k(\mathbf{y}_k - \mathbf{H}\mathbf{x}^{\mathbf{f}}{}_k)\right] \\
&= \mathbf{x}^{\mathbf{t}}{}_k - \left[\mathbf{x}^{\mathbf{f}}{}_k + \mathbf{K}_k(\mathbf{H}\mathbf{x}^{\mathbf{t}}{}_k + \varepsilon_k^m - \mathbf{H}\mathbf{x}^{\mathbf{f}}{}_k\right] \\
&= \mathbf{x}^{\mathbf{t}}{}_k - \mathbf{x}^{\mathbf{f}}{}_k - \mathbf{K}_k\mathbf{H}\mathbf{x}^{\mathbf{t}}{}_k - \mathbf{K}_k\varepsilon_k^m + \mathbf{K}_k\mathbf{H}\mathbf{x}^{\mathbf{f}}{}_k \\
&= (\mathbf{1} - \mathbf{K}_k\mathbf{H})(\mathbf{x}^{\mathbf{t}} - \mathbf{x}^{\mathbf{f}}) - \mathbf{K}_k\varepsilon_k^m .
\end{aligned} \tag{4.62}$$

Hence, $\mathbf{B}^{\mathbf{a}}{}_k$ is calculated as

$$\begin{aligned}
\mathbf{B}^{\mathbf{a}}{}_k &= \left\langle\left[(\mathbf{1} - \mathbf{K}_k\mathbf{H})(\mathbf{x}^{\mathbf{t}} - \mathbf{x}^{\mathbf{f}}) - \mathbf{K}_k\varepsilon_k^m\right]\left[(\mathbf{1} - \mathbf{K}_k\mathbf{H})(\mathbf{x}^{\mathbf{t}} - \mathbf{x}^{\mathbf{f}}) - \mathbf{K}_k\varepsilon_k^m\right]^T\middle|\mathcal{Y}_k\right\rangle \\
&= \left\langle(\mathbf{1} - \mathbf{K}_k\mathbf{H})(\mathbf{x}^{\mathbf{t}} - \mathbf{x}^{\mathbf{f}})(\mathbf{x}^{\mathbf{t}} - \mathbf{x}^{\mathbf{f}})^T(\mathbf{1} - \mathbf{K}_k\mathbf{H})^T|\mathcal{Y}_k\right\rangle \\
&\quad + \left\langle\mathbf{K}_k\varepsilon_k^m(\varepsilon_k^m)^T\mathbf{K}_k^T\middle|\mathcal{Y}_k\right\rangle \\
&\quad - \left\langle(\mathbf{1} - \mathbf{K}_k\mathbf{H})(\mathbf{x}^{\mathbf{t}} - \mathbf{x}^{\mathbf{f}})(\varepsilon_k^m)^T\mathbf{K}_k^T\middle|\mathcal{Y}_k\right\rangle \\
&\quad - \left\langle\mathbf{K}_k\varepsilon_k^m(\mathbf{x}^{\mathbf{t}} - \mathbf{x}^{\mathbf{f}})^T(\mathbf{1} - \mathbf{K}_k\mathbf{H})^T\middle|\mathcal{Y}_k\right\rangle \\
&= (\mathbf{1} - \mathbf{K}_k\mathbf{H})\left\langle(\mathbf{x}^{\mathbf{t}} - \mathbf{x}^{\mathbf{f}})(\mathbf{x}^{\mathbf{t}} - \mathbf{x}^{\mathbf{f}})^T\middle|\mathcal{Y}_k\right\rangle(\mathbf{1} - \mathbf{K}_k\mathbf{H})^T + \mathbf{K}_k\left\langle\varepsilon_k^m(\varepsilon_k^m)^T\middle|\mathcal{Y}_k\right\rangle\mathbf{K}_k^T \\
&= (\mathbf{1} - \mathbf{K}_k\mathbf{H})\mathbf{B}^{\mathbf{f}}{}_k(\mathbf{1} - \mathbf{K}_k\mathbf{H})^T + \mathbf{K}_k\mathbf{R}_k\mathbf{K}_k^T .
\end{aligned} \tag{4.63}$$

This expression is indeed equivalent to (4.12), as we can see by some restructuring,

$$\begin{aligned} & (\mathbf{1}-\mathbf{K}_k\mathbf{H})\mathbf{B^f}_k(\mathbf{1}-\mathbf{K}_k\mathbf{H})^T+\mathbf{K}_k\mathbf{R}_k\mathbf{K}_k^T \\ = \;& \mathbf{B^f}-\mathbf{KHB^f}-\mathbf{B^f}\mathbf{H}^T\mathbf{K}^T+\mathbf{KHB^f}\mathbf{H}^T\mathbf{K}^T+\mathbf{KRK}^T \\ = \;& \mathbf{B^f}-\mathbf{KHB^f}+\alpha\ , \end{aligned} \tag{4.64}$$

with

$$\alpha \equiv -\mathbf{B^f}\mathbf{H}^T\mathbf{K}^T+\mathbf{KHB^f}\mathbf{H}^T\mathbf{K}^T+\mathbf{KRK}^T\ . \tag{4.65}$$

For (4.64) to equal (4.12), α must be zero, which is easily shown by applying the definition of $\mathbf{K}$, (4.60):

$$\begin{aligned} \alpha \;=\; & \left[-\mathbf{B^f}\mathbf{H}^T+\mathbf{B^f}\mathbf{H}^T(\mathbf{HB^f}\mathbf{H}^T+\mathbf{R})^{-1}\mathbf{HB^f}\mathbf{H}^T+\mathbf{B^f}\mathbf{H}^T(\mathbf{HB^f}\mathbf{H}^T+\mathbf{R})^{-1}\mathbf{R}\right]\mathbf{K}^T \\ = \;& \left[\mathbf{B^f}\mathbf{H}^T(\mathbf{HB^f}\mathbf{H}^T+\mathbf{R})^{-1}\right]\left(-\mathbf{HB^f}\mathbf{H}^T-\mathbf{R}+\mathbf{HB^f}\mathbf{H}^T+\mathbf{R}\right)\mathbf{K}^T \\ = \;& \mathbf{0}\ . \end{aligned} \tag{4.66}$$

Finally, reintroducing the time index k, we have

$$\mathbf{B^a}_k = \mathbf{B^f}-\mathbf{KHB^f}\ , \tag{4.67}$$

which is indeed identical to (4.12). Hence, we have derived the whole set of equations that constitute the Kalman filter from estimation theory, namely the forecast step, (4.35) and (4.36), and the analysis step, (4.60) and (4.67).

4.3.3. The χ^2 diagnostic

It remains to derive the characteristics of the χ^2 diagnostic given in (4.13). First we note that the χ^2 diagnostic can be introduced in a number of ways (Dee, 1995). The formulation chosen here is the one used by Ménard et al. (2000), Chipperfield et al. (2002), and Khattatov et al. (2000), which is designed to compare the covariance of the *observation minus forecast* (OmF) residual ν_k,

$$\begin{aligned} \nu_k \;\equiv\; & \mathbf{y}_k-\mathbf{H}_k\mathbf{x^f}_k \\ = \;& (\mathbf{y}_k-\mathbf{H}_k\mathbf{x^t}_k)-\mathbf{H}_k(\mathbf{x^f}_k-\mathbf{x^t}_k) \\ = \;& \varepsilon_k^m-\mathbf{H}_k\varepsilon_k^f\ , \end{aligned} \tag{4.68}$$

to the Kalman filter covariance $\mathbf{S}_k$,

$$\begin{aligned} \mathbf{S}_k \;\equiv\; & \left\langle \nu_k\nu_k^{\mathrm{T}}|\mathcal{Y}_{k-1}\right\rangle \\ = \;& \left\langle \left(\varepsilon_k^m-\mathbf{H}_k\varepsilon_k^f\right)\left(\varepsilon_k^m-\mathbf{H}_k\varepsilon_k^f\right)^T\Big|\mathcal{Y}_{k-1}\right\rangle \qquad (4.69) \\ = \;& \left\langle \varepsilon_k^m(\varepsilon_k^m)^T|\mathcal{Y}_{k-1}\right\rangle+\left\langle \mathbf{H}_k\varepsilon_k^f(\varepsilon_k^f)^T\mathbf{H}_k^T|\mathcal{Y}_{k-1}\right\rangle \\ & -\left\langle \varepsilon_k^m(\varepsilon_k^f)^T\mathbf{H}_k^T\mathcal{Y}_{k-1}\right\rangle-\left\langle \mathbf{H}_k\varepsilon_k^f(\varepsilon_k^m)^T|\dagger_{k-1}\right\rangle \\ = \;& \mathbf{R}_k+\mathbf{H}_k\mathbf{B^f}_k\mathbf{H}_k^T\ , \qquad (4.70) \end{aligned}$$

where we have used the definitions of $\mathbf{R}_k$ and $\mathbf{B^f}_k$, (4.18) and (4.24), plus (4.27) by which the cross-terms vanish.
Note that by (4.42), ν_k can also be written as

$$\nu_k = \mathbf{y}_k - \langle \mathbf{y}_k | \mathcal{Y}_{k-e} \rangle \ , \tag{4.71}$$

which means that ν_k is the part of the current observation that is not already explained by past observations, also called the innovation (Ménard et al., 2000).
The quantity χ^2 is then defined as

$$\begin{aligned} \chi^2 &\equiv \nu_k^T \mathbf{S}_k^{-1} \nu_k \\ &= (\mathbf{y}_k - \mathbf{H}_k \mathbf{x^f}_k)^T \left[\mathbf{H}_k \mathbf{B^f}_k \mathbf{H}_k^T + \mathbf{R}_k \right]^{-1} (\mathbf{y}_k - \mathbf{H}_k \mathbf{x^f}_k) \ . \end{aligned} \tag{4.72}$$

The conditional expectation of χ^2 is

$$\begin{aligned} \langle \chi_k^2 | \mathcal{Y}_{k-1} \rangle &= \langle \nu_k^T \mathbf{S}_k^{-1} \nu_k | \mathcal{Y}_{k-1} \\ &= \langle \mathrm{trace}(\mathbf{S}_k^{-1} \nu_k \nu_k^T) | \mathcal{Y}_{k-1} \rangle \\ &= \mathrm{trace} \left(\mathbf{S}_k^{-1} \langle \nu_k \nu_k^T | \mathcal{Y}_{k-1} \rangle \right) \\ &= \mathrm{trace} \left(\mathbf{S}_k^{-1} \mathbf{S}_k \right) \qquad (4.73) \\ &= n_{obs,k} \qquad (4.74) \end{aligned}$$

since $\mathbf{S}_k \in \mathbb{R}^{n_{obs,k} \times n_{obs,k}}$. Here we have used the identity

$$\mathbf{x}^T \mathbf{A} \mathbf{x} = \mathrm{trace}(\mathbf{A} \mathbf{x} \mathbf{x}^T) \ , \tag{4.75}$$

with $\mathbf{x}$ an n-vector and $\mathbf{A}$ an $n \times n$ matrix. (4.75) is best verified using a component-wise notation, $\mathbf{x} \equiv x^i$, $\mathbf{x}^T \equiv x_i$, and $\mathbf{A} \equiv A^i{}_j$, in which an upper index is a column index and a lower index is a row index. Using the Einstein summation convention, we get

$$\begin{aligned} \mathbf{x}^T \mathbf{A} \mathbf{x} &= x_i \, A^i{}_j \, x^j \\ &= A^i{}_j \, x^j \, x_i \\ &= \delta^i{}_k \, A^k{}_j \, x^j \, x_i \\ &= \mathrm{trace}(A^i{}_j \, x^j \, x_k) \\ &= \mathrm{trace}(\mathbf{A} \mathbf{x} \mathbf{x}^T) \ , \end{aligned} \tag{4.76}$$

with $\delta^i{}_k$ the Kronecker Delta, $\delta^i{}_k = 1$ if $i = k$ and 0 otherwise.
So far we have not used any assumptions about error statistics in this section. If, however, Gaussian errors are assumed (as usual), then it can be shown that the quantity χ^2 defined in (4.72) actually obeys a χ^2 distribution with n_{obs} degrees of freedom (Ménard et al., 2000).

5. SBUV satellite observations

As described in Chapter 4, the basic idea of assimilation is combining observational data with an atmospheric model. Chapter 3 gave details on the chemistry transport model used, and this short chapter is dedicated to providing some information about the satellite data used for assimilation. In order to analyse the research questions posed in the introduction, a consistent observation based data set of stratospheric ozone with longest possible extent is needed. Thus, the Solar Backscatter UV (SBUV and SBUV/2) instruments (Heath et al., 1975; McPeters et al., 1984) are selected as observational data source, which provide the longest continuous time series of global profile ozone observations since the late 1970s. Due to their viewing geometry, the vertical resolution of SBUV instruments is rather coarse ($\sim 7\,\mathrm{km}$). In the meantime, several satellite instruments with better vertical resolution have been developed and flown (e.g. the SCIAMACHY and MIPAS instruments used later in this thesis); however, their available time spans of data are incomparably shorter. Data from these instruments with high vertical resolution are used for validation of the assimilated data set in Chapter 6.

5.1. Overview

SBUV instruments have been flown on a series of NASA and NOAA satellites since 1978. Essential features of the instruments are identical. This study uses the data version 8 provided on the v8 DVD released 2006 (Bhartia et al., 2004), and an update (v8b) released 2008 (L. Flynn, personal communication, 2008), which extends to the end of 2007. An overview of the periods of satellite instruments available on the v8b release is given in table 5.1. In some periods, more than one satellite instrument was operational. In these cases, the choice of instruments used for assimilation is based on the idea of maximising continuous time series of each satellite instrument, while avoiding phases where known problems occurred – e.g., S. Frith strongly suggested using NOAA-11 data instead of NOAA-9 in 1992 and 1993, as NOAA-9 profiles seemed more affected by Pinatubo aerosol (personal communication, 2010).

SBUV instruments are nadir-viewing double monochromators, measuring backscattered radiation from the atmosphere in 12 discrete wavelengths from 255 nm to 340 nm with a 1 nm bandpass (Heath et al., 1975; McPeters et al., 1984). Despite the vertical light path, it is possible to derive information not only about the integrated ozone column, but also about the vertical distribution. This is based on the fact that in a pure single-scattering Rayleigh regime, as it is approximately realised for light with wavelengths $< 290\,\mathrm{nm}$, light with a given wavelength originates mostly in a limited vertical region of the atmosphere. Hence, altitude information can be derived from the wavelength distribution – at least to a certain degree. Understandably, a nadir viewing instrument cannot achieve a vertical resolution as high as provided by limb or occultation observations, as the latter methods observe the atmosphere in tangential geometry. Since this study is focused on a long-term analysis of polar ozone, and the only continuous long-term series of observations stems from nadir-viewing SBUV

Satellite instrument	Data available	Data used
Nimbus-7	30 Oct 1978 – 21 Jun 1990	30 Oct 1978 – 31 Dec 1989
NOAA-9 (asc)	2 Feb 1985 – 31 Dec 1989	—
NOAA-9 (desc)	1 Jan 1992 – 19 Feb 1998	1 Jan 1994 – 31 Jul 1997
NOAA-11 (asc)	1 Dec 1988 – 31 Mar 1995	1 Jan 1990 – 31 Dec 1993
NOAA-11 (desc)	15 Jul 1997 – 27 Mar 2001	1 Aug 1997 – 31 Dec 2000
NOAA-16	3 Oct 2000 – 31 Dec 2007	1 Jan 2001 – 31 Dec 2007
NOAA-17	1 Jan 2003 – 31 Dec 2007	—

Table 5.1.: *SBUV satellite instruments, periods of data availability, and periods in which they are used for assimilation. (asc) and (desc) denote ascending (north-bound equator crossing) and descending (south-bound equator crossing) nodes, respectively.*

instruments, the viewing geometry of the observations used for assimilation in this project is no question of choice.
The SBUV retrieval algorithm is based on the principle of increasing backscattering albedos with increasing air density, while transmission decreases with the penetration depth. The balance between exponentially increasing backscattering sources and exponentially decreasing transmission initially allows to determine the scattering height within a layer of $\sim 14\,\mathrm{km}$ half width. The full retrieval process including error correlation estimates yields profile information in layers with an approximate altitude resolution of 7 km. Since the transmission does not decrease strongly below the maximum of the ozone layer ($\sim 25\,\mathrm{km}$), SBUV sensitivity is low in this region.
All SBUV satellites were intended for a sun-synchronous orbit at about 850 km altitude, crossing the equator close to noon (13:30 – 14:30) every day. 14 orbits are flown per day, with an orbit spacing of 26° at the equator. In practise, most of the instruments experienced orbit drifts that decreased the usability of their data due to high solar zenith angles during certain periods (Price, 1991). NOAA-9 and NOAA-11 even crossed the terminator during the drift; in these cases, data are available in two different periods corresponding to ascending (northbound equator crossing) and descending (southbound equator crossing) nodes. The selection of satellites in this study is similar to that in a previous version of the assimilated data set published by Kiesewetter et al. (2010a).

A quality error flag is included in the SBUV data product for every single ozone profile measurement. In order to avoid assimilation of erroneous profiles, only profiles with an error flag of 0 or 10 are selected for assimilation. This rules out profiles with high solar zenith angles or known problems in the retrieval process.
Standard errors of the retrieved version 8 profiles are provided by Bhartia et al. (1996) and listed in Table 5.2. These are taken as observational errors in the assimilation.

5.2. Bias adjustment coefficients

Although SBUV does provide a long continuous ozone time series, observations stem from individual instruments whose consistency is not necessarily guaranteed. Naturally, a consis-

Layer	P at bottom [hPa]	Observational error [%]	Representativeness error [%]
1	1013 hPa	N/A	N/A
2	63.93 hPa	12	7
3	40.33 hPa	10	7
4	25.45 hPa	7	7
5	16.06 hPa	6	5
6	10.13 hPa	5	5
7	6.393 hPa	5	5
8	4.034 hPa	5	5
9	2.545 hPa	5	5
10	1.606 hPa	6	5
11	1.013 hPa	6	5
12	0.639 hPa	6	5
13	0.403 hPa	10	5

Table 5.2.: *Observational and representativeness errors for the SBUV partial columns. Observational errors as provided by Bhartia et al. (1996).*

tent time series is desirable for any long-term study, and several efforts have been made in this direction.
Ozone time series from different SBUV instruments have been basically inter-calibrated from the available overlap times in the version 8 profiles. In some cases, however, the overlap time is short, and comparisons to other instruments exhibit jumps between instruments and drifts of individual instruments. In an effort to compensate these jumps and drifts, adjustment coefficients have been developed which were used by Stolarski and Frith (2006) to generate a homogenised merged data set of global column ozone. In the meantime, adjustment coefficients have also been calculated for profile ozone from a more stringent evaluation of instrument overlaps and inter-comparisons to other instruments. These adjustment coefficients have been made available to this study by S. Frith (personal communication, 2010) and are used in one of the two versions of the assimilated data set.

These adjustment terms were derived by the following procedures [1]:

- NOAA-7: adjusted to NOAA-11 using the overlap
- NOAA-9: drift correction using observations from the Microwave Limb Sounder (MLS) instrument (Barath et al., 1993)
- NOAA-9: absolute calibration using NOAA-11 overlap
- NOAA-11 1989–1995: no corrections applied, used as standard
- NOAA-11 1997–2000: drift correction using observations from the SAGE-II instrument
- NOAA-16: adjusted to corrected NOAA-11 using SAGE-II as the transfer standard

[1] see the SBUV MOD home page for details,
http://acdb-ext.gsfc.nasa.gov/Data_services/merged/prof_external_cal.html

Layer	bottom pressure	Nimbus 7	NOAA-9 1992–1998		NOAA-11 1989–1994	NOAA-11 1997–2001		NOAA-16
			slope	offset		slope	offset	
1	1013 hPa	N/A	N/A	N/A	0	N/A	N/A	N/A
2	63.93 hPa	-0.19	-0.08	0.15	0	0.63	-2.55	-0.62
3	40.33 hPa	-0.37	-0.06	0.48	0	0.17	-0.92	1.09
4	25.45 hPa	-0.02	-0.40	1.31	0	-0.20	-0.13	-0.02
5	16.06 hPa	0.56	-0.64	1.48	0	-0.33	0.32	-2.21
6	10.13 hPa	0.58	-0.19	0.35	0	-0.17	0.58	-1.36
7	6.393 hPa	0.19	-0.08	0.02	0	-0.08	0.61	0.08
8	4.034 hPa	-0.03	-0.09	-0.05	0	-0.11	0.48	0.19
9	2.545 hPa	-0.03	-0.07	-0.03	0	-0.08	0.12	-0.07
10	1.606 hPa	0.01	-0.02	0.01	0	-0.02	-0.02	-0.03
11	1.013 hPa	0.01	0.00	0.01	0	0.00	-0.03	0.01
12	0.639 hPa	0.01	0.00	0.00	0	0.01	-0.02	0.00
13	0.403 hPa	N/A	N/A	N/A	0	N/A	N/A	N/A

Table 5.3.: *Partial column ozone bias adjustment coefficients for the different SBUV satellites. Bias adjustments for NOAA-9 and NOAA-11 (1997–2001) are linearly time-dependent, with offsets corresponding to decimal years 1993.0 for NOAA-9 and 1997.5 for NOAA-11. Units are DU (offsets) or DU/yr (slopes).*

Values for the adjustment coefficients are provided in Table 5.3.

Practise shows that applying adjustment coefficients to the SBUV v8 single profile observations before assimilation has both advantages and disadvantages.
Firstly, the adjustment coefficients are insensitive to geographical latitude, while in fact there are reasons to include a latitude dependence (Frith, 2010). This leads to small artifacts in polar ozone as mentioned in Section 8.2.
Secondly, while adjustment coefficients ensure consistency of assimilated profile ozone with the SBUV merged profile ozone data set in the SBUV vertical range, they also affect ozone mixing ratios in regions outside the SBUV domain through transport, due to the very long lifetime of ozone in the lower stratosphere. In particular, the large negative corrections of the NOAA-11 1997–2000 record in the lower stratosphere affect ozone in model levels below SBUV coverage. Negative ozone anomalies induced by the adjustment coefficients are transported downward and lead to low ozone in the bottom levels of the CTM, which are decisive for column ozone. Hence, column ozone in the assimilated data set using SBUV adjustments shows negative anomalies at the end of the 1990s. As the long-term trend study presented in Chapter 8 uses a piecewise linear trend model with a trend change at the turn of 1999/2000, the final years of the 20^{th} century have an increased weight for the determination of the 1979–1999 trend. Low TO3 anomalies during this period lead to a stronger negative TO3 trend in the assimilated data set with adjustment coefficients, thereby in fact reducing consistency of assimilated TO3 with the TOMS/SBUV merged TO3 data set.
In view of the ambivalence of advantages and disadvantages of SBUV adjustment coefficients for the assimilated data set, both versions of the assimilated data set are discussed in this

thesis. Concerning short-term variability of polar ozone, as discussed in Chapter 7, both data sets give virtually identical results. Differences are only encountered in the analysis of long-term ozone evolution in Chapter 8, and are discussed explicitly wherever necessary.

Part III.

Results and Discussion

6. The 29 year assimilated data set and sensitivity runs: Characterisation and validation

This chapter provides an overview of the model runs analysed in this thesis. In addition to the assimilation runs, which are described and validated in Section 6.1, also sensitivity calculations without assimilation were performed. These are described and validated in Section 6.3.

6.1. Assimilation runs: Setup

The assimilation scheme described in Chapter 4 was used to generate an assimilated data set of stratospheric ozone covering the period of 1979–2007, corresponding to the period for which SBUV satellite data are available. The assimilated data sets used in this thesis are slightly improved versions of the data set published by Kiesewetter et al. (2010a). Modifications concern the selection of satellites, the setup, and the assumed observational and representativeness errors.

The assimilation in Kiesewetter et al. (2010a) was set up as a concatenation of two-year runs, of which the first year was discarded as overlap. This method was chosen mainly for practical reasons, as a continuous run of the then unparallelised CTM would have consumed too much wall clock time. Thanks to recent improvements in high performance computing facilities at the Institute of Environmental Physics, and a shared-memory (OpenMP) parallelisation of the code, a significant speed-up was achieved in the meantime. Hence, all runs used in this thesis were performed in one piece for each assimilation period.

Two assimilation runs were conducted, differing only in the SBUV satellite data used. One configuration uses the original v8 profiles as published by Bhartia et al. (2004) (referred to as "standard" assimilated data set wherever distinction is needed), while the second setup uses the SBUV data with bias adjustment coefficients as obtained from S. Frith (personal communication, 2010) and discussed in Chapter 5. Differences between these assimilation runs are small for most periods, as the adjustment coefficients are. Notable differences are only encountered in the long-time evolution of ozone, owing to the strong corrections in lower stratospheric ozone during the NOAA-11 period 1997–2001. If not otherwise specified, the assimilated data set without adjustment coefficients is used throughout this thesis. For results concerning short-time variability of ozone as presented in Chapter 7, both versions of the assimilated data set give nearly identical results unless periods of satellite instrument switches are explicitly addressed.

6.1.1. General considerations regarding SBUV data and meteorological reanalyses

Both assimilation runs were performed in two periods separately, using the ERA-40 meteorological fields (1978–1999) or the ERA-Interim fields (1989–2007). The overlap allows for the characterisation of differences introduced by the different meteorological reanalysis data sets. Although ozone mixing ratios throughout most of the stratosphere are mainly determined by satellite information, some differences arise in areas not or little influenced by satellite observations. Below the limit of satellite coverage (64 hPa in the SBUV data version used here), ozone is mainly controlled by transport. Since some differences exist in the general circulation characteristics of ERA-40 and ERA-Interim fields, also lower stratospheric ozone exhibits some differences. Furthermore, ozone depletion in polar areas is mainly controlled by the volume of polar stratospheric clouds present during winter (Rex et al., 2004), which is itself very sensitive to temperatures in the polar lower stratosphere. Since these are not identical in ERA-40 and ERA-Interim data, amounts of springtime polar ozone loss – in particular those attained with the simple polar chemistry scheme used here – differ when the meteorological fields are switched. The upper limit of SBUV coverage, on the other hand, is not specified – theoretically, the uppermost partial column extends to the top of the atmosphere. Nonetheless there are limitations to the extent to which the SBUV data influence the final assimilated data set in the upper stratosphere, owing to the short lifetime of ozone in this region. This is basically a sampling issue universally valid for any kind of assimilation. For the assimilation to make sense, the interval between inputs of external information should be at maximum of the order of the internal lifetime of the processes that influence the assimilated quantity, in this case the photochemical lifetime of ozone. In the upper stratosphere, ozone lifetime can be as low as minutes to hours, while the satellite orbit covers the globe approximately every six days. Even though each observation influences not only the exact measurement location but also its surroundings with a certain correlation length, an ozone anomaly introduced by satellite data in the upper stratosphere has at least partly relaxed against its photochemical equilibrium steady state before the next observation is available. In practise, the transition between mainly satellite determined ozone vmr and mainly internally controlled ozone vmr is a gradual one, starting above around 1000–1200 K dependent on latitude and season. Although the assimilated 3D field is not realistic in the upper stratosphere since artificial ozone structures are introduced near the satellite tracks, the zonally averaged field still contains parts of the satellite information. Examples are discussed in Section 7.3 (signatures of solar particle precipitation in the assimilated data set) and Section 8.3.3 (gradual transition of assimilated ozone profile trends from SBUV to CTM chemistry).

6.2. Assimilation runs: Validation

In this section, the assimilated data set is validated against independent observations. Given the extent of the assimilated data set covering the whole stratosphere daily for three decades, it is impossible to provide a validation – or even characterisation – that is sufficiently complete. Thus, several aspects are taken out in order to provide a consistent picture. These aspects are the agreement of the ozone profile with observations, the capture of ozone structures during polar night, the consistency of high-latitude ozone profiles and variability with in-situ balloon borne sonde observations, and the distribution and variability of global col-

umn ozone.

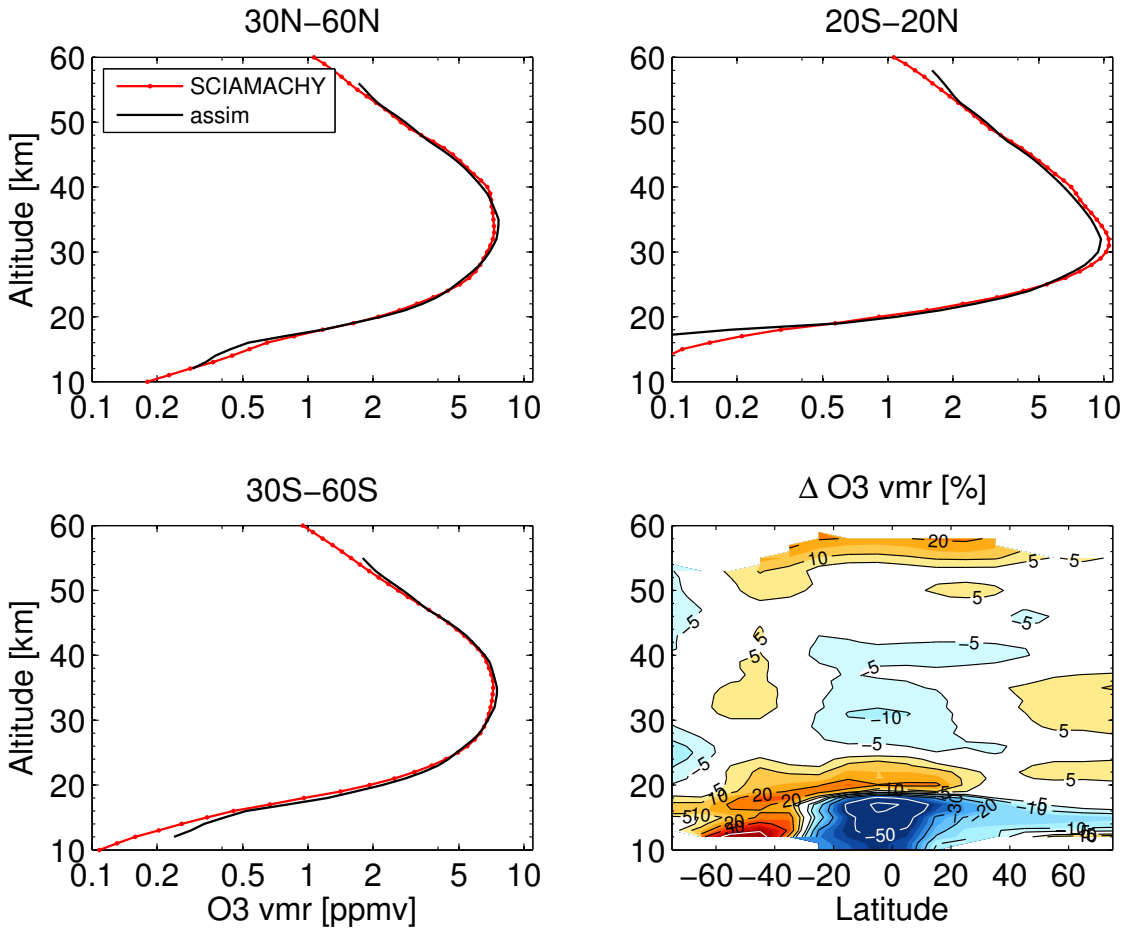

Figure 6.1.: *Mean ozone profiles in the assimilated data set as compare to SCIAMACHY limb observations during the period of overlap. (A) – (C): Mean profiles for different latitude bands (NH mid-latitudes, tropics, SH mid-latitudes). (D): Zonal mean differences between SCIAMACHY and the assimilated data set, calculated as* $(O3_{assim} - O3_{SCIAMACHY})/O3_{SCIAMACHY}$ *and expressed as percent values of SCIAMACHY observations.*

6.2.1. Global ozone profiles: SCIAMACHY

In order to evaluate the mean global ozone distribution in the assimilated data set, a comparison to independent satellite observations with near-global coverage is useful. The Scanning Imaging Absorption Spectrometer for Atmospheric Chartography (SCIAMACHY) (Bovensmann et al., 1999) is a satellite instrument flown on ENVISAT since 2002, which has the capability of measuring ozone both in nadir (for columns) as well as in limb sounding mode (for profiles). The limb viewing geometry provides for a significantly better vertical resolution of ozone profiles than achieved by the SBUV instruments, which use the nadir-looking geometry (perpendicular to Earth's surface) for obtaining both column and profile ozone in-

formation. The temporal overlap between SCIAMACHY measurements and the assimilated data set is only five and a half years; consequently, an evaluation of long-term behaviour of assimilated ozone is not possible with this instrument. Instead, I compare the average ozone profile in the tropics and mid-latitudes to that obtained from SCIAMACHY limb observations (Rozanov et al., 2007) during 2003-2007. Figure 6.1 shows mean profiles from SCIAMACHY and the assimilated data set for three latitude bands (tropics and mid-latitudes of both hemispheres), and zonal mean differences between them expressed as fractions of the SCIAMACHY values. From panels (A) to (C), it is obvious that the assimilated data set agrees very well to SCIAMACHY throughout most of the stratosphere, with differences mostly below 5% (see panel D). Considerable differences are only encountered in the lowermost stratosphere below the limit of SBUV data coverage (which is around 20 km), and in the uppermost model levels, where ozone is mainly controlled by the Linoz chemistry scheme. Differences vary with latitude, and there are different possible causes for offsets in the lowermost stratosphere. The upper boundary of the negative ozone anomaly in the tropics and the Northern Hemisphere (NH) follows rather closely the average tropopause height, indicating that the very simple treatment of the troposphere in the CTM may play a major role for offsets. Below the tropopause, which is calculated from a potential vorticity limit of 2 PVU (potential vorticity units, $1\,\mathrm{PVU} = 10^{-6}\mathrm{K\,m^2/(kg\,s)}$), complete mixing is applied every time step in order to account for the well-mixed nature of the troposphere. However, complete mixing of all tropospheric boxes is not very realistic. In the Southern Hemisphere (SH), positive offsets may indicate that the rather simple polar chemistry scheme applied here is not able to fully reproduce the extent of Antarctic ozone depletion. However, coverage of polar areas by SCIAMACHY is limited, and there are better options for validating polar ozone in the assimilated data set, which are presented in Sections 6.2.2 and 6.2.3 below.

6.2.2. Ozone structures during polar night: MIPAS

As a first test of how well structures are captured during polar night, I compare the assimilated data set to observations from the Michelson Interferometer for Passive Atmospheric Sounding (MIPAS) instrument on ENVISAT (von Clarmann et al., 2009). Due to its observational technique as an infrared sounder relying on emitted rather than backscattered radiation, MIPAS can operate without sunlight and is thus one of the few sources for satellite ozone data covering the polar regions during winter. Since ENVISAT was launched in 2002, MIPAS data are available only for a relatively short period, and not every day is fully sampled. Figure 6.2 shows maps of assimilated ozone together with MIPAS observations interpolated to several stratospheric pressure levels at 17 Nov 2007. As a reference, the corresponding SBUV ozone vmr at these pressures (as provided in the SBUV profiles data set and used for assimilation) are shown.

Despite minor (expected) biases between the assimilated ozone and MIPAS observations, it is obvious that many structures of the ozone field are captured well in the assimilated data set, even though SBUV observations from NOAA-16 terminate around 58°N and no external information enters the assimilation beyond this latitude. As a quantitative measure of the good agreement even at high latitudes, Figure 6.3 shows the spatial correlation between MIPAS and assimilated ozone poleward of 58°N (A) and their mean difference (B) for the whole vertical range available. The correlation is higher than 0.6 for the whole stratosphere; values of 0.9 and more are attained throughout the middle stratosphere between 25 and 5 hPa ($\sim 24 - 34$ km). The assimilated data set shows no or small positive biases in the

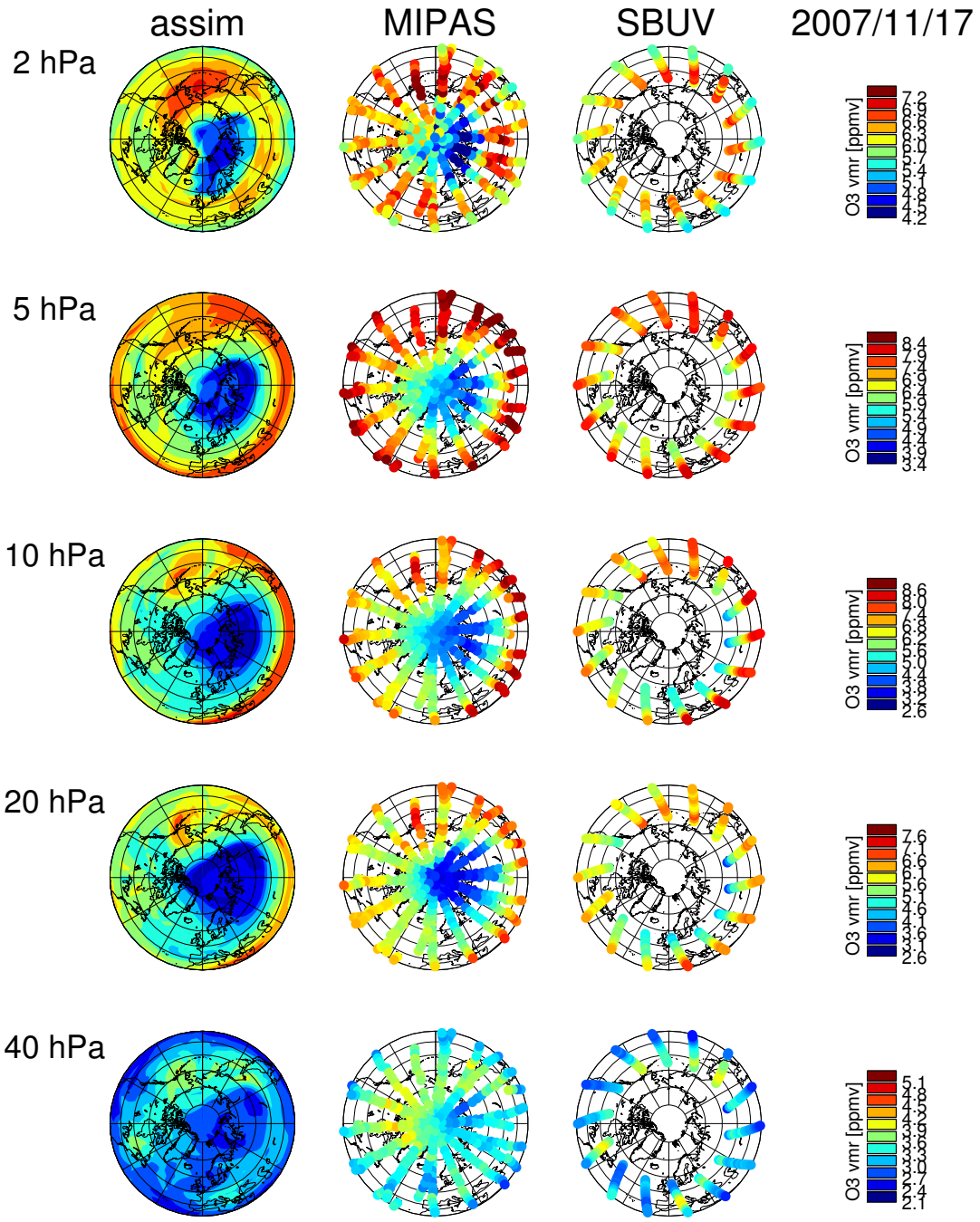

Figure 6.2.: *Comparison of assimilated ozone field with MIPAS and SBUV observations for a single day (17 Nov 2007), interpolated to SBUV pressure levels. Although minor differences to MIPAS are visible, ozone structures are represented well in the assimilated data set, despite the lacking coverage of high latitudes by the SBUV data used for assimilation (NOAA-16). Plot updated from Kiesewetter et al. (2010a).*

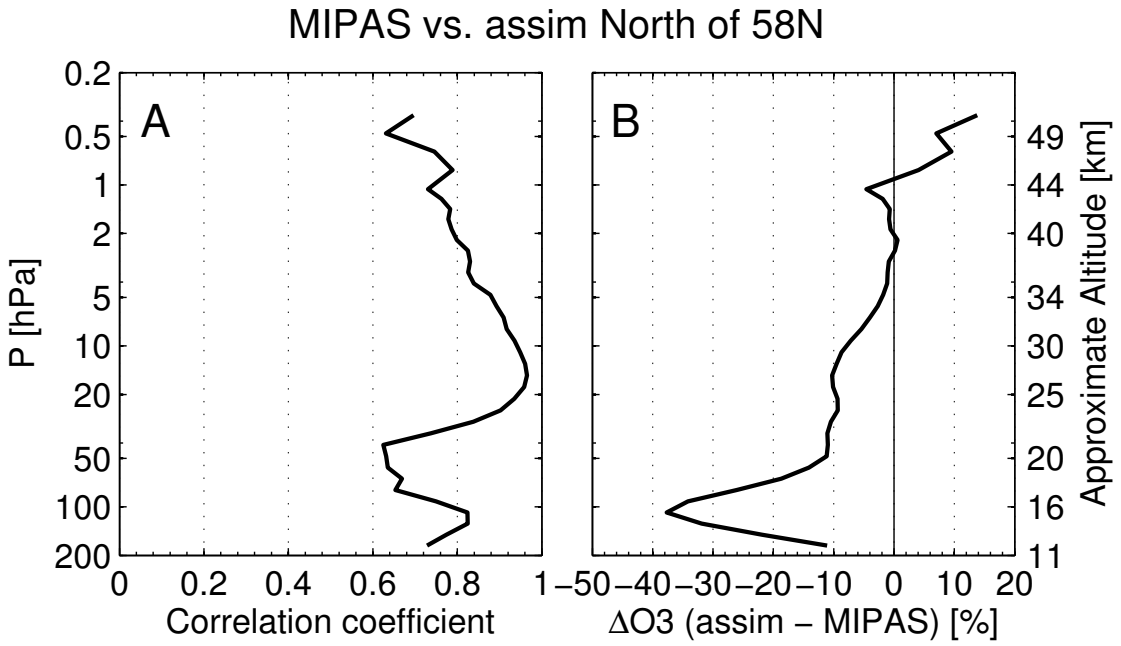

Figure 6.3.: *Correlation (A) and mean difference (B) between assimilated ozone and MIPAS observations at high latitudes, for the same day as in Figure 6.2 (17 Nov 2007).*

upper stratosphere, small negative biases in the middle stratosphere, and moderate biases of up to −35% in the lower stratosphere. The biases are at least partly attributable to biases between SBUV and MIPAS outside polar latitudes, where the CTM is constrained by the SBUV observations (see Figure 6.2). It should be emphasised that this comparison takes into account only a single day and can mainly demonstrate that spatial ozone patterns are represented well in the assimilated data set. Due to the relatively short period for which MIPAS data are available, MIPAS is not ideally suited for evaluating the long-term variability of polar ozone in the assimilated data set. This is done by sonde observations in the following section.

6.2.3. High-latitude ozone variability: Sondes

Balloon borne sondes provide highly accurate in-situ measurements of ozone which are, however, only available at single locations. As chemical detectors, sondes are able to measure during polar night and are thus ideal for evaluating the assimilated data set at high latitudes. Here sonde data from four different polar stations are used: Ny-Ålesund (79°N, 12°E), Sodankylä (67°N, 27°E), Neumayer (71°S, 8°W), South Pole (90°S).
Sonde data are available starting in 1992 (Ny-Ålesund, Neumayer), 1991 (South Pole) and 1989 (Sodankylä), respectively. Altogether, more than 1600 sondes are used from Ny-Ålesund, more than 1400 from Sodankylä, more than 1250 from Neumayer and more than 1150 from South Pole.
In order to provide a qualitative impression of the agreement between assimilated data set and sondes, Figure 6.4 displays the ozone time series from the assimilated data set at the

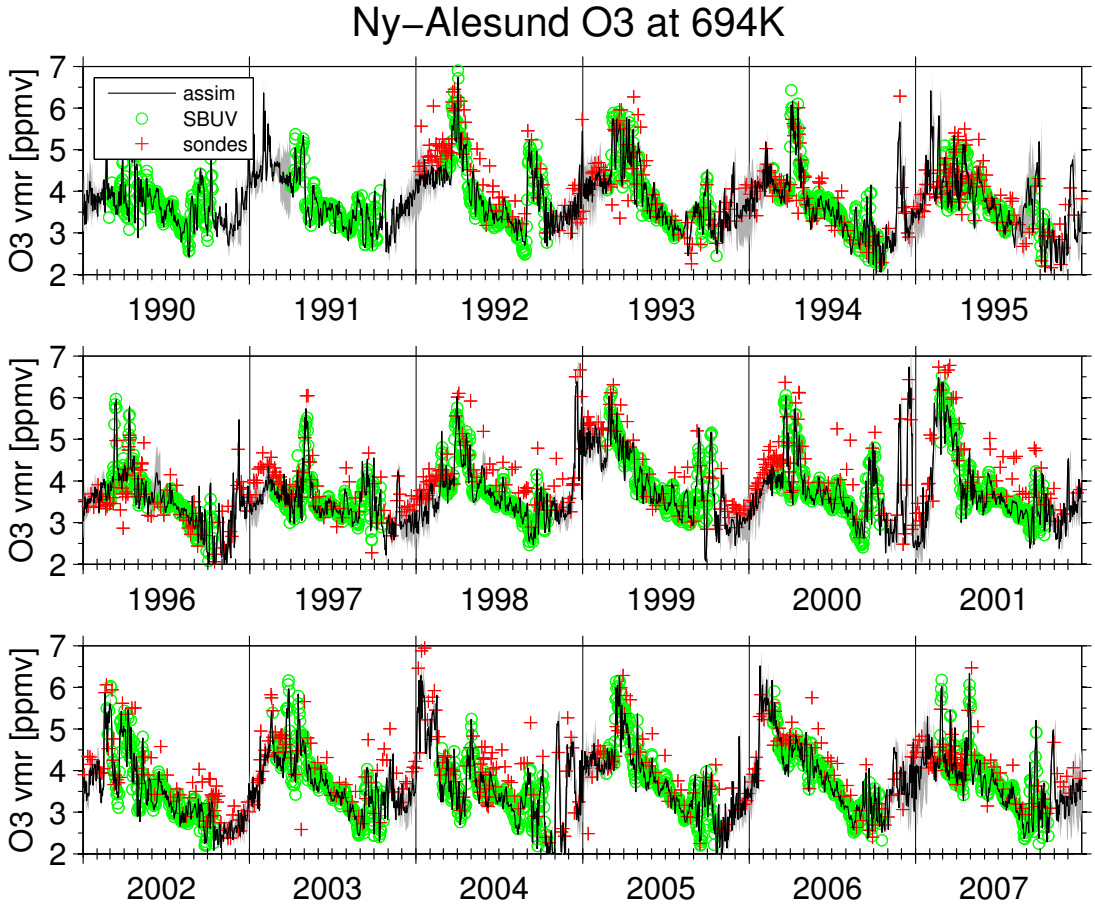

Figure 6.4.: *Ozone time series at Ny-Ålesund, Spitsbergen, at 694 K (27 km). Assimilated data set (black line), sondes (red +), and SBUV satellite observations located within 250 km (green ∘) are shown. Uncertainty estimates as provided by the ozone error tracer for the assimilated data set are indicated as grey shadow.*

Arctic site of Ny-Ålesund, Spitsbergen, as compared to sonde observations (shown as red +) and SBUV satellite observations (green ∘) at the mid-stratospheric potential temperature level of 694 K ($\sim$ 27 km). Sondes are vertically averaged to the respective potential temperature level (since both pressure and temperature are measured by the sondes, potential temperature is a native, if unusual, vertical coordinate of the sondes). SBUV observations located within 250 km from the Ny-Ålesund station are averaged and interpolated to CTM potential temperature levels. In addition, the ozone error tracer contained in the assimilated data set is shown as grey shadow around the assimilated time series (black line). Given the several empirical parameters that enter the assimilation process, the error tracer should be treated with caution and not taken as 1σ error bars. Nonetheless, it can provide a qualitative estimate for the uncertainty in the assimilated data set. The overall agreement between assimilated ozone and sonde observations is striking. During summer months, assimilated ozone is dominated by the SBUV observations, which provide a very dense coverage during

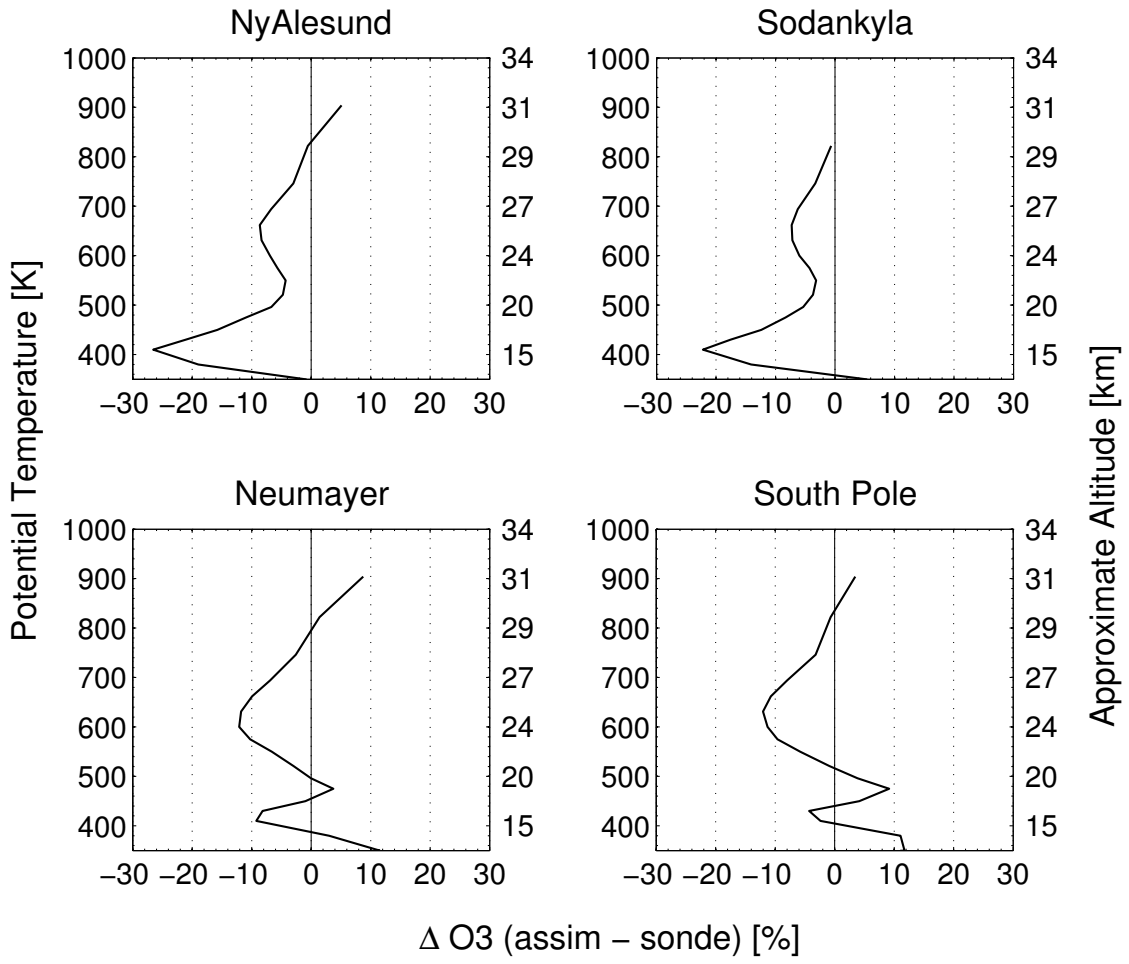

Figure 6.5.: *Relative differences between assimilated ozone vmr and sonde observations from high-latitude stations (NH: Ny-Ålesund, Sodankylä; SH: Neumayer, South Pole), averaged over the whole available sonde time series (1991–2007) and expressed as fraction of sonde ozone vmr.*

these months. Differences to the sonde time series are mainly caused by offsets between sondes and SBUV. However, also during the winter periods without input of satellite data, assimilated ozone is in good agreement with sonde observations. The growth of assimilated ozone uncertainty in the absence of SBUV data during winter is well visible. Most sonde observations lie within the error margins given by the ozone error tracer, indicating that its order of magnitude is reasonable.

In order to quantify the agreement of the assimilated data set with sondes, the mean offset and correlation is calculated. Figure 6.5 shows the mean overall difference between monthly means of assimilated ozone and sonde ozone at the four stations. Differences are averaged over the whole period for which sonde observations are available, and expressed as fractions of sonde ozone vmr. In the middle and lower stratosphere, assimilated ozone is generally lower than sonde ozone; above 850 K, differences reverse. Deviations are less than 10% throughout most of the stratosphere; only in the Arctic lower stratosphere, differences of up to around

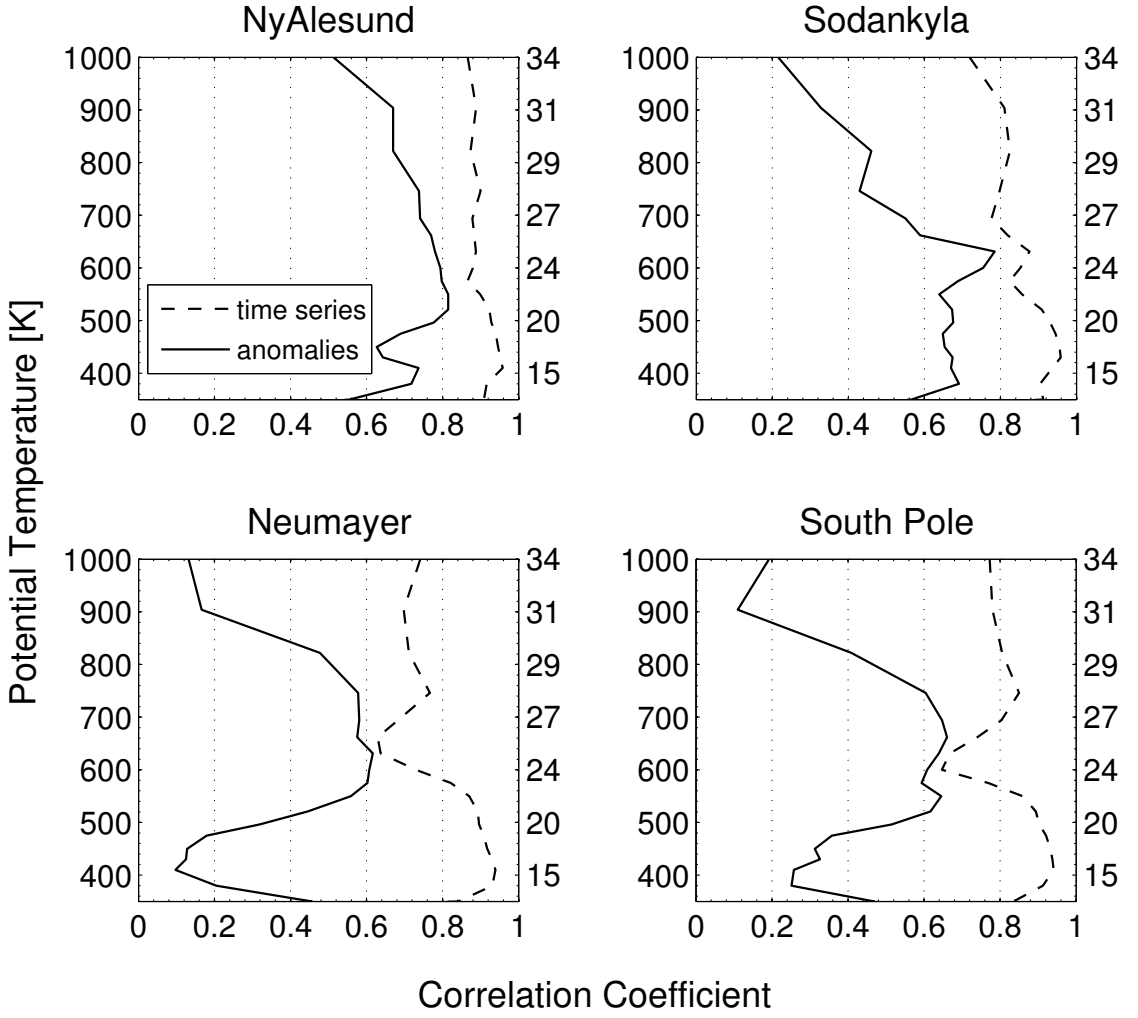

Figure 6.6.: *Correlation between assimilated data set and sonde ozone time series and anomalies (deviations from the annual cycle) at four high-latitude stations (NH: Ny-Ålesund, Sodankylä; SH: Neumayer, South Pole) during the whole period for which sonde observations are available (1991–2007).*

25% are found. This picture is consistent with the differences to MIPAS observations for a single day (Figure 6.3). Note the different vertical ranges of Figures 6.5 and 6.3 as sonde measurements terminate around 1000 K (32 km/5 hPa).

In the middle stratosphere, differences are mainly attributable to differences between SBUV and sonde ozone. Below about 450 K, however, SBUV is not sensitive and thus assimilation does not influence modelled ozone directly; differences there arise from the downward transport of low-ozone air from higher altitudes. Here we experience a rather fundamental difficulty with assimilation: while constraining the CTM leads to more realistic ozone values in atmospheric regions where data are available, and propagation of information beyond the limits of observational coverage is generally an advantage of the assimilation scheme, it may

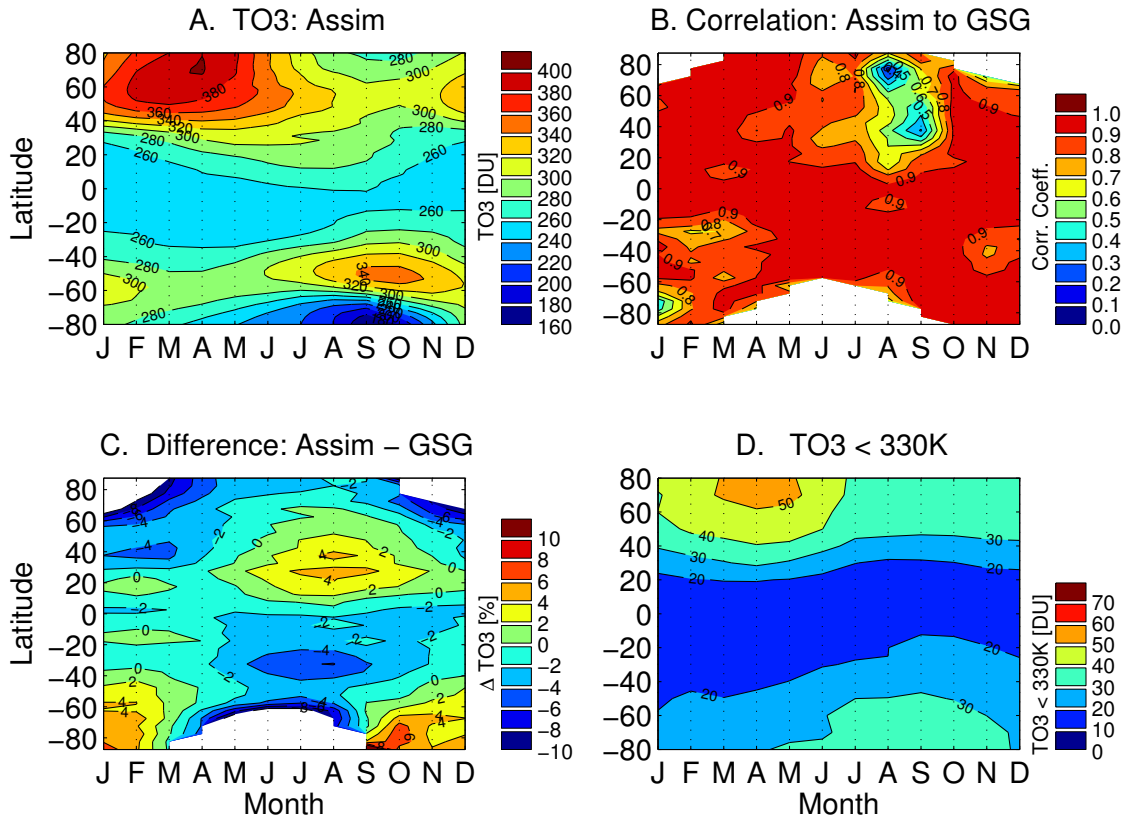

Figure 6.7.: *Comparison of column ozone in the assimilated data set to GOME/SCIAMACHY/GOME2 ("GSG") observations, using monthly zonal means: A. TO3 in the assimilated data set (including climatological column below the CTM boundary), B. Correlation of the assimilated TO3 to GSG TO3, C. Difference between TO3 in the assimilated data set and GSG, expressed as fractions of GSG TO3, D. Climatological Ozone Column below 330 K (lower model boundary) that has been included in (A)-(C).*

induce unrealistic effects if e.g. the transport in the CTM is not perfectly realistic. However, the offsets are of an order of magnitude which is not detrimental to studying ozone variability. Indeed, one could consider implementing an external correction of lower stratospheric ozone vmrs in the assimilated data set, if realistic absolute values are needed in the lower stratosphere.

A quantitative measure of how well the variability of polar ozone is captured in the assimilated data set is the correlation of time series and anomalies for the whole period of sonde observations. These are shown in Figure 6.6 for the same four sonde stations as used before. Anomalies are calculated as monthly means of daily deviations from the annual cycle. The correlation of time series is higher than 0.9 in the lower stratosphere and decreases only slightly in the middle stratosphere, demonstrating how well the overall ozone time series is captured in the assimilated data set. Parts of this high correlation are of course due to the annual cycle signal. However, even if the latter is removed to show correlations of ozone anomalies, the correlation coefficients decrease only moderately except for the Antarctic

lower stratosphere. Correlations of anomalies exceed 0.6 throughout the middle stratosphere at all stations, i.e. at altitudes where the sondes have their best accuracy, showing that the assimilated data set gives a realistic representation of ozone variability throughout most of the stratosphere. It is noteworthy that the realism of ozone anomalies, as depicted by this correlation, is independent of absolute offsets; even at altitudes where the relative vmr offset is large (up to −25% around 16 km), ozone variability is still captured well. Only in the SH does the correlation and hence the realism of ozone variability in the assimilated data set decay rapidly below 500 K. This effect is at least partly attributable to the lower quality of the meteorological analyses used in driving the CTM in the SH, and – as this is the vertical region of polar heterogeneous chemistry – may also be related to the simplicity of the polar chemistry scheme used.
The evaluation of polar ozone in the assimilated data set against sondes and MIPAS satellite observations shows that although absolute biases exist in the lower stratosphere, the assimilated data set is very well suited to study polar (especially Arctic) ozone variability.

6.2.4. Column ozone: GOME/SCIAMACHY/GOME2

Global total column ozone data are available from a number of satellite instruments. A continuous data set of satellite-observed TO3 is provided by the TOMS/SBUV merged data set using TOMS (Total Ozone Mapping Spectrometer) and SBUV observations from 1979-2009, which is used at other stages throughout this thesis for comparison of long-term evolution of column ozone. However, in order to validate the assimilated data set against completely independent observations, I use data from the Global Ozone Monitoring Experiment (GOME) (Burrows et al., 1999; Coldewey-Egbers et al., 2005; Weber et al., 2005), the Scanning Imaging Absorption Spectrometer for Atmospheric Chartography (SCIAMACHY) (Bovensmann et al., 1999; Bracher et al., 2005), and GOME2 (Callies et al., 2000) instruments, which have been operating on ESA research satellites. Although GOME, SCIAMACHY and GOME/2 are independent instruments, their retrieval algorithms and observations are sufficiently similar to allow merging their data sets into one, covering 1995 to present. This GOME/SCIAMACHY/GOME2 ("GSG") merged data set has been compiled by Weber et al. (2007), who adjusted the SCIAMACHY (2002–present) and GOME2 (2007–present) data to the GOME data record (1995–present) by determining a mean scaling factor (GOME2 and SCIAMACHY) and trend (SCIAMACHY only) in the monthly mean zonal mean ratios. Using the criterion of optimum global sampling the GSG data set is then composed of GOME1 until 2003, SCIAMACHY (2003–2006), and GOME2 after 2006[1]. In order to enable comparison of modelled column ozone to observations, the ozone column below the lower boundary of the CTM has to be taken into account. This is taken here from the climatology of Fortuin and Kelder (1998) and added to TO3 calculated from the assimilated data set. Monthly zonal mean TO3 in the assimilated data set is shown in Figure 6.7 (A). Inter-annual variations correlate very well with the GSG data set (B), with correlation coefficient exceeding .9 for most latitudes and months. Note that due to the decomposition in monthly time series, effects of the annual cycle are automatically left out in this depiction. Panel (C) shows differences between the assimilated data set and the GSG data set, expressed as percent values of GSG TO3. Both data sets show excellent agreement within less than 5% at most latitudes and months, with maximal offsets of +6% in Antarctic spring and

[1]The GSG data set is described at and available from
http://www.iup.uni-bremen.de/gome/wfdoas/wfdoas_merged.html.

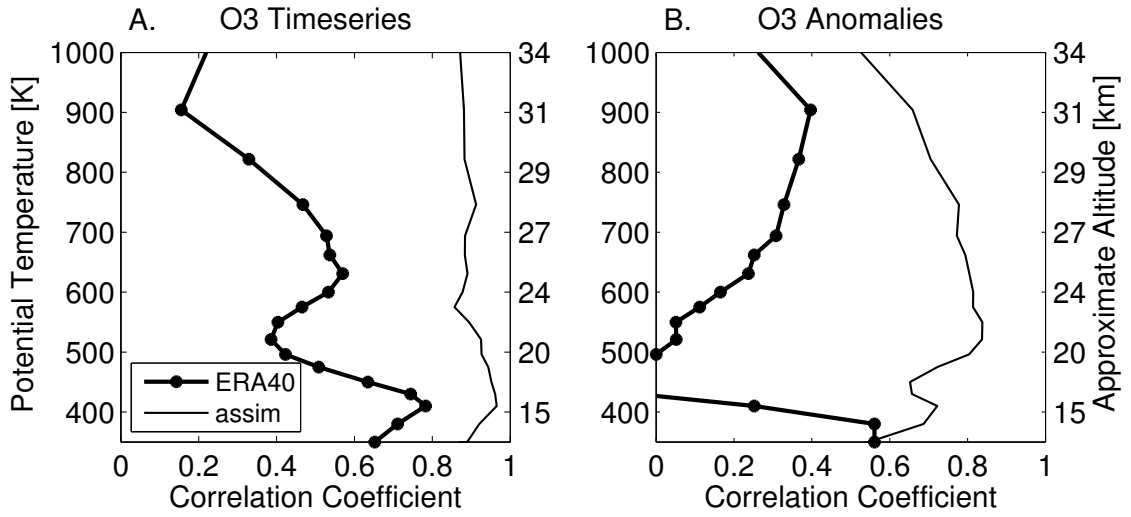

Figure 6.8.: *Correlation of ERA-40 ozone to Ny-Ålesund sondes (solid black line with symbols), as compared to assimilated ozone (thin black line) for the period of 1991-2001. (A): Correlation of monthly mean ozone vmr (including the annual cycle), (B): Correlation of monthly mean ozone vmr anomalies (annual cycle subtracted). High-latitude profile ozone variability seems underrepresented in the ERA-40 data set. Plot updated from Kiesewetter et al. (2010a).*

−8% in Arctic spring. Taking into account that column ozone itself is not assimilated, the agreement to independent observations is remarkable. The long-term evolution of column ozone in the assimilated data set is discussed extensively in Chapter 8.

6.2.5. Comparison to ERA-40 Ozone

The ERA-40 reanalysis (Dethof and Hólm, 2004; Uppala et al., 2005) conducted by ECMWF provides a 3D global long-term data set of stratospheric ozone obtained from a 3DVAR (intermittent variational) assimilation of SBUV profile ozone and TOMS column ozone. While ERA-40 total column ozone has been reported to agree well with independent observations (Dethof and Hólm, 2004; Jiang et al., 2008b), studies involving ERA-40 ozone profiles seem scarce. A comparison of ERA-40 ozone to sonde observations at high latitude shows that although the annual cycle is captured reasonably well, polar ozone variability is severely under-represented in this data set. Figure 6.8 displays correlations between ERA-40 ozone and Ny-Ålesund sonde observations for the years 1991-2001 (solid line with symbols). (A) compares the monthly mean ozone vmr time series including the annual cycle, while (B) only shows correlations between monthly mean anomalies (mean annual cycle subtracted). At the Arctic site of Ny-Ålesund, the ERA-40 correlation to sondes is generally rather low and in particular much lower than the correlation of the assimilated data set to the sondes (thin black line). Thus the ERA-40 ozone data set does not seem suited to address the research objectives of this thesis.

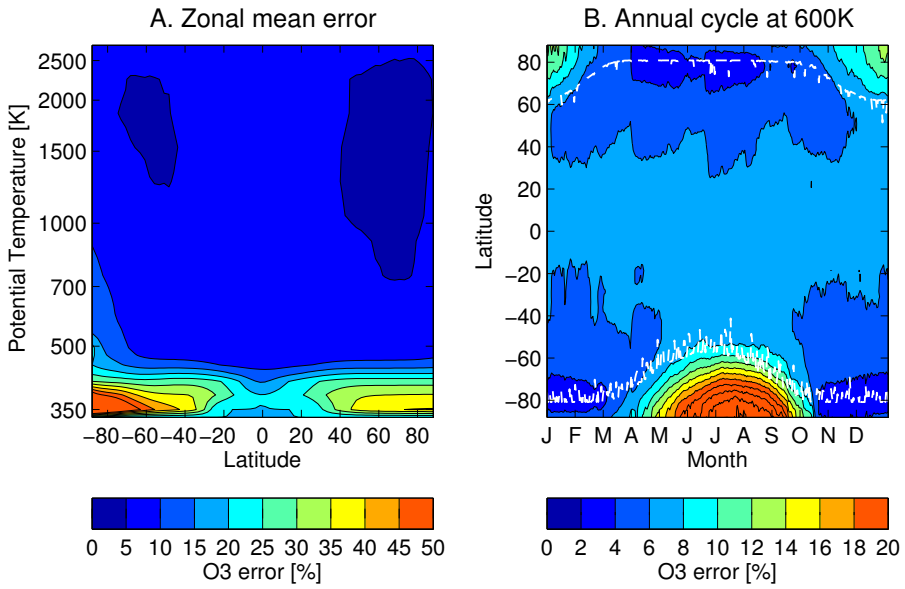

Figure 6.9.: *Relative ozone analysis error, averaged over the whole assimilation period. (A): Meridional cross-section. (B): Annual cycle in the middle stratosphere (600 K / ∼24 km). In addition to the colour coded error tracer, also the average extent of SBUV satellite coverage is shown in (B) as white dashed line. While in oversampled areas (high latitudes during summer) the error decreases distinctly below the measurement error, the error increases strongly in the absence of data during polar night.*

6.2.6. Error statistics and χ^2 validation

As discussed in Chapter 4, an ozone error tracer is included in the assimilation scheme, which grows in the absence of observations and can thus also be used to evaluate the content of observed information present in the assimilated field.

Figure 6.9 shows averages of the ozone error tracer during the whole assimilation period. The meridional cross-section of the error tracer (panel A) shows that the error tracer is between 5 and 10% in the vertical range of SBUV coverage, while it increases sharply below 450 K, reaching values of 30 to 50% in the polar lower stratosphere. Hence, although offsets to sondes and satellite observations are present in the lower stratosphere (see Figures 6.1 and 6.5), these are mostly below the analysis error. The annual cycle of the error tracer in the middle stratosphere is displayed in panel (B). In the absence of data during polar night, the error increases to $\sim$ 20% in both hemispheres, while in high-latitude summer it decreases distinctly below the observational error due to oversampling. It is interesting to note that although satellite coverage is meridionally symmetric (with a 6 month time shift between the hemispheres), the error tracer is not. The increase of error in the Antarctic polar night is smoother and stronger than in the Arctic, related to the greater stability of the Antarctic polar vortex. In the Antarctic, the error starts to grow monotonously as soon as SBUV coverage retreats in April. The Arctic polar vortex is less secluded from mid-latitude influences, and air masses which have shortly been in contact to satellite data and thus carry a low ozone error are able to enter the Arctic vortex until winter, leading to a less

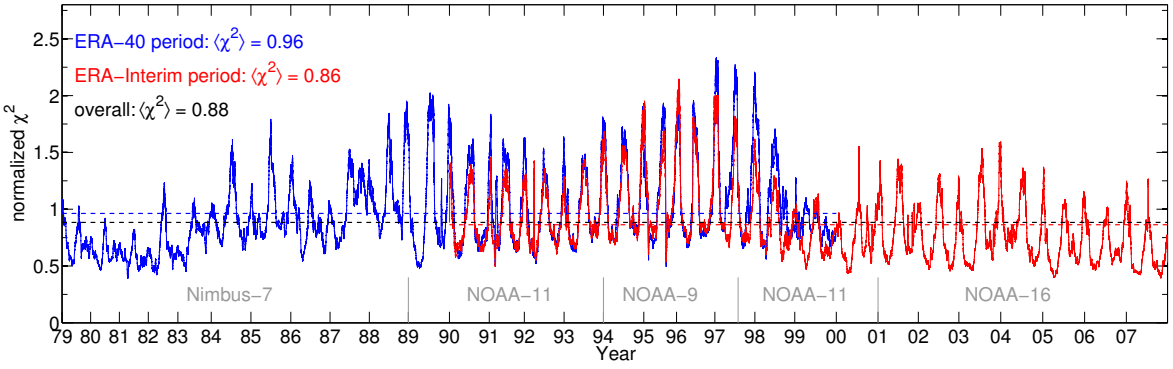

Figure 6.10.: *Normalised χ^2 diagnostic for the whole assimilated data set, smoothed with a running mean of approximately 5 days. Assimilation periods driven by different meteorological data are shown including their overlap (ERA-40 period in blue and ERA-Interim period in red). The overall mean is shown as black dashed line. Periods of different satellites used for assimilation are marked in grey. Note that the distances between ticks corresponding to individual years vary, as only model time steps with observations are taken into account here, the absolute amount of which varies between different years.*

uniform and weaker error growth. Again, it should be noted that the absolute magnitude of the error must be treated with caution, as it is intrinsically determined by the choice of empirical parameters, which are tuned based on reasonable assumptions and a χ^2 diagnostic. Hence the ozone error tracer is best viewed as a tool needed in the assimilation process and a qualitative estimate of uncertainty rather than a precise quantification of error. Nonetheless, it is a strong point of the Kalman filter algorithm used in sequential data assimilation that a reasonable error estimate is provided, as it does not rely on the assumption of a perfect dynamical model applied in other assimilation techniques like 4DVAR.

In addition to the error tracer, a goodness-of-fit diagnostic χ^2 is calculated every analysis step, which provides a quantitative measure of – depending on the point of view – either the quality of the analysis as a fit to the data, or the suitability of chosen assimilation parameter values. Both perceptions are true, but since empirical assimilation parameters such as the model error growth rate are tuned by hand, one may prefer to view χ^2 as a diagnostic of whether these parameters are chosen reasonably. Figure 6.10 shows the long-term evolution of normalised χ^2, that is, $\tilde{\chi}^2 \equiv \chi^2/n_{obs}$. Since the expectation value of χ^2 is n_{obs}, the number of available observations, the expectation value of $\tilde{\chi}^2$ is 1. As seen in Figure 6.10, the long-term mean of $\tilde{\chi}^2$ in the assimilated data set is indeed close to 1 (0.96 for the ERA-40 driven period, 0.86 for the ERA-Interim period, and 0.88 for the overall data set if overlapping periods are averaged first). Also visible are some differences in χ^2 evolution between different satellite instrument periods, which may raise the question whether the assumption of uniform observation errors for all different SBUV(/2) satellite instruments involved is optimal. Since published uncertainties (Bhartia et al., 1996) refer to the whole retrieval scheme and no detailed information about ozone profile errors for individual instruments is available, this is the only option so far. In addition, periodical oscillations in $\tilde{\chi}^2$ with a period of about six months are observed for some instruments. This

Run	Linoz	Polar chemistry
tt	changing ODS levels	$1/\tau_{pc} \propto$EESC
tT	changing ODS levels	$1/\tau_{pc} \propto$ EESC^2
ct	constant ODS levels (2000)	$1/\tau_{pc} \propto$EESC
tc	changing ODS levels	constant EESC (2000)
cc	constant ODS levels (2000)	constant EESC (2000)
tn	changing ODS levels	off

Table 6.1.: *CTM sensitivity simulations, distinguished by the gas phase chemistry (Linoz) mode and polar chemistry mode used. For details see text.*

feature is most likely related to the periodically changing latitudinal coverage and sampling of the satellite instruments (symmetric near equinox vs. strongly asymmetric near solstice). In this regard, it might be desirable to develop a local χ^2 diagnostic which would provide a goodness-of-fit diagnostic for different areas. On the other hand, assimilation parameters are set globally, and thus a global diagnostic is the only suitable one to evaluate their selection. It does not make sense to choose e.g. the model error growth differently for different latitudes, as it represents uncertainties inherent to the CTM advection and chemistry scheme. Hence, the best one can do is to ensure that in spite of oscillations, the long-term mean of $\tilde{\chi}^2$ is near 1 when parameters are chosen reasonably.

6.3. Sensitivity runs

In addition to the assimilation runs, sensitivity runs of the unconstrained CTM without assimilation were conducted, which allow for the attribution of ozone changes in the assimilated data set and satellite observations to natural and anthropogenic contributions. For this purpose, chemistry and polar chemistry schemes were used that either do or do not take into account the temporal variation of ozone depleting substances. Details on the setup for the sensitivity runs are provided in Section 6.3.1, and a characterisation and validation is attempted in Section 6.3.2.

6.3.1. Setup

Six CTM sensitivity runs without assimilation were performed in total, which are labelled here according to the gas-phase chemistry (Linoz) and the polar chemistry scheme used. Five runs use polar chemistry; in these, Linoz and polar chemistry were switched between changing atmospheric composition (as described in Chapter 3) and constant year 2000 conditions. The sixth run, which does not use polar chemistry, is used only as a reference for diagnosing the effects of polar ozone loss.
Labels consist of two letters, the first of which corresponds to the gas phase chemistry scheme used and the second corresponds to polar chemistry scheme. Letters referring to gas-phase chemistry are "*t*" for time-dependent (varying stratospheric chlorine loadings) and "*c*" for constant year 2000 conditions. Letters describing the polar chemistry mode are "*t*" for time-dependent with $1/\tau_{pc} \propto$ EESC, "*T*" for time-dependent with $1/\tau_{pc} \propto$ EESC^2, "*c*" for constant year 2000 conditions ($\tau_{pc} = 10\,\mathrm{d}$), and "*n*" for none. The *tt* run uses the same model setup as the assimilation, and is thus suited best to diagnose effects of the assimilation by direct comparison to assimilated ozone. Run *tT* was included to provide an upper estimate on

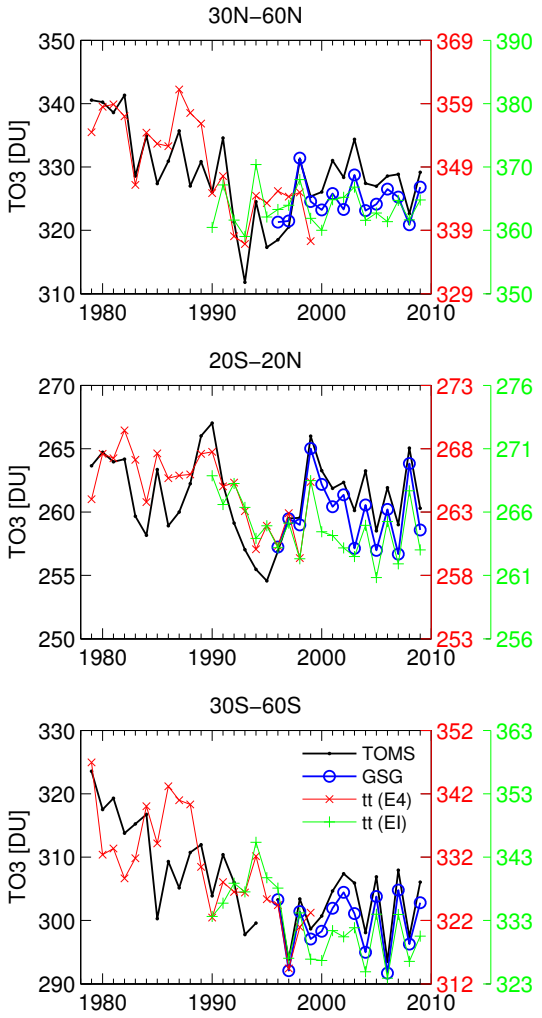

Figure 6.11.: *Total ozone (TO3) annual means from the tt (EESC-dependent chemistry) model run, as compared to the TOMS/SBUV merged TO3 data set ("TOMS") and the GOME/SCIAMACHY/GOME2 merged data sets ("GSG"). Model runs are performed in two periods, 1979–1999 using ERA-40 data (red ×), and 1990–2009 using ERA-Interim (green +). Modelled TO3 shows different offsets to satellite observations during the two periods, which are accounted for by the shifted right-hand axes (middle axis 1979–1999, far right axis 1990–2009). The climatological ozone column below 330 K has been added to modelled TO3. Plot updated from Kiesewetter et al. (2010b).*

the effects of changing stratospheric chlorine levels on polar heterogeneous ozone depletion. As discussed in Section 3.3, the dependence of the ozone decay rate on EESC is usually estimated to be close to linear (Harris et al., 2010); however, there are also arguments in favour of a stronger dependence (Searle et al., 1998; Hsu and Prather, 2009). Reasonable values for the exponent n ($1/\tau_{pc} \propto \mathrm{EESC}^n$) range between 1 and 2 (Searle et al., 1998). Model runs analysed in this paper generally use either $1/\tau \propto$ EESC or constant year 2000 conditions. In order to obtain an upper estimate for trends related to polar chemistry, one

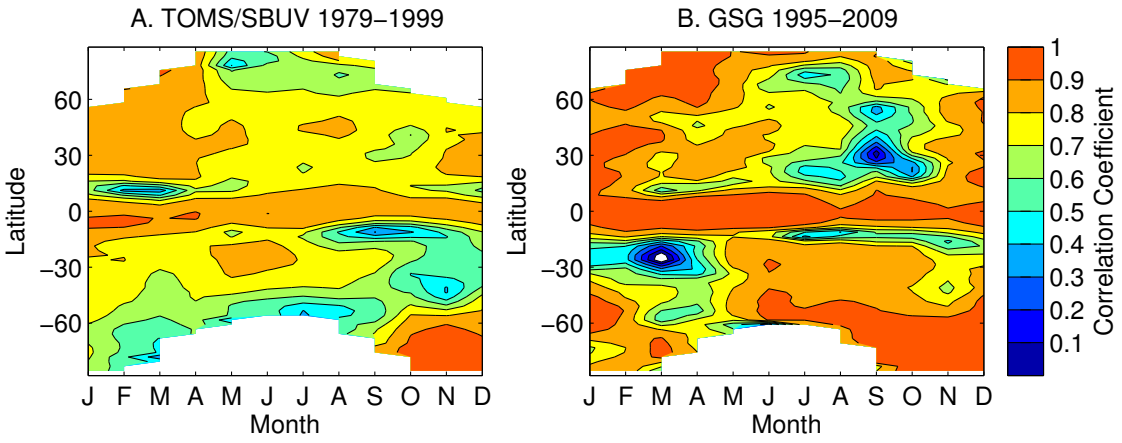

Figure 6.12.: *Correlation between observed and modelled TO3 (sensitivity run tt) monthly mean time series: TOMS/SBUV vs. CTM in the period 1979–1999 (A, model driven by ERA-40) and GOME/SCIAMACHY/GOME2 vs. CTM in the period 1990–2009 (B, model driven by ERA-Interim). Plot modified from Kiesewetter et al. (2010b).*

additional model run uses a higher scaling of $1/\tau \propto \mathrm{EESC}^2$.
An overview of the sensitivity runs is provided in Table 6.1. As with the assimilation, all model runs were performed in two parts, 1978–1999 (using ERA-40 data) and 1989–2009 using ERA-Interim. A long overlap period using different meteorological analyses is desirable in order to quantify offsets in ozone concentrations due to differences in the meteorological data sets, such as discussed by e.g. Chipperfield (2003) and in this thesis. The runs starting in January 1978 are initialised from an ozone climatology (Fortuin and Kelder (1998)), while the runs starting January 1989 use output from the corresponding ERA-40 driven run as initial conditions. In both cases, the first year is discarded in the analysis in order to account for initial spin-up and meteorological transition.

6.3.2. Validation

If the CTM sensitivity runs are to provide useful information about trend contributions, it is first necessary to establish that the CTM with EESC-varying chemistry (*tt*) represents the evolution of observed ozone well. Naturally, a validation of sensitivity runs which explicitly do not include a realistic evolution of ODS levels or no polar chemistry at all is not possible – for these, it can only be determined that their deviations from observed data have the expected sign and magnitude. Here the Total Ozone Mapping Spectrometer and Solar Backscatter UV (TOMS/SBUV) merged data set (Stolarski and Frith, 2006) as well as the GSG merged data set introduced in Section 6.2.4 are used to evaluate the performance of the unconstrained *tt* model run.
Figure 6.11 shows time series of annual mean *tt* TO3 (ERA-40 and ERA-Interim driven runs separately) together with TOMS/SBUV and GSG for mid-latitudes and tropics. The climatological ozone column below 330 K (see Figure 6.7 D) is added to modelled TO3. Modelled ozone is generally higher than the observations, with a distinct difference between the different meteorological data sets (note the shifted right-hand side axes regarding mod-

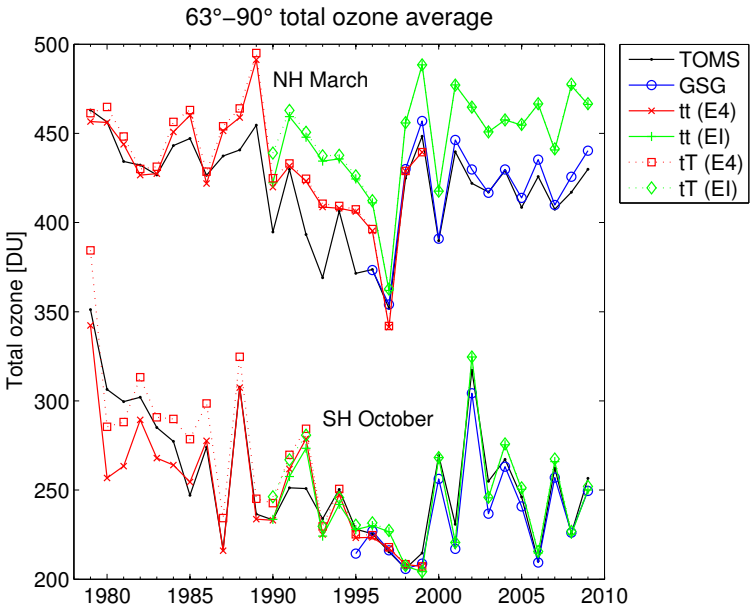

Figure 6.13.: *Springtime polar TO3 from model runs with EESC-dependent chemistry, as compared to the TOMS/SBUV merged TO3 data set ("TOMS") and the GSG merged data set. In addition to the tt model runs using a polar chemistry scheme scaled linearly with EESC, also the tT model runs are shown, which use a polar chemistry reaction rate* $\propto \mathrm{EESC}^2$*. Model runs are performed in two periods, 1979–1999 using ERA-40 wind fields ("E4"), and 1990–2009 using ERA-Interim ("EI"). The climatological ozone column below 330 K has been added to modelled TO3. From Kiesewetter et al. (2010b).*

elled TO3 during the different periods). The fact that switching meteorological analyses during a model run can lead to considerable artifacts in modelled ozone has been reported before, e.g. by Feng et al. (2007) at the transition between ERA-40 and operational analyses. Mean offsets are +19 DU (+40 DU) in NH mid-latitudes, +3 DU (+6 DU) in the tropics, and +22 DU (+33 DU) in SH mid-latitudes for ERA-40 (ERA-Interim) meteorological fields. Ozone variability is represented well especially in the 1979–1999 period (ERA-40) and in the 2000–2009 period (ERA-Interim), while the ERA-Interim driven run does not fully capture the increase in TO3 around the turn of the century.

In order to quantify how well the inter-annual variability of column ozone is captured in the unconstrained CTM, Figure 6.12 shows correlation coefficients in the same manner as Figure 6.7 (B) for TOMS/SBUV in the interval 1978–1999 (CTM driven by ERA-40 data) and for GSG in the 1995–2009 interval (CTM driven by ERA-Interim data). Although correlation coefficients are lower than for assimilated ozone, also the *tt* reference run captures inter-annual variability well. For the ERA-40 driven run, correlation coefficients are higher than 0.7 for most of the NH, higher than 0.8 in the inner tropics, and higher than 0.9 for Antarctic spring. In the case of the ERA-Interim driven run, most features are similar; correlation in the inner tropics and throughout mid-latitude/polar spring and summer is excellent, while the model seems to have some difficulties in reproducing subtropical ($\sim 30°$) ozone variability for single months (mainly September in the NH and March in the SH).

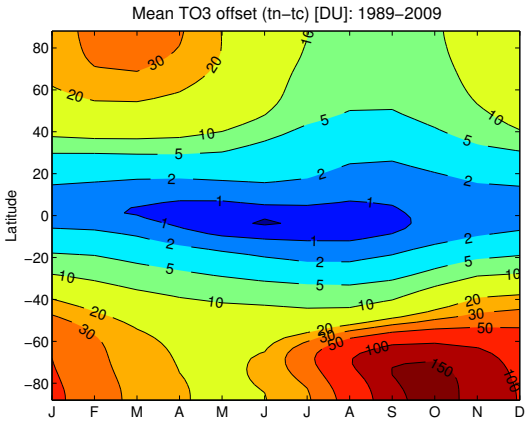

Figure 6.14.: *TO3 offsets due to polar ozone depletion and its export (differences of runs $tn - tc$), given in DU for all latitudes and months, averaged over the ERA-Interim driven period (1989-2009). From Kiesewetter et al. (2010b).*

Figure 6.13 illustrates the variability of springtime polar ozone in the CTM sensitivity runs. Here TO3 time series from the *tt* CTM run averaged poleward of 63° (corrected for TO3 below 330 K) are compared to satellite observations from the TOMS/SBUV and GSG merged ozone data sets. In addition, the *tT* run is shown, which has a polar ozone decay rate $\propto$ EESC^2. The observed polar ozone variability is captured very well by the CTM. Considerable differences in Antarctic TO3 evolution are visible between *tt* and *tT*. While *tt* slightly underestimates the magnitude of the observed trend, *tT* shows a stronger negative trend than the observations. Hence, the two runs *tt* and *tT* may be regarded as both extremes of Antarctic TO3 evolution and are both analysed in this thesis.

6.3.3. Polar ozone depletion

Polar heterogeneous chemistry leads to severe ozone depletion during late winter and spring, after which the air masses are exported to lower latitudes. The magnitude of polar ozone loss in the CTM can be diagnosed by comparing model runs with polar chemistry (*tt*, *tc*, *tT*) to the reference run *tn* without polar chemistry.

The year-to-year variability of polar ozone losses depends largely on the meteorological conditions during winter (Rex et al., 2004; Weber et al., 2003). A comparison of polar ozone losses in the CTM to observed results reported by World Meteorological Organization (2007) shows that the model generally captures the variability of NH losses well, but underestimates the magnitude. Values obtained in cold Arctic winters with strong ozone depletion such as 1995/96 or 1999/2000 are around 50-60 DU in the CTM, while the WMO assessment gives values of 80-100 DU in these years. The main reason for this behaviour lies in the construction of the simple polar chemistry scheme, which destroys ozone only while temperatures are cold enough for PSC formation. Thus effects of activated chlorine are missed which may occur after PSCs themselves are no longer formed, especially in a denitrified vortex (Waibel et al., 1999).

A climatology of the TO3 offsets induced globally by polar ozone depletion and dilution

of ozone-depleted air is shown in Fig. 6.14. A residual effect of polar chemistry persists at all latitudes and does not vanish throughout the year. The minimal values of these offsets amount to around 2-5 DU in the northern mid-latitudes (June – September), 1-2 DU in the tropics, and 8-15 DU in the southern mid-latitudes (March – August).

7. Variability of high-latitude ozone in the assimilated data set

In this chapter, the variability of high-latitude ozone is investigated in the assimilated 29-year data set. Although polar ozone depletion caused by heterogeneous activation of chlorine during winter and spring has been extensively investigated and is well understood (e.g., World Meteorological Organization, 2011, and references therein), studies about the variability of polar ozone during fall and winter are scarce due to the lack of consistent data sets. Some observation-based studies (Kawa et al., 2005; Sinnhuber et al., 2006) have shown that Arctic ozone exhibits unexpectedly high variability during fall and winter. While polar ozone is largely in photochemical equilibrium during summer, wintertime ozone is mainly dynamically controlled. As a result of geostrophically balanced thermal winds, the polar vortex forms in autumn, constituting a relatively secluded zone of low ozone mixing ratios. The winter stratosphere is subject to intense dynamical perturbations, and the polar column ozone change during winter has been shown to depend mainly on the deposition of energy by upward propagating planetary waves in the mid-latitudes (Weber et al., 2003). The advantage of the assimilated data set as compared to satellite observations is that it offers a daily coverage of the three-dimensional ozone field also during polar night. Hence, it is ideally suited to study the evolution of ozone in the polar winter stratosphere.

This chapter focuses on the question, how polar ozone anomalies develop during winter, how they evolve and how long they last, and how their evolution is connected to meteorological variability. The focus here is on Arctic ozone, although some references are also made to Antarctic ozone.

Throughout this chapter, zonal means over high geographical latitudes are used as well as zonal means over high equivalent latitudes, distinguishing high-latitude ozone and polar vortex ozone. Especially in the northern hemisphere, the polar vortex is usually not exactly pole-centred, and hence these coordinate systems differ. Equivalent latitude is defined as an isoline of potential vorticity enclosing the same area as the corresponding latitude coordinate would. Equivalent latitude values were calculated from the potential vorticity for every model grid point and ozone vmr then averaged in 5° bins. Averages over an equivalent latitude range denote area-weighted means of these bins. The three-dimensional nature of the assimilated data set allows to switch between these reference systems and to separate geographical effects due to displacement of the polar vortex, which are contained in the zonal picture, from processes within the polar vortex. Although a distinct polar vortex does not exist during summer months, equivalent latitude is a useful reference system for distinguishing polar air masses throughout the year.

Key findings of this chapter have been published in the peer reviewed literature as: Kiesewetter, G., Sinnhuber, B.-M., Vountas, M., Weber, M., and Burrows, J. P.: A Long-term Stratospheric Ozone Dataset from Assimilation of Satellite Observations: High-Latitude Ozone Anomalies, J. Geophys. Res., 115, D10307, 2010 (listed in the bibliography as Kiesewetter et al., 2010a). This mainly concerns Sections 7.1 and 7.2.2 – 7.2.3.

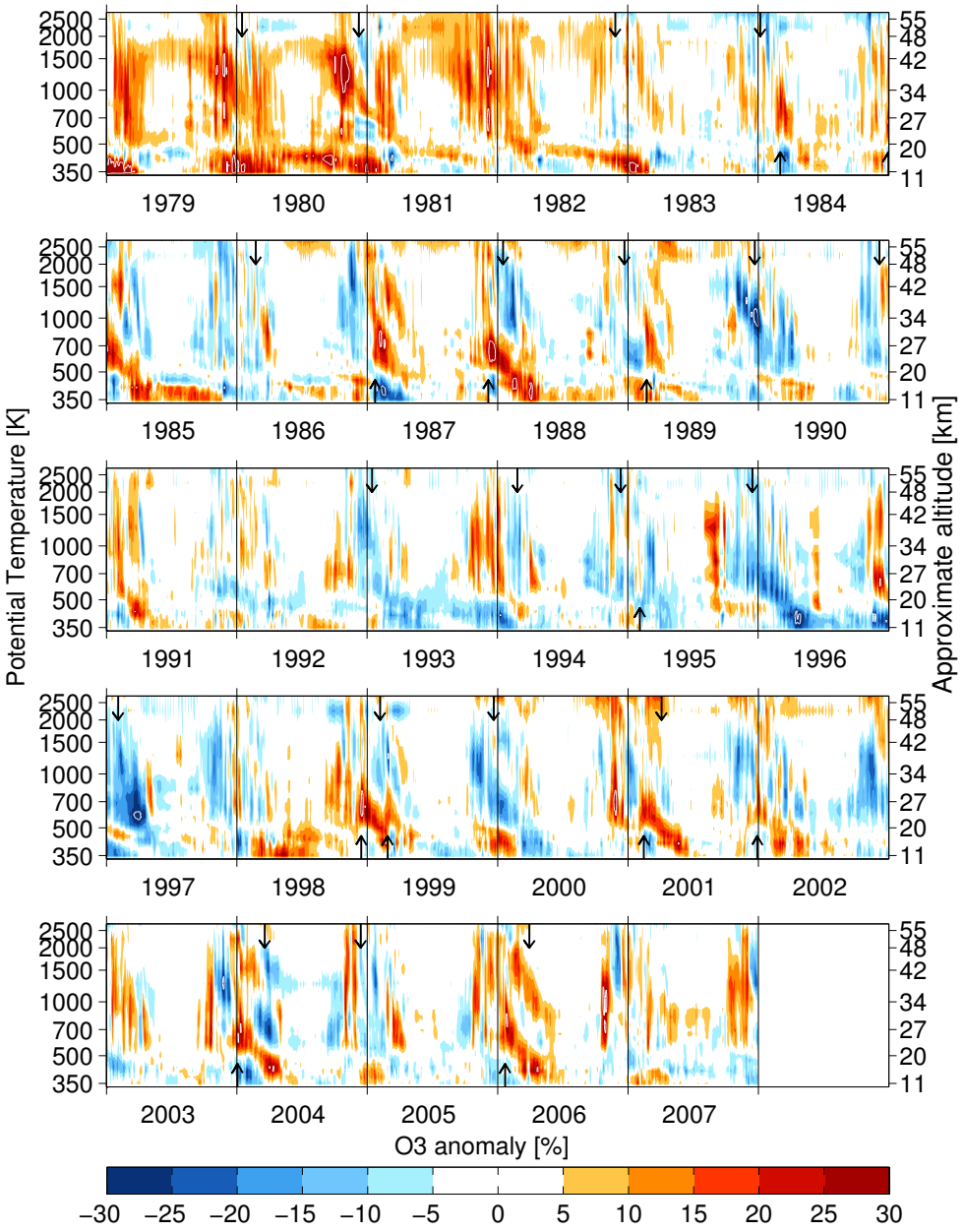

Figure 7.1.: *Ozone anomalies relative to the annual cycle, area weighted zonal average over 70–90° N. In addition to potential temperature (left axis), also the corresponding approximate geopotential height values are given (right axis). Values exceeding the colour scale are indicated by the white contour lines. A 3 day running mean has been applied for smoother viewing. Dates of weak and strong vortex events are indicated as arrows near the lower (weak vortex events) and the upper (strong vortex events) boundaries of the vertical range (for definition and details see Section 7.2). Plot updated from Kiesewetter et al. (2010a).*

7.1. Polar ozone anomalies in the assimilated data set

Owing to its daily global coverage of the stratosphere for 29 years, the assimilated data set is an ideal tool for studying in detail the development and propagation of polar ozone anomalies. *Ozone anomalies* are defined throughout this thesis as deviations of the ozone field from its long-term annual cycle. At many stages, *relative anomalies* are used, that is, ozone anomalies divided by the annual cycle value for the respective day. All analyses are conducted on a daily basis. Overlaps of ERA-40 and ERA-Interim driven periods of the assimilated data set are averaged before the analysis.

A first attempt to characterise Arctic ozone anomalies in the assimilated data set is taken in Figure 7.1, which shows relative polar ozone anomalies averaged zonally over $70 - 90°$N for the whole assimilation period. A three-day running mean has been applied in order to smooth the image. Alternating patterns of positive and negative anomalies (up to $\pm 40\%$, exceeding the colour scale) are visible, many of which seem to develop in the middle to upper stratosphere during winter months and then descend downwards to the lower stratosphere, where they remain for a long time. Prominent examples for this downward propagation of large positive anomalies are, for example, visible during the early 1980's (1979/80,1980/81,1981/82). These structures develop in October to December at potential temperatures of ~ 1000 K (~ 34 km) and more, and then descend to ~ 500 K (~ 20 km) during the winter months, from where they slowly descend further and remain visible for up to a year. Similar, if somewhat weaker, anomalies are observed in the winters of 1987/88, 2003/2004, and 2005/06. Conversely, also examples for negative anomalies with long lifetimes are easily found, most prominently during the 1990's. Note, e.g., the large negative anomalies in the winters 1995/96 and 1996/97. The negative anomalies developing end of 1995 or end of 1992 remain in the lower stratosphere almost unchanged for more than a year. Other examples for descending negative anomalies may be found in 1989/90, 1994/95, 2002/03, 2004/05, 2003/04, some of which also show long residence times in the lower stratosphere. In addition to the ozone anomalies, dates of weak and strong vortex events are shown as arrows in Figure 7.1. These indicate strong anomalies in the stratospheric circulation and are discussed in detail in Section 7.2.3. For the moment, the reader may notice that around the dates marked by the arrows, often similar ozone anomaly patterns are observed.

The descending behaviour of high-latitude ozone anomalies is expected from the stratospheric circulation (see Chapter 2). Under the influence of the Brewer-Dobson circulation, air masses are transported poleward and downward, and in the absence of light during winter the anomalies remain chemically undisturbed. However, the lifetime of anomalies is remarkable as they seem to appear also during months when the lifetime of ozone is short (less than a month above 700 K, less than a week above 850 K, less than a day above 1200 K during summer).

The picture for the Southern Hemisphere is very similar (Figure 7.2). Positive and negative anomalies typically develop at the beginning of winter in the middle to upper stratosphere and then descend to the lower stratosphere. In spring, the picture is disrupted by heterogeneous ozone depletion after 1982, which is by far more important in the SH. Throughout the first years of the assimilation, outstandingly large positive anomalies are visible in the lower stratosphere in spring, which are statistical artifacts due to the persistent formation of the Antarctic ozone hole in later years.

In recognition of the qualitative view of ozone anomalies obtained in Figures 7.1 and 7.2, their behaviour is now investigated in a systematic way. A suitable measure for the temporal

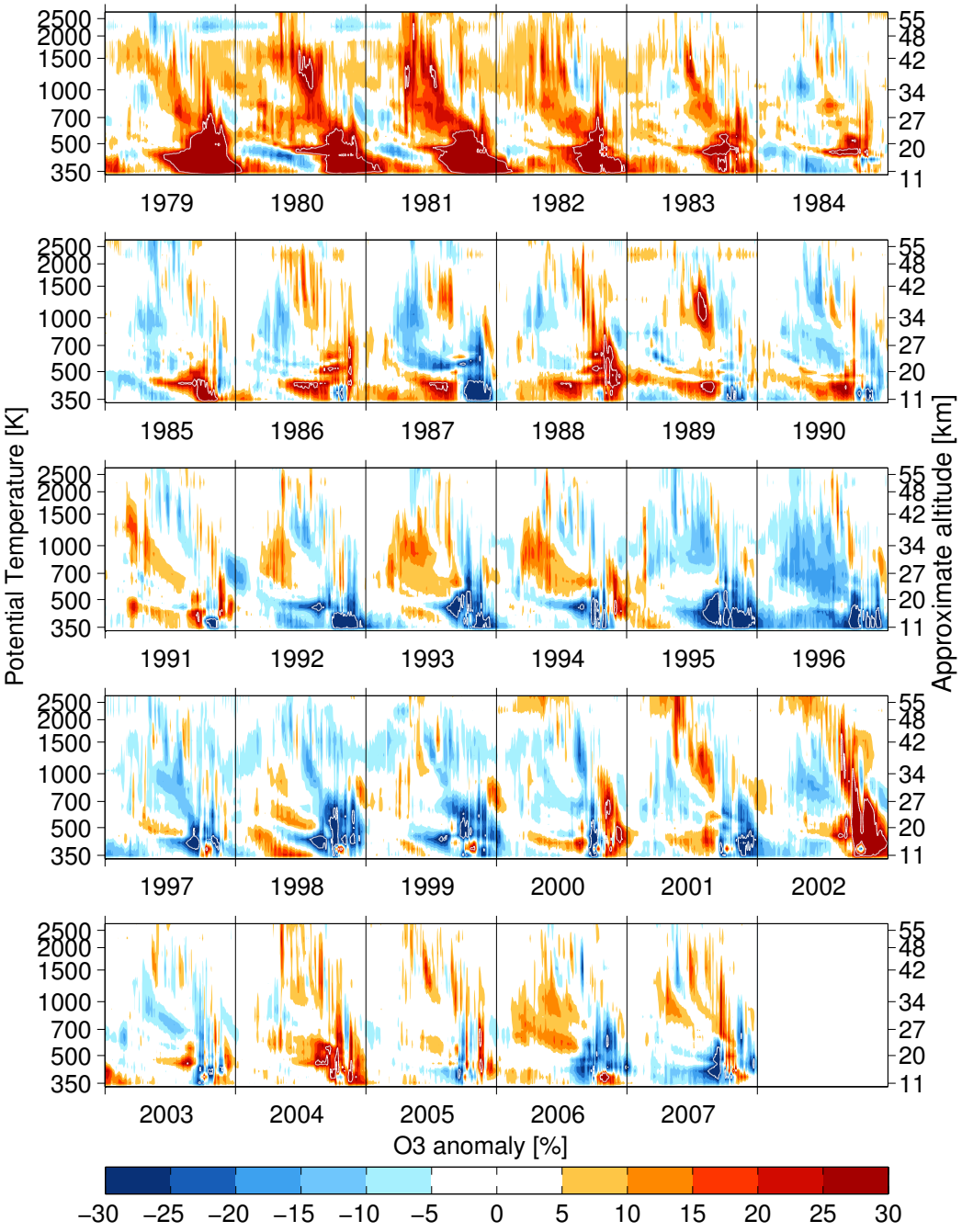

Figure 7.2.: *Ozone anomalies relative to the annual cycle, area weighted average over 70-90° S equivalent latitude. The whole available time series is shown, split up into six year stretches. In addition to the model-inherent potential temperature (left axis), also the corresponding approximate geopotential height values are given (right axis). Values exceeding the colour scale are indicated by the white contour lines. A 3 day running mean has been applied for smoother viewing. Plot updated from Kiesewetter et al. (2010a).*

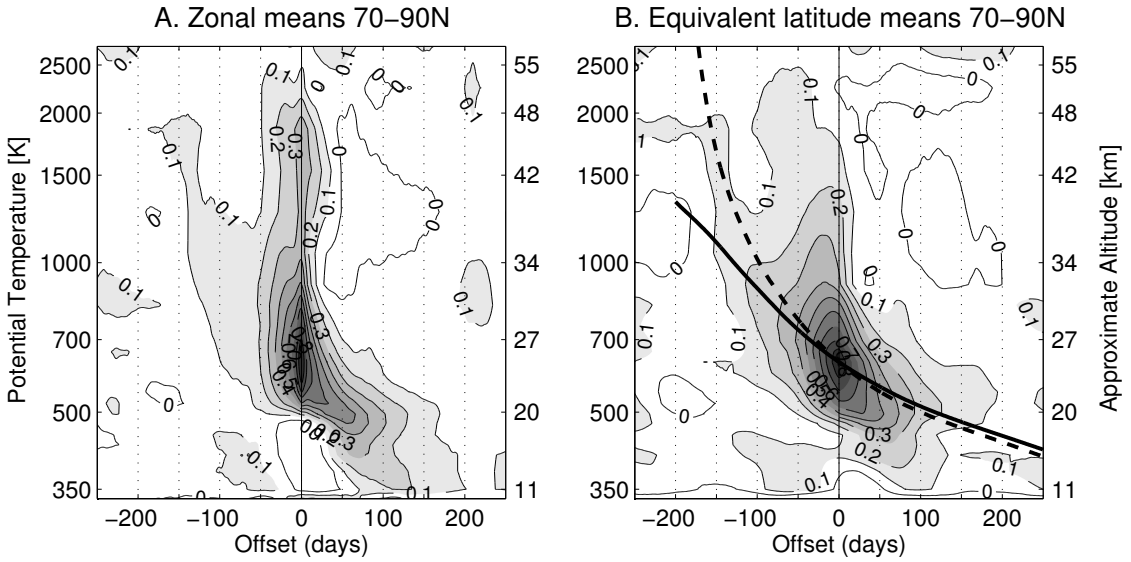

Figure 7.3.: *Autocorrelation of relative ozone anomalies, with respect to relative ozone anomalies at 631 K (∼25 km). Area weighted averages over 70–90° N latitude (A) as well as equivalent latitude (B) are shown. The solid black line in (B) displays the diabatic descent path expected from heating rates averaged year-round; the dashed line represents the same but following winter (DJF) heating rates.*

and vertical evolution of ozone anomalies is their autocorrelation. As the structures seen in Figure 7.1 often originate in the upper stratosphere and then propagate to the lower stratosphere, a mid-stratospheric model level (631 K, ∼ 25 km) is selected as point of origin. Varying this potential temperature level between 600 and 700 K does not change results significantly. Figure 7.3 shows the correlation coefficient for 250 days backward and forward in time, for ozone anomalies averaged polewards of 70°N latitude (A), as well as polewards of 70°N equivalent latitude. Although the autocorrelation does show a consistent pattern extending roughly 150 days into the past and into the upper stratosphere, most of the variability in the zonal frame is short-lived. This short-lived component of variability is partly due to geographical effects – in particular, displacement of the polar vortex – and is significantly reduced in the vortex-centred equivalent latitude coordinate system. In order to provide a feeling for the approximate path an air parcel follows during winter, the descent curve obtained by integrating the heating rates forward and backward in time from 631 K is shown (dashed line). In the equivalent latitude domain, the autocorrelation shows a two-fold pattern, with one branch following closely the descent curve given by the wintertime (December to February) heating rates. For comparison, the descent curve following the year-round average heating rates is shown (solid line), which does not match the evolution of the autocorrelation pattern well. Another branch of the autocorrelation descends much faster than allowed by the heating rates, covering the range from 1500 K to 630 K within ∼ 30 days. This fast-descending branch may be interpreted as a signature of large anomalies in stratospheric dynamics, that will be discussed further in Section 7.2.

Having studied the general statistical properties of ozone anomalies in the polar middle stratosphere, we may ask the question how the “average anomaly” evolves and propagates.

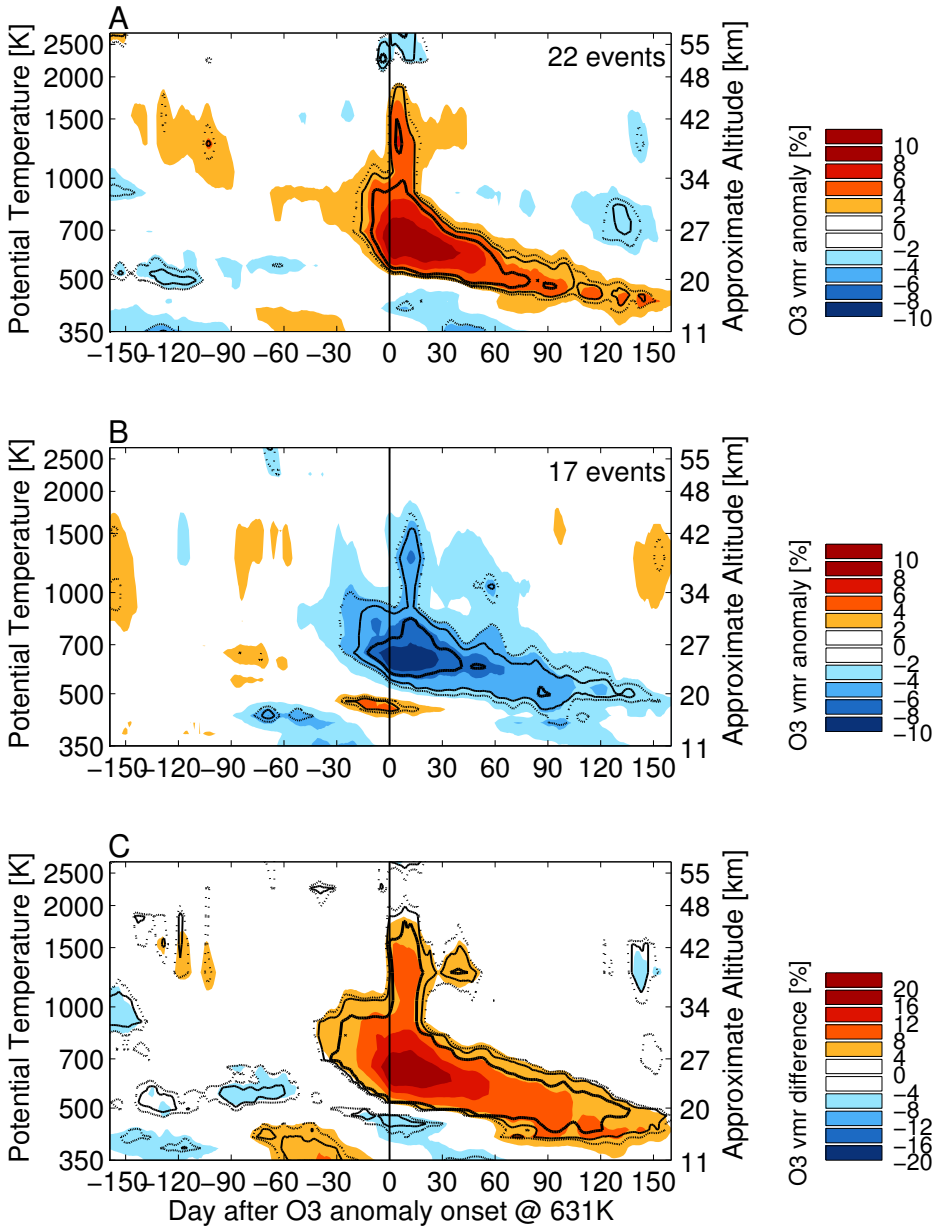

Figure 7.4.: *Composite plots of positive (A) and negative (B) ozone anomalies, and their difference (C). Day 0 indicates the day when the ozone anomaly at 631 K exceeds ±10% for the first time in 60 days; the evolution of ozone anomalies is shown for 150 days before and after day 0. All values are area-weighted averages north of 65° N equivalent latitude. (A),(B): Composites of ozone anomalies. Colours indicate anomalies in terms of fractions of annual cycle values, black contour lines indicate the significance of the anomalies: dotted 90%, thin black 95%, solid black 99%. (C): Difference of (A) and (B), divided by the mean ozone at the respective day. Black contour lines indicate the significance of the difference: dotted 90%, thin black 95%, solid black 99%. Plot modified from Kiesewetter et al. (2010a).*

For this purpose, I now analyse composites of high and low ozone anomalies. These are identified here by the date when the ozone anomaly at 631 K exceeds 10% of the annual cycle value. As before, the level is selected as a transition between upper and lower stratosphere, while the threshold values are selected so that the number of events provides for a reasonable statistics while ensuring that effects do not get blurred due to a weak threshold. In order to ensure that anomalies are not counted more than once, it is required that the ozone anomaly stays below the threshold value for 60 days before the onset date. With these parameters, 22 positive and 17 negative ozone anomaly events are detected. Figure 7.4 shows the composite plots of positive (A) and negative (B) ozone anomalies, and their difference (C), as equivalent latitude means northwards of 65°N. 65° is chosen as southern boundary since this area contains the polar vortex; moving the boundary to 70° or 75° does not change the picture significantly. The time axis is drawn with respect to the onset day of the ozone anomaly at 631 K and extends from 160 days before to 160 days after the event. In addition to the anomalies as fractions of annual cycle values (colours), the levels of significance are shown as black contour lines. These values are obtained from a Student's t-test, in a very similar fashion as by Kodera (2006). In (A) and (B), the levels of significance indicate the significance of a departure from the multi-year annual cycle values at the respective days, while in (C) the contours indicate the statistical significance of the difference between ozone mixing ratios under positive and negative anomaly conditions, as obtained from a Student's t-test for significantly different means.

Both positive anomalies (A) and negative anomalies (B) seem to originate at ~ 1000 K about 1–2 months before day 0, and then descend slowly to the lower stratosphere. The anomaly stays significant for around 5 months, resulting in an average overall lifetime of up to seven months from appearance at 1000 K to decay at 450 K. This is confirmed by the difference plot (C), which shows a compact and highly significant anomaly signature propagating from 1000 to 450 K in seven months. The positive anomaly in (A) seems to have some origins in the upper stratosphere going back as far as 5 months before day 0 (the patches of positive ozone anomalies preceding day 0 form a contiguous, although not significant, structure if the threshold value for the ozone anomaly at 631 K is increased to 15%).

Although the focus of this study is mainly on Arctic ozone, the question how the situation looks like in the Antarctic is legitimate. Figure 7.5 shows an overlay plot of ozone anomalies poleward of 65°S equivalent latitude, distinguished by the day when the ozone anomaly at 700 K first exceeds 10% of the annual cycle. Here the decisive potential temperature level is chosen somewhat higher in order to rule out effects caused by ozone hole chemistry, which dominates ozone variability in Antarctic lower stratospheric spring on scales that exceed those of mid-stratospheric ozone variability by far. The picture is similar to the Arctic; positive and negative ozone anomalies show a lifetime of 5–6 months. In the Antarctic, the anomalies indeed reach the bottom boundary of the CTM. Due to the greater stability of the southern polar vortex as compared to the Arctic polar vortex, the long lifetime of ozone anomalies is more expected here. Nonetheless it is remarkable as ozone lifetime is short in the middle atmosphere during most of the year. Since 29 positive and 29 negative events are detected in the 29 year data set, each anomaly surviving for almost half a year, the Antarctic middle stratosphere is indeed characterised by alternating patterns of positive and negative ozone anomalies continuously propagating from the middle to the lower stratosphere.

The long persistence of ozone anomalies is remarkable in two aspects. Firstly, it shows that ozone anomalies are dynamically stable in polar areas of both hemispheres, unless a large dynamical disruption of the whole polar vortex sets in (for more details on this latter issue,

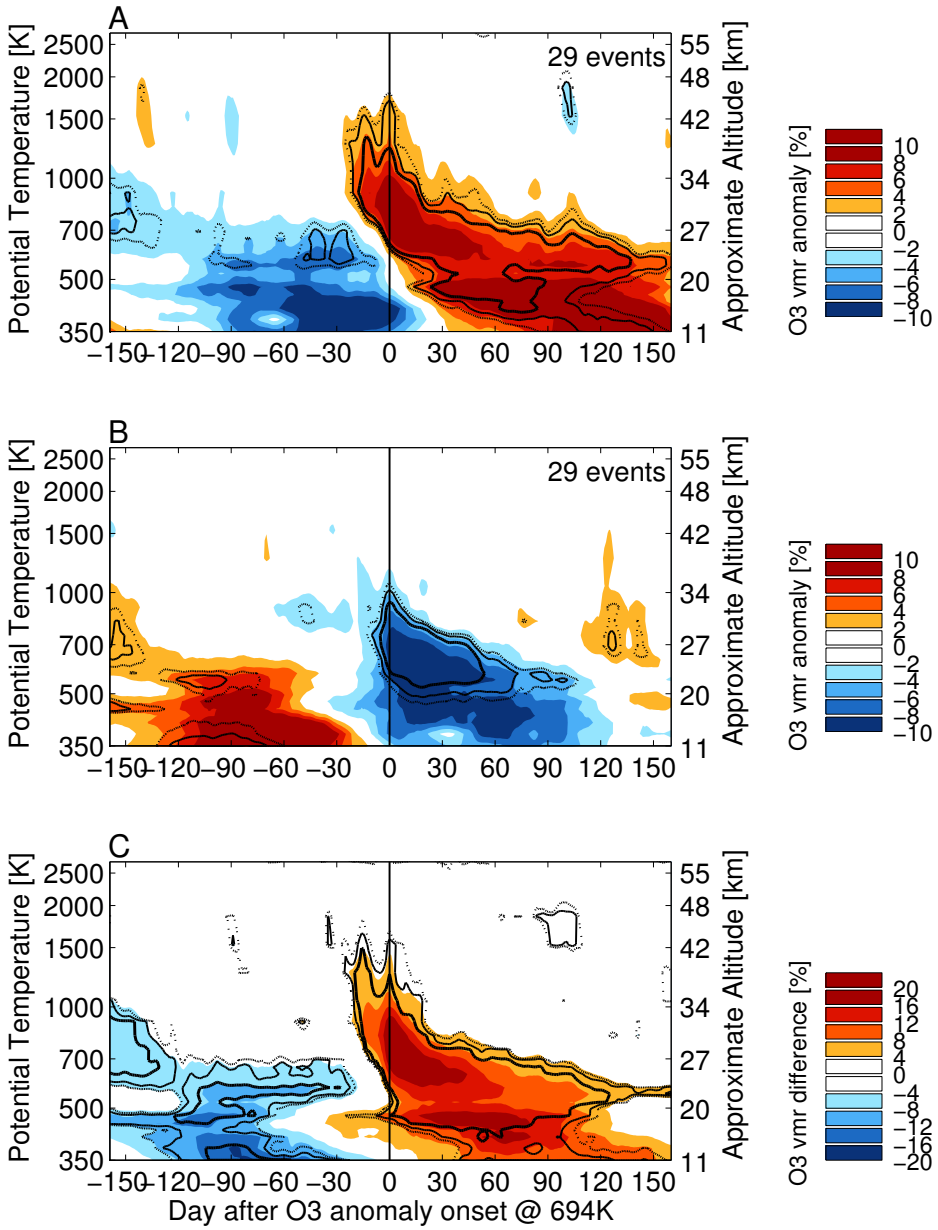

Figure 7.5.: *As Figure 7.4, but for the southern hemisphere. Composite plots of positive (A) and negative (B) ozone anomalies, and their difference (C). Day 0 indicates the day when the ozone anomaly at 700 K exceeds ±10% of the annual cycle mean for the first time in 60 days; the evolution of ozone anomalies is shown for 150 days before and after day 0. All values are area-weighted averages south of 65° S equivalent latitude. (A),(B): Composites of ozone anomalies. Colours indicate anomalies in terms of fractions of annual cycle values, black contour lines indicate the significance of the anomalies: dotted 90%, thin black 95%, solid black 99%. (C): Difference of (A) and (B), divided by the mean ozone at the respective day. Black contour lines indicate the significance of the difference: dotted 90%, thin black 95%, solid black 99%.*

see Section 7.2.2 below). Secondly, it raises the question how anomalies evolving in $\sim 1000\,\mathrm{K}$ may persist on timescales of several months when the photochemical lifetime of ozone at these altitudes is below 10 days for most of the year. Although large anomalies occur more frequently during winter months (when the photochemical lifetime of ozone is long), the onset dates are not restricted to winter and also include events during summer (8 out of 22 positive and 4 out of 17 negative events in the Arctic occur from April to September). In a recent paper, Tegtmeier et al. (2008) have reported an unexpectedly long persistence of ozone anomalies in the middle stratosphere at mid-latitudes, and hypothesised that this persistence is connected to transport-induced anomalies in odd nitrogen (NO_y). Since NO_y, which has a lifetime exceeding 1 year at $30 - 40\,\mathrm{km}$, acts as an ozone destructing agent through the NO_x cycle, NO_y anomalies perturb the chemical ozone balance on timescales far beyond the photochemical lifetime of ozone itself. Transport-induced NO_y anomalies may play a role for the long persistence of polar ozone anomalies observed here, but this assumption is speculative at the moment and requires further testing.

Around day 0 in Figure 7.4, a vertical "branch" of the anomaly is visible both in the positive as in the negative case, reaching into the upper stratosphere. This is at least partly a contribution from strong anomalies in the stratospheric circulation, which are discussed further in Section 7.2.

Polar ozone anomalies can basically arise from two different processes, anomalies in dynamics (stronger or weaker transport of ozone into the vortex) and anomalies in chemistry. Anomalous chemical production or loss of ozone may be induced by different factors or a combination of them: Anomalies in concentrations of important reaction partners, e.g. NO_x, or shifts of reaction equilibria due to variations in temperature or incident UV radiation. As anomalies in temperature are connected to dynamical anomalies, the possibilities for interactions between chemistry and dynamics are numerous. In fact, feedbacks also exist between ozone anomalies at different altitudes: Lower ozone concentrations in the upper stratosphere will allow more UV radiation to penetrate the lower stratosphere and foster enhanced ozone production at lower altitudes (hence the direct dependence of ozone production on the overhead column in the Linoz scheme). While all these mechanisms are important, the approach chosen in this study allows only a limited attribution of anomalies to their underlying processes, due to the limitations of the simple ozone chemistry; e.g., effects of varying concentrations of reaction partners can only be introduced indirectly through the assimilation. Two special cases of processes leading to ozone anomalies in the assimilated data set are analysed in the remaining part of this chapter: anomalies in the stratospheric circulation (Section 7.2) and solar particle precipitation events (Section 7.3).

7.2. Influence of the Northern Hemisphere Annular Mode

It is well known that the dynamical variability of the Arctic stratosphere plays a large role for the variability of the Arctic ozone layer. In this section, the impact of anomalies in the stratospheric circulation on the formation and persistence of polar ozone anomalies is investigated using the Northern Hemisphere Annular Mode (NAM) as a dynamical proxy. The NAM is introduced in Section 7.2.1, and its relation to ozone is discussed in Section 7.2.2.

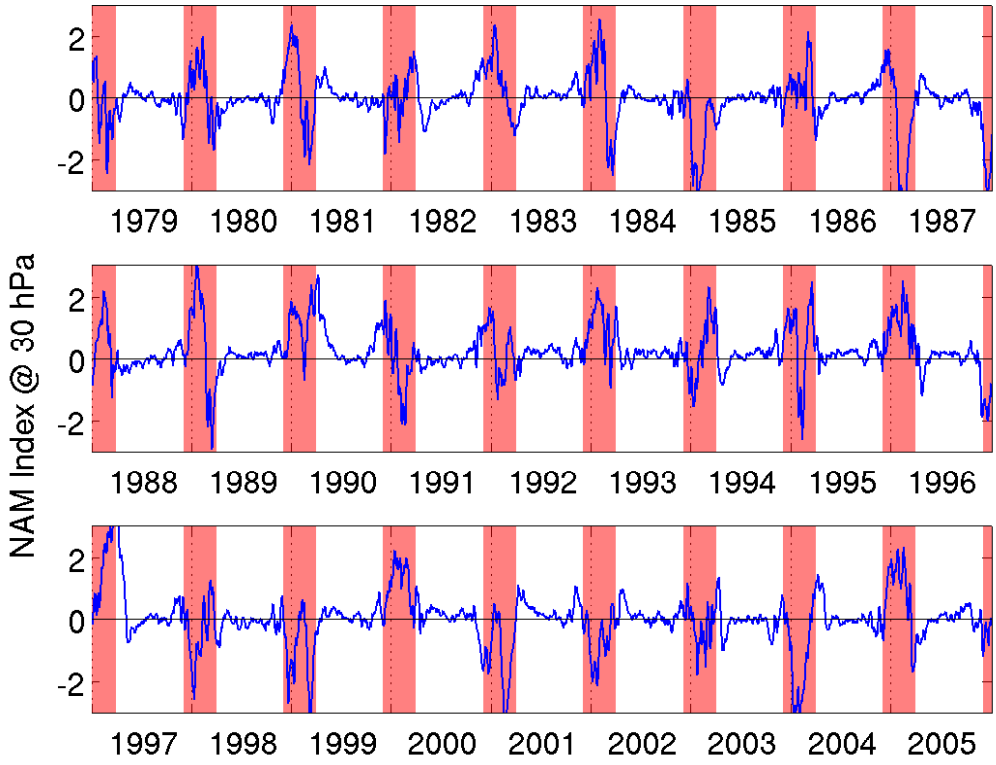

Figure 7.6.: *The NAM index at 30 hPa (∼ 25 km). The largest variability of the NAM index is observed during the winter months (December to March), which are highlighted in red.*

7.2.1. The Northern Hemisphere Annular Mode

The Northern Hemisphere Annular Mode (NAM) (Thompson and Wallace, 2000; Baldwin and Dunkerton, 2001) constitutes the dominating pattern of climate variability in the Northern Hemisphere, on a seasonal as well as on a daily base (Thompson and Wallace, 2001). The NAM represents the pressure difference between polar regions and lower latitudes and is closely related to other climate patterns such as the North Atlantic Oscillation (NAO) and the Arctic Oscillation (AO, Thompson and Wallace, 2000).
Annular modes have been introduced in a number of slightly differing ways, usually as the leading empirical orthogonal function of either sea-level pressure (Thompson and Wallace, 2000; Hurrel et al., 2003) or extra-tropical zonal-mean zonal wind (Kidson, 1988; Hartmann and Lo, 1998; Lorenz and Hartmann, 2001, 2003). Contrary to these studies, which use only one index independent of a vertical coordinate, Baldwin and Dunkerton (2001) introduced a definition of the NAM depending on pressure height, which is applied in this study. On each pressure level, the NAM is then defined as the first empirical orthogonal function (EOF) of 90-days low-pass filtered geopotential anomalies north of 20°N. The NAM data set used in this study is an update from the data set introduced and analysed by Baldwin

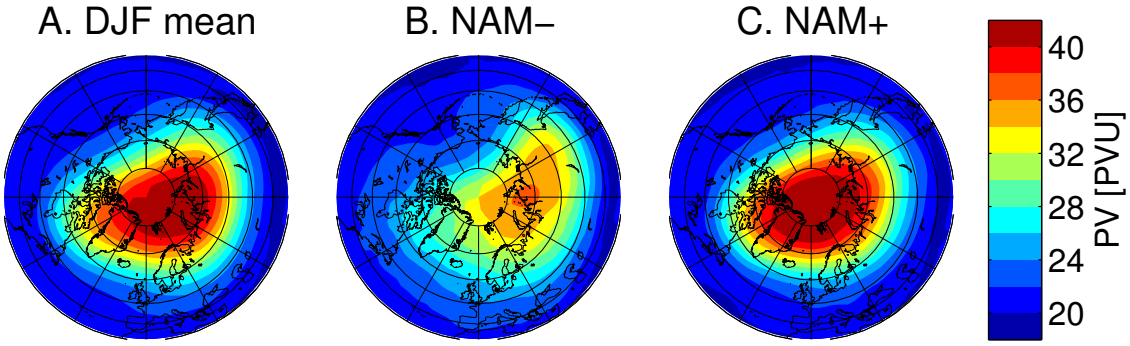

Figure 7.7.: *Potential vorticity (PV) at 475 K potential temperature. A. winter average (DJF), B. negative NAM phases (weak vortex events), C. positive NAM phases (strong vortex events). Units:* $1\,\mathrm{PVU} = 10^{-6}\mathrm{K\,m^2/(kg\,s)}$.

and Dunkerton (2001) and was obtained from M. Baldwin (personal communication). It is provided on 23 pressure heights from 1000 to 0.3 hPa on a daily basis for the whole assimilation period (1979–2007). Figure 7.6 shows the NAM index at the 30 hPa level for the 1979–2005 period. As an indicator of the dynamical state of the stratosphere, the NAM index shows its largest excursions during winter months (December to March, indicated by red shading), when the stratosphere is dynamically active. A positive index corresponds to a strong meridional pressure gradient, strong westerly winds and a strong, more isolated polar vortex, while a negative index indicates a weak or even reversed meridional pressure gradient, weak westerlies or easterly winds, and hence a weaker vortex.
The effects of large NAM excursions on the stratospheric circulation are displayed in Figure 7.7 in terms of potential vorticity (PV) at 475 K ($\sim$ 18 km), an altitude which is often used to characterise the extent of the polar vortex. Figure 7.7 (A) shows the average PV configuration during winter (December – February). The centre of the polar vortex is slightly displaced from the north pole towards Europe. In Figure 7.7 (B), the average is taken over low-NAM phases only, while in Figure 7.7 (C) the average is taken over episodes of high NAM index [1]. For negative NAM excursions (B), the vortex is significantly weakened (maximum PV is lower) and displaced towards Russia. High NAM excursions (C), on the contrary, lead to a strengthened (maximum PV is slightly increased) and pole-centred vortex.

7.2.2. NAM influence on polar ozone

Annular modes have been shown to have a profound influence on total column ozone variability (Orsolini and Doblas-Reyes, 2003; Jiang et al., 2008a,b). A straightforward analysis of the assimilated data set shows that also the variability of polar profile ozone is directly connected to the NAM phase. This is demonstrated qualitatively in Figure 7.8, which shows the average ozone configuration in the polar middle stratosphere (631 K / 25 km) in the same way as Figure 7.7 – the average winter (December to February) ozone field (A), and

[1] More exactly, the episodes were selected as 1-month means after weak and strong vortex events. These events are defined by NAM < -3.0 and NAM > 1.5 for the first time in 60 days for weak and strong vortex events, respectively. Vortex events are discussed further in Section 7.2.2.

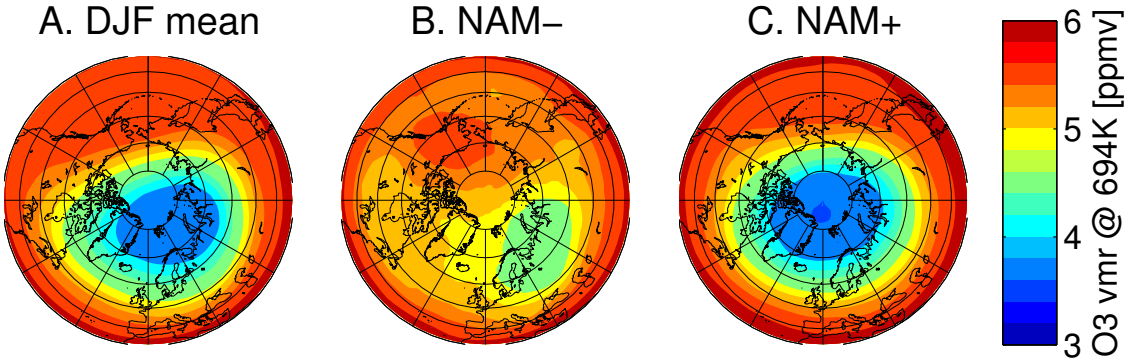

Figure 7.8.: *Assimilated ozone at 694 K potential temperature (27 km). A. winter average (DJF), B. negative NAM phases (weak vortex events), C. positive NAM phases (strong vortex events).*

the ozone field during episodes of strongly negative (B) and positive NAM phase. The ozone field follows inversely the PV configuration shown in Figure 7.7, regions of high PV (the polar vortex) correspond to low ozone mixing ratios and vice versa. Strong effects of the NAM on the ozone field are visible: in the negative NAM (weak vortex) case, high-latitude ozone mixing ratios are significantly enhanced, while in the positive NAM (strong vortex) period ozone mixing ratios are lower in the more pole-centred vortex.

The qualitative picture in Figure 7.8 suggests a strong influence of the NAM phase on polar ozone, that is now quantified further by directly comparing the NAM index and ozone anomalies at a given altitude. Figure 7.9 shows ozone anomalies in the mid-stratosphere (600 K, $\sim$ 24 km), here for the years 2000-2001. For comparison, zonal means (solid black line) and equivalent latitude means (grey line) north of 75°N are shown, both referring to the left axis. The NAM index at 30 hPa, corresponding to roughly the same altitude, is superimposed (red line, right axis). It is obvious that zonal anomalies follow the inverted NAM index closely and instantaneously. This is understandable, as a weakening of the vortex during low NAM phases usually coincides with a geographical displacement of the vortex (see Figures 7.7 and 7.8), leading to higher than average ozone mixing ratios at high latitudes, while a strengthening of the vortex also induces a stricter confinement to high latitudes, thus lowering average mixing ratios there. However, even though the correlation coefficient itself shows a weaker anticorrelation, also equivalent latitude ozone anomalies appear well related to the NAM phase, especially to strong NAM excursions, albeit with a time lag. This time-lag is quantified more precisely in Figure 7.10, where the correlation between the whole ozone anomaly time series to the NAM index is plotted as a function of offset time between the curves (same model and NAM levels as before). While the anticorrelation to zonal anomalies shows a distinctive peak at day zero (no offset), the peak of the equivalent latitude anticorrelation is broader and shifted by about ten days (indicated by an arrow). Dynamical variations of the stratosphere as expressed by the NAM phase affect zonal high-latitude ozone anomalies immediately and strongly (for the 29 year time series, the correlation coefficient reaches values < -0.7), which is in part due to geographical effects. If such geographical effects are eliminated by the transition to equivalent latitude coordinates, the anticorrelation is weaker but still present (maximum anticorrelation -0.4), and the effects are shifted in

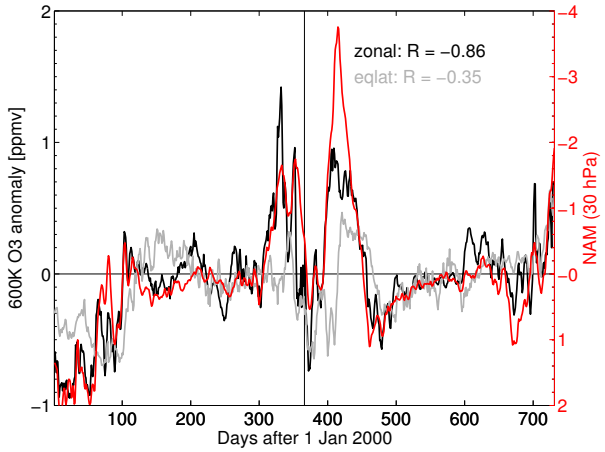

Figure 7.9.: *Polar ozone anomalies in the mid-stratosphere (600 K, ∼ 24 km) versus NAM phase (at 30 hPa, ∼ 24 km), for the years 2000 and 2001. Black solid line, left axis: Zonal ozone anomalies averaged over 75 − 90° N. Gray line, left axis: Equivalent latitude ozone anomalies averaged over 75 − 90° N. Red line, right axis: NAM index. Note the reversed NAM axis. Plot updated from Kiesewetter et al. (2010a).*

time. This result may be explained as a distinct sign of the mixing in of ozone rich air during a disturbed vortex (negative NAM), and conversely the suppression of mixing during a stronger, more secluded vortex (positive NAM) when transport of ozone rich air into the vortex is largely inhibited.

7.2.3. Vortex events and their influence on polar ozone

Figure 7.9 suggests that especially large NAM anomalies play an important role for polar ozone variability. In order to study this aspect further, I investigate the evolution of ozone anomalies especially in connection with such large NAM anomalies, or “strong” and “weak” vortex events. In order to distinguish such events, the definition of Baldwin and Dunkerton (2001) is followed here, who classified “strong” and “weak” vortex events by the date when the NAM index at 10 hPa first exceeds threshold values of +1.5 in positive and -3.0 in negative direction, after 60 days of not exceeding these limits. The event indicates the onset of an anomaly in stratospheric circulation.
With the above definition applied, 20 strong and 12 weak vortex events are found in the time span covered by the assimilation. Dates of weak and strong vortex events are indicated as arrows in Figure 7.1. The temporal distributions of weak and strong vortex events are shown in Figure 7.11. NAM events occur from December to February; in order to make the ozone changes due to NAM events comparable in spite of the different background values of ozone at the respective onset times, the anomalies are normalised by the annual cycle. The evolution of the NAM index in the vicinity of such vortex events is shown in Figure 7.12.
By comparing episodes around weak and strong vortex events in Figure 7.1, the reader may find striking similarities in the evolution of ozone anomalies around these dates. In Figure 7.13, the average evolution of vortex ozone anomalies is shown as composites of anomalies during weak and strong vortex events, and their difference, in the same way as

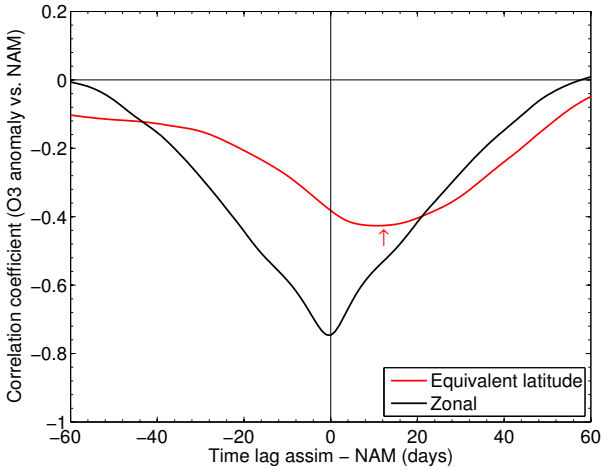

Figure 7.10.: *Correlation between polar ozone anomalies at 600 K (north of 75° N) and 30 hPa NAM index for the whole time series 1979-2007, in dependence of the temporal shift between the time series. Black line: zonal means, red line: equivalent latitude. The minimum of the correlation to equivalent latitude ozone anomalies is shifted by 11 days (indicated by the arrow). Plot updated from Kiesewetter et al. (2010a).*

in Figure 7.4. In panels (A) and (B), colours represent relative ozone anomalies averaged poleward of 65° equivalent latitude during weak and strong vortex events, respectively; black contour lines indicate the significance of deviations from the mean annual cycle. For better viewing, a five day running mean has been applied to smooth the figure. As already mentioned, the 10 hPa surface is chosen to distinguish NAM events, which corresponds to a mid-stratospheric potential temperature of around 840 K.

Before the weak vortex event (panel A), only little changes in ozone are visible, which are hardly significant. Concurrent with the onset of the vortex event, a downward propagating positive ozone anomaly appears in the upper stratosphere, which intensifies to $\sim 10\%$ almost instantaneously after the threshold day and rapidly descends to the middle stratosphere. There, the ozone anomaly intensifies to $\sim 15\%$ around 600 K. Since the descent of the upper stratospheric anomaly is by far too fast to be attributable to actual physical transport of air, it is rather the dynamical information which propagates downward and induces the ozone anomaly. After a vortex event, the NAM index relaxes on a timescale of ~ 60 days (Baldwin and Dunkerton, 2001, see also Figure 7.12). Nonetheless the ozone anomaly in the middle stratosphere continues to exist long after the NAM index has returned to standard values, slowly descending to the lower stratosphere. The ozone anomaly is well visible as $\sim 5\%$ enhancement up to 120-150 days after the event, when it reaches the lower boundary of the CTM.

In the case of strong vortex events (panel B), the observed anomalies are smaller, but the general picture is strikingly similar with inverse signs. A negative anomaly arises first in the middle to upper stratosphere, but the propagation of the anomaly to the lower stratosphere is less compact than in the weak vortex case. Inverse to the weak vortex composite picture, a negative anomaly develops in the middle stratosphere at 500-700 K, accompanied by a positive anomaly below 500 K. The negative anomaly reaches values of -5% and slowly

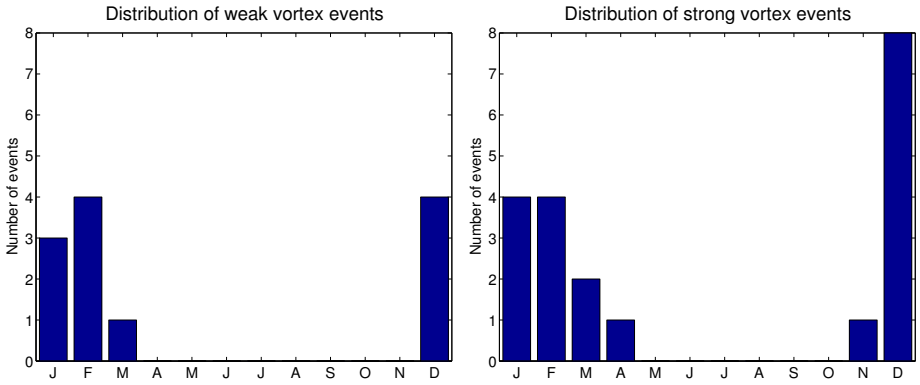

Figure 7.11.: *Monthly distribution of weak (left) and strong (right) vortex events 1979-2007.*

propagates downward on the same timescale as the positive anomaly in the weak vortex case. A strong negative anomaly appears in the upper stratosphere delayed by almost a month, intensifying to about -10% and slowly descending to 1200 K, where it vanishes after 50 days. Also in this case, the lower stratospheric ozone anomaly remains visible for months after the NAM phase has relaxed to "normal" values, and reaches the lower boundary of the CTM 120-160 days after the NAM event onset.

Panel (C) shows the difference of A and B, demonstrating the high significance of ozone changes induced by strong versus weak vortex events. As in Figure 7.4, the difference is normalised by the mean ozone vmr during strong and weak vortex events. The significance of the ozone difference between strong-vortex and weak-vortex phases was established by a Student's t-test for significantly different means. Since the shape of positive and negative ozone anomalies during weak and strong vortex events is similar, but with inverted sign, also their difference closely resembles the weak vortex (or inverted strong vortex) pattern, with differences reaching 20% of mean ozone in the mid-stratosphere. The significance of the anomaly exceeds 95% from the onset of the NAM event to the passage of the anomaly through the lower model boundary five months later. In the middle stratosphere, significance levels exceed 99% for more than three months.

The behaviour of such purely dynamics-induced anomalies is distinctly different from that of "average" anomalies shown in Figure 7.4, as the former transit most of the stratosphere almost instantaneously and then slowly descend to the lower boundary of the CTM in five months. However, it should be emphasised that 7 out of 22 positive ozone anomalies shown in Figure 7.4 occur within 30 days of weak vortex events, and 7 out of 17 negative events occur within 30 days of strong vortex events.

A closer look at the mechanisms that cause the distinct shape of ozone anomalies occurring in connection with vortex events is taken in Section 7.2.4. For the moment, we can ask the legitimate question for independent confirmation of the shape of ozone anomalies, which is not straightforward to obtain. The significant advantage of the assimilated data set for the investigation of polar ozone anomalies and their connection to NAM events is the daily global coverage for almost 30 years, which allows us to study in detail processes taking place during the dynamically active winter time. The only long time observational record of polar ozone during winter comes from sonde measurements, which provide data at single

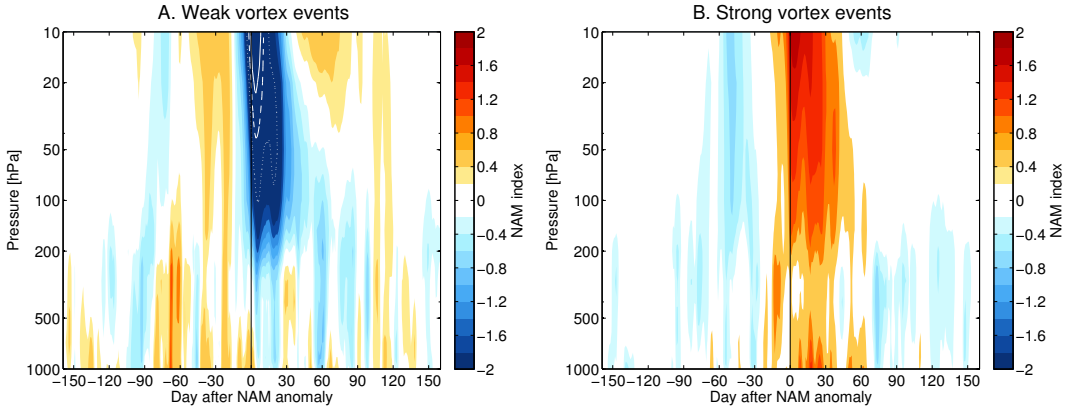

Figure 7.12.: *NAM index in the vicinity of weak (left) and strong (right) vortex events. The colour scales are equal; in order to depict the strong negative values around weak vortex events, additional contour lines have been drawn corresponding to –2 (dotted), –2.5 (dashed), and –3 (solid). The large circulation anomaly relaxes to usual values after 1–2 months. Anomalies propagate down to the surface in weakened form.*

high-latitude sites up to around 30 km altitude ($\sim$ 900 K), typically on a weekly base for the period 1990–present. Thus the possibilities to obtain a confirmation of the ozone anomaly picture shown in Figure 7.13 are limited; nonetheless the attempt is made here anyway. Figure 7.14 shows ozone anomalies in the NyÅlesund sonde record around weak (left) and strong (right) vortex events, in the same way as in Figure 7.13. In order to account for the limited temporal coverage, the annual cycle is computed in five-day intervals (as for the validation in Chapter 6), and ozone anomalies are averaged on a weekly base around the NAM events. The resulting picture is – understandably – more blurred than Figure 7.13, but essentially very similar. In the weak vortex event case, the resulting ozone anomalies are stronger than those in the assimilated data set, which may also be related to the shorter record. Since ozone anomalies are generally weaker around strong vortex events, the onset criterion had to be raised to NAM > 2.0 in order to obtain a clear signal. Especially around vortex breakup at the end of the winter, Ny-Ålesund is not necessarily located within the vortex (or its remains), while anomalies in Figure 7.13 represent vortex averages. Episodes of extra-vortex air parcels above the station lead to discontinuities in the ozone anomaly such as observed after day 80 in both cases.

The picture of Arctic ozone anomalies in the vicinity of NAM events is quite consistent. Two obvious questions arise: Firstly, what are the mechanisms causing the ozone anomalies, and secondly, what about the other hemisphere – are ozone anomalies of similar magnitude and shape induced by the Southern Annular Mode (SAM), counterpart of the NAM? For the first question, the reader is deferred to the following section. Regarding the influence of the SAM on Antarctic ozone, we have to take into account several factors that distinguish the Antarctic from the Arctic, as mentioned in Chapter 2. Due to its geographical situation, the Antarctic polar vortex is less disturbed by planetary waves, more stable and less susceptible to dynamic anomalies. Until today, only a single stratospheric sudden warming has been recorded in the Antarctic in 2002 (Sinnhuber et al., 2003). Similarly, large anomalies in the SAM index are scarce. Using the same onset criteria as for NAM events, only six weak

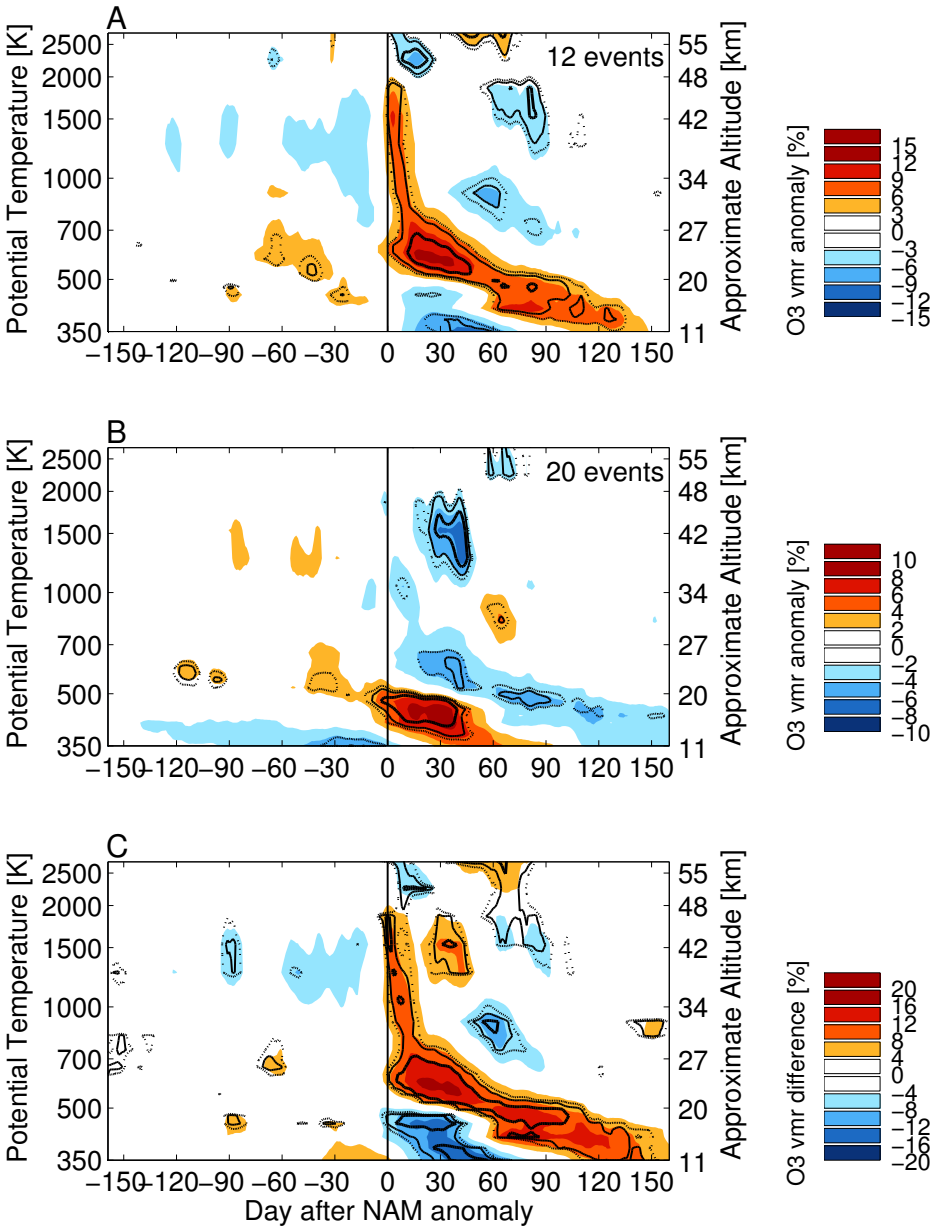

Figure 7.13.: *Ozone anomalies after weak (A) and strong (B) vortex events, and their difference (C). Day 0 is the onset of the vortex event, i.e. the day when the NAM index at 10 hPa first crosses the threshold value (+1.5/-3.0) in 60 days. All values are equivalent latitude averages north of* 65° *N. (A),(B): Composites of ozone anomalies. Colours indicate anomalies in terms of fractions of annual cycle values, black contour lines indicate the significance of the anomalies: dotted 90%, thin black 95%, solid black 99%. (C): Difference of (A) and (B), divided by the mean ozone at the respective day. Black contour lines indicate the significance of the difference: dotted 90%, thin black 95%, solid black 99%. Plot updated from Kiesewetter et al. (2010a).*

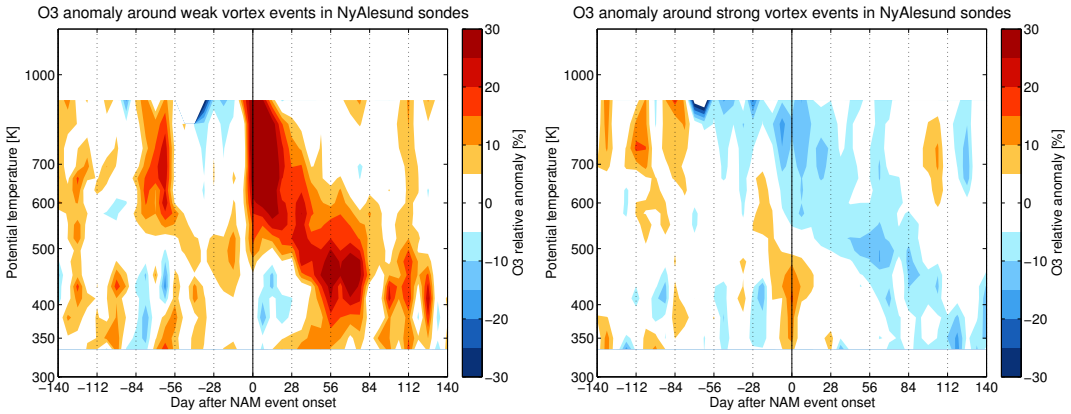

Figure 7.14.: *Composite plots of ozone anomalies in the Ny-Ålesund sonde record (1990–2007), averaged around weak (left) and strong (right) vortex events. Ozone anomalies are expressed as relative deviations from annual cycle values and averaged around weak and strong vortex event dates, as in Figure 7.13. In order to cover for data gaps, sonde observations have been binned in 7 day intervals around the vortex event onset date before averaging. The onset criterion for strong vortex events is* NAM > 2.0 *here, as compared to* NAM > 1.5 *in Figure 7.13 (B). Note also the different vertical coverage in Figure 7.13.*

and 13 strong vortex events are detected in the southern hemisphere. Around weak vortex events a positive ozone anomaly occurs in assimilated ozone, which exhibits a similar shape as the anomaly in Figure 7.13 (A), but is significant only in the middle stratosphere for a few days after day 0 and disappears after 3 months. Strong vortex events are accompanied by a negative ozone anomaly, which is, however, not significant and already present long before the vortex event. In this sense, a SH strong vortex event does mark a cold and stable SH polar vortex, leading to a negative ozone anomaly, but it does not occur as suddenly as in the NH. The distinct shape of dynamically induced ozone anomalies as observed in Figure 7.13 seems a uniquely Arctic phenomenon.

7.2.4. Chemistry-dynamics interaction in the vicinity of NAM events

Section 7.2.2 demonstrated that strong anomalies in the stratospheric circulation, so-called NAM events, exert a strong and distinctive influence on Arctic ozone throughout the whole stratosphere. In this section, the attempt is made to investigate the mechanisms that lead to the distinct shape shown in Figure 7.13 in detail.

A NAM vortex event is accompanied by strong anomalies in both circulation and temperature. Hence, ozone anomalies may be caused by either transport anomalies, or changes in photochemical reaction equilibria by changed temperatures. In fact, a deeper analysis will show that it is an interplay of both factors that is responsible for the shape of the ozone anomaly.

The CTM offers the possibility to directly separate ozone changes (called *increments* in the following) due to the different processes involved – in particular transport, gas-phase chemistry, and polar chemistry. With the simple chemistry scheme employed here, this method of attribution has to be used with caution, as the chemical increment directly depends on

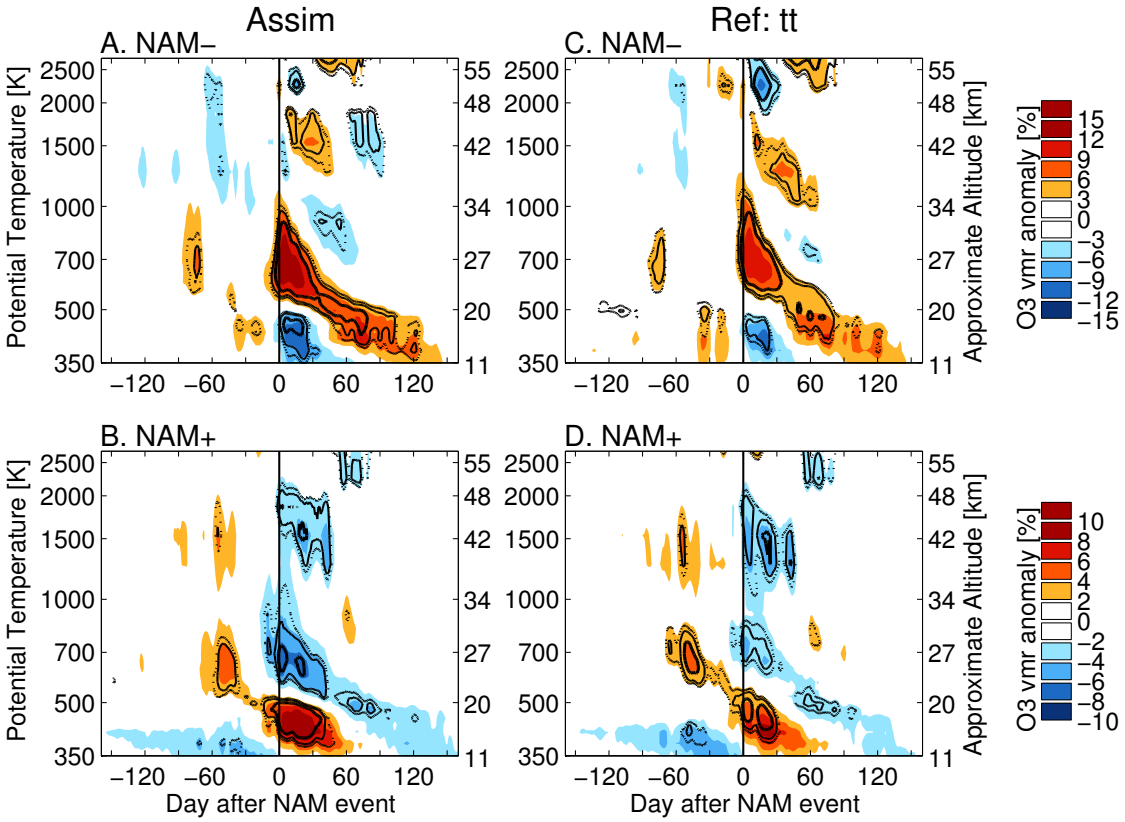

Figure 7.15.: *Ozone anomalies around weak and strong vortex events in the assimilated data set and in the tt reference run, zonally averaged poleward of 65° N. Left column (A, B): assimilated ozone, right column (C, D): reference run tt. Top row (A, C): weak vortex events, bottom row (B, D): strong vortex events.*

the ozone mixing ratio: the Linoz scheme calculates a steady state ozone vmr depending on ozone vmr, temperature, and overhead column ozone (see Section 3.2). Ozone is forced to relax against its steady state, and thus the chemistry scheme tends to cancel strong ozone excursions (provided that the ozone lifetime is short enough). Hence, to first order, chemistry "counteracts" transport, and the chemical increment is not very useful to diagnose the influence of actual chemical processes involved. Nonetheless, the chemistry scheme provides useful information, in particular the chemical lifetime of ozone, and a measure for the influence of temperature anomalies on the photochemical ozone equilibrium. Both of these will be used in an analysis of the causes of NAM-related ozone anomalies.

The advective increment, on the other hand, is directly related to the external wind fields and not subject to any simplifications. In the CTM run with assimilation, the assimilation increment constitutes the third source of ozone changes that has to be taken into account. In practise, the assimilation increment complicates attribution, as it partly counteracts chemistry (if the observed ozone does not agree with the Linoz steady state ozone). In order to be self-consistent, increments from a model run without assimilation are analysed here (the *tt* reference run introduced in Chapter 6). Furthermore, the analysis presented in this section

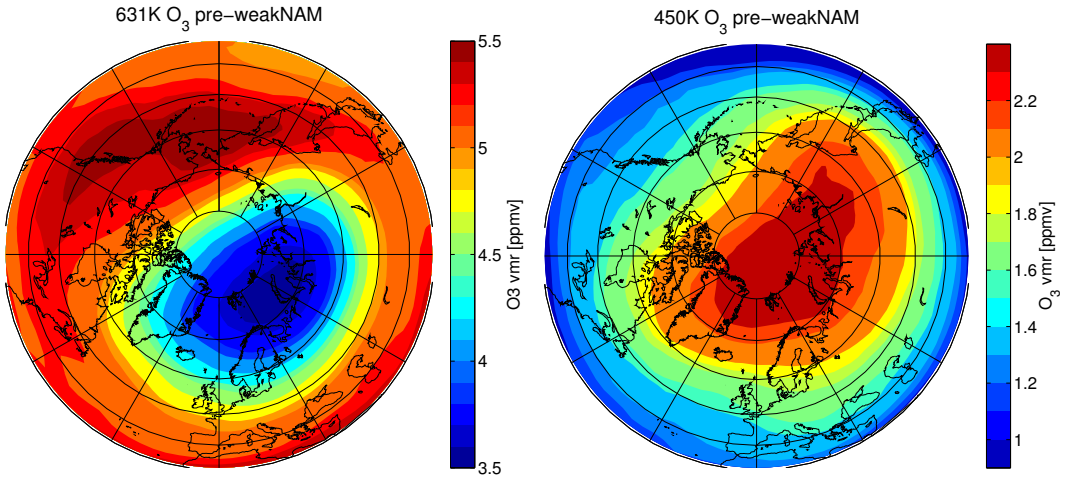

Figure 7.16.: *Typical ozone configuration in the assimilated data set in the middle and lower Arctic stratosphere during winter. Left: Middle stratosphere (631 K / 25 km), Right: lower stratosphere (450 K / 16 km). The fields shown here are averages of 1 month before weak vortex events.*

is restricted to zonal means, as the moving equivalent latitude coordinate system induces ozone changes due to the changing vortex position that complicate attribution. Zonal mean ozone anomalies in the vicinity of NAM events are shown in Figure 7.15, which compares assimilated ozone (panels A and B, left column) to ozone from the *tt* reference run (C and D, right column) around weak (top row) and strong (bottom row) vortex events. The conclusion to be drawn from Figure 7.15 is two-fold: Firstly, NAM-related ozone anomalies follow a similar pattern in the zonal frame as in the equivalent latitude frame (with the notable exception of the delayed upper-stratospheric ozone signal being present in the case of weak vortex events as well). Secondly, the ozone anomaly patterns in the *tt* run are weaker than in assimilated ozone (assimilation obviously enhances NAM-related ozone variability), but qualitatively resemble those in the assimilated run quite well. Hence the increments from the *tt* run can be used to obtain a qualitative view of what happens. Indeed, ozone differences between the assimilated run and the free model are useful to identify small shortcomings of the chemistry scheme, as explained below.

The average initial ozone configuration in the month before a vortex event (which is actually the typical ozone configuration during winter) in the assimilated data set is shown in Figure 7.16 for two different potential temperature levels. Ozone is mainly generated in the tropical middle stratosphere and then transported poleward and downward by the Brewer-Dobson circulation (see Chapter 2). During winter months, the polar vortex forms a rather secluded area of relatively low ozone concentrations into which transport is decreased, and hence an often crescent-shaped zone of high ozone concentration builds up around the vortex during winter. Consequently, the ozone field typically exhibits a negative poleward gradient in the middle stratosphere (Figure 7.16, left). Below about 500 K potential temperature ($\sim$ 20 km), the accumulated ozone in polar regions exceeds the low ozone mixing ratios in the extra-polar lower stratosphere, and the meridional ozone gradient reverses (Figure 7.16, right).

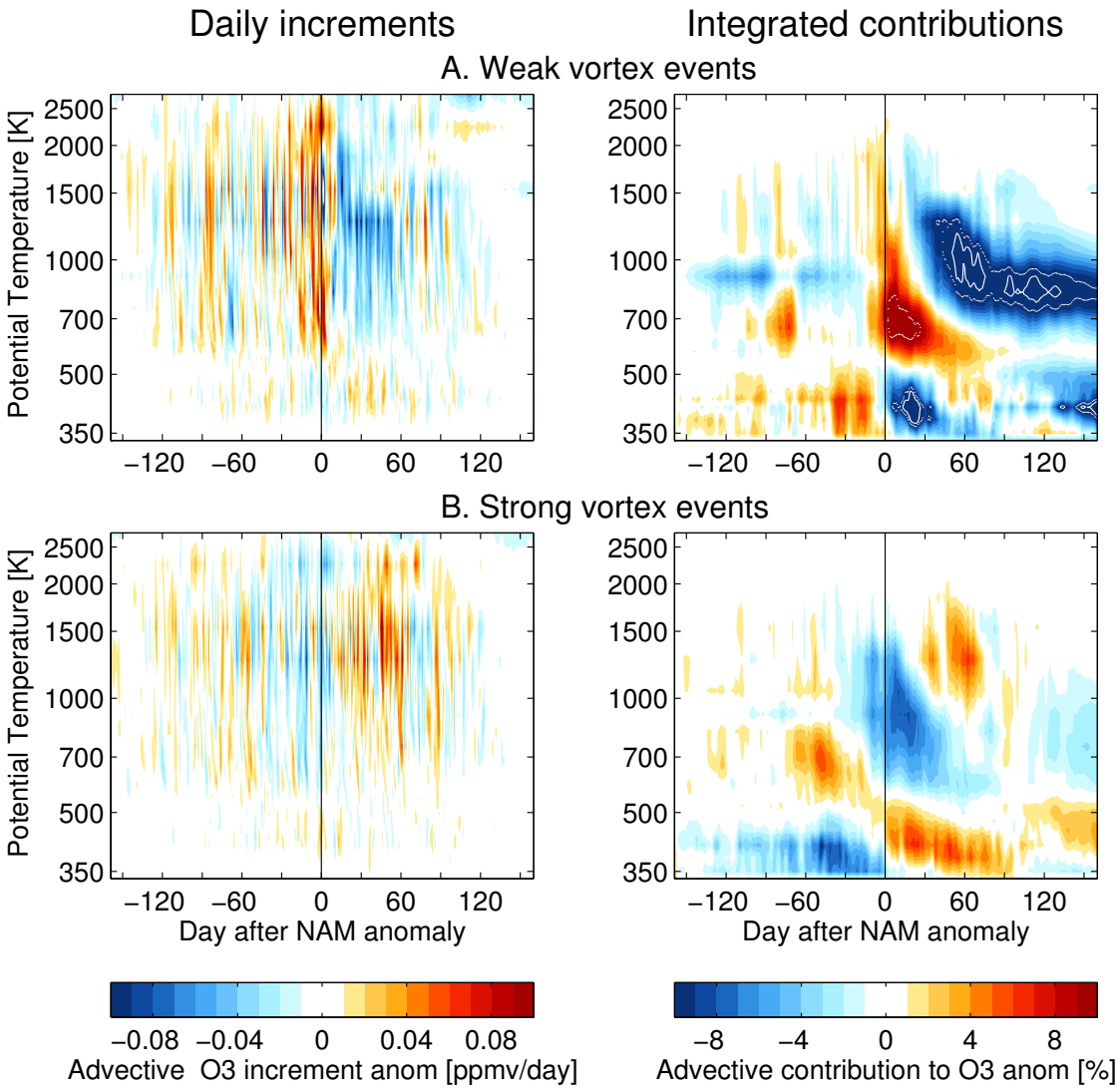

Figure 7.17.: *Anomaly in ozone transport to high latitudes around weak (A) and strong (B) vortex events, and its contribution to the observed ozone anomalies. Left column: Anomaly in daily ozone increment caused by advection. Right: Integrated contributions to the overall ozone anomaly. Here, all increments from the beginning of the depicted time slot to day x are summed up, weighted exponentially with the average ozone lifetime. To enhance comparability with Figure 7.15, the integrated contributions are shown as fractions of the mean ozone field. Values exceeding the colour scale are indicated by white contour lines (dotted line ±10%, solid line ±12%).*

First and foremost, a "vortex event" is characterised by a strong anomaly in meridional advection of air. In case of a weak vortex event, advection of air into the vortex is greatly enhanced and the vortex may even experience an early breakup, while during a strong vortex event, the vortex is more secluded and transport is suppressed. The dynamical information, or phase, of the transport anomaly propagates from the uppermost stratosphere or even mesosphere down to the tropopause, opposite to the direction of the energy flux of the

atmospheric waves that caused it. Baldwin and Dunkerton (2001) have shown that the dynamical anomaly, albeit weakened, even penetrates the troposphere and eventually reaches the surface (see Figure 7.12). These anomalies in transport are indeed responsible for the shape of the ozone anomaly around day 0 up to a month afterwards. The first impression of the ozone anomaly in Figure 7.13 might be that it traverses most of the stratosphere quickly, then suddenly stops around 500 K and descends from there at a much decreased speed, implying that there might be two different processes involved that lead to the characteristic "hockey stick" shape. However, as detailed below, a closer inspection shows that no second process is needed to explain the shape of the curve. Indeed, the ozone anomaly follows closely the dynamical anomaly across the whole stratosphere and does not stop at 500 K. What does change at 500 K is the sign of the meridional ozone gradient, and hence the effect of the transport anomaly. Above 500 K, enhanced transport in a weak vortex event leads to higher ozone vmr, while below 500 K low-ozone air is advected into the vortex.

The anomaly in ozone advection into the polar vortex is displayed in Figure 7.17, which shows both the anomaly in ozone transport itself expressed as additional vmr increase per day (left), as well as its integrated counterpart, the total contribution of advection anomalies to the ozone anomaly up to day t (right), for both weak (A) and strong (B) vortex events. Contributions $\mathrm{O}_{3_{advect}}$ are calculated as sum of daily increments $\Delta\mathrm{O}_{3_{advect}}$ from day -160 to day t, weighted exponentially with the mean ozone lifetime τ,

$$\mathrm{O}_{3_{advect}}(t) = \sum_{t'=-160}^{t} \Delta\mathrm{O}_{3_{advect}} \exp\left(-\frac{t-t'}{\tau}\right) \ . \tag{7.1}$$

Due to the noisiness of the daily increments, the integrated quantity is more useful for attribution. For better comparability to Figure 7.15, the advective contributions are shown here in relation to the mean ozone. Note the (natural) delay between advection anomaly and its integrated effect.

As obvious from Figure 7.17, the transport anomaly accounts for much of the initial shape of the ozone anomaly. Shortly after the vortex event, the anomaly in daily advective increment reverses, resulting in a transport-related ozone anomaly with opposite sign (especially in the weak vortex case), that does however not show up in the overall effect (compare Figure 7.15) and must hence be cancelled by chemical effects. The average photochemical lifetime of ozone in the vicinity of vortex events is shown in Figure 7.18. Figure 7.18 uses dates of strong vortex events, but is almost identical if weak vortex event dates are used, as their average time of occurrence does not differ significantly (see Figure 7.11). In the upper stratosphere, where the ozone lifetime is short, ozone mixing ratios quickly return to their undisturbed state. In the lower stratosphere, the ozone lifetime is long, and from a chemical viewpoint the anomaly can survive for months while following a slow descent curve that is governed by diabatic cooling rates.

Although much of the initial shape of the ozone anomaly is related to dynamics, chemistry plays a distinct role for the overall picture. Advection alone can neither explain the second ozone anomaly arising in the upper stratosphere roughly a month after the vortex event, nor the very long persistence of the ozone anomaly in the lower stratosphere. Shortly after the onset of the vortex event, the respective anomaly in advection reverses in the middle stratosphere, and consequently a strong ozone signal of opposing sign sets in that is not present in the overall anomaly picture. While chemistry generally tends to counteract transport-induced anomalies, in this case the chemical effect dominates. After weak vortex

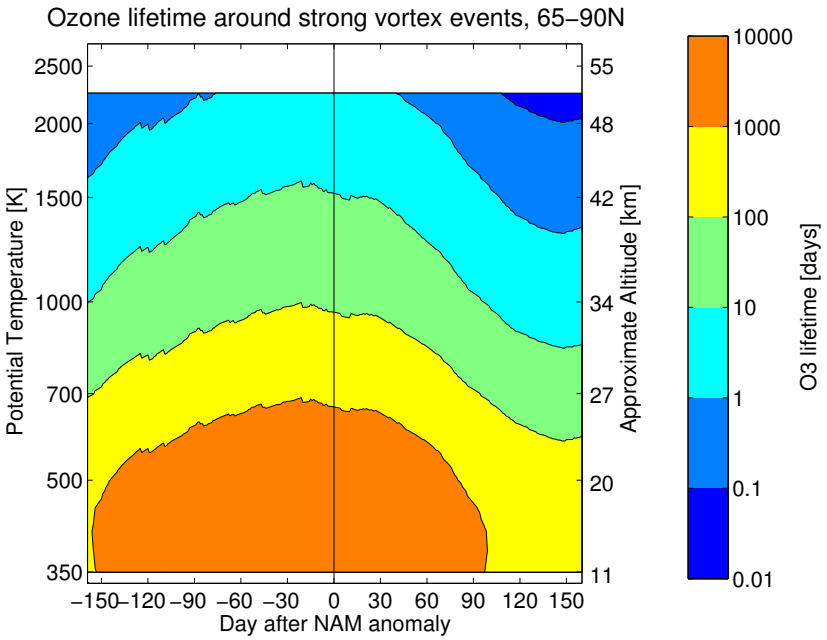

Figure 7.18.: *Photochemical ozone lifetime in the vicinity of strong vortex events, zonal average from 65° N poleward. The picture is very similar if weak vortex event dates are used.*

events, a positive ozone anomaly is observed in the middle stratosphere 1–2 months after the NAM event *in spite of* decreased ozone transport to high latitudes, and vice versa after strong vortex events (see Figure 7.15). A deeper investigation shows that this dominance of chemical against transport effects is related to the shift of photochemical reaction equilibria due to temperature anomalies.

Figure 7.19 shows the Arctic temperature anomaly (deviation from the annual cycle) in the vicinity of weak and strong vortex events. Again, as for ozone anomalies, the pictures are strikingly similar with an inverse sign. Alternating patterns of temperature anomalies are observed: Around day 0 of a weak (strong) vortex event, a strong positive (negative) temperature anomaly (around +10 K/-5 K) occurs which proceeds from the upper to the lower stratosphere quickly (somewhat slower in a strong vortex event) and lasts up to a month. Concerning the smaller absolute values of the temperature anomalies, it should be noted that following Baldwin and Dunkerton (2001), the criterion used here to identify a strong vortex event is incomparably weaker than that for a weak vortex event (NAM index values of 1.5 vs. −3.0, respectively). If a stricter threshold value for strong vortex events is used, such as NAM > 3.0, the temperature anomaly is stronger, but the number of events is then so far decreased that the ozone anomalies are hardly significant.

After this initial anomaly, an even stronger second anomaly with reversed sign occurs, also propagating down from the top of the stratosphere and terminating in the middle stratosphere around 700 K. A third anomaly phase with the same sign as the initial one becomes barely visible at the upper boundary of the model. During the weak vortex event, the ozone anomaly around day 0 seems to follow the descending path of the temperature anomaly. However, this relation is only indirect – both effects are triggered by the large dynamical

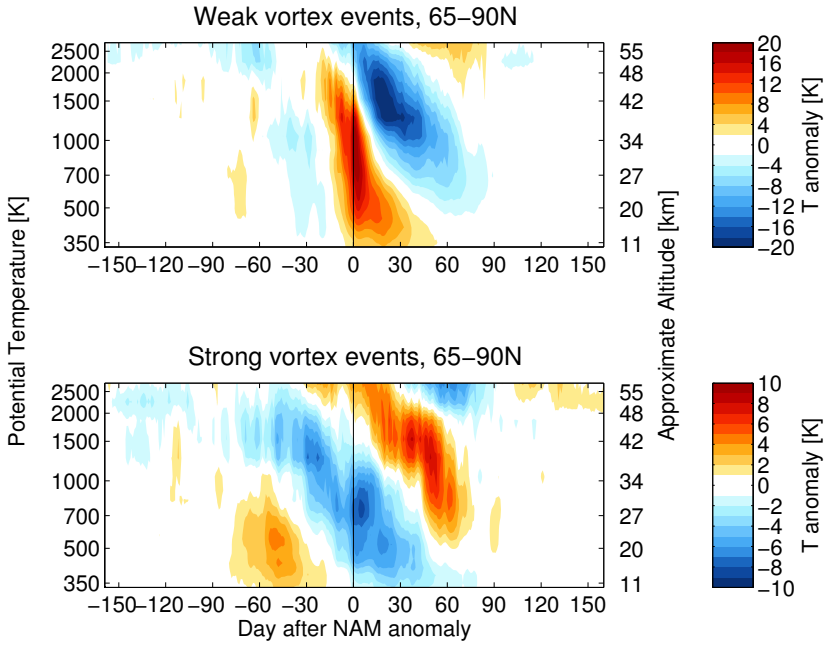

Figure 7.19.: *Temperature anomaly in the vicinity of vortex events, averaged poleward of 65° N: (A) weak vortex events, (B) strong vortex events.*

anomaly, and there is no direct causal relation between a positive T anomaly and a positive ozone anomaly, as the dependency of ozone production on temperature is negative (see below). On the other hand, a striking correlation is seen between the positive temperature anomaly in the upper stratosphere a month after the strong vortex event, and the negative ozone anomaly at the same time and height. To investigate this more quantitatively, the temperature dependence of photochemical steady state net ozone production on temperature, $\partial(P-L)/\partial T$, needs to be considered, which is tabulated in the Linoz chemistry scheme in dependence of latitude, month and altitude. Figure 7.20 shows $\partial(P-L)/\partial T$ as ppmv additional production per day, per K temperature increase for Arctic latitudes around strong vortex events (again, the picture is very similar for weak vortex events, as their average occurrence dates do not differ much). The sign is always negative, and the dependency is largest in the upper stratosphere, increasing roughly exponentially with height. Combining this information with the average lifetime shown in Figure 7.18, it is obvious that in the upper stratosphere, a temperature anomaly of a few Kelvins can very well trigger a large ozone anomaly within a few days.
The effect of temperature changes on net ozone production $P-L$ can be expressed by

$$\Delta(P-L)_T = \frac{\partial(P-L)}{\partial T}(T-T_0) \tag{7.2}$$

where $T-T_0$ is the temperature deviation from the Linoz climatology or long-term average temperature in the assimilated data set (which need not be exactly identical). Figure 7.21 shows the net effect of the temperature anomalies displayed in Figure 7.19 on ozone production for weak and strong vortex events. Due to the long lifetime of ozone in the lower

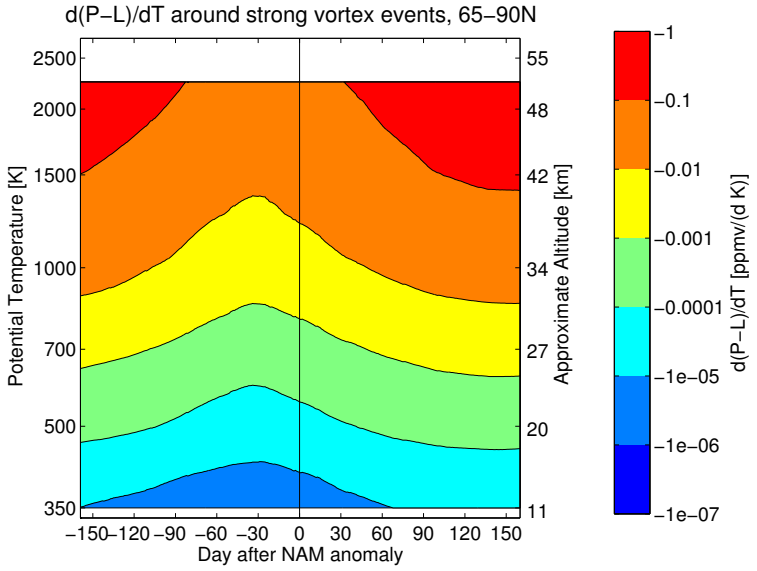

Figure 7.20.: *Dependence of the net ozone production rate on temperature in the Linoz chemistry scheme, averaged for Arctic latitudes around strong vortex event occurrence dates.*

stratosphere, chemistry is only relevant above $\sim 1000\,\mathrm{K}$ / $35\,\mathrm{km}$. In the weak vortex event, the temperature anomaly around day zero causes a small effect on ozone that is adverse to the overall ozone anomaly seen in Figure 7.15, and hence acts to damp the ozone anomaly. A much larger effect with opposite sign – that is, enhanced photochemical ozone production – is caused by the low temperatures a month after the weak event. Conversely, the high temperatures a month after the strong vortex event onset lead to a negative anomaly in ozone production. In both cases, it is the secondary temperature anomaly – opposite to that commonly associated with the respective vortex event – that leads to significant chemical effects on ozone. Thus, it is a close interplay of chemistry and dynamics that leads to the characteristic shape of ozone anomalies during vortex events in the middle stratosphere. In the equivalent latitude coordinate system (Figure 7.13), the shift of chemical reaction equilibria only shows up as negative ozone anomaly 1 month after strong vortex events, while for weak vortex events it is exactly balanced by advection (in this case, a re-strengthening of the vortex).

Interaction between chemistry and dynamics is also relevant in another context for NAM-related ozone anomalies. As explained above, activation of chlorine via gas-phase reactions is almost absent in the lower stratosphere, and hence there is no temperature dependency of gas phase chemistry in the lower stratosphere. However, the formation of polar stratospheric clouds, and heterogeneous activation of chlorine on PSC particle surface, does indeed depend strongly on temperature. In particular, a distinct threshold temperature exists ($T_{NAT} \approx 195\,\mathrm{K}$, depending on pressure), above which no PSC formation is possible. Thus, temperature anomalies present during weak and strong vortex events can have a profound influence on lower stratospheric ozone. In the same manner as Figures 7.17, Figure 7.22 shows the anomaly in polar chemistry contribution to ozone budgets around weak (A) and strong (B) vortex events, showing both changes in daily increments (left) and their integrated

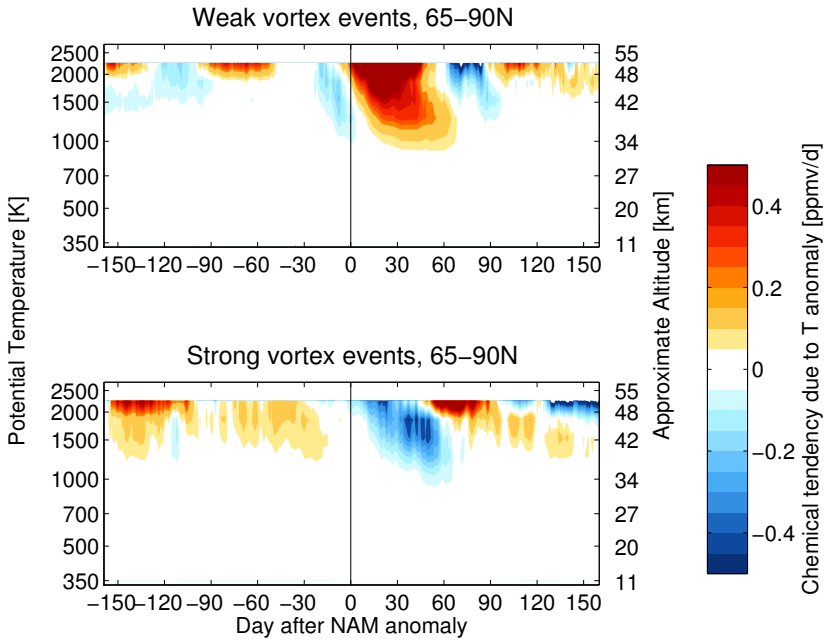

Figure 7.21.: *Effect of temperature anomalies on ozone net photochemical production P–L, in the vicinity of weak (top) and strong (bottom) vortex events.*

contributions (right). In both cases, the daily increments show distinct anomalies in the polar chemical contribution to ozone changes in line with the temperature anomaly occurring at the same time – positive temperature anomalies during weak vortex events inhibiting the occurrence of heterogeneous ozone destruction, while lower temperatures during a strong vortex event lead to enhanced chlorine activation and ozone destruction. Interaction with polar chemistry hence acts to strengthen the dynamically induced ozone anomalies in both cases. The integrated effect becomes most pronounced with a delay of 1-2 months, thus contributing significantly to the long lifetimes of the ozone anomalies seen in Figure 7.13.
A slightly different view on the same issue is provided in Figure 7.23, which shows the probability for strong polar ozone depletion to occur in the vicinity of vortex events. The magnitude of ozone depletion is derived as differences between reference runs *tt* (with polar chemistry scheme) and *tn* (without polar chemistry scheme). The threshold criterion used here to distinguish "strong polar chemistry" is rather arbitrary: As soon as the offset $tt - tn$ is larger than 0.3 ppmv at any model level, the day is counted in the strong polar chemistry category. Varying the threshold value up to 1.0 ppmv offset does not change the overall picture much, but decreases the number of occurrence times and hence the significance of the effect. For weak vortex events, only a few occurrences of strong polar chemistry are counted (20% of all strong vortex events), lasting for a few days around day 0. These are caused by a small negative temperature anomaly before the vortex event. Hereafter, strong polar chemistry is practically absent after weak vortex events. On the contrary, in more than 50% of strong vortex events, strong effects of heterogeneous chemistry are experienced. Due to the accumulation of daily ozone losses, the peak of the distribution is delayed from the vortex event by about two months.

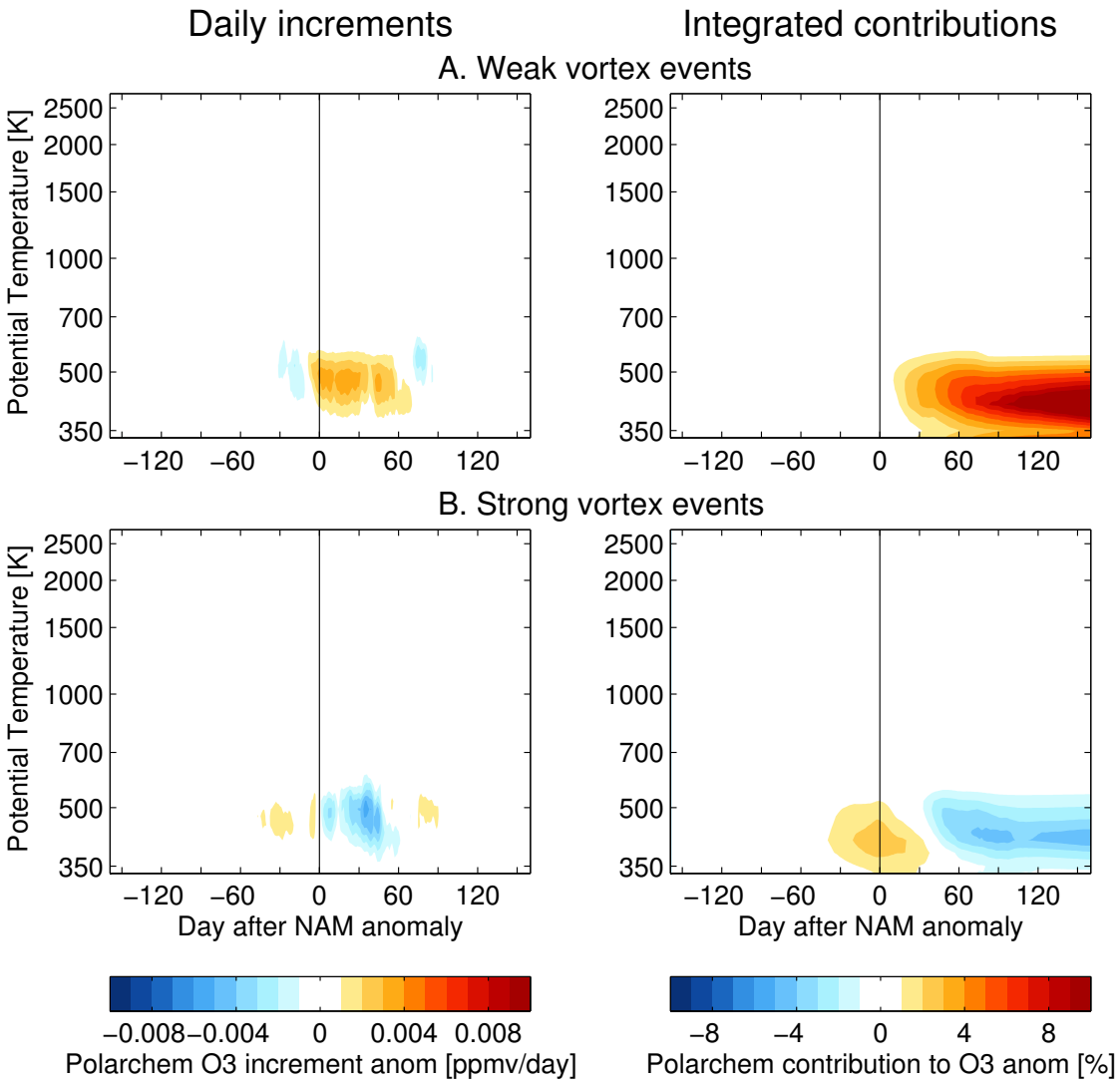

Figure 7.22.: *As Figure 7.17, but for heterogeneous polar chemistry. Left column: Anomaly of daily ozone increment due to heterogeneous chemistry around weak (A) and strong (B) vortex events, and its contribution to the observed ozone anomalies. Right: Integrated contributions to the overall ozone anomaly. Here, all increments from the beginning of the depicted time slot to day x are summed up, weighted exponentially with the average ozone lifetime. To enhance comparability with Figure 7.15, the integrated contributions are shown as fractions of the mean ozone field.*

7.3. Effects of energetic particle precipitation

The foregoing section focused on the influence of dynamical anomalies on polar ozone, and illustrated effects of chemistry-dynamics interaction. As an example for a completely different process by which anomalies in mid-stratospheric polar ozone may arise, a short discussion of the effects of energetic solar proton precipitation is presented here. Since chemical effects of particle precipitation are in no way included in the CTM chemistry scheme, the discussion

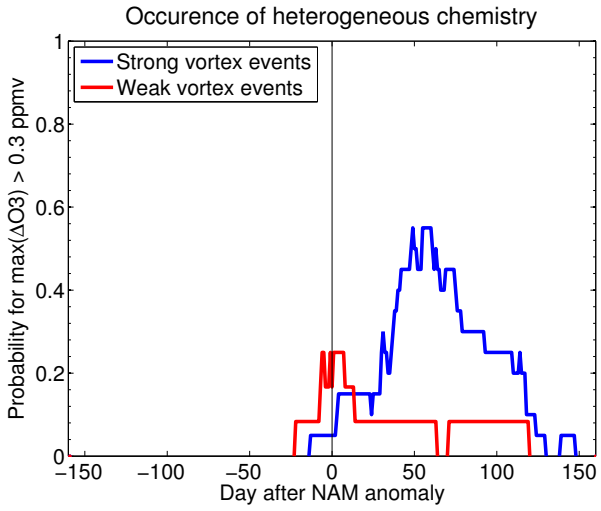

Figure 7.23.: *Probability distribution for the occurrence of strong heterogeneous chemistry around weak and strong vortex events, as obtained fro differences of CTM reference runs tt (with polar chemistry) and tn (without polar chemistry).*

cannot be as detailed as in the case of dynamical effects. Nonetheless, useful insights can be gained on the creation and persistence of polar ozone anomalies induced by solar particle precipitation from an analysis of the assimilated data set in conjunction with reference runs without assimilation.

Large outbursts of high energetic particles from the sun have been measured for several decades (for early records see e.g., Freier and Webber, 1963), and their effects on the Earth's atmosphere have become a field of intense research. Crutzen et al. (1975) first showed that solar proton events (SPEs) are a source of nitric oxides in the upper stratosphere, which in turn lead to ozone destruction via the NO_x catalytic cycle (Crutzen, 1970). While also odd-hydrogen compounds are produced (Swider and Keneshea, 1973; Frederick, 1976) that impact ozone via the HO_x cycle (Hunt, 1966), the lifetime of HO_x is too short to lead to a longer-lasting effect. NO_x, on the other hand, is long-lived during winter, especially when the inactive odd-nitrogen species are included and the whole NO_y family is considered ($NO_y = \{N, NO, NO_2, NO_3, N_2O_5, HNO_3, HO_2NO_2, ClONO_2, BrONO_2\}$). NO_y produced in the upper stratosphere may propagate downward into the polar vortex, contributing to ozone destruction on timescales of months (Jackman et al., 2000). This mechanism has been observed several times in connection with SPEs during the last solar cycles, and is consistent with model calculations (Jackman et al., 2001).

Figure 7.24 shows the time series of high energetic proton fluxes ($E > 100\,\mathrm{MeV}$) as measured by the Geostationary Earth Observing Satellites (GOES). Several SPEs are visible as large excursions of the proton flux from an otherwise low baseline. Solar particle eruptions are positively correlated with the solar cycle. During high solar activity, SPEs occur typically in groups with only weeks or months interval, whereas years of low proton flux (and lower solar activity) lie between these groups. Dates of the most important SPEs during the period of the assimilated data set are given in Table 7.1.

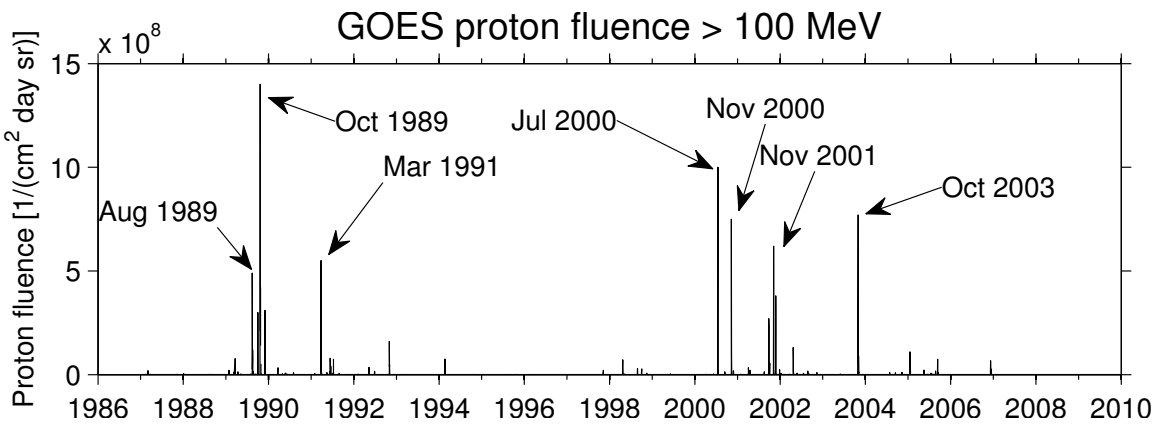

Figure 7.24.: *High energetic proton flux as measured by the GOES satellites.*

Short name	Date
Aug 1989	1989/08/13 – 1989/08/26
Sep 1989	1989/09/29 – 1989/09/30
Oct 1989	1989/10/19 – 1989/10/27
Jul 2000	2000/07/14 – 2000/07/16
Nov 2000	2000/11/09 – 2000/11/11
Sep 2001	2001/09/24 – 2001/09/30
Nov 2001 (1)	2001/11/05 – 2001/11/07
Nov 2001 (2)	2001/11/23 – 2001/11/25
Oct 2003	2003/10/28 – 2003/10/31
Nov 2003	2003/11/03 – 2003/11/05

Table 7.1.: *Periods of some large SPEs during recent decades.*

It is straightforward to test the assimilated data set for effects of these large SPEs by simply analysing polar ozone anomalies in the vicinity of SPE dates. This is done in Figure 7.25 for the SPE in October 1989, showing the deviation from the annual cycle of assimilated ozone northwards of 70°N. A negative ozone anomaly is observed, which seems to originate around day 0 in the uppermost levels of the CTM and then gradually propagates downward, reaching its largest amplitude of $\sim -30\%$ at 1000 K after about two months. Due to the large variability of the assimilated ozone field during winter months, the propagation of the anomaly is not easy to follow with the naked eye, and the approximate trajectory of an air parcel starting at ~ 2000 K at day 0 is plotted as a guide to the eye. The descent of the ozone anomaly is in good agreement to the expected pathway. Although Figure 7.25 suggests a strong negative ozone anomaly as a result of the SPE, the analysis can go beyond this point. Firstly, it is useful to take the transition to equivalent latitudes in order to eliminate some of the short-timescale variability and concentrate on polar vortex ozone. Secondly, an analysis of assimilated ozone alone is not the optimal method to specifically distill anomalies related

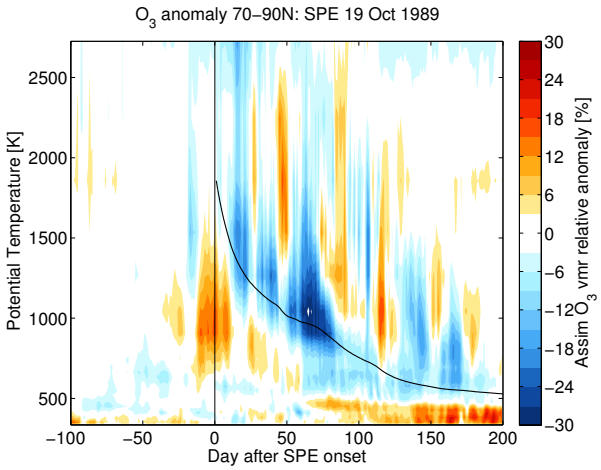

Figure 7.25.: *Anomaly in assimilated ozone around the SPE in October 1989. Ozone anomalies are calculated as deviations from the annual cycle and averaged north of 70°N. As a guide to the eye, the approximate trajectory of an air parcel starting at day 0 at 1900 K is shown, calculated from the diabatic heating rates.*

to SPE-related chemistry. Information of any ozone anomaly in assimilated ozone induced by solar proton precipitation can only stem from the SBUV satellite data, as the CTM itself is completely unaware of energetic particle precipitation and the chemistry associated with it. Hence, a rather than analysing ozone anomalies in the assimilated data set itself, it is more instructive to analyse the difference between assimilated ozone and ozone in the *tt* reference run without assimilation, which is defined here as $\Delta O_3 \equiv \text{assim} - tt$. Such an analysis shows changes in assimilated ozone that are not explained by the CTM chemistry and dynamics.

Both CTM and assimilated ozone contain annual cycles, which do not fully agree. Thus the annual cycle of ΔO_3 has to be subtracted in order to arrive at increments that are not caused by seasonal variations in either data set. Since biases exist between different SBUV satellite instruments that are not completely cancelled by the SBUV adjustment coefficients[2], the annual cycle of ΔO_3 needs to be calculated separately for each SBUV satellite instrument.

The analysis of SPE effects is focused on autumn and winter events, for good reasons: Firstly, lifetimes of photochemically active NO_x species are short during summer; secondly, downward transport and influences on the middle and lower stratosphere are only efficient during winter months, and thirdly, the sampling of SBUV satellite observations is not high enough to guarantee a high level of satellite information in the assimilated data set in the summertime upper stratosphere. With increasing altitude, the ozone lifetime gradually becomes too short to bridge the gap between satellite observations, as the ozone field relaxes against its photochemical equilibrium state given by the Linoz chemistry scheme. Hence high-stratospheric summer ozone anomalies present in the SBUV observations may be lost in the assimilated data set.

Figure 7.26 shows anomalies of ΔO_3, calculated as described and averaged poleward of 70°N

[2]In some cases, biases towards the CTM are even enlarged by the adjustment coefficients (NOAA 11 1997–2001).

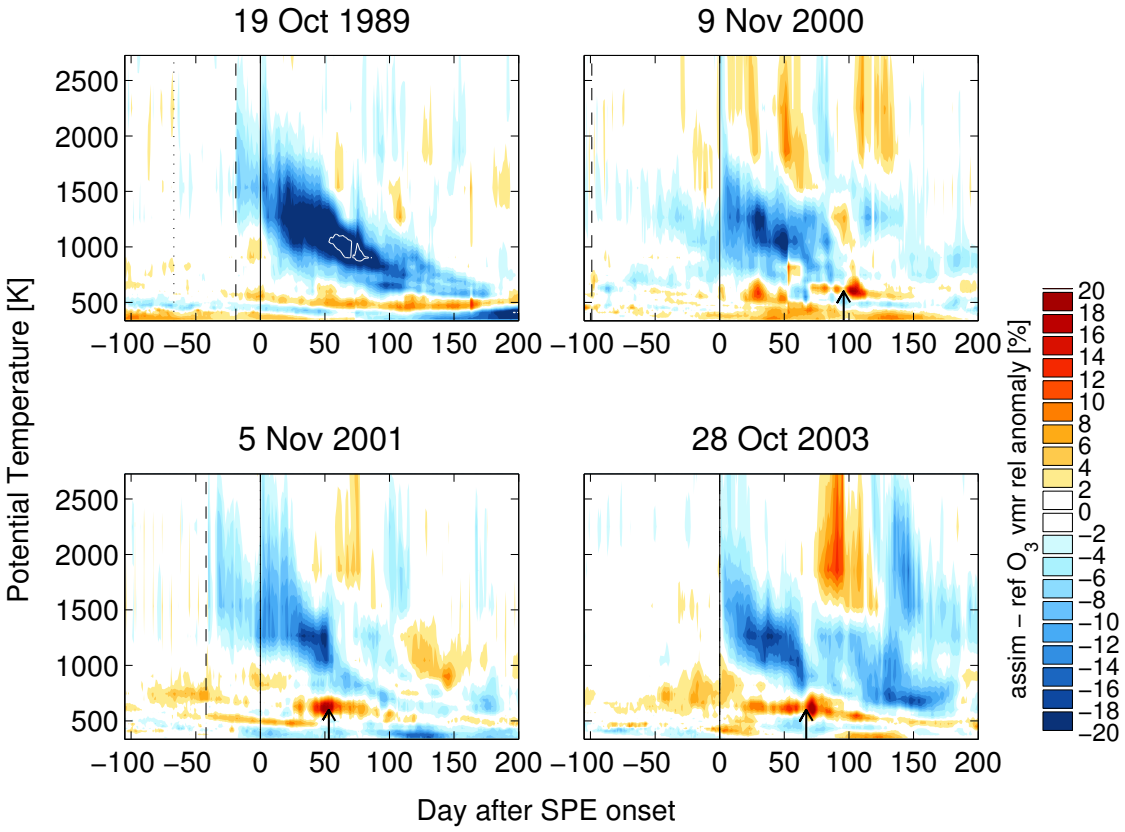

Figure 7.26.: *Anomalies of ozone differences between assimilated ozone and reference runs, averaged poleward of 70°N equivalent latitude in the vicinity four different SPEs and expressed as percentages of assimilated ozone vmr. An additional contour line at the −25% level is shown as while line (only relevant for winter 1989/90). Day zero (black line) corresponds to the first day of enhanced proton fluxes measured by GOES. For October 1989, November 2000, and November 2001, also the preceding events of 30 September 1989, 2 August 2000, and 24 Sep 2001 are shown as dashed lines. In 1989, the SPE in August is shown as dotted line. Weak vortex events occurring in three of the winters are shown as black arrows near the abscissa (13 Feb 2001, 28 Dec 2001, 3 Jan 2004).*

equivalent latitude in the vicinity of four large SPEs that occurred during fall or winter. In addition, preceding events are marked by dashed lines. In all cases, strong negative anomalies in ΔO_3 are visible which originate exactly at the onset date of the SPE, covering the region from the top of the CTM down to $\sim$ 1200 K. While upper-stratospheric ΔO_3 returns to normal values within a few days to weeks (mainly depending on the duration of enhanced ionisation rates), the negative anomaly in mid-stratospheric ozone intensifies, reaching values of up to −20% of *tt* reference run ozone (in absolute values around 1 ppmv) after a few weeks. In accordance with the negative heating rates in Arctic mid-stratospheric winter, the anomaly slowly descends to the lower stratosphere, remaining visible for at least 2–3 months. For the reasons explained above, preceding events during summer (August 1989, July 2000) cause only weak anomalies in the assimilated data set which are slightly better visible in the middle stratosphere after the July 2000 event. After the very strong SPEs 1989 and

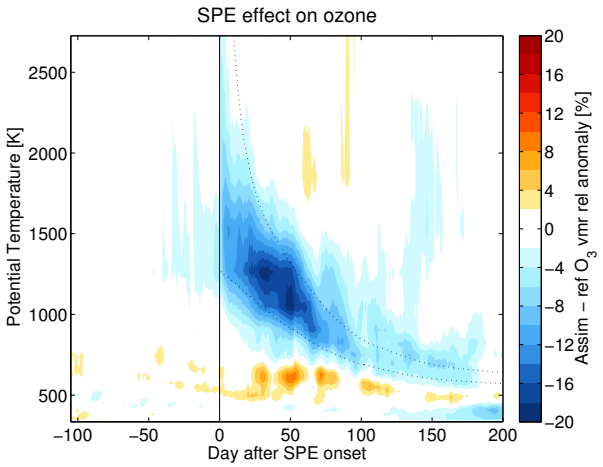

Figure 7.27.: *Average effect of solar proton precipitation on Arctic ozone. Composite plot of anomalies in* $\Delta O_3 = assim - tt$ *around wintertime SPEs, averaged poleward of* $70°N$ *equivalent latitude and expressed as fraction of the tt reference run ozone vmr. Black dotted lines indicate the descent pathways of air parcels initiated at 1265 K at day 0 and at 2730 K ten days later, approximately framing the evolution of the initial ozone anomaly.*

2003 low ozone anomalies seem to survive for about six months before disappearing. Three out of the four winters displayed in Figure 7.26 experienced weak vortex events, leading to the effects discussed in detail in Section 7.2.2 – in particular, mixing in of extra-vortex air masses and thus a weakening of the negative ozone anomaly resulting from the SPE-induced NO_x production. Dates of the weak vortex events are marked as arrows near the abscissa in Figure 7.26. In all three cases, the descending negative ozone anomaly is disrupted at the date of the weak vortex event. The winters of 2000/2001, 2001/2002, and 2003/2004 display the full range of possible effects of weak vortex events: early vortex breakup and total dispersion of the negative ozone anomaly (2000/2001), disruption of the vortex and weakening of the ozone anomaly (2001/2002), or re-establishment of a strong polar vortex and continuation of the negative ozone anomaly after a short disturbance (2003/2004). All weak vortex events are followed by positive ozone increment anomalies in the upper stratosphere, confirming the conclusion in Section 7.2.4 that the assimilation acts to correct the CTM chemistry in the upper stratosphere after NAM events.

The very long lifetime of SPE-induced ozone anomalies may be attributed to ongoing ozone destruction through the NO_x catalytic cycle along the descent path of the air parcel. Since NO_x is not included explicitly in the CTM version used for generating the assimilated data set, this explanation cannot be tested at present.

The long lifetimes observed in single events give rise to the question for the average duration of the SPE-related ozone anomaly. Figure 7.27 displays a composite of Arctic ΔO_3 anomalies in the vicinity of winter-time SPE events. In order to avoid double counts of neighbouring events, only the four events in Figure 7.26 (day 0) are used in the composite. For this reason, the impression arises of a small anomaly preceding the event, which is caused by the preceding events shown in Figure 7.26. The average ozone anomaly is around -20% in the middle stratosphere and displays a compact pattern in its descent to the lower stratosphere,

where it remains visible as $\sim 5\%$ anomaly up to six months after the SPE. As a guide to the eye, descent pathways of air parcels starting from 1265 K at day 0 and from 2700 K ten days after the SPE are plotted in Figure 7.27, as calculated from the ERA-Interim heating rates. The descent of the average ozone anomaly is in good agreement with the expected path. Small positive ozone anomalies in the middle stratosphere below the negative anomaly may be related to the so-called self-healing effect of the ozone layer: in the event of a negative ozone anomaly in the upper stratosphere, more short-wave UV radiation is able to penetrate to the lower stratosphere, enhancing ozone production there via the Chapman cycle.
Qualitatively, these results are in good agreement to model results by Jackman et al. (2005) and Jackman et al. (2009), who reported statistically significant ozone decreases of $2-10\%$ in the middle stratosphere for a few months after the SPEs of 2000, 2001, and 2003, deduced from calculations of the 2D GSFC CTM and the WACCM3 climate model including SPE-related ion chemistry. Jackman et al. (2005) calculated ozone losses of more than 50% in the uppermost stratosphere for single days after the July 2000 SPE owing to HO_x chemistry. Although the assimilated data set does show a distinct signal in the upper stratosphere directly after the SPE, the magnitude of this effect is far lower in the assimilated data set, which is understandable given the short lifetime of ozone there and the decreased content of SBUV satellite information in the assimilated data set. On the other hand, the long-lived effects related to NO_x chemistry in the middle stratosphere are well represented in the assimilated data set.
Although SPEs are relatively rare events and only contribute to a few of the Arctic ozone anomalies in Figure 7.1, results presented in this section show that large influences of energetic particle precipitation are contained in the assimilated ozone data set. Negative ozone anomalies of more than 20% occur in the middle stratosphere after solar proton events, and these ozone anomalies exhibit a long lifetime of up to six months. As a side remark, it should be noted that large proton events are often accompanied by strong precipitation of energetic electrons, which cannot penetrate as deeply into the stratosphere as protons but may also contribute to the observed ozone anomalies via downward transport of NO_y formed in the mesosphere (Sinnhuber et al., 2006).
Influences of solar variability affect the ozone layer in several ways. As shown here, solar particle precipitation constitutes an important influence to polar ozone variability on a timescale of a few months. In contrast, direct influences of solar radiation variation on the long-term variability of global ozone are discussed in Chapter 8 below.

7.4. Summary of Chapter 7

Chapter 7 was concerned with the analysis of polar ozone variability. With a daily three-dimensional coverage of the whole stratosphere over three decades, and a realistic representation of high-latitude ozone variability (see Section 6.2.3), the assimilated data set is the ideal tool for studying the formation and evolution of ozone anomalies during polar night, when satellite observations are scarce.
The analysis of polar ozone anomalies in Chapter 7 shows that in general, Arctic ozone anomalies develop in the middle to upper stratosphere around 30 km altitude, mostly during autumn or winter, and then descend from there following the diabatic cooling rates. An average lifetime of ozone anomalies of around 6 months is inferred both from the auto-correlation and composite plots of ozone anomalies. Hence, ozone anomalies in the lower

stratosphere often have precursors in the mid- to upper stratosphere a few months earlier. Two mechanisms for the creation of ozone anomalies are investigated in detail: anomalies in the stratospheric circulation, which are expressed here by the NAM index, and effects of solar energetic particle precipitation.
While the NAM phase is generally anti-correlated with mid-stratospheric polar ozone, pointing to its role as a proxy for poleward transport of mid-latitudinal air masses, strong excursions of the NAM phase lead to strong and distinct patterns of polar ozone anomalies (around 20%) of opposite sign. These so-called weak and strong vortex events induce ozone anomalies that follow characteristic patterns in space and time, which – as discussed in Section 7.2.4 – are the result of different interactions of stratospheric dynamics and chemistry at different altitudes. While the anomaly in meridional ozone transport dominates the immediate shape of the ozone pattern a few days after the vortex event, effects of temperature changes due to the vortex event have a profound influence on the shape of the medium-term evolution of the ozone anomaly: in the upper stratosphere, temperature anomalies lead to ozone anomalies with inverse sign (lower temperatures shift photochemical gas-phase reaction equilibria towards higher ozone mixing ratios), while in the lower stratosphere the opposite is true (lower temperatures enable more formation of polar stratospheric clouds leading to stronger chlorine activation and more ozone depletion).
In addition to the dynamical effects on ozone, influences of solar particle precipitation have been discussed in Section 7.3 as a completely different mechanism by which ozone anomalies are created. Large outbursts of high energetic protons from the sun – so-called SPEs – are well known to cause negative ozone anomalies in the upper and middle stratosphere through enhanced production of nitrogen oxides, which in turn deplete ozone through the catalytic NO_x cycles. Some of the long-lived ozone anomalies in the assimilated data set are indeed related to large solar proton events. By analysing differences between ozone from the assimilation run and the unconstrained CTM, it is shown that in the wake of SPEs significant ozone decreases occur in the middle and even lower stratosphere that are visible up to 6 months later – as long as the vortex is not completely disrupted as a consequence of a weak vortex event.
The long lifetimes of such solar particle induced ozone anomalies point out the importance of NO_x chemistry for understanding mid-stratospheric ozone, and raises the question whether NO_y anomalies may also play a role for other long-lived polar ozone anomalies that are not directly related to solar activity.

8. Long-term trends in the assimilated data set, and their attribution to chemistry and dynamics

While Chapter 7 was focused on the short-time variability of polar ozone and its interaction with stratospheric dynamics, this chapter is dedicated to an analysis of the long-term evolution of ozone in the assimilated data set and the sensitivity simulations. The long-term evolution of the ozone layer is of great interest for both science and society, as it regulates the UV irradiation at the earth surface, and has thus been an intense field of research during the last decades (e.g., World Meteorological Organization, 2007, 2011, and references therein).

Strong decreases of stratospheric ozone were first observed over the Antarctic with the discovery of the ozone hole in 1985 (Farman et al., 1985; Stolarski et al., 1986). Antarctic ozone depletion was soon attributed to anthropogenic inputs of chlorofluorocarbons (Solomon et al., 1986), leading to the regulation of the responsible substances by the Montreal Protocol in 1987 and subsequent amendments. Some years after the discovery of the Antarctic ozone hole, significant decrease of the ozone layer was observed outside of polar areas as well (Stolarski et al., 1991), the reasons for which became then subject of an active scientific debate. A large number of studies has since dealt with the detection and attribution of ozone trends, both from observational (e.g., Stolarski et al., 1991; Harris et al., 1998; World Meteorological Organization, 2003), and modelling (e.g., Chipperfield, 2003; Hadjinicolaou et al., 2005; Fleming et al., 2007) perspectives. Several factors have been identified as contributors to mid-latitude ozone decreases, of both natural and anthropogenic origin. While anthropogenic contributions mainly concern the mentioned increases in ozone depleting substances, acting via in-situ gas phase activation of chlorine or through export of ozone depleted air masses from polar latitudes after the polar vortex breakup, also natural factors contribute, mainly through changes in stratospheric dynamics and temperature. In addition to gas-phase activation in the upper stratosphere, chlorine from CFCs may be activated in the mid-latitudes by heterogeneous reactions on sulfate aerosols (Solomon et al., 1996), similar to those on polar stratospheric clouds. Consequently, strong mid-latitude ozone losses were observed for several years in the wake of the major volcanic eruptions of El Chichón in 1981 (Adriani et al., 1987) and Pinatubo in 1991 (Gleason et al., 1993; Toohey, 1995, and references therein).

As a consequence of the Montreal Protocol, stratospheric concentrations of ozone depleting substances (ODSs) have peaked in the late 1990s (World Meteorological Organization, 2007) and since begun to slowly decrease. An onset of recovery of the ozone layer is thus expected to happen within the present years (Newman et al., 2006), and most climate models expect a return to 1980 conditions around the mid-21st century, at different timescales at the mid-latitudes and the poles (Eyring et al., 2007).

For a reliable prediction of the future evolution of the ozone layer, it is essential to analyse

past ozone changes in detail in order to interpret them as effects of chemistry or dynamics. This is of special importance for the question whether a trend reversal has already happened as a result of the turnaround in ODS concentrations. Some studies have reported the first stage of recovery, a positive change in trend, for Antarctic springtime column ozone (Newchurch et al., 2003), as well as mid-latitude upper stratospheric ozone (Yang et al., 2008) and mid- to high latitude column ozone (Reinsel et al., 2005). On the other hand, several studies on observed ozone trends have shown that increases in NH total ozone since the middle 1990s were mainly driven by changes in transport and dynamics and to a lesser extent by changes in ozone depleting substances (Dhomse et al., 2006; Wohltmann et al., 2007; Harris et al., 2008). Only few modelling studies have so far explicitly addressed the effects of the reversal of chlorine trends. Hadjinicolaou et al. (2005) analysed modelled ozone trends in two periods (1979–1993 and 1994–2003) and showed that most of the apparent recovery trend in 1994–2003 is attributable to meteorology.

Together with the sensitivity simulations introduced in Section 6.3, the assimilated data set is a mighty tool for analysing and attributing global ozone changes since the late 1970s, in particular relating to the question of a possible turnaround at the end of the 20$^{\text{th}}$ century. Therefore, a trend analysis of the assimilated data set and sensitivity calculations is presented in this chapter. While the assimilated data set allows the extension of observed trends into data sparse areas, the sensitivity calculations can be used to diagnose the origin of trends present in satellite observations and the assimilated data set. As described in Section 6.3, polar chemistry and gas-phase chemistry are switched between the states "constant chemical composition" and "time-dependent chemical composition" in the sensitivity runs. Differences in ozone amounts and trends between the different runs thus allow us to distinguish the influences of anthropogenic influences – ODS changes acting through polar heterogeneous chemistry and its export, or gas phase chemistry – and natural influences which affect the ozone layer through changes in temperature and transport. The CTM sensitivity runs include most of the agents influencing ozone variability, while ignoring direct influence of the solar cycle and volcanic eruptions (indirectly, these effects are partly included through the external wind fields that drive the CTM). This simplicity of the chemistry scheme may initially be perceived as a disadvantage, but on the other hand it provides the possibility to diagnose solar and volcanic effects in the assimilated data set and satellite data by analysing ozone differences.

In the analysis, a linear regression model with piecewise linear trends is applied which allows for a trend change between the "increasing stratospheric halogen loading" phase (1979–1999) and the "decreasing stratospheric halogen loading" phase (2000–2007[1]). By this method, it is possible to explicitly address the question of how ozone (column and profile) trends are affected by the reversal of ODS trends, and whether observed changes in stratospheric ozone over this period are consistent with known changes in the chemical composition of the stratosphere and meteorology.

Key findings of this chapter have been published in the peer reviewed literature as: Kiesewetter, G., Sinnhuber, B.-M., Weber, M., Burrows, J. P.: Attribution of stratospheric ozone trends to chemistry and dynamics: a modelling study, Atmos. Chem. Phys., 10, 12073–12089, 2010, listed in the bibliography as Kiesewetter et al. (2010b). In Kiesewetter et al. (2010b), the trend study and attribution is presented using only the CTM sensitivity calculations, while it has been extended to include the assimilated data set here. The discussion of

[1] In some cases, the unconstrained CTM runs are used until 2009.

volcanic and solar cycle effects has been expanded. Also, Kiesewetter et al. (2010b) show only seasonal trends in case of TO3 trends, while also annual average TO3 trends are discussed in this chapter.

8.1. The long-term perspective: Global ozone during the past 32 years

Figure 8.1 shows the evolution of column ozone in the assimilated data set and reference run *tt* in the tropics and mid-latitudes over the whole period of model integrations. ERA-40 and ERA-Interim periods are drawn separately. For comparison, also column ozone from the TOMS/SBUV merged data set (MOD, Stolarski and Frith, 2006) and the GSG data set (Weber et al., 2007) is shown. Differences owing to meteorological data switching are mainly encountered in the CTM sensitivity simulations (note the shifted right-hand side axes), while assimilated ozone shows only moderate differences of around 5–10 DU (less than 4%) in the NH mid-latitudes. Long-term evolution of TO3 in both assimilated data set and the *tt* sensitivity run is largely consistent with the satellite record. A distinct TO3 minimum is encountered in the NH during mid-1990s, partly attributable to aerosol inputs from the Pinatubo eruption, and after an apparent recovery phase TO3 levels off after 2000. The ERA-Interim driven model runs do not fully capture the decadal-scale increase in TO3 around the turn of the century. Since the assimilated run shows a comparable behaviour, changes in observed ozone must originate in a vertical range where the satellite observations do not influence assimilated ozone strongly. This is discussed in Section 8.3.3 below.
In tropical ozone no long-term trend is visible. Absolute TO3 variability in the tropics is much lower than in the mid-latitudes or polar regions. Southern mid-latitudes show a strong decrease of column ozone values throughout the 1980s and 1990s, followed by a levelling off around the turn of the century.
Since the long-term evolution of column ozone is well represented in the CTM also without assimilation, the sensitivity runs may be used to distill different contributions to TO3 trends. This is done in Section 8.3.4.

8.2. Solar and aerosol effects

A significant part of the decadal-scale deviations between observations and the unconstrained CTM in Figure 8.1 is related to variations in solar irradiation, which are not directly accounted for in the CTM. Concurrent with variations in the magnetic field of the sun, solar irradiation varies with a period of ~ 11 years. Changes in solar activity lead to variations in column ozone in the order of $2-3\%$ (solar maximum – solar minimum, in phase with the solar cycle) due to an enhancement of odd-oxygen production via the Chapman cycle in the middle to upper stratosphere during high solar activity (Chipperfield, 2003; Soukharev and Hood, 2006). The 10.7 cm radio flux (also referred to as "F10.7") is often used as a proxy for solar activity (see Figure 8.2, left).
In addition to variations in solar irradiation, also variations in volcanic aerosols, which are not included in the Linoz scheme, contribute to differences between modelled and observed ozone. Solar and aerosol influences may be quantified by regressing the monthly deseasonalised time series of ΔTO3 := TOMS/SBUV $-$ *tt* (ERA-40 and ERA-Interim driven part concatenated) against a simple model,

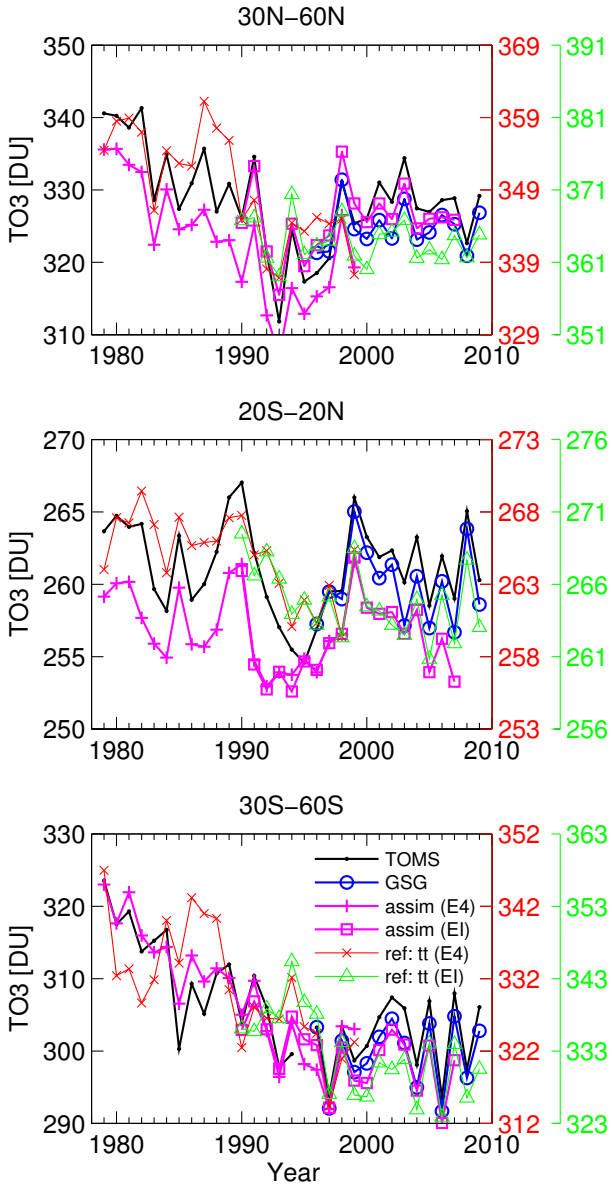

Figure 8.1.: *Total ozone (TO3) annual means from the assimilated data set (standard run), the tt (EESC-dependent chemistry) sensitivity run, the TOMS/SBUV merged TO3 data set ("TOMS") and the GSG data set. Shifted right-hand axes correspond to the tt run in ERA-40 (middle axis) and ERA-Interim periods. Assimilated ozone and TOMS/SBUV use the left axis. The climatological ozone column below 330 K has been added to modelled TO3. Plot adapted from Kiesewetter et al. (2010b).*

$$\Delta\mathrm{TO3}(t) = \mu + \omega_0\Theta(t - t_0) + \chi_1 F_{10.7}\,(t - \Delta t) + \chi_2 AOD(t) + N(t)\ , \tag{8.1}$$

which contains a constant offset μ, a step function $\Theta(t-t_0)$ at the transition of the meteorological data sets ($\Theta = 0$ for the ERA-40 period and $\Theta = 1$ for the ERA-Interim period), the

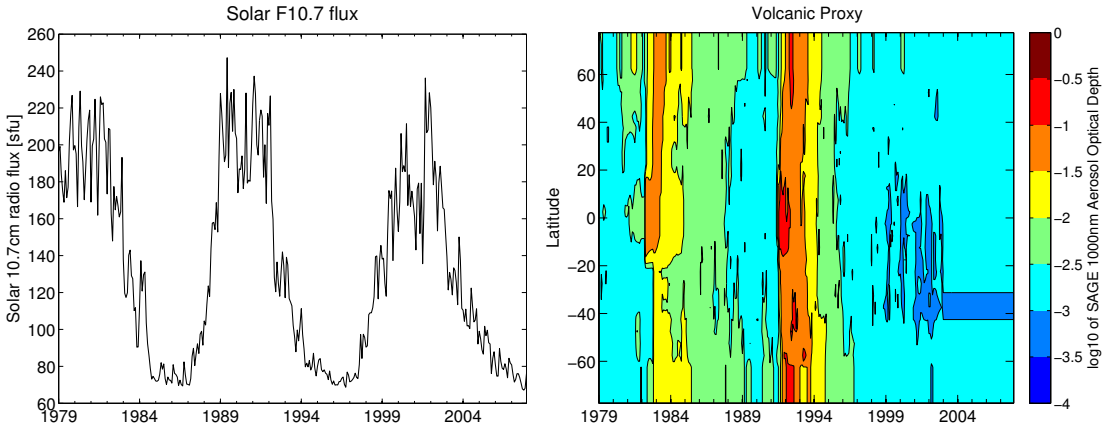

Figure 8.2.: *Solar (left) and aerosol (right) proxies used in the multiple linear regression model. Left: Solar spectral irradiance at 10.7 cm ("F10.7 radio flux"), which is commonly used as a proxy for solar activity. Units are solar flux units;* $1\,\mathrm{sfu} = 10^{-22}\mathrm{Wm}^{-2}\mathrm{Hz}^{-1}$*. Right: Logarithm of aerosol optical depth as observed from the Stratospheric Aerosol and Gas Experiment (SAGE) satellite instrument series. Data set obtained from the SPARC data centre. As the SAGE record ends in 2002, the record has been extended at constant levels (2002 mean) to cover the full time span of the assimilation. The eruptions of El Chichón (1981) and Pinatubo (1991) are visible as major signals.*

solar 10.7 cm radio flux time series ($F_{10.7}$), and the 1000 nm aerosol optical depth $AOD(t)$ (taken from the SAGE data set provided through the SPARC data centre, see Thomason and Peter, 2006, and extended at constant levels after 2002). $N(t)$ represents the unexplained noise. In the regression analysis it became obvious that TO3 responses to aerosol inputs decay more slowly than the AOD time series. Since it is not clear why the ozone response should depend linearly on the AOD, and since no exact quantification of aerosol chemistry is attempted here, as a practical compromise the logarithm of the AOD is used as a regression function. Basic features are then similar to the AOD itself, but the decay after El Chichón and Pinatubo eruptions is slower. Solar and aerosol proxies used in the regression model are shown in Figure 8.2. The solar proxy is used with a latitude-dependent time shift that is determined by maximum correlation of fit residuals without $F_{10.7}$ to the $F_{10.7}$ flux. This correlation is shown in Figure 8.3 as a function of latitude and time delay between ΔTO3 and F10.7. A distinct pattern of high correlation in phase with the solar cycle is visible, which maximises near the equator at little or no delay (correlation coefficient > 0.6), and is delayed by a few months and up to one year in the mid-latitudes. Correlation coefficients are higher than 0.4 up to 50°N and 60°S. A linear fit of the delay for which the correlation maximises is shown as black dashed line, calculated between 0 and 50° latitude at each hemisphere and extrapolated polewards from there. These delays are used as the time shift in (8.1).

Results from the regression of ΔTO3 (Equation 8.1) are presented in Figure 8.4, which shows the ΔTO3 time series (black line), along with the regressed $\widehat{\Delta\mathrm{TO3}}$ obtained from Equation 8.1 (red dashed line). In addition, Figure 8.4 shows $\widehat{\Delta\mathrm{TO3}}$ obtained from regression without a solar term (blue line), as well as without the aerosol proxy (green line). The full regression model reproduces the TO3 observed offsets well, particularly in the tropics. Best

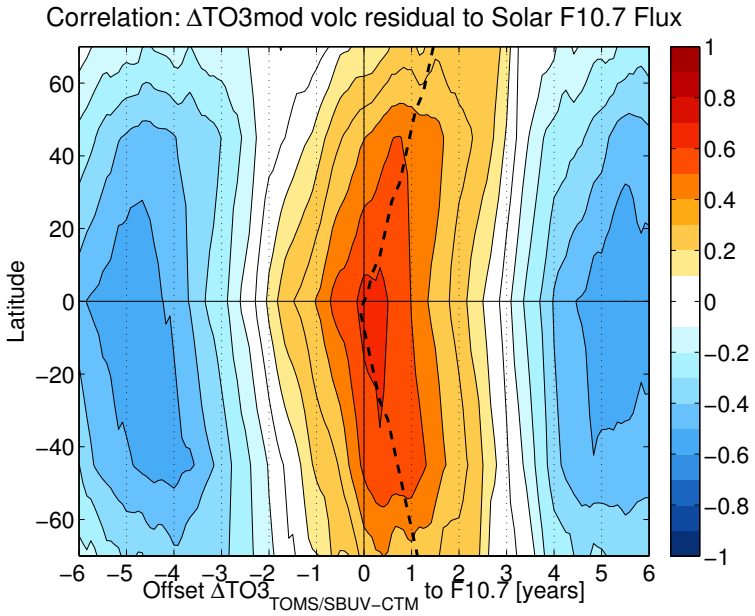

Figure 8.3.: *Correlation between the residuals of the "aerosol only" regression of* TOMS/SBUV − *tt offsets and the Solar F10.7 radio flux, for different time delays between* ΔTO3 *and F10.7. A linear fit to the delays for which the correlation maximises is shown as black dashed line (calculated between 0 and 50° latitude in each hemisphere and extrapolated from there). The regressed time series is shown in Figure 8.4. Plot modified from Kiesewetter et al. (2010b).*

agreement is observed when the full regression with both aerosol and solar terms is used. In the tropics, residuals are considerably smaller in the reduced model without F10.7 flux as compared to the reduced model without aerosols, pointing to the possibly larger relative importance of volcanic aerosols.

The same regression analysis presented here for differences between reference run *tt* and TOMS/SBUV observations can also be applied to differences between *tt* and the assimilated data set in order to diagnose possible solar and volcanic signals in the assimilated data set, and to generate a solar signal free version of it. Since SBUV satellite observations are assimilated, which are per definition quite consistent with the TOMS/SBUV merged TO3 data set, one could assume that the solar and aerosol terms in the assimilated data set are largely identical to those in the TOMS/SBUV MOD. However, variations in solar irradiation affect predominantly the middle to upper stratosphere where the ozone lifetime is short, ozone relaxes against its predicted photochemical equilibrium state fast and the information contained in the satellite data becomes gradually less important. Indeed, the correlation shown in Figure 8.3 becomes weaker and more blurred for assimilated ozone, and in the tropics it decreases to almost zero. Nevertheless, to be consistent, the assimilated time series with solar and aerosol signal removed is used in the trend analysis presented below.

The regression coefficients from Equation 8.1 can be used to quantify the absolute magnitude of solar and volcanic signals in observations or the assimilated data set:

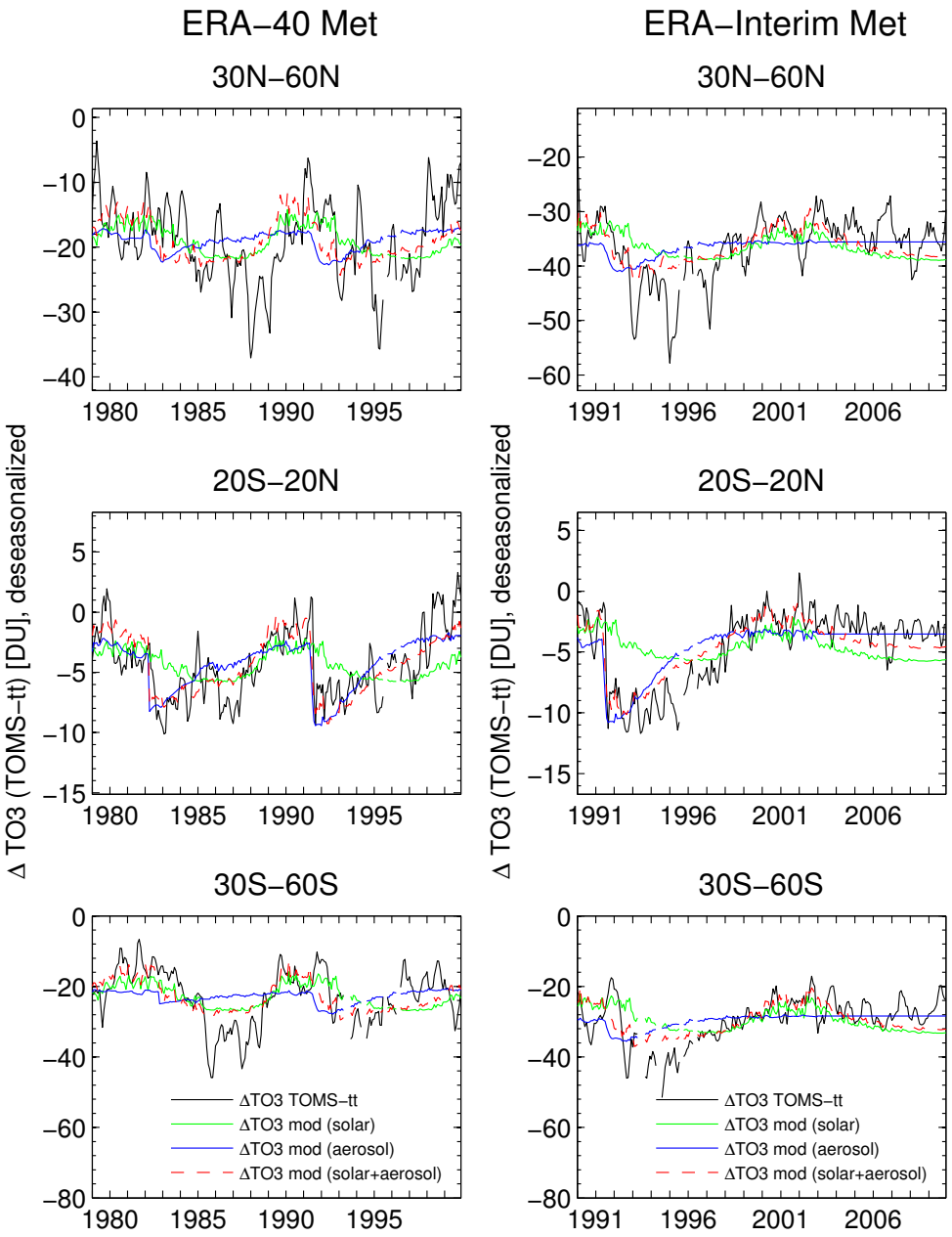

Figure 8.4.: *Regression analysis of TO3 deviation (TOMS/SBUV observations – CTM run tt) to solar F10.7 radio flux and SAGE aerosol optical depth. Left: ERA-40 driven model run (1979–1999), right: ERA-Interim driven model run (1990–2009). The actual deseasonalised TO3 offset between observations and model (black line) is shown along with the best fit including the solar proxy only (green line), the aerosol proxy only (blue line), and both (red dashed line). From Kiesewetter et al. (2010b).*

$$\begin{aligned} \Delta\mathrm{TO3}_{solar} &= \chi_1(\Delta\mathrm{TO3}_{solmax} - \Delta\mathrm{TO3}_{solmin}) \\ \Delta\mathrm{TO3}_{volc} &= \chi_2(\Delta\mathrm{TO3}_{Pinatubo} - \Delta\mathrm{TO3}_{background}), \end{aligned} \tag{8.2}$$

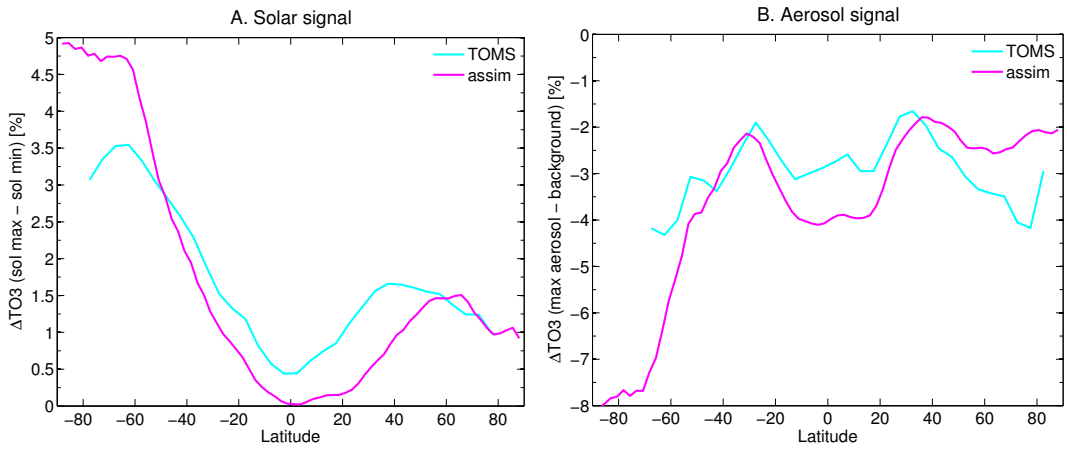

Figure 8.5.: *Solar and aerosol signal in the TOMS/SBUV MOD and in assimilated TO3, as obtained from the regression of* TOMS $- tt$ *and* assim $- tt$*. Left: TO3 difference between solar minimum and solar maximum, expressed as fraction of the average TO3 field. Right: TO3 difference between the 12 months of highest aerosol loading and clean background conditions (2007), expressed as fraction of 2007 TO3.*

where $\Delta\text{TO3}_{solmax}$, $\Delta\text{TO3}_{solmin}$, $\Delta\text{TO3}_{Pinatubo}$, and $\Delta\text{TO3}_{background}$ indicate TO3 differences between TOMS/SBUV $- tt$ and assim $- tt$ at times of solar maximum, solar minimum, shortly after the Mt. Pinatubo eruption, and at clean aerosol background conditions. χ_1 and χ_2 are the regression coefficients introduced in Equation 8.1. The resulting TO3 signals are shown in Figure 8.5 in dependence of latitude (A: solar cycle signal, B: Pinatubo signal). Outside high southern latitudes, the latitudinal structure of both solar and volcanic signals is quite consistent in the TOMS/SBUV MOD and the assimilated data set. The solar signal is somewhat diminished in the assimilated data set in the tropics. Although the tropics receive most intense solar irradiation, the solar signal is more intense in the mid-latitudes, where it reaches 3% ($6 - 8$ DU) in the SH and 1.5% (~ 4 DU) in the NH, which is in the order of the expected magnitude (Soukharev and Hood, 2006). The obvious zonal asymmetry may be related to inter-hemispheric differences in stratospheric dynamics, which lead to different mixing patterns. In the Antarctic, the solar signal reaches values of up to 5% in the assimilated data set, which is unexpected. At the same time, the volcanic signal of -8% in the Antarctic is indeed consistent with estimates by Randel et al. (1995) and Rosenfield et al. (1997) obtained from observations and model studies. Since a correspondingly large signal is not present in Arctic ozone, however, both solar and aerosol signals may be over-estimated in the assimilated data set at high southern latitudes, partly cancelling each other. In tropical and mid-latitudes, both observed and assimilated ozone show a column ozone decline of $3 - 4$% compared to undisturbed conditions as a result of the Pinatubo eruption, in excellent agreement to values obtained by Randel et al. (1995) and Rosenfield et al. (1997).

Again, it should be emphasised that the regression analysis based on TO3 differences can only diagnose ozone changes directly related to changed chemistry and misses indirect effects which are present in the meteorological fields used.

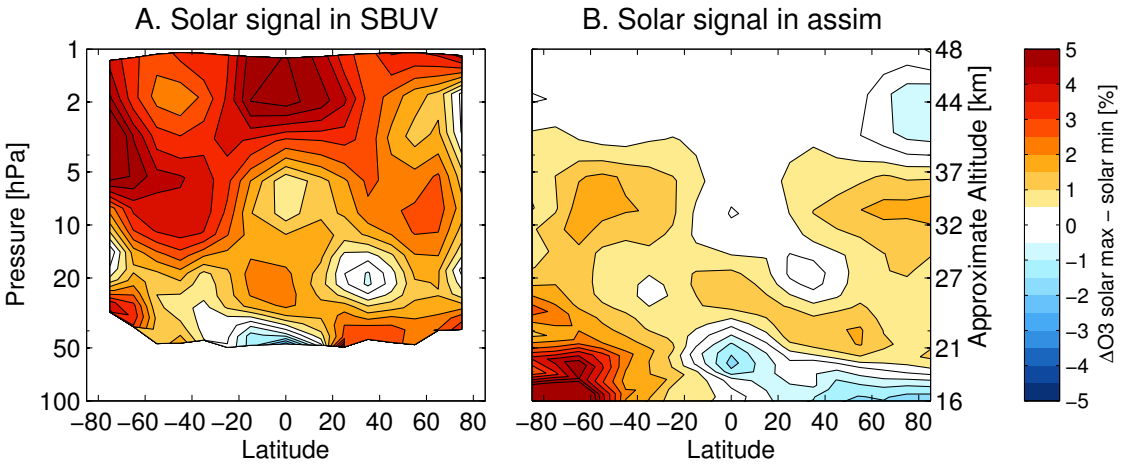

Figure 8.6.: *Vertical distribution of the solar signal in the SBUV merged data set (A) and the assimilated data set (B). Contour levels correspond to differences in regressed solar-related ozone vmr offset (SBUV-tt or assim-tt), taken between solar maximum and solar minimum and expressed as fractions of the mean ozone vmr.*

Applying Equation 8.1 accordingly to profile ozone in every model level, we can diagnose the vertical distribution of solar and aerosol contributions to differences between observations and CTM sensitivity runs. In this case, Equation 8.1 reads

$$\Delta\mathrm{O3}(t) = \mu + \omega_0\Theta(t - t_0) + \chi_1 F_{10.7}(t) + \chi_2\, SAD(t) + N(t)\ , \tag{8.3}$$

where $SAD(t)$ represents the aerosol surface area density at this pressure, taken from the SAGE data set (Thomason and Peter, 2006).

Figure 8.6 shows the vertical distribution of the relative solar signal in SBUV and assimilated ozone, obtained as amplitude of regressed solar-related ozone vmr differences between solar minimum and solar maximum. Although it was shown above that the solar cycle effect in TO3 appears delayed by a few months in the mid-latitudes, for simplicity no delay of the solar flux time series is used in the regression of profile ozone discussed here. In principle, the delay should be latitude- and height-dependent. In the upper stratosphere, the ozone lifetime is short, and a possible ozone signal must follow the solar flux immediately, while in the lower stratosphere (which is dominant for TO3) ozone lifetime is long enough to allow for delays due to transport of solar-induced ozone anomalies from the tropics to higher latitudes. Variability of profile ozone on single levels and latitude bins is too high to allow for a reasonable detection of optimal delay from a mere correlation analysis – as it was done for column ozone –, and therefore no delay is used. In order to be consistent with the original SBUV data set, the vertical axes in Figure 8.6 are in units of pressure. In the SBUV MOD (left), a strong solar cycle effect is visible throughout the middle to upper stratosphere. As expected, ozone changes between solar minimum and solar maximum are positive (higher solar activity corresponds to stronger UV irradiation and more ozone production).

The SBUV MOD shows a prominent solar signal up to 5% in the upper stratosphere. Remarkably, the signal is not diminished in the extra-tropics but rather appears intensified

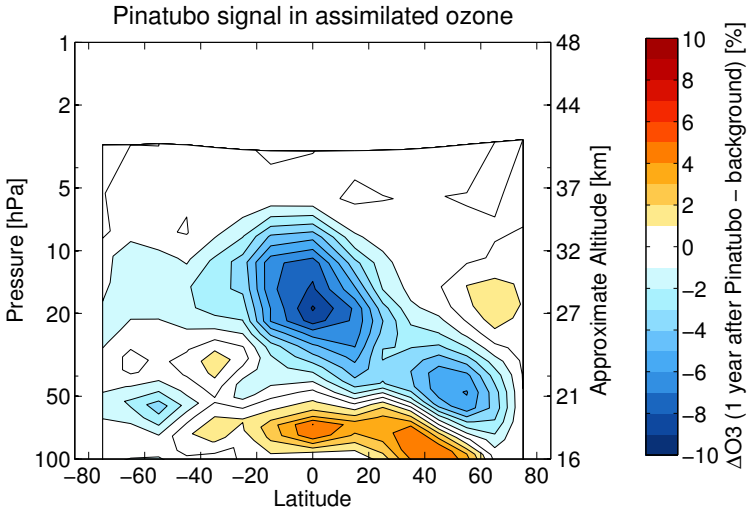

Figure 8.7.: *Vertical-zonal distribution of the amplitude of the Pinatubo signal in assimilated ozone. Contour levels correspond to differences in regressed aerosol-related ozone vmr offsets (assim-tt) taken between the 12 months after the Pinatubo eruption and clean background conditions in 2007, expressed as fractions of 2007 ozone vmr.*

around 10 hPa / 32 km at high latitudes of both hemispheres. The solar cycle signal in assimilated ozone (Figure 8.6, right) is similar but considerably weaker; nearly all signal is lost in the upper stratosphere. This is expected, as the assimilation does not work optimally at high altitudes due to the short lifetime of ozone. An unexpectedly strong solar signal seems present in the lower stratosphere in the Antarctic in the assimilated data set.

Qualitatively, the solar cycle effect in the middle stratosphere is in good agreement with findings of Sinnhuber et al. (2006), who detected a large solar signal in mid-stratospheric polar ozone in both hemispheres from an analysis of sonde data and SBUV satellite observations. Vertical regions of peak solar signal in polar latitudes in Figure 8.6 correspond well to those analysed by Sinnhuber et al. (2006). They related the presence of a solar cycle in mid-stratospheric polar ozone to influences of high energetic electron precipitation. Effects of precipitating electrons on polar ozone have been described by several studies, e.g. Callis et al. (1991), Gaines et al. (1995), Kazeminejad (2009), Sinnhuber et al. (2010). Similar to the solar proton events analysed in Section 7.3, precipitation of energetic electrons leads to the formation of NO_x species in the polar mesosphere, which are then transported downward during winter and lead to enhanced ozone destruction in the stratosphere through the catalytic NO_x cycle. Since the emission of energetic particles from the sun is modulated by solar activity, this mechanism may present an explanation for the presence of an 11 year solar cycle signal in polar ozone other than just export of additional ozone from tropical latitudes. However, a more stringent test of this mechanism is beyond the scope of this study.

In the same way as for the solar cycle, the vertical distribution of volcanic aerosol effects can be analysed, in particular regarding the catastrophic eruption of Mt. Pinatubo in June 1991. A reasonable analysis is only possible for assimilated ozone, as for the critical months after the Pinatubo (and also El Chichón) eruption, large gaps exist in the merged SBUV data

set (whereas the single profile measurements which are used for assimilation are available). Figure 8.7 shows the magnitude of ozone vmr decreases owing to the Pinatubo eruption, diagnosed as differences of regressed aerosol-related ozone vmr offset (assim − tt) between the first 12 months after the Pinatubo eruption and clean background conditions in 2007. Strong decreases in ozone are observed particularly in the tropical middle stratosphere (up to −10%) and in the NH mid-latitudes (−5%), while effects are smaller in the SH. Interestingly, a positive signal is observed below the strong decreases, pointing to some sort of indirect effects, which may act either via dynamical changes, or demonstrate a self-healing effect due to increased solar UV radiation as a result of the ozone decreases in higher altitudes.
Due to the mentioned coincidence of two large volcanic eruptions at the decreasing phase of the solar cycle, the attribution of ozone changes to solar and volcanic effects is not unambiguous. Parts of the very large Pinatubo signal in tropical assimilated ozone might also be a misinterpretation and make up for the decreased strength of the solar signal in tropical assimilated ozone.

8.3. Trend Analysis

After discussing the solar and volcanic contributions to global ozone variability from an analysis of offsets between CTM sensitivity runs and observations (or the assimilated data set), I now proceed to an analysis of trends in the assimilated data set and sensitivity runs. After a mathematical introduction to the regression analysis in Section 8.3.1, the regression model is applied to column ozone in Section 8.3.2, and to profile ozone in Section 8.3.3. An attribution of column ozone trends to contributing factors is conducted in Section 8.3.4 from an analysis of differences between individual CTM sensitivity runs. Finally, in Section 8.3.5, the question is addressed whether a significant positive change in trends is detected.

8.3.1. Methodology

The ozone trend analysis presented here follows largely the method of piecewise linear trends as described by Reinsel et al. (2002). In this method, linear trend functions are explicitly included in the design matrix of the least squares regression model. Several studies during recent years have used a different approach which does not include linear trend functions but rather diagnoses the trend from a fit to EESC. While this EESC-fitting method has its advantages, it is not followed here for several reasons: Firstly, the linearised ozone chemistry assumes a piecewise linear source gas evolution (see Section 3.2), and hence a piecewise linear trend regression seems ideal for the model setup chosen. It is not clear which EESC curve should be taken as a regressor, as the shape of the curve should differ for different latitudes corresponding to different age of air. Furthermore, the same regression model is applied to time series of different sensitivity runs with changing or constant chemical conditions, in order to derive e.g. trends driven by dynamical variability. There is no reason to assume that these should follow an EESC curve.
The whole time series of modelled ozone is analysed in one piece, concatenating the ozone time series obtained with ERA-40 and ERA-Interim data. Let $Y(t)$ be a time series of monthly mean ozone, t running from 1 to 42 (21 years ERA-40 driven CTM output plus 21 years ERA-Interim driven CTM output). Y may either be continuous in time in order to derive *overall* trends, or a series of monthly mean ozone values with spacing of one year in

between, the latter being used for the calculation of *seasonal* trends (from seasonal averages of monthly trends). Then the piecewise linear regression model used is of the form

$$Y(t) = \mu + \omega_0 \Theta(t - t_0) + \omega_1 X_1(t) + \omega_2 X_2(t) + N(t), \tag{8.4}$$

where μ is a constant (mean level), Θ is a Heaviside step function that accounts for the offset ω_0 induced by different meteorological fields ($\Theta = 0$ for the ERA-40 period and $\Theta = 1$ for the ERA-Interim period), $X_1(t)$ is a linear trend function, and $X_2(t)$ represents an additional linear trend function ("change in trend") that is zero before Jan 2000 and increases linearly afterwards.
The exact values of trend functions $X_1(t)$ and $X_2(t)$ must, of course, be adapted according to whether seasonal or annual trends are calculated. The trend function $X_1(t)$ is replicated during the overlap period of the different meteorological fields. Hence, the same linear trend is used for both ERA-40 and ERA-Interim driven fields, while allowing for an offset due to the changed meteorology.
$N(t)$ represents the unexplained noise term. In case of annually spaced Y, $N(t)$ is not auto-correlated, simplifying the calculation of uncertainties of the fit coefficients. Equation 8.4 may be rewritten in matrix form as

$$\mathbf{Y} = \mathbf{A} \cdot \omega + \mathbf{N}, \tag{8.5}$$

where $\mathbf{A}$ is the fit matrix, and $\mathbf{Y}$, ω and $\mathbf{N}$ denote the vectors of ozone time series, fit coefficients, and residual noise. Inversion of Equation 8.5 yields the generalised least squares estimator $\hat{\omega}$ for the fit coefficients ω. For a non-autocorrelated noise N, the standard error of the fit coefficients, $\sigma_{\hat{\omega}}$, is then derived from the standard error of the fit residuals, $\sigma_{\mathbf{N}} = (\mathrm{Var}(\mathbf{N}))^{1/2}$, as

$$\sigma_{\hat{\omega}}{}^2 = \left(\mathbf{A}^T \mathbf{A}\right)^{-1} \cdot \sigma_{\mathbf{N}}{}^2. \tag{8.6}$$

Standard errors of the different fit coefficients are obtained as the diagonal elements of $\sigma_{\hat{\omega}}$, whereas the standard error of the combined trend $\hat{\omega}_{12} := \hat{\omega}_1 + \hat{\omega}_2$ is given by

$$\sigma_{\hat{\omega}_{12}} = \left[\mathrm{Var}(\hat{\omega}_1) + \mathrm{Var}(\hat{\omega}_2) + 2\mathrm{Cov}\left(\hat{\omega}_1, \hat{\omega}_2\right)\right]^{1/2}. \tag{8.7}$$

Equations 8.6 and 8.7 provide a correct quantification of standard errors only if the regressed time series is not auto-correlated, such as a series of yearly ozone values for a particular month, and is thus usable for the error analysis of seasonal trends. If the trend model (8.4) should be applied to the whole deseasonalised ozone time series, to derive overall trends without distinguishing seasonal differences, the autocorrelation of the time series has to be taken into account. Following Reinsel et al. (2002), a 1 month autocorrelation model is assumed, which is indeed a common assumption for ozone trend regression. $\sigma_{\hat{\omega}_1}$ and $\sigma_{\hat{\omega}_{12}}$ are then calculated as

$$\sigma_{\hat{\omega}_1} \approx \frac{\sigma_N}{n^{3/2}} \sqrt{\frac{1+\phi}{1-\phi}}, \tag{8.8}$$

$$\sigma_{\hat{\omega}_2} \approx \frac{\sigma_N}{2} \sqrt{\frac{1+\phi}{1-\phi}} \left(\frac{n}{n_0 n_1}\right)^{3/2}, \tag{8.9}$$

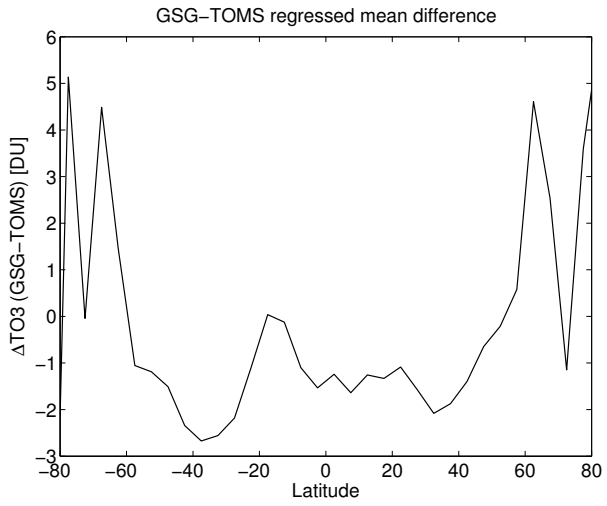

Figure 8.8.: *Offset between TOMS/SBUV and GSG TO3 data sets, as obtained from the TO3 trend regression analysis.*

and

$$\sigma_{\hat{\omega}_{12}} \approx \frac{\sigma_N}{n_1^{3/2}} \sqrt{\frac{1+\phi}{1-\phi}} \sqrt{\frac{n_0 + 4n_1}{4n}}, \tag{8.10}$$

where n_0 is the number of years of data prior to the change in trend, n_1 is the number of years from the trend change, and $n = n_0 + n_1$. ϕ is the autocorrelation of the residuals $N(t)$ with a time lag of 1 month,

$$\phi = \mathrm{Corr}\left[N(t), N(t-1)\right]. \tag{8.11}$$

The trend regression of the whole time series including the auto-correlation is applied to column ozone in Section 8.3.2 and to profile ozone in Section 8.3.3.

8.3.2. Column ozone trends

In this section, trends of the column ozone time series are analysed as described in Section 8.3.1. The observational TO3 time series used here is a concatenation of TOMS/SBUV and GSG data sets with volcanic and solar signal removed as described in Section 8.2. TOMS/SBUV and GSG data sets are regressed here in one piece, in a similar fashion as the ERA-40 and ERA-Interim driven parts of the model runs, to arrive at a single trend estimate for "TOMS/GSG". While the same trend regression functions are used during the period of overlap, an offset between the two satellite time series is allowed, which is less than $\sim 2\,\mathrm{DU}$ outside polar areas. The regressed offset between TOMS/SBUV and GSG is shown in Figure 8.8.
Here the piecewise linear trend model with 1 month autocorrelation is applied to the complete time series of assimilated TO3 (including overlaps of different meteorological data sets), TOMS/SBUV and GSG observations, and the CTM sensitivity runs without assimilation.

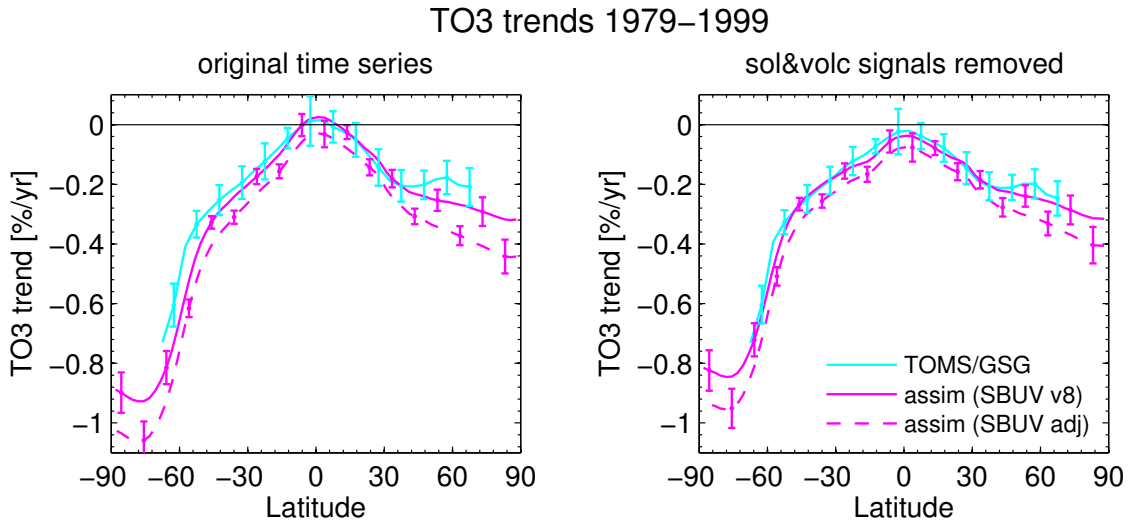

Figure 8.9.: *TO3 trends in the assimilated data set and the TOMS/SBUV MOD for the period 1979–1999, as obtained from the piecewise linear trend regression model. Both versions of the assimilated data set are shown, using either the unmodified SBUV v8 profiles (solid purple line), or with SBUV adjustment coefficients applied (dashed purple line). Left panel: unmodified time series, right: time series with solar cycle and aerosol signals removed.*

Figure 8.9 shows TO3 trends of assimilated ozone and TOMS/GSG observations in the period 1979–1999, expressed as percent of 1980 values per year. Here, both versions of the assimilated data set are shown: In addition to the standard data set (solid magenta line), also the data set using SBUV adjustment coefficients as described in Section 5.2 is displayed (dashed line). While the unmodified time series have been used in the left panel, the right panel shows trends for the time series after removal of regressed solar and volcanic signals as described in Section 8.2.

Column ozone trends in the standard assimilated data set during this period are in good agreement to satellite observations, while using SBUV adjustment coefficients leads to too strong column ozone decreases in both hemispheres. This effect is mainly attributable to rather large negative adjustment coefficients in the lower stratosphere for the NOAA-11 satellite in the period 1997–2000, thus severely influencing 1979–1999 trend calculations. For the unmodified time series, also the standard assimilated data set shows a small negative bias against TOMS/GSG trends, which vanishes when solar and volcanic signals are removed as described in Section 8.2 (right panel). Strong negative TO3 trends are present in the extra-tropics, peaking around -0.9%/yr (-2.5 DU/yr) at high southern latitudes and around -0.3%/yr (-1 DU/yr) in the Arctic. Around the equator, trends are zero or at least not significantly negative. Figure 8.9 shows that using SBUV profile ozone adjustment coefficients in fact decreases agreement of column ozone trends to observations (while the agreement to profile ozone trends is slightly improved in the vertical range of SBUV observations, see Section 8.3.3 below). This seemingly paradox effect can be explained by the downward transport of low ozone mixing ratios (from negative adjustment coefficients in lower SBUV levels) beyond the lower limit of SBUV profile coverage, where they then

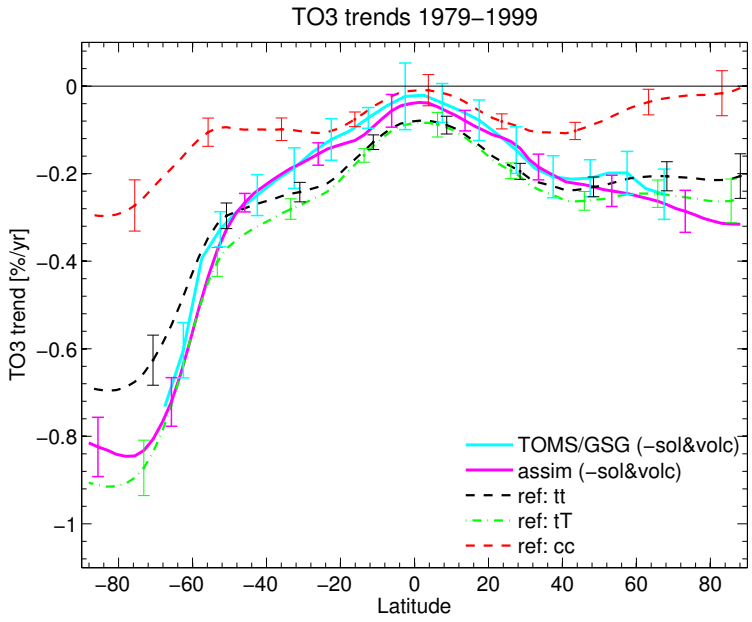

Figure 8.10.: *As Figure 8.9 (right), but including the tt, tT, and cc sensitivity runs without assimilation.*

strongly influence column ozone abundance. Consequently, either the NOAA-11 adjustment coefficients in the lower stratosphere are problematic, or vertical transport in the CTM has deficiencies. Clarification of this issue is beyond the scope of this study. Since the long-term evolution of TO3 in the standard assimilated data set is in excellent agreement to observations, this version is used if not otherwise indicated.

The CTM sensitivity runs without assimilation reproduce the trend pattern observed in satellite data and assimilated data set well. Figure 8.10 shows TO3 trends for the 1979-1999 period as before, but including the *tt* and *tT* sensitivity runs. The *tt* run with linear scaling of polar heterogeneous ozone depletion with EESC slightly under-estimates the magnitude of the negative trend at high latitudes of both hemispheres, while the *tT* run with $r_{pc} \propto \mathrm{EESC}^2$ slightly over-estimates it at least in the Antarctic. As argued above, the true proportionality of the polar ozone decay rate to EESC is between 1 and 2, and thus *tt* and *tT* are assumed to represent the plausible range of effects of polar chemistry changes due to increased chlorine loading. Both runs are analysed in parallel in the analysis below.

The analysis may be extended to discriminate seasonal differences in ozone changes. For example, ozone changes induced by changing polar chemistry exhibit a pronounced seasonality, while there is no reason to expect a strong seasonal cycle in gas phase chemistry (except that during polar night ozone lifetime is long). Figure 8.11 shows the linear trend $\hat{\omega}_1$ of the EESC-varying model runs (*tt*, *tT*), expressed here as percent change of 1980 values per year, along with its standard error. For comparison, the *cc* trend is shown, which represents the effects of changing meteorology. In most latitudes and seasons, TO3 trends observed in model run *tt* are in excellent agreement with trends in TOMS/GSG. In the seasons when polar ozone depletion is relevant, TOMS/GSG shows slightly stronger TO3 losses, indicating

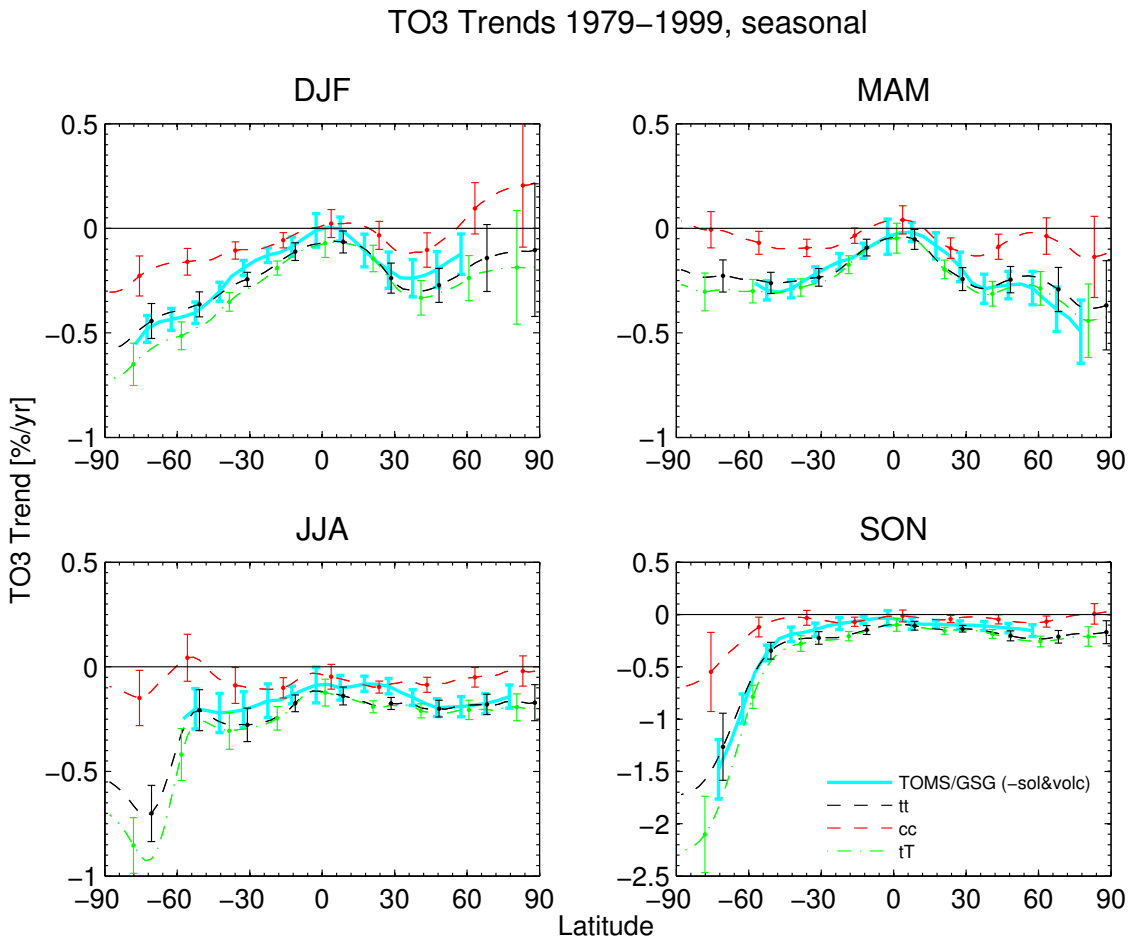

Figure 8.11.: *Seasonal TO3 trends for the period 1979–1999 in the CTM sensitivity runs, as compared to the combined TOMS/SBUV + GOME/SCIAMACHY/GOME2 TO3 trends ("TOMS/GSG"), obtained from the time series with solar and aerosol signals removed. Trends from model runs with EESC-dependent chemistry and polar chemistry (tt, black line, and tT, green dash-dotted line) and constant chemical conditions (cc, red line) are shown. In tT, the polar ozone destruction rate is scaled* $\propto \mathrm{EESC}^2$ *while in tt, it is scaled* $\propto$ EESC. *Trends are given in percent of 1980 values per year. Error bars represent the* 1σ *variance as obtained from the residuals of the linear fit. Plot updated from Kiesewetter et al. (2010b).*

that the linear scaling of the ozone destruction rate with EESC in *tt* may underestimate the real proportionality, while it is over-estimated in the *tT* run. It is remarkable that also run *cc* shows a distinct downward trend in TO3 in the SH polar spring (almost 50% of the *tt* trend), which is a clear indication of a feedback of ozone depletion on the stratospheric circulation as discussed before by, e.g., Randel and Wu (1999).

Figure 8.12 shows TO3 trends relative to 1980 values for the post peak-EESC 2000–2007 period. The values shown here are total trends $\hat{\omega}_{12} = \hat{\omega}_1 + \hat{\omega}_2$. In the TOMS/GSG satellite observations, positive trends are present in mid-latitudes of both hemispheres up to around +0.3%/yr; however, no trends are significant at more than 1σ. Both assimilated ozone and

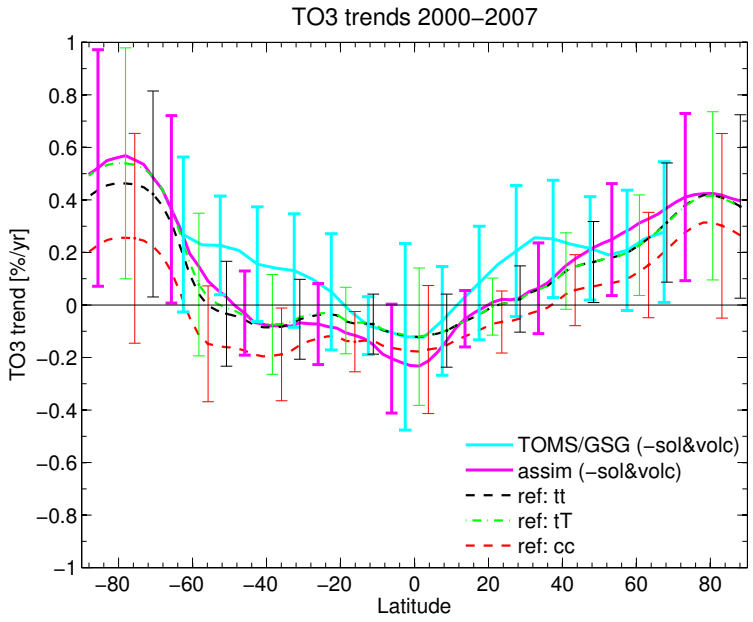

Figure 8.12.: *As Figure 8.10, but for the post peak-EESC period 2000–2007. TO3 trends in the assimilated data set, the TOMS/GSG merged data sets, and CTM sensitivity runs, as obtained from the piecewise linear trend regression model for the whole time series.*

CTM reference runs are not able to reproduce the magnitude of positive trends in the mid-latitudes, while they agree well with the satellite data around the equator and at high latitudes. The tendency of observed TO3 towards more positive trends is related to very strong increases in observed upper stratospheric ozone that are not fully reproduced by the Linoz chemistry scheme (see Section 8.3.3), and thus are not transferred into the assimilated data set which relies more on the CTM chemistry than on observations in the upper stratosphere due to the short chemical lifetime of ozone. However, these large ozone increases in observations are possibly in parts related to instrument inter-calibration issues. The increase seen in the TOMS/SBUV merged ozone data set alone is larger than that in the GSG record, and in particular in the latitudinal region where the largest trend offsets between observations and CTM are present, in the northern an southern mid-latitudes, there is also disagreement between TOMS/SBUV and GSG column ozone, as shown in Figure 8.13. While the offset of GSG to modelled TO3 remains largely constant or shows only a small trend, the offset between TOMS/SBUV and modelled TO3 displays a relatively sudden increase around the turn of the year 2001 in NH mid-latitudes. The TOMS/SBUV merged data set is compiled from TO3 records of different satellites, with the turn of the year 2000/2001 marking the transition between NOAA-11 and NOAA-16 satellites (see Chapter 5). Although the different instruments have been inter-calibrated as far as possible from the record overlap, one can speculate that the sudden increase in TOMS/SBUV offset, which accounts for a significant part of the trend offset in mid-latitudes, might be related to difficulties in instrument inter-calibration.

Given the similarity of modelled trends with and without changing EESC, it is obvious that

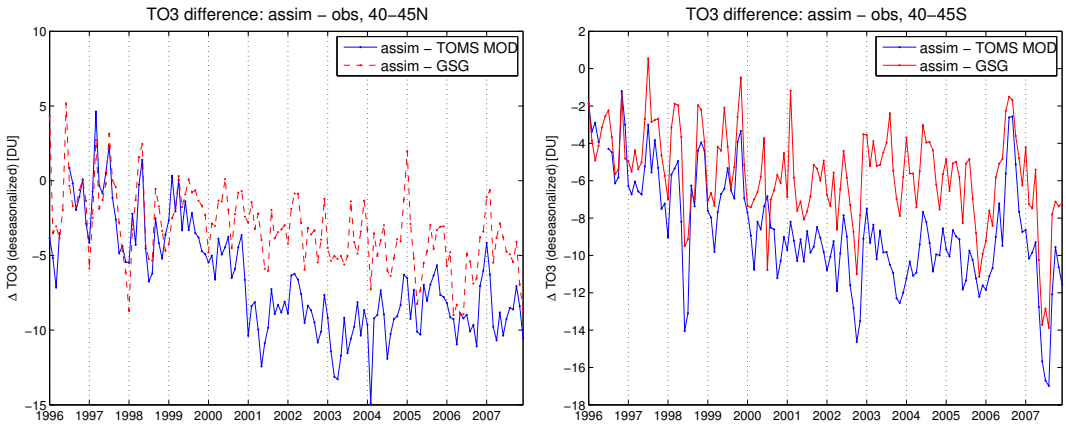

Figure 8.13.: *Offsets of TOMS/SBUV and GOME/SCIAMACHY/GOME2 to assimilated TO3, in the northern (left) and southern (right) mid-latitudes (40°).*

most of the trends during this period are a result of changing meteorological conditions rather than changing chemical composition. This is expected, as the decrease of EESC during the period 2000–2007 is only around 10% of previous increases, and even less at high latitudes. Nonetheless, it is instructive to test which seasons the trends originate at – if any in particular – and whether positive effects of ODS changes are seen in polar spring.
A decomposition into seasonal trends (here for the full period 2000–2009, hence without assimilated TO3) is shown in Figure 8.14. Trends are generally close to zero, with the notable exception of relatively strong positive trends in the polar areas during winter and spring. However, none of the observed seasonal trends is significant at more than $\sim$ 1 standard deviation. Trend patterns from *tt* and *cc* runs show only small differences and are both in good agreement with observations; *tT* is almost indistinguishable from *tt* except for Antarctic spring. Obviously, the positive trends seen in Figure 8.12 stem mainly from the winter and spring seasons at polar latitudes of both hemispheres. Although this might point to possible recoveries in polar ozone chemistry, from the small differences between *tt* and *cc* it is clear that in the Arctic, the positive TO3 trends are almost entirely related to meteorological changes. Only in the Antarctic is it possible to distinguish a clear difference in springtime TO3 evolution between *tt*/*tT* and *cc*, and hence a direct effect of decreasing ODS concentrations.
While the piecewise linear regression model applied here to the entire time span allows for useful statements about whether changes in long-term trends have occurred, it is not necessarily the most appropriate description of the TO3 evolution itself during the second analysis period (2000–2009). In several latitude regions, the TO3 time series rather follows a steep ascent during the late 1990s and then levels off during the 2000–2009 decade (see Figure 8.1).

8.3.3. Profile ozone trends

This section provides a short analysis of ozone profile trends to investigate to what extent the CTM and the assimilated data set reproduce observed vertical distributions of ozone

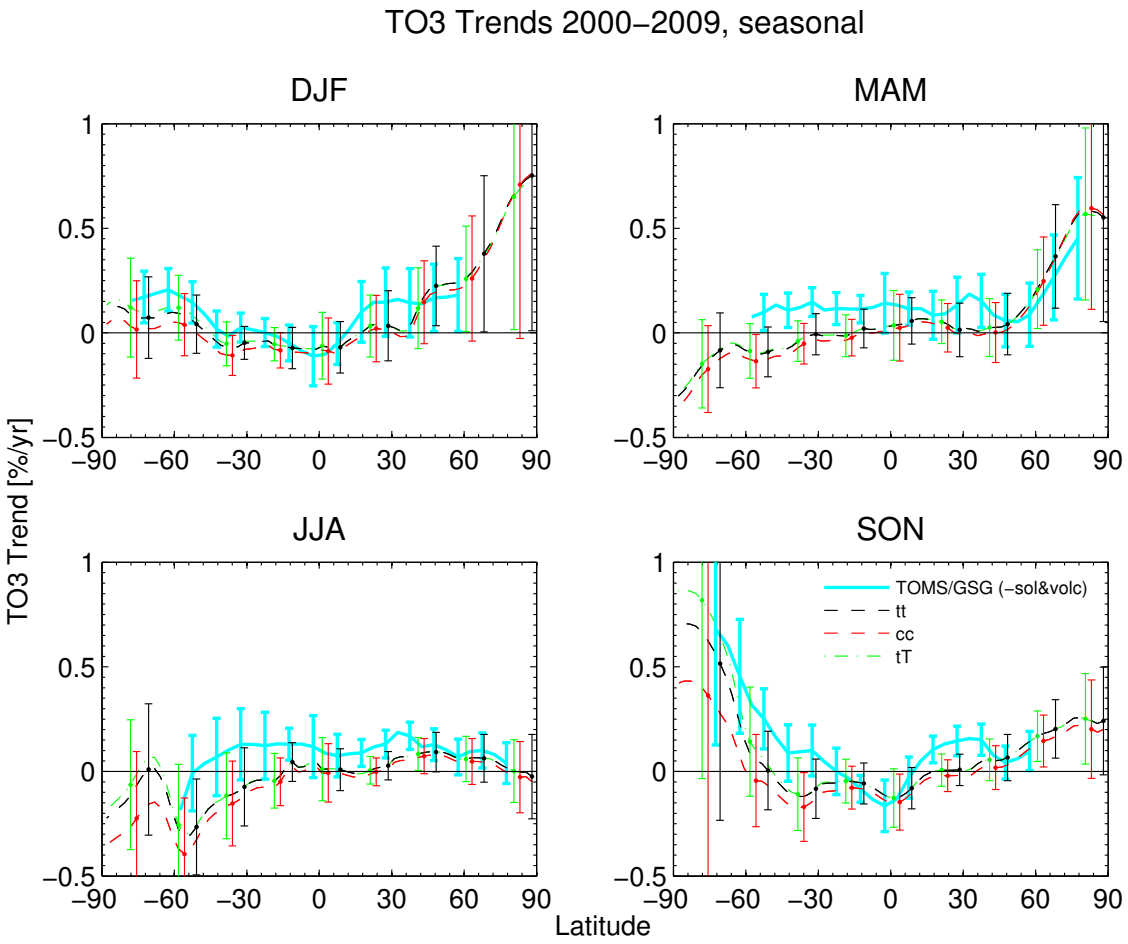

Figure 8.14.: *Like Figure 8.11, but for the 2000–2009 period. Seasonal decomposition of TO3 trends in the 2000–2009 period. Modelled TO3 trends are compared to the combined TOMS/GSG TO3 trends, obtained from the time series with solar and aerosol signals removed. Total trends ($\hat{\omega}_{12} = \hat{\omega}_1 + \hat{\omega}_2$) from model runs with EESC-dependent chemistry and polar chemistry (tt, tT) and constant chemical conditions (cc) are shown. Trends are given in percent of 1980 values per year. Error bars represent the 1σ variance as obtained from the residuals of the linear fit. Plot updated from Kiesewetter et al. (2010b).*

trends. In order to increase significance of profile trends, the whole monthly time series is analysed by taking into account its autocorrelation, as described in Section 8.3.1.

Figure 8.15 shows annual mean trends in ozone number density for mid-latitudes and tropics in the two analysis periods. Trends from the assimilated data set (both versions with and without SBUV adjustment coefficients) are compared to the SBUV merged ozone data set and the CTM reference simulations *tt* and *cc*. Error bars represent 1σ uncertainty of the trend estimate as calculated from Equation 8.8.

General features of observed trends are similar for all considered latitude bands. In the 1979–1999 period, a large negative trend is observed in the upper stratosphere with peak values

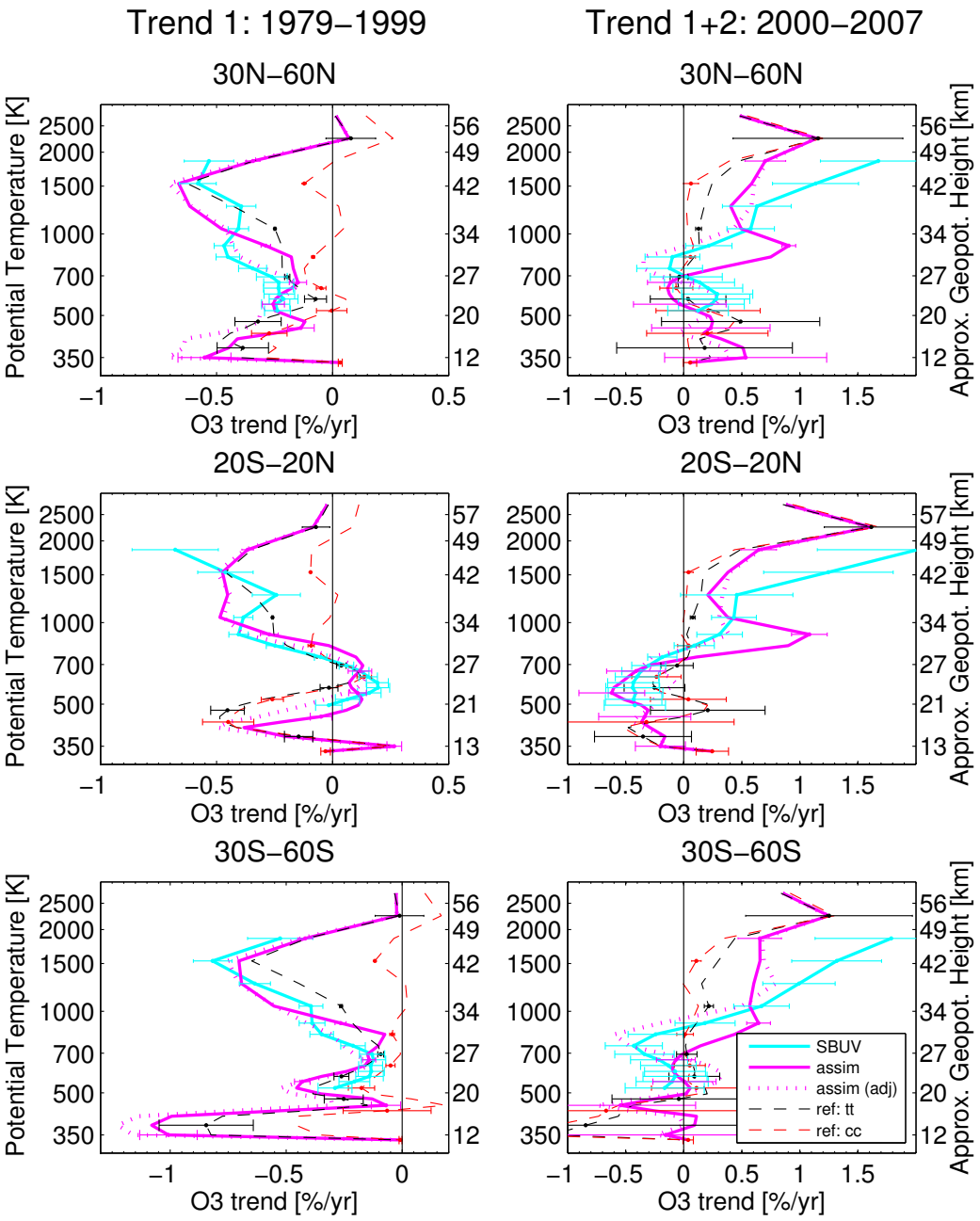

Figure 8.15.: *Ozone number density trends in the assimilated data set, as compared to the SBUV merged data set, and CTM sensitivity runs tt and cc, in the period 1979 – 1999 (left) and 2000 – 2007 (right). Trends are given in percent of 1980 values per year. Error bars represent the 1σ standard error as obtained from the residuals of the linear fit, accounting for the auto-correlation of the time series. Plot adapted from Kiesewetter et al. (2010b).*

about −0.6%/yr at ∼ 1500 K / ∼ 42 km, a near-zero or even positive trend around 600 K / 25 km, and a strong negative trend in the lower stratosphere around 380 K / 16 km, which resembles or slightly exceeds the upper stratospheric trend in magnitude. Both assimilated data sets as well as the *tt* reference simulation agree well with SBUV in the upper and middle stratosphere. In the assimilated data set using SBUV adjustment coefficients, the agreement is slightly better in the middle stratosphere, but larger negative trends are seen in the lower stratosphere.

The strong upper stratospheric decrease in *tt* is mostly attributable to changing atmospheric composition (in particular, increasing EESC acting through the ClO_x cycle) since the *cc* trend is either close to zero or positive there.

In the 2000–2009 analysis period, profile ozone trends are mostly close to zero or positive, particularly in the middle and upper stratosphere. Again, the assimilated data set using SBUV adjustment coefficients is in better agreement to SBUV. Generally, agreement between modelled and SBUV trends is reasonable only below ∼ 35 km. Above this altitude, the SBUV data set shows a large positive trend that exceeds the preceding negative trend by a factor of 2 in absolute numbers. This strong positive trend is not reproduced by the CTM ozone chemistry scheme. As ozone lifetime decreases, so does the content of SBUV information in the assimilated data set, and the ozone trend in the assimilated data set approaches the trend in the CTM sensitivity runs near the top of the CTM domain.

Since the decrease in EESC during recent years is considerably slower than its increase before, the magnitude of the observed trend is puzzling and can hardly be explained by changing ODS loading of the stratosphere. Effects of changing ODS loading are indeed seen in the CTM as differences between the *tt* and *cc* simulations, which are yet small. For technical reasons, the temperature dependency in the Linoz chemistry scheme is switched off in the two uppermost CTM levels, and thus a potential effect of decreasing temperature in the uppermost stratosphere is missed in the CTM.

While it is common to analyse relative ozone profile trends, this is not ideal for attributing TO3 trends vertically. For this purpose, an analysis using absolute trends of partial column ozone is more suitable as it accounts for the different contributions to the total column according to the different air densities at different altitudes. Figure 8.16 shows trends of the assimilated data set including adjustment coefficients, integrated to the SBUV partial column grid, along with the SBUV merged ozone data set and the SBUV v8 single profile observations that were used in the assimilation. The latter are included in this figure to demonstrate that some discrepancies between assimilated ozone trends and SBUV MOD trends are in fact related to inconsistencies between the single SBUV profile observations and the MOD. Adjustment coefficients used in the MOD are applied to the SBUV observations shown here, as used in the assimilation. Both trends for the 1979–1999 (left) and the 2000–2007 period (right) are shown, in the same manner as in Figure 8.15. It is obvious that trends in the assimilated data set agree well to the SBUV profiles used in the assimilation, while SBUV single profile trends do not agree perfectly to trends in the SBUV MOD, which may be due to selection of SBUV profiles for the MOD. The lower stratosphere is dominant for negative column ozone trends in the mid-latitudes during the period of increasing ODS concentrations 1979–1999, due to the high air density there. During the 2000–2007 period, the picture is not uniform. In the extra-tropics, significant positive contributions to TO3 trends stem from both the lower and the upper stratosphere, while a small, marginally significant negative trend is present in the middle stratosphere. If it is not an artifact due to the shortness of the analysis period, this negative trend may be related to the large

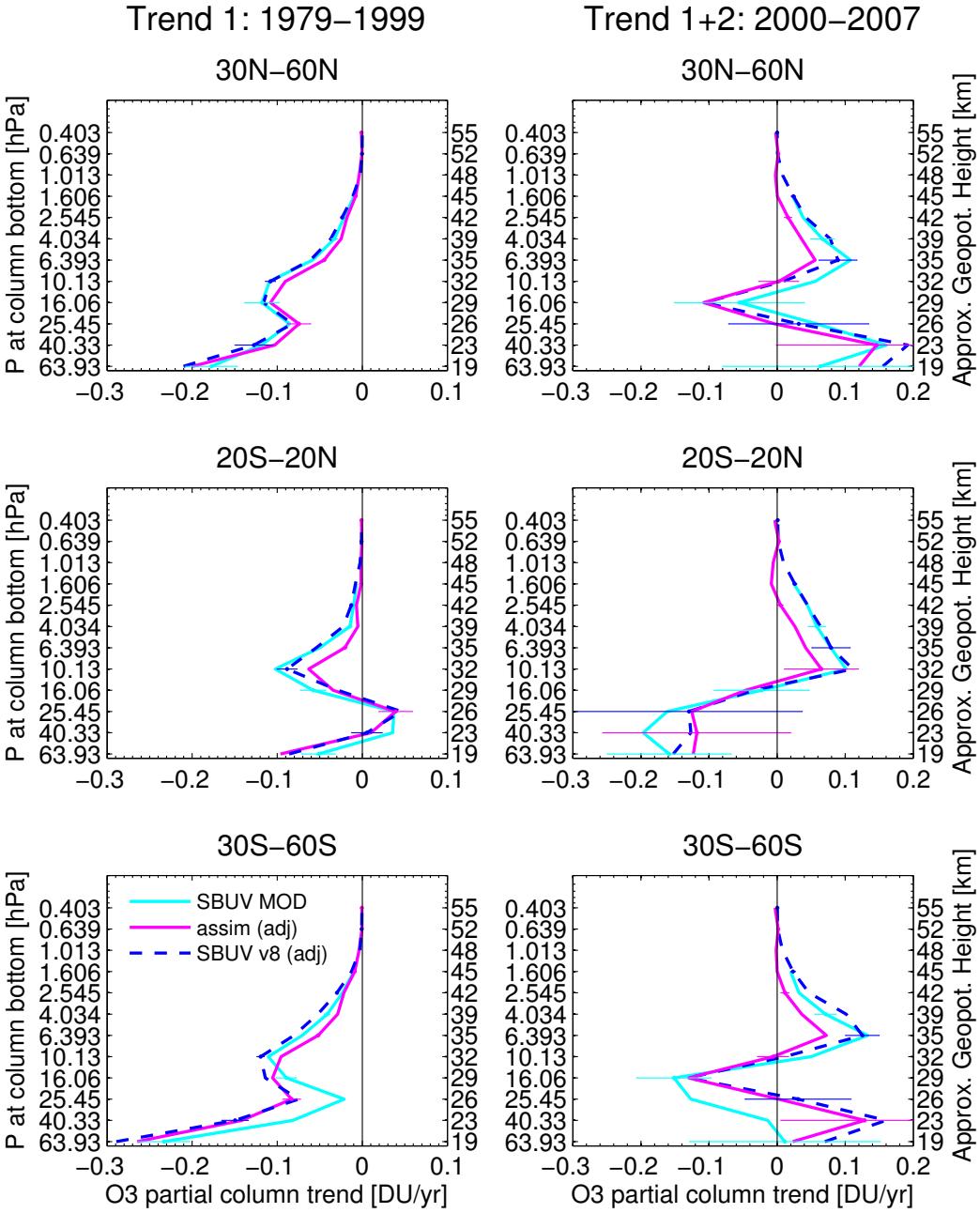

Figure 8.16.: *Trends in partial column ozone, comparing the assimilated data set (including SBUV adjustment coefficients) to the SBUV merged ozone data set and the SBUV single profile observations (including SBUV adjustment coefficients) which are used for assimilation. Vertical spacing refers to the native SBUV partial column grid.*

ozone increases in the upper stratosphere during this period, which lead to a decrease in UV radiation in the middle stratosphere that in turn decreases ozone production via the Chapman cycle.
Ozone trends in the tropical lower stratosphere have been subject of a scientific debate during recent years, as a negative ozone trend would indicate acceleration of the Brewer-Dobson circulation, which is predicted by most climate models (Butchart et al., 2006; Li et al., 2008; Oman et al., 2009). However, observational studies have so far not shown any significant acceleration (Engel et al., 2009). The trend study presented here does not provide a conclusive result in this regard. While there are indications for negative trends in the tropical lower stratosphere from both Figures 8.15 and 8.16, the trend patterns differ during the two time periods, and trends in the 2000–2007 period are not significant. Moreover, results for the lowermost model levels (below 18 km) must be viewed with caution in the tropics, as the upper troposphere is not resolved realistically in the CTM.

8.3.4. Contributions to column ozone trends

With the help of the different CTM sensitivity runs, it is possible to distinguish contributions of different processes to the overall modelled trends – effects of atmospheric composition changes, acting through polar heterogeneous chemistry and export of ozone-depleted air, as well as through gas phase chemistry, as opposed to meteorological (transport and temperature) effects. Since trends in the *tt* and *tT* runs resemble observed trends well, it is reasonable to assume that an attribution undertaken in this way provides a realistic picture of actual contributions. Unfortunately, there is no way of quantifying these contributions experimentally – we have no chance to turn back time and observe what would have happened without the increase in anthropogenic ODS.
In order to separate effects of changing atmospheric composition on gas-phase chemistry and polar chemistry, differences of runs *tt*, *tc*, and *cc* are analysed. While the *tt* run takes into account the changes of ODSs acting via gas phase and heterogeneous chemistry (polar ozone decay rate $1/\tau_{pc} \propto$ EESC), *cc* assumes constant chemical conditions and underlies only meteorological variability. Thus, trends of $tc - cc$ can be interpreted as the contribution of changing ODS levels to ozone trends through gas phase activation, while $tt - tc$ yields the fraction of ozone trends attributable to the effect of ODS changes on polar chemistry.
Figure 8.17 shows the trend decomposition for the period of rising ODS loadings. Effects of both linear and quadratic scaling of the polar ozone destruction rate with EESC are shown. For comparison, also the total (*tt*, *tT*) trends are displayed, which are obtained by adding the different contributions. Note that trends are given as absolute values in DU/year here, contrary to Figures 8.9 – 8.12 which relate them to 1980 TO3 values. The reason is the obvious ambiguity as to which reference value trends of ΔTO3 should be related to. While heterogeneous activation of ODSs dominates the negative trends in polar areas, in particular for $r_{pc} \propto \mathrm{EESC}^2$, in the mid-latitudes of both hemispheres contributions from ODS changes acting through polar chemistry (export of ozone depleted air masses), ODS changes acting through in-situ gas phase activation, and meteorological changes are roughly equal. Surprisingly, gas phase chemistry contributes almost as much as heterogeneous chemistry to negative trends in the Arctic.
A seasonal decomposition of trend contributions during this period is shown in Figure 8.18. Again, also the effect of a stronger scaling of the polar ozone decay rate $1/\tau \propto \mathrm{EESC}^2$ is shown by the $tT - tc$ difference, and the *tT* trend is included as a reference for the overall

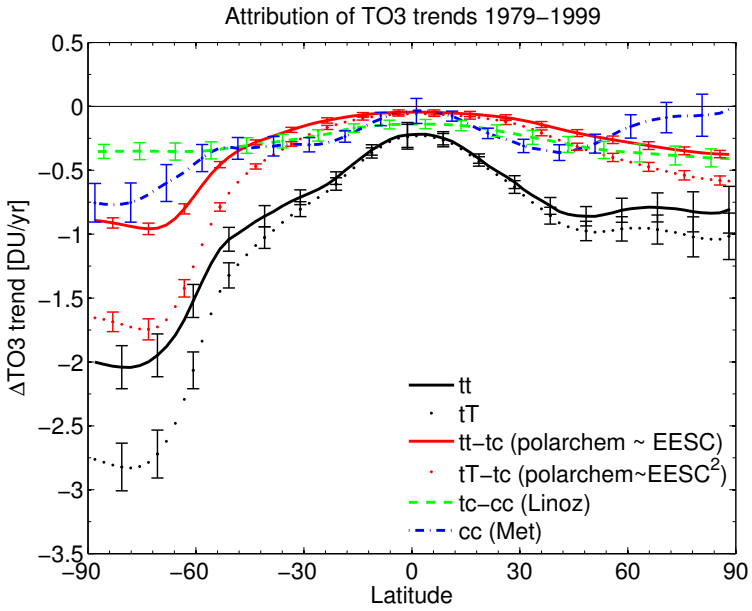

Figure 8.17.: *Attribution of TO3 trends in the period 1979–1999 to effects of ODS changes on gas phase chemistry ($tc-cc$), ODS changes on polar chemistry ($tt-tc$ for polar chemistry $\propto$ EESC, and $tT-tc$ for polar chemistry $\propto EESC$), and natural variability under constant atmospheric composition (cc). Total trends are shown as a reference (tt and tT).*

trend in this case.

Error bars in Figure 8.18 correspond to 1σ standard variation of the residuals. Since the differences $tc - cc$ and $tt - tc$ originate in information that is put into the CTM (i.e., the linear interpolation of Linoz tables and the linear scaling of the polar chemistry with EESC), it is not surprising that the residuals of a linear fit to these differences are small. For this reason, the error bars of $tc-cc$ and $tt-tc$ should be viewed with caution, as they are directly influenced by the model setup and do not correspond to actual uncertainty.

Due to its low seasonal variation, gas phase activation of ODSs ($tc - cc$) constitutes the main contribution to observed trends outside the polar latitudes in most seasons. Gas phase chemistry alone contributes about $-0.7\,\mathrm{DU/yr}$ during polar winter and spring, and a slightly weaker trend in polar summer and autumn ($-0.5\,\mathrm{DU/yr}$).

Effects of heterogeneous chlorine activation, on the other hand, are first observed during SH winter, peaking at high southern latitudes during spring ($-1.8\,\mathrm{DU/yr}$ for $tt - tc$ and $-3.5\,\mathrm{DU/yr}$ for $tT - tc$), after which the ozone depleted air masses are distributed to mid-latitudes where they cause a trend of about $-0.35\,\mathrm{DU/yr}$ ($\sim -0.7\,\mathrm{DU/yr}$) or $\sim 35\%$ ($\sim 50\%$) of the total observed tt (tT) trend during SH spring and summer. Due to the warmer Arctic vortex and its earlier breakup, heterogeneous ozone depletion is generally less severe in the NH and peaks earlier in spring than in the Antarctic, resulting in a peak trend at high northern latitudes of $-0.35\,\mathrm{DU/yr}$ ($-0.7\,\mathrm{DU/yr}$) for linear (quadratic) polar chemistry during DJF and MAM. Through dispersion to lower latitudes this effect contributes around $-0.15\,\mathrm{DU/yr}$ ($-0.3\,\mathrm{DU/yr}$) or $\sim 15\%$ ($\sim 30\%$) to tt (tT) column ozone trends in NH mid-latitude spring. Note that due to the large uncertainties in total (tt) trends, percentages of

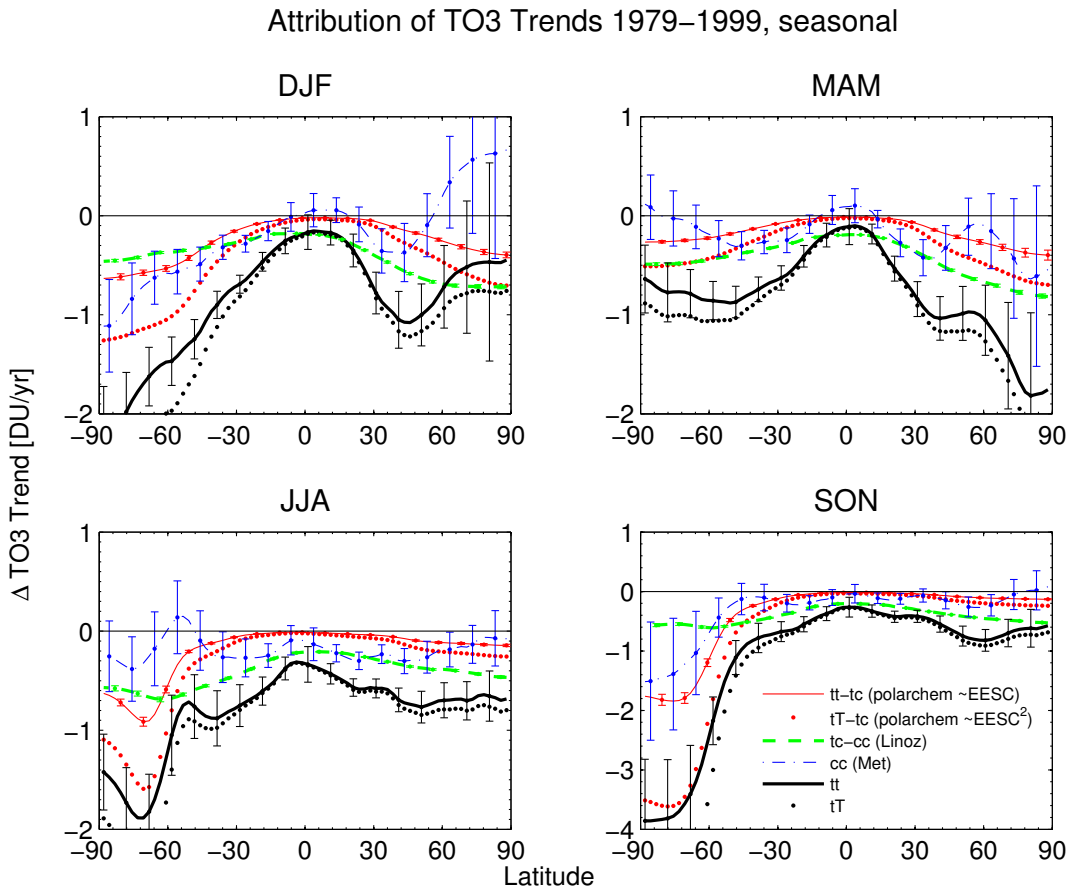

Figure 8.18.: *Seasonal decomposition of Figure 8.17. Attribution of seasonal TO3 trends in the 1979–1999 period to effects of ODS change on gas phase chemistry ($tc - cc$), ODS change on polar chemistry ($tt - tc$ for polar chemistry $\propto$ EESC, and $tT - tc$ for polar chemistry $\propto EESC$), and natural variability under constant atmospheric composition (cc). Total trends are shown as a reference (tt and tT). Plot updated from Kiesewetter et al. (2010b).*

different contributions are also subject to large uncertainties although the absolute trend differences are well determined (see error bars in Figure 8.18).

The values of trends caused by ODS changes described here compare well to results obtained by Chipperfield (2003), who reported that in a similar modelling study for the 1979–1998 period, the total effect of increasing halogen loadings (in-situ chemistry + polar chemistry) led to a reduction of mid-latitude column ozone of around 20 DU in the SH and 10 DU in the NH. The corresponding amounts obtained from the simulations analysed in this thesis are 15 DU (20 DU) in the SH and 11 DU (14 DU) in the NH in the tt (tT) runs.

The absolute magnitude of seasonal trends caused by meteorological changes alone (cc) is comparable to those of the trends caused by changing ODSs; however, meteorological changes have also had a positive influence on TO3 trends during some seasons. Their contributions range between -0.5 and $+0.5$ DU/yr except for polar latitudes during the ozone hole season in spring. As noted before, the strong negative trend in SH polar latitudes during SON ($\sim$

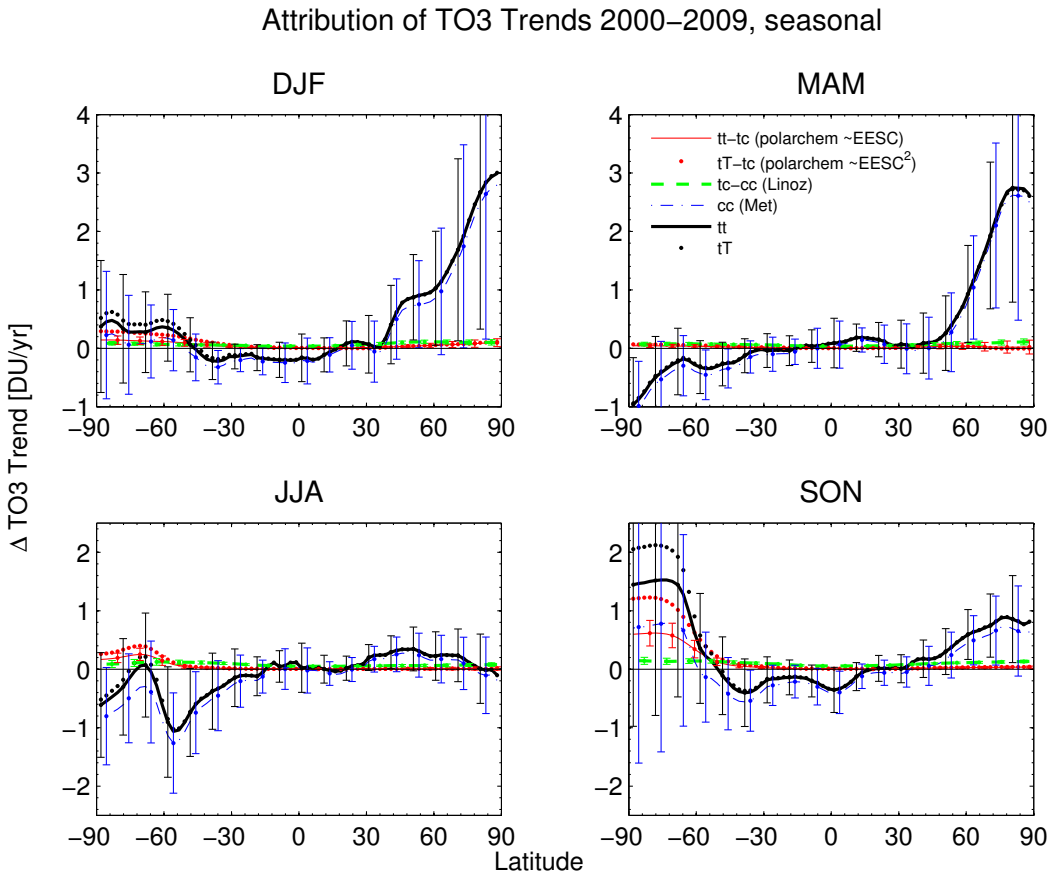

Figure 8.19.: *Attribution of TO3 trends in the 2000–2009 period to effects of ODS change on gas phase chemistry ($tc - cc$), ODS change on polar chemistry ($tt - tc$ for polar chemistry $\propto$ EESC, and $tT - tc$ for polar chemistry $\propto EESC$), and natural variability under constant atmospheric composition (cc). Total trends are shown as a reference (tt and tT). Plot updated from Kiesewetter et al. (2010b).*

$-1.5\,\mathrm{DU/yr}$) may clearly be interpreted as a sign of dynamical feedback of ozone depletion, as are most probably also the strong negative *cc* trends at high southern latitudes during DJF ($\sim -1\,\mathrm{DU/yr}$) and high northern latitudes during MAM ($\sim -0.9\,\mathrm{DU/yr}$). In the NH mid-latitudes, meteorological changes contribute around $30-40\%$ year-round to total ozone trends, while in the SH mid-latitudes their contribution is comparably large during summer but negligible during winter months. The value of $\sim 35\%$ meteorological contribution to NH mid-latitude trends in spring lies between values reported by Randel et al. (2002) ($20-30\%$), Steinbrecht et al. (1998) (30%), Hadjinicolaou et al. (2002) ($\gtrsim 30\%$), and those obtained by Hood et al. (1997) (up to 50%).

Figure 8.19 shows the trend decomposition as in Figure 8.18 but for the 2000–2009 period. From Figure 8.12 it was obvious that effects of ODS decreases are yet small, and hence only the question is raised here whether these small effects originate in a recovery trend in springtime polar ozone depletion, or in gas-phase activation. Figure 8.19 shows that a

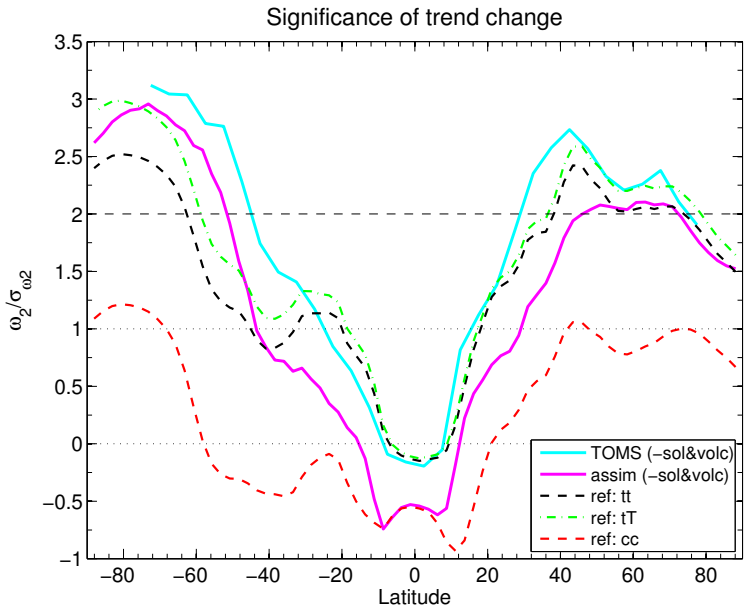

Figure 8.20.: *Change in linear trend between the two analysis periods of increasing and decreasing halogen loading, 1979–1999 and 2000–2009, expressed in terms of standard deviations in order to provide a measure of its significance. Model runs with changing chemical composition (tt, tT) are compared to the model run with constant chemical composition (cc) and TOMS/SBUV satellite observations. Contrary to Figures 8.9 and 8.12, only the TOMS/SBUV merged data set is used as a reference, in order to avoid artificial lengthening of the satellite time series in the trend-change period.*

distinct positive trend in column ozone is visible as an effect of ODS decreases on gas phase chemistry ($tc-cc$), while effects of a changing polar chemistry ($tt-tc$ and $tT-tc$) are not yet significant except for the ozone hole season (Antarctic SON), where they are directly imposed by the EESC scaling. During Antarctic spring, linear response of polar ozone depletion to EESC ($tt-tc$) causes a trend of $+0.6\,\mathrm{DU/yr}$, while quadratic dependency causes a trend of $+1.1\,\mathrm{DU/yr}$. However, these positive trends cannot yet be distinguished from the expected evolution in case of constant chemical conditions (cc) due to high inter-annual variability. Overall, meteorological variability dominates, showing strong variations with season and latitude. In particular, the strong and marginally significant trend in NH TO3 during winter and spring is almost entirely caused by meteorology and to less than 10% attributable to ODS changes.

8.3.5. Is there a significant trend change?

Even though no significantly (2σ) positive column ozone trend in the post peak-EESC period is detectable yet, changes in trends between the two analysis periods are evident from Figures 8.10 and 8.12. A significant change in TO3 trend may be regarded as a first step towards ozone recovery (Reinsel et al., 2002; Newchurch et al., 2003; Yang et al., 2008), and has already been reported for different atmospheric regions (Newchurch et al., 2003; Yang et al., 2008). Figure 8.20 shows the level of significance of a presumed change in TO3 trend, which

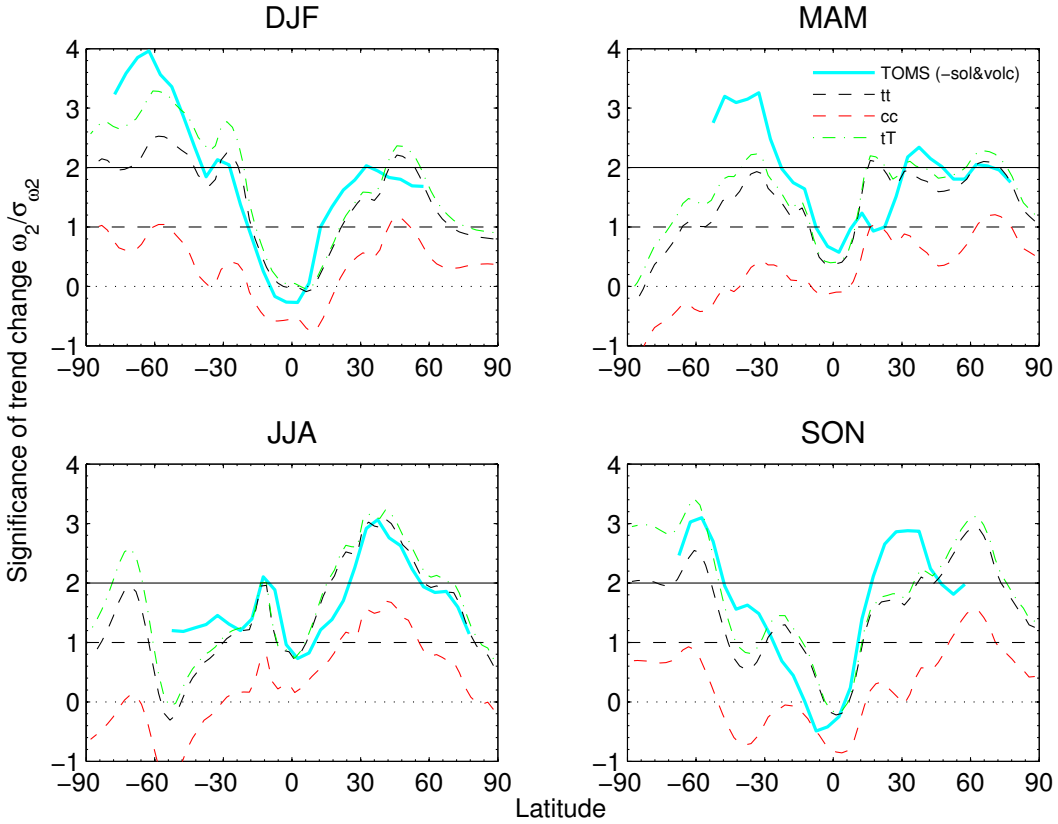

Figure 8.21.: *Significance of change in seasonal trends between the periods of increasing and decreasing halogen loading, 1979–1999 and 2000–2009, expressed in terms of standard deviations in order to provide a measure of its significance. Model runs with changing chemical composition (tt, tT) are compared to the model run with constant chemical composition (cc) and TOMS/SBUV satellite observations.*

is calculated here as $\hat{\omega}_2/\sigma_{\hat{\omega}_2}$. A trend change may be regarded significant at the 95 (99.7, 66.7)% level if $\hat{\omega}_2/\sigma_{\hat{\omega}_2} > 2(3,1)$. Contrary to Figures 8.9 and 8.12, only the TOMS/SBUV merged data set (aerosol and solar signal removed) is used here as a reference, in order not to artificially lengthen the relevant trend change period.

TOMS/SBUV, the assimilated data set, and the unconstrained CTM with changing stratospheric chlorine loading (in particular tT) generally agree well on the level of significance of the trend change. Significant changes are present in the mid to high latitudes of both hemispheres, reaching up to 3σ in the Antarctic. The positive trend changes observed in TOMS/SBUV are consistent with results presented by Reinsel et al. (2005), indicating that a significant trend change has occurred. However, also the CTM run with constant stratospheric composition (cc) does show a trend change in the NH mid-latitudes and the Antarctic that is significant at the 1σ level.

The seasonal decomposition of trend changes is shown in Figure 8.21 (the whole period

of CTM sensitivity simulations is used). Also in the seasonal domain, observed patterns of trend change are in excellent agreement to the CTM simulations with changing stratospheric chlorine. The positive change in tt (tT) trend is significant at 2σ (3σ) in Antarctic spring and summer. Without following the detailed treatment by Yang et al. (2008), this confirms that a positive trend change has occurred in Antarctic spring, and is reproduced in the assimilated data set as well as in the unconstrained CTM with varying chemistry. However, part of this positive change in trend is related to changing meteorology, as indicated by the marginally significant ($\sim 1\sigma$) change in cc trend. In addition, tt and tT runs show significant changes in NH mid-latitude trends during summer and fall, which are equally related to marginally significant positive changes in the cc trends.
These results suggest that although a significant change in column ozone trend has occurred in mid- to high latitudes of both hemispheres, it must yet be treated with caution, as it is not entirely a result of declining ODS concentrations but in parts caused by changes in meteorological conditions.

8.4. Summary of Chapter 8

Chapter 8 has been concerned with an in-depth analysis of long-term ozone variability, in particular trends, and their attribution to contributing factors. Column ozone showed large decreases during the 1980s and early 1990s in both hemispheres, which are very well reproduced in the assimilated data set as well by in the unconstrained CTM accounting for ODS variations. Large offsets in ozone are seen in the unconstrained model at the transition between different meteorological data sets, which are possibly enhanced here due to the use of a simple linearised ozone chemistry. Similar effects have been reported by previous studies (Chipperfield, 2003; Hadjinicolaou et al., 2005), but then treated rather arbitrarily or ignored. This study follows a more stringent approach by extending the regression analysis to treat stretches which use different meteorological data in parallel, forcing the same trend upon them but allowing for an offset.
Ozone differences between the unconstrained CTM and observations are used to quantify effects of the 11 year solar cycle as well as effects of volcanic eruptions in the observations and the assimilated data set. Variations in solar irradiation and volcanic eruptions are to a large degree able to explain differences between modelled and observed (and assimilated) ozone. The distinction between solar cycle and volcanic signals is not unambiguous, due to the occurrence of two major volcanic eruptions during similar periods of the solar cycle. Nonetheless, diagnosed solar cycle and aerosol signals are of the expected order of magnitude in column and profile ozone observations. Interestingly, the solar cycle signal in column ozone appears delayed by a few months in the extra-tropics consistently in both hemispheres, which may be due to transport via the Brewer-Dobson circulation. The strong upper stratospheric solar signal is not well reproduced in the assimilated data set, as ozone lifetimes are too short there to ensure a high content of satellite information in the assimilated data set.
A standard least squares multi-linear regression model with piecewise linear trends is used to derive column and profile ozone trends during periods of increasing and decreasing stratospheric chlorine loadings separately, while analysing the whole time series including overlaps in one piece. Negative TO3 trends during the 1979–1999 period are very well reproduced by the EESC-dependent CTM reference run. In polar latitudes, the observed trend is between modelled trends with linear and quadratic dependence of polar chemistry on EESC, suggest-

ing that the real dependency of the chemical ozone destruction rate constant on EESC is between 1 and 2, as suggested also by previous studies (Searle et al., 1998; Hsu and Prather, 2009).

Both versions of the assimilated data set – with and without application of SBUV adjustment coefficients in the assimilation – are tested in the trend analysis. Conclusions regarding the benefits of adjustment coefficients are ambiguous. While the adjustment coefficients increase consistency of profile ozone trends between assimilated ozone and the SBUV MOD, the consistency of column ozone trends is diminished by the adjustment coefficients, mainly due to negative adjustments in the lower stratosphere for single satellite instruments (particularly NOAA-11) which subsequently propagate downward and decrease column ozone. Whether this points to issues in the CTM advection, or in the calculation of the adjustment coefficients is not clear at present. For most applications, the standard version of the assimilated data set (without SBUV adjustment) is preferred.

Sensitivity calculations with the unconstrained CTM are used to attribute the observed trends to their processes of origin, distinguishing the effects of ODS changes acting via gas-phase chemistry or via heterogeneous chemistry (polar chemistry), from changes in transport or temperature. The decline of column ozone in the NH mid-latitudes during the 1980s and 1990s is attributed roughly equally to polar chemistry, gas-phase chemistry and meteorology, while in the SH polar chemistry played a larger role. During the period of declining stratospheric halogen loadings after 1999, no significantly (2σ) positive trends are present yet, although positive trends are seen at polar regions of both hemispheres consistently in the observational record, assimilated ozone and sensitivity runs. Assimilated ozone and the TOMS/SBUV merged data set disagree about mid-latitude trends in this period, as the SBUV merged data set shows exceptionally strong ozone increases in the upper stratosphere where assimilated ozone is not very sensitive to satellite information, even exceeding in absolute magnitude the decreases seen before. This may be either connected to shortcomings of the simple Linoz chemistry scheme used here or, since the GSG merged column ozone data set is in better agreement to assimilated and modelled TO3, point to difficulties in the TOMS/SBUV MOD set. Overall, column ozone changes during the last decade are dominated by natural (meteorological) variability, and only small effects of ODS decreases can be seen as differences between the EESC-dependent and constant chemistry sensitivity runs. These are yet too small to enable a significant distinction from the expected evolution in case of constant ODS loadings.

Cessation of decline of column ozone has been characterised as a first stage of ozone recovery, and has been reported for some areas (Reinsel et al., 2005; Newchurch et al., 2003; Yang et al., 2008). Consistently with these studies, I find that a significant trend change has occurred in the observational TO3 record that is reproduced well in the CTM sensitivity runs with changing ODS levels, and in the assimilated data set. However, extreme caution must be exercised when attributing this trend change to decreasing ODS concentrations, as the CTM sensitivity run with constant chemical conditions also shows a positive trend change that is marginally significant at the 1σ level.

9. Conclusions and Outlook

The research work described in this thesis was concerned with an analysis of stratospheric ozone variability on different timescales. As explained in the introduction, in spite of the long history of ozone research, several questions regarding stratospheric ozone have so far remained unresolved. These concern in particular the question of inter-annual variability of polar ozone, and the question of how the ozone layer will evolve in a future climate. In order to reliably predict a possible recovery of stratospheric ozone, a thorough analysis and interpretation of past ozone changes is necessary, asking whether our understanding of dynamics and chemistry is sufficient to explain the observed evolution of stratospheric ozone during past decades.

In order to address these questions, a consistent long-term observation based data set of stratospheric ozone is needed, which cannot be provided by measurements alone due to technical limitations. Within the research work described in this thesis, a sequential assimilation scheme has been implemented in the Bremen 3D Chemistry Transport Model, and a 29 year data set of stratospheric ozone has been generated from sequential assimilation of SBUV satellite observations. The data set, which is an update from Kiesewetter et al. (2010a), is in excellent agreement with independent observations from other satellite instruments – GOME, SCIAMACHY and MIPAS – as well as sonde data throughout most of the stratosphere. Although some biases remain in areas beyond the limits of SBUV coverage, particularly in the lowermost stratosphere, ozone variability is represented very well throughout the stratosphere. Hence, the data set may be viewed as a global extension of an observation based record of profile ozone, into zones where satellite observations are not available. The error tracer that is provided with the assimilated ozone itself gives an estimate of the information content in – or the uncertainty associated with – the assimilated ozone field.

The extent of the data set, offering 3D fields of $3.75° \times 2.5°$ resolution on 24 isentropic levels covering the whole stratosphere on a daily base, invites a wealth of analyses that may be conducted regarding ozone variability, and there are good reasons to hope that this data set – or further improved versions of it – may contribute to several issues of research interest. In this thesis, two particular topics have been investigated in detail: the development, propagation, and lifetimes of anomalies in Arctic mid-stratospheric ozone on the one hand, and global ozone trends over the last three decades on the other hand. Although these two studies had distant origins and research questions of their own, they actually provide good examples for the extent of different domains on which the data set may be investigated: daily regional 3D fields (also taking the transition from zonal means to vortex means is only possible in a three-dimensional data set) grouped and analysed for periods of a few months, and global zonal monthly means of column-integrated ozone for the whole three decades the data set covers. Sensitivity calculations with the unconstrained CTM provide additional information on contributions to trends, besides demonstrating that the CTM itself is able to reproduce many features of ozone variability during the 30 year periods very well even with the simplified ozone chemistry used here – a property that is essential also for the quality

of the assimilated data set.
With its excellent representation of polar ozone variability, especially in the NH, the assimilated data set is very well suited to address the research question of how ozone anomalies develop and evolve during winter. The assimilated data set shows large patterns of ozone anomalies in the Arctic that develop in the middle to upper stratosphere around 30 km altitude, mostly during autumn or winter, and subsequently descend from there following the diabatic cooling rates. Hence ozone anomalies in the lower stratosphere, which determine the variability of column ozone, often have precursors in the mid to upper stratosphere a few months earlier. Two different mechanisms by which polar ozone anomalies are formed in the assimilated data set have been selected and investigated in detail: anomalies in the stratospheric circulation, and solar proton events. Mid-stratospheric ozone anomalies develop in particular in connection with anomalies in the stratospheric circulation, which are expressed in this study by the Northern Hemisphere Annular Mode. As an indicator of poleward transport, the NAM phase is generally anti-correlated with mid-stratospheric polar ozone. Strong excursions of the NAM phase – so-called weak and strong vortex events – lead to strong and distinctively shaped patterns of polar ozone anomalies (around ±15%). These ozone anomalies traverse most of the stratosphere within days to weeks, remaining significant in the lower stratosphere for almost five months after the dynamic anomaly, when the NAM index itself has long relaxed to average values. In the vicinity of vortex events, an interesting interplay of chemistry and dynamics takes place at different altitudes, the result of which is the unique shape of the ozone anomaly. Vortex events – in particular weak vortex events – can almost immediately be identified with the naked eye from a vertically resolved daily time series of Arctic ozone anomalies.
It should be emphasised that this study can only go as far as describing the effects of dynamical anomalies on the ozone field. It is tempting to take a step further and ask the question what exactly causes these events and whether there are feedbacks from the previous ozone field on dynamics; however interesting, this question is beyond the scope of this study for fundamental reasons: In the CTM, all dynamics (except vertical transport through radiative heating) are prescribed externally, and no feedback from ozone is possible.
As an example for a distinctly different cause of anomalies in polar ozone, effects of solar particle precipitation are investigated. It is shown that some of the long-lived ozone anomalies in the assimilated data set are indeed related to solar proton events, which lead to strong increases in odd-nitrogen production in the upper stratosphere and mesosphere. During winter, these NO_y anomalies are efficiently transported downward into the polar vortex and lead to significant ozone decreases in the middle and even lower stratosphere that are visible up to 6 months later – as long as the vortex is not completely disrupted as a consequence of a weak vortex event. The long lifetimes of such solar particle induced ozone anomalies point out the importance of NO_x chemistry for understanding mid-stratospheric ozone, and raise the question whether NO_y anomalies may also play a role for other long-lived polar ozone anomalies that are not directly related to solar activity. This question cannot be addressed reasonably in an assimilation scheme with only one tracer, and a follow-up project is underway to investigate this issue further, as described below.
Global ozone trends are the second focus of this thesis. In the introduction, the question was raised whether ozone changes during the last three decades are consistent with our understanding of ozone chemistry and stratospheric dynamics, and whether signs of an onset of chemical ozone recovery are already detectable as a consequence of decreasing ODS levels since the turn of the century. The analysis presented in Chapter 8 shows that both

assimilated ozone as well as ozone from an unconstrained CTM run with varying stratospheric chlorine loading reproduce very well the strong declines in polar and mid-latitude column ozone that were observed during the 1980s and 1990s, and the subsequent levelling-off. Again, the assimilated ozone can be used as an extension of existing observational data sets into areas where satellite data are not available. In the period of decreasing ODS loadings, small positive trends are present at polar regions of both hemispheres consistently in the observational record, assimilated ozone and sensitivity runs. However, one has to be very careful when attributing these positive trends to chemical recovery, as demonstrated by the CTM sensitivity calculations analysed together with the assimilated data set. These leave either gas-phase chemistry or polar chemistry constant at year 2000 conditions and hence allow for an attribution of ozone changes to contributing factors, namely ODS changes acting via gas phase chemistry, ODS changes acting via heterogeneous polar chemistry, and natural variability which influences ozone through changes in transport or temperatures. In this context, the simple chemistry scheme employed in the CTM is an advantage, as it ensures that any ozone changes in the model must stem from one of these factors. While the decline of column ozone in the NH mid-latitudes can be attributed roughly equally to polar chemistry, gas-phase chemistry and meteorology, column ozone changes during the last decade are dominated by natural (meteorological) variability. Only small effects of ODS decreases can be seen as differences between the EESC-dependent and constant chemistry sensitivity runs, in particular during the Antarctic ozone hole season. Positive trends in Arctic springtime column ozone are almost entirely caused by meteorological variability.

Assimilated ozone and the TOMS/SBUV merged data set disagree on mid-latitude trends in this period, as the SBUV merged data set shows exceptionally strong ozone increases in the upper stratosphere where assimilated ozone is not very sensitive to satellite information, even exceeding in magnitude the decreases seen before. This may be either connected to shortcomings of the simple Linoz chemistry scheme used in our CTM or, since the GOME/SCIAMACHY/GOME2 merged column ozone data set is in better agreement with assimilated and modelled TO3, point to difficulties in the TOMS/SBUV merged ozone data set.

Cessation of decline of column ozone has been characterised as a first stage of ozone recovery, and has been reported for some areas (Reinsel et al., 2005; Newchurch et al., 2003; Yang et al., 2008). Consistently with these studies, I find that a significant trend change has occurred in the observational TO3 record that is reproduced well by the CTM sensitivity runs with changing EESC and the assimilated data set. However, extreme caution must be exercised when attributing this trend change to decreasing ODS concentrations, as the CTM sensitivity run with constant chemical conditions also shows a positive trend change that is marginally significant at the 1σ level.

No data set will ever be perfect and no study final. Although the assimilated data set presented in this thesis generally shows good agreement to observations and does capture ozone variability well, there are certainly possibilities for improvements, e.g. concerning negative biases in the lowermost stratosphere, or a more realistic treatment of the upper stratosphere. Furthermore, this study only focused on two distinct aspects of ozone variability, while there are many possibilities for applications and further studies with this data set. An application may be in the retrieval of tropospheric ozone from satellite observations. Due to the optical thickness of the troposphere, limb observation of tropospheric ozone is impossible. Thus, tropospheric ozone columns have to be diagnosed indirectly by subtracting the stratospheric column from the total column derived from a nadir-viewing geometry. With its daily global

coverage, the assimilated data set could be a source for stratospheric ozone columns (in this case, biases in the lower stratosphere should be corrected first).
New research questions have arisen from the current work, e.g. concerning the role of NO_y anomalies for the persistence of ozone anomalies. In particular the long persistence of SPE-induced ozone anomalies, which are in fact related to NO_y enhancements, provides indications for a strong contribution of NO_y to the lifetimes of ozone anomalies. In order to investigate this aspect further, odd-nitrogen species are needed as tracers in the CTM. This can either be achieved by the full chemistry scheme containing the JPL 2006 reaction kinetics, or with a simplified N_2O/NO_y chemistry scheme. As a follow-up project to the DACCS (Data Assimilation and model calculations to study Chemistry Climate interactions and solar influence in the polar Stratosphere) project under which this PhD work was funded, an extension project has been granted which will explicitly focus on the synchronous assimilation of odd-nitrogen and ozone (which were both observed by the HALOE instrument onboard the Upper Atmosphere Research Satellite 1991–2005) in the Bremen 3D CTM, in order to investigate whether the long lifetimes of ozone anomalies are connected to NO_y anomalies.
Also in a broader sense, implementation of the full 57 tracer chemistry scheme for assimilation is an interesting option. Thanks to recent increases in the institute's high performance computing clusters and to parallelisation applied to both the CTM core as well as the assimilation algorithm, this is possible now within a reasonable time. An important point that has to be taken into account in this regard has been raised by Chipperfield (2003), who reported that assimilation of single tracers into a full chemistry scheme severely impacts the inter-correlations between different tracers that are prominent in the stratosphere. Measures have to be taken in order to avoid unrealistic chemical balances.
Also assimilation from other satellite instruments into the Bremen 3D CTM is an interesting option, as the Institute of Environmental Physics is one of the leading satellite retrieval centres in which data are available first-hand, in particular from the GOME and SCIAMACHY instruments. First steps in this directions have been made in a bachelor thesis partly under the supervision of the author. In this preliminary study, the sequential assimilation scheme described here and in Kiesewetter et al. (2010a) was slightly modified to enable assimilation of SCIAMACHY observations, and a first version of an assimilated data set has been produced.
In the current study, the attempt has been made to pursue a balanced approach between creating a long-term data set of stratospheric ozone and analysing it along the lines of basic scientific research questions. Many more open questions remain – with the big issues of global climate change and recovery of the ozone layer facing mankind, and a growing recognition of interconnections between atmospheric chemistry and climate, the need for further research on the Earth's atmosphere, the drivers of its variability and interactions is higher than ever. Observational data sets and theoretical computer models will continue to constitute the basis of atmospheric research, with data assimilation offering the unique possibility for their synthesis.

Appendix

A. List of abbreviations

AOD	Aerosol Optical Depth, a measure for the integrated column amount of aerosol
BrO_x	chemical family of reactive bromine species, $BrO_x = \{Br, BrO\}$
CFCs	Chlorofluorocarbons, a group of chemically inert chlorine containing compounds that were widely used as refrigerants and propellants in the 1960s to 1980s, and have a large potential of depleting stratospheric ozone. Regulated under the Montreal Protocol.
ClO_x	chemical family of reactive chlorine species, $ClO_x = \{Cl, ClO\}$
CTM	Chemistry Transport Model
DU	Dobson unit, unit for measuring column ozone. 1 DU is equivalent to a total column height of 0.001 cm at sea level pressure, or $2.69 \cdot 10^{16}$ ozone molecules per square centimetre.
ECMWF	European Centre for Medium Range Weather Forecasts (Reading, UK)
EECl	Equivalent Effective Chlorine, a measure for the overall chlorine and bromine contained in tropospheric CFCs, weighted by their potential to deplete stratospheric ozone (see Section 2.3.2)
EESC	Equivalent Effective Stratospheric Chlorine, essentially equal to EECl but with a delay accounting for the stratospheric age of air.
ERA-40	a reanalysed data set of various meteorological parameters for the period 1957–2002 prepared by ECMWF
ERA-Interim	a reanalysed data set of various meteorological parameters since 1989 (continuously updated) prepared by ECMWF
F10.7	solar 10.7 cm radio flux, often used as a proxy for solar activity
GOME	Global Ozone Monitoring Experiment, a nadir viewing remote sensing instrument onboard the ERS/2 satellite (operational since 1995).
GOME2	Global Ozone Monitoring Experiment 2, a nadir viewing remote sensing instrument onboard the MetOp-A satellite (operational since 2006).
GSG	GOME/SCIAMACHY/GOME2 merged column ozone data set
HALOE	Halogen Occultation Experiment onboard the Upper Atmosphere Research Satellite (UARS), a satellite remote sensing instrument observing in solar occultation geometry. Operational 1991 – 2005.
HO_x	chemical family of odd-hydrogen, $HO_x = \{H, OH, HO_2\}$
JPL	NASA Jet Propulsion Laboratory, Pasadena, California
MIPAS	Michelson Interferometer for Passive Atmospheric Sounding, a remote sensing instrument onboard ENVISAT measuring atmospheric emissions in limb geometry. In orbit since 2002.
MOD	Merged Ozone Dataset. Refers to the TOMS/SBUV MOD (column ozone) or SBUV MOD (profile ozone) if not otherwise specified.
NAO	North Atlantic Oscillation, a principal pattern of surface pressure in the NH, closely related to the NAM at the surface.

NAM	Northern Hemisphere Annular Mode, see Section 7.2.1.
NASA	(U.S.) National Aeronautics and Space Administration
NH	Northern Hemisphere
NOAA	(U.S.) National Oceanic and Atmospheric Administration
NO_x	chemical family of reactive nitrogen oxides, $NO_x = \{N, NO, NO_2\}$
NO_y	chemical family of odd-nitrogen species, same as NO_x but including chemically inactive reservoir species, $NO_y = \{N, NO, NO_2, NO_3, N_2O_5, HNO_3, HO_2NO_2, ClONO_2, BrONO_2\}$
ODS	Ozone depleting substance, usually of anthropogenic origin. Often used synonymously with CFCs, although also N_2O has the potential to destroy stratospheric ozone via the catalytic NO_x cycle.
PV	Potential Vorticity, see Section 2.2.3
PVU	Potential Vorticity Unit, $1\,\mathrm{PVU} = 10^{-6}\,\mathrm{K\,m^2/(kg\,}s)$
SAD	Aerosol Surface Area Density, a measure of the available aerosol surface at a given location and altitude
SAGE	Stratospheric Aerosol and Gas Experiment, a series of satellite remote sensing instruments (SAGE I – III) that were in orbit 1984 – 2006. Several trace gases and aerosol abundances were measured.
SAM	Southern Hemisphere Annular Mode, see Section 7.2.1.
SBUV	Solar Backscatter UltraViolet, the satellite instruments used for assimilation in this study. SBUV instruments have been flown on different NASA and NOAA satellites (see Chapter 5)
SCIAMACHY	Scanning Imaging Absorption Spectrometer for Atmospheric Chartography, a remote sensing instrument onboard the ENVISAT satellite, measuring backscattered sunlight in nadir as well as in limb mode, and transmitted sunlight in solar occultation mode. In orbit since 2002.
SH	Southern Hemisphere
SPE	Solar Proton Event, large outburst of high energetic protons from the sun
TO3	Total column ozone abundance, the total number of ozone molecules over a unit surface. Usually measured in Dobson Units (DU).
TOMS	Total Ozone Mapping Spectrometer, a series of nadir viewing satellite instruments flown on NASA and NOAA satellites continuously since 1978.
vmr	volume mixing ratio

B. List of publications

Peer-reviewed articles

Kiesewetter, G., Sinnhuber, B.-M., Vountas, M., Weber, M., and Burrows, J. P.: A Long-term Stratospheric Ozone Dataset from Assimilation of Satellite Observations: High-Latitude Ozone Anomalies, J. Geophys. Res., 115, D10307, doi: 10.1029/2009JD013362, 2010a

Kiesewetter, G., Sinnhuber, B.-M., Weber, M., Burrows, J. P.: Attribution of stratospheric ozone trends to chemistry and dynamics: a modelling study, Atmos. Chem. Phys., 10, 12073–12089, doi: 10.5194/acp-10-12073-2010, 2010b

Conference contributions

Talks

Kiesewetter, G., et al.: What controls the inter-annual variability of Arctic ozone? DPG Frühjahrstagung, Darmstadt, Germany, March 2008.

Kiesewetter, G., et al.: Assimilation of ozone observations to study chemistry climate interactions and solar impact in the polar stratosphere – First results from the DACCS project. CAWSES meeting, Berlin, Germany, September 2008.

Kiesewetter, G., et al.: Investigation of polar ozone anomalies with a Chemistry Transport Model: The role of stratospheric dynamics and heterogeneous chemistry. EGU General Assembly, Vienna, Austria, May 2010.

Kiesewetter, G., et al.: Attribution of modelled column ozone trends to chemistry and dynamics: Signs of recovery? EGU General Assembly, Vienna, Austria, April 2011.

Posters

Kiesewetter, G., et al.: What controls the inter-annual variability of Arctic ozone? EGU General Assembly, Vienna, Austria, April 2008.

Kiesewetter, G., et al.: Evolution of polar ozone anomalies during winter as derived from satellite data assimilation. Quadrennial Ozone Symposium, Tromsø, Norway, July 2008.

Kiesewetter, G., et al.: Investigation of Arctic ozone anomalies by assimilation of satellite data. EGU General Assembly, Vienna, Austria, April 2009.

Kiesewetter, G., et al.: The roles of stratospheric dynamics and chemistry for polar ozone anomalies in a 29 year assimilated ozone dataset. Scientific Comittee on Solar Terrestria Physics (SCOSTEP) Conference, Berlin, Germany, July 2010.

Kiesewetter, G., et al.: A long-term stratospheric ozone dataset from assimilation of satellite observations. 38^{th} COSPAR Scientific Assembly, Bremen, Germany, July 2010.

Invited talks

Kiesewetter, G., et al.: The roles of stratospheric dynamics and chemistry for polar ozone anomalies and global ozone trends. Seminar talk at the Institute for Meteorology, Free University of Berlin, Berlin, Germany, December 2010.

Bibliography

Adriani, A., Fiocco, G., Gobbi, G. P., and Congeduti, F.: Correlated behaviour of aerosol and ozone contents of the stratosphere after the El Chichón eruption, J. Geophys. Res., 92, 8365–8372, 1987.

Andrews, D. G., Holton, J. R., and Leovy, C. B.: Middle atmosphere dynamics, International Geophysics Series, volume 40, Academic Press, New York. 489 pp., ISBN: 978-0-12-058576-2, 1987.

Aschmann, J., Sinnhuber, B.-M., Atlas, E. L., and Schauffler, S. M.: Modeling the transport of very short-lived substances into the tropical upper troposphere and lower stratosphere, Atmos. Chem. Phys., 9, 9237–9247, doi:10.5194/acp-9-9237-2009, 2009.

Aschmann, J., Sinnhuber, B., Chipperfield, M. P., and Hossaini, R.: Impact of deep convection and dehydration on bromine loading in the upper troposphere and lower stratosphere, Atmos. Chem. Phys., 11, 2671–2687, doi:10.5194/acp-11-2671-2011, 2011.

Atkinson, R., Baulch, D. L., Cox, R. A., Crowley, J. N., Hampson, R. F., Hynes, R. G., Jenkin, M. E., Rossi, M. J., and Troe, J.: Evaluated kinetic and photochemical data for atmospheric chemistry: Volume I - gas phase reactions of O_x, HO_x, NO_x and SO_x species, Atmos. Chem. Phys., 4, 1461–1738, 2004.

Baldwin, M. P. and Dunkerton, T. J.: Stratospheric Harbingers of Anomalous Weather Regimes, Science, 294, 581–584, 2001.

Barath, F. T., Chavez, M. C., Cofield, R. E., Flower, D. A., Frerking, M. A., Gram, M. B., Harris, W. M., Holden, J. R., Jarnot, R. F., and Kloezeman, W. G.: The Upper Atmosphere Research Satellite microwave limb sounder instrument, J. Geophys. Res., 98, 10 751–10 762, doi:10.1029/93JD00798, 1993.

Bates, D. R. and Nicolet, M.: Atmospheric Hydrogen, Publ. Astron. Soc. Pacific, 62, 105–110, doi:10.1086/126244, 1950.

Bhartia, P. K., McPeters, R. D., Mateer, C. L., Flynn, L. E., and Wellemeyer, C.: Algorithm for the estimation of vertical ozone profiles from the backscattered ultraviolet technique, J. Geophys. Res., 101, 18 793 – 18 806, 1996.

Bhartia, P. K., Wellemeyer, C. G., Taylor, S. L., Nath, N., and Gopolan, A.: Solar backscatter ultraviolet (SBUV) version 8 profile algorithm, in: Proceedings of the XX Quadrennial Ozone Symposium, 1 - 8 June 2004, Kos, Greece, edited by Zerefos, C. S., pp. 295–296, University of Athens, Athens, Greece, 2004.

Bovensmann, H., Burrows, J. P., Buchwitz, M., Frerick, J., Noel, S., Rozanov, V. V., Chance, K. V., and Goede, A. P. H.: SCIAMACHY: Mission objectives and measurement modes, J. Atmos. Sci., 56, 127–150, 1999.

Bracher, A., Lamsal, L., Weber, M., Bramstedt, K., Coldewey-Egbers, M., and Burrows, J. P.: Global satellite validation of SCIAMACHY O3 columns with GOME WFDOAS, Atmos. Chem. Phys., 5, 2357–2368, 2005.

Brasseur, G. P. and Solomon, S.: Aeronomy of the Middle Atmosphere: Chemistry and Physics of the Stratosphere and Mesosphere, Atmospheric and Oceanographic Sciences Library, volume 32, Springer, 3rd edn., ISBN: 978-1-4020-3284-4, 2005.

Brasseur, G. P., Orlando, J. J., and Tyndall, G. S., eds.: Atmospheric Chemistry and Global Change, Oxford University Press, ISBN: 0-19-510521-4, 1999.

Brewer, A. W.: Evidence for a world circulation provided by the measurements of helium and water vapour distribution in the stratosphere, Quart. J. Roy. Meteor. Soc., 75, 351–363, doi:10.1002/qj.49707532603, 1949.

Burrows, J. P., Weber, M., Buchwitz, M., Rozanov, V., Ladstätter-Weißenmayer, A., Richter, A., Debeek, R., Hoogen, R., Bramstedt, K., Eichmann, K.-U., Eisinger, M., and Perner, D.: The Global Ozone Monitoring Experiment (GOME): Mission Concept and First Scientific Results., J. Atmos. Sci., 56, 151–175, doi:10.1175/1520-0469(1999)056<0151:TGOMEG>2.0.CO;2, 1999.

Butchart, N., Scaife, A. A., Bourqui, M., de Grandpré, J., Hare, S. H. E., Kettleborough, J., Langematz, U., Manzini, E., Sassi, F., Shibata, K., Shindell, D., and Sigmond, M.: Simulations of anthropogenic change in the strength of the Brewer-Dobson circulation, Clim. Dyn., 27, 727–741, doi:10.1007/s00382-006-0162-4, 2006.

Butler, J. H., King, D. B., Lobert, J. M., Montzka, S. A., Yvon-Lewis, S. A., Hall, B. D., Warwick, N. J., Mondeel, D. J., Aydin, M., and Elkins, J. W.: Oceanic distributions and emissions of short-lived halocarbons, Global Biogeochemical Cycles, 21, GB1023, doi:10.1029/2006GB002732, 2007.

Callies, J., Corpaccioli, E., Eisinger, M., Hahne, A., and Lefebvre, A.: GOME-2 - METOP's second-generation sensor for operational ozone monitoring, ESA Bull., 102, 28–3–6, 2000.

Callis, L. B., Boughner, R. E., Baker, D. N., Blake, J. B., and Lambeth, J. D.: Precipitating relativistic electrons - Their long-term effect on stratospheric odd nitrogen levels, J. Geophys. Res., 96, 2939–2976, doi:10.1029/90JD02184, 1991.

Chapman, S.: On ozone and atomic oxygen in the upper atmosphere, Phil. Mag., 10, 369, 1930.

Charney, J. G. and Drazin, P. G.: Propagation of planetary-scale disturbances from the lower into the upper atmosphere, J. Geophys. Res., 66, 83–109, doi:10.1029/JZ066i001p00083, 1961.

Charney, J. G. and Stern, M. E.: On the Stability of Internal Baroclinic Jets in a Rotating Atmosphere., J. Atmos. Sci., 19, 159–172, doi:10.1175/1520-0469(1962)019<0159:OTSOIB>2.0.CO;2, 1962.

Chipperfield, M. P.: A three-dimensional model study of long-term mid-high latitude lower stratosphere ozone changes, Atmos. Chem. Phys., 3, 1253–1265, 2003.

Chipperfield, M. P. and Jones, R. L.: Relative influences of atmospheric chemistry and transport on Arctic ozone trends, Nature, 400, 551–554, 1999.

Chipperfield, M. P., Khattatov, B. V., and Lary, D. J.: Sequential assimilation of stratospheric chemical observations in a three-dimensional model, J. Geophys. Res., 107, 4585, doi:10.1029/2002JD002110, 2002.

Cicerone, R. J., Stolarski, R. S., and Walters, S.: Stratospheric Ozone Destruction by Man-Made Chlorofluoromethanes, Science, 185, 1165–1167, doi:10.1126/science.185.4157.1165, 1974.

Cohn, S. E.: An introduction to estimation theory, J. Meteor. Soc. Japan, 75, 257–288, 1997.

Coldewey-Egbers, M., Weber, M., Lamsal, L. N., de Beek, R., Buchwitz, M., and Burrows, J. P.: Total ozone retrieval from GOME UV spectral data using the weighting function DOAS approach, Atmos. Chem. Phys., 5, 5015–5025, 2005.

Crutzen, P. J.: The influence of nitrogen oxides on the atmospheric ozone content, Q. J. R. Meteorol. Soc., 96, 320–325, 1970.

Crutzen, P. J., Isaksen, I. S., and Reid, G. C.: Solar Proton Events: stratospheric sources of Nitric Oxide, Science, 8, 457–459, 1975.

Cutchis, P.: Stratospheric Ozone Depletion and Solar Ultraviolet Radiation on Earth, Science, 184, 13–19, doi:10.1126/science.184.4132.13, 1974.

Dee, D. P.: Testing the perfect-model assumption in variational data assimilation, in: Proc. Second Int. Symp. on Assimilation of Observations in Meteorology and Oceanography, Tokyo, Japan, pp. 225–228, World Meteorological Organization, 1995.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N., and Vitart, F.: The ERA-Interim reanalysis: configuration and performance of the data assimilation system, Quarterly Journal of the Royal Meteorological Society, 137, 553–597, doi:10.1002/qj.828, 2011.

Dessens, O., Zeng, G., Warwick, N., and Pyle, J.: Short-lived bromine compounds in the lower stratosphere; impact of climate change on ozone, Atmos. Sci. Lett., 10, 201–206, doi:10.1002/asl.236, 2009.

Dethof, A. and Hólm, E. V.: Ozone assimilation in the ERA-40 reanalysis project, Q. J. R. Meteorol. Soc., 130, 2851–2872, doi:10.1256/qj.03.196, 2004.

Dhomse, S., Weber, M., Wohltmann, I., Rex, M., and Burrows, J.: On the possible causes of recent increases in northern hemispheric total ozone from a statistical analysis of satellite data from 1979 to 2003, Atmos. Chem. Phys., 6, 1165–1180, 2006.

Dobson, G. M. B., Kimball, H. H., and Kidson, E.: Observations of the Amount of Ozone in the Earth's Atmosphere, and Its Relation to Other Geophysical Conditions. Part IV, Proc. Roy. Soc. London A, 129, 411–433, 1930.

Engel, A., Möbius, T., Bönisch, H., Schmidt, U., Heinz, R., Levin, I., Atlas, E., Aoki, S., Nakazawa, T., Sugawara, S., Moore, F., Hurst, D., Elkins, J., Schauffler, S., Andrews, A., and Boering, K.: Age of stratospheric air unchanged within uncertainties over the past 30years, Nature Geoscience, 2, 28–31, doi:10.1038/ngeo388, 2009.

Errera, Q., Daerden, F., Chabrillat, S., Lambert, J. C., Lahoz, W. A., Viscardy, S., Bonjean, S., and Fonteyn, D.: 4D-Var assimilation of MIPAS chemical observations: ozone and nitrogen dioxide analyses, Atmos. Chem. Phys., 8, 6169–6187, 2008.

Ertel, H.: Ein neuer hydrodynamischer Erhaltungssatz, Naturwissenschaften, 30, 543–544, 1942a.

Ertel, H.: Ein neuer hydrodynamischer Wirbelsatz, Meteor. Z., 59, 277–281, 1942b.

Eyring, V., Waugh, D. W., Bodeker, G. E., Cordero, E., Akiyoshi, H., Austin, J., Beagley, S. R., Boville, B. A., Braesicke, P., Brühl, C., Butchart, N., Chipperfield, M. P., Dameris, M., Deckert, R., Deushi, M., Frith, S. M., Garcia, R. R., Gettelman, A., Giorgetta, M. A., Kinnison, D. E., Mancini, E., Manzini, E., Marsh, D. R., Matthes, S., Nagashima, T., Newman, P. A., Nielsen, J. E., Pawson, S., Pitari, G., Plummer, D. A., Rozanov, E., Schraner, M., Scinocca, J. F., Semeniuk, K., Shepherd, T. G., Shibata, K., Steil, B., Stolarski, R. S., Tian, W., and Yoshiki, M.: Multimodel projections of stratospheric ozone in the 21st century, J. Geophys. Res., 112, D16 303, doi:10.1029/2006JD008332, 2007.

Eyring, V., Shepherd, T. G., and Waugh, D. W., eds.: SPARC Report on the Evaluation of Chemistry-Climate Models, SPARC Report No. 5, WCRP-132, WMO/TD-No. 1526, 2010.

Farman, J. C., Gardiner, B. G., and Shanklin, J. D.: Large losses of total ozone in Antarctica reveal seasonal ClOx/NOx interaction, Nature, 315, 207–210, 1985.

Feng, W., Chipperfield, M. P., Dorf, M., Pfeilsticker, K., and Ricaud, P.: Mid-latitude ozone changes: studies with a 3-D CTM forced by ERA-40 analyses, Atmos. Chem. Phys., 7, 2357–2369, 2007.

Fioletov, V. E., Kerr, J. B., Wardle, D. I., Davies, J., Hare, E. W., McElroy, C. T., and Tarasick, D. W.: Long-term ozone decline over the Canadian Arctic to early 1997 from ground-based and balloon observations, Geophys. Res. Lett., 24, 2705–2708, doi: 10.1029/97GL52829, 1997.

Fisher, M. and Lary, D. J.: Lagrangian four-dimensional variational data assimilation of chemical species, Q. J. R. Meteorol. Soc., 121, 1681–1704, doi:10.1002/qj.49712152709, 1995.

Fleming, E. L., Jackman, C. H., Weisenstein, D. K., and Ko, M. K. W.: The impact of interannual variability on multidecadal ozone simulations, J. Geophys. Res., 112, D10 310, doi:10.1029/2006JD007953, 2007.

Fortuin, J. F. P. and Kelder, H.: An ozone climatology based on ozonesondes and satellite measurements, J. Geophys. Res., 103, 31 709–31 734, 1998.

Frederick, J. E.: Solar corpuscular emission and neutral chemistry in the earth's middle atmosphere, J. Geophys. Res., 81, 3179–3186, doi:10.1029/JA081i019p03179, 1976.

Freier, P. S. and Webber, W. R.: Exponential Rigidity Spectrums for Solar-Flare Cosmic Rays, J. Geophys. Res., 68, 1605–1629, doi:10.1029/JZ068i006p01605, 1963.

Frith, S.: personal communication, 2010.

Gaines, E. E., Chenette, D. L., Imhof, W. L., Jackman, C. H., and Winningham, J. D.: Relativistic electron fluxes in May 1992 and their effect on the middle atmosphere, J. Geophys. Res., 100, 1027–1033, 1995.

Gillett, N. P. and Thompson, D. W. J.: Simulation of Recent Southern Hemisphere Climate Change, Science, 302, 273–275, doi:10.1126/science.1087440, 2003.

Gleason, J. F., Bhartia, P. K., Herman, J. R., McPeters, R., Newman, P., Stolarski, R. S., Flynn, L., Labow, G., Larko, D., and Seftor, C.: Record low global ozone in 1992, Science, 260, 523–526, doi:10.1126/science.260.5107.523, 1993.

Hadjinicolaou, P., Jrrar, A., Pyle, J. A., and Bishop, L.: The dynamically driven long-term trend in stratospheric ozone over northern middle latitudes, Q. J. R. Meteorol. Soc., 128, 1393–1412, 2002.

Hadjinicolaou, P., Pyle, J. A., and Harris, N. R. P.: The recent turnaround in stratospheric ozone over northern middle latitudes: A dynamical modeling perspective, Geophys. Res. Lett., 32, 12 821, doi:10.1029/2005GL022476, 2005.

Haigh, J. D. and Pyle, J. A.: A two-dimensional calculation including atmospheric carbon dioxide and stratospheric ozone, Nature, 279, 222–224, doi:10.1038/279222a0, 1979.

Harris, N., Hudson, R., and Phillips, C., eds.: Assessment of trends in the vertical distribution of ozone. SPARC Rep. 1, WMO-Ozone Research and Monitoring Project Rep. 43, 1998.

Harris, N., Kyrö, E., Staehelin, J., Brunner, D., Andersen, S.-B., Godin-Beekmann, S., Dhomse, S., Hadjinicolaou, P., Hansen, J., Isaksen, G., Jrrar, A., Karpetchko, A., Kivi, R., Knudsen, B., Krizan, P., Lastovicka, J., Maeder, J., Orsolini, Y., Pyle, J., Rex, M., Vanicek, K., Weber, M., Wohltmann, I., Zanis, P., and Zerefos, C.: Ozone trends at northern mid- and high latitudes – a European perspective, Ann. Geophys., 26, 1207–1220, 2008.

Harris, N. R. P., Lehmann, R., Rex, M., and von der Gathen, P.: A closer look at Arctic ozone loss and polar stratospheric clouds, Atmos. Chem. Phys., 10, 8499–8510, doi:10.5194/acp-10-8499-2010, 2010.

Hartmann, D. L. and Lo, F.: Wave-driven zonal flow vacillation in the Southern Hemisphere, J. Atmos. Sci., 55, 1303–1315, 1998.

Heath, D. F., Krueger, A. J., Roeder, H. A., and Henderson, B. D.: The Solar Backscatter Ultraviolet and Total Ozone Mapping Spectrometer (SBUV/TOMS) for Nimbus G, Opt. Eng., 14, 323–331, 1975.

Holton, J. R.: An introduction to dynamic meteorology, International geophysics series, volume 88, Elsevier Academic Press, Burlington, San Diego, London, 3rd edn., ISBN: 978-0-12-354015-7, 2004.

Hood, L. L., McCormack, J. P., and Labitzke, K.: An investigation of dynamical contributions to midlatitude ozone trends in winter, J. Geophys. Res., 102, 13 079–13 093, 1997.

Hsu, J. and Prather, M. J.: Stratospheric variability and tropospheric ozone, J. Geophys. Res., 114, D06102, doi:10.1029/2008JD010942, 2009.

Hunt, B. G.: Photochemistry of Ozone in a Moist Atmosphere, J. Geophys. Res., 71, 1385–1398, 1966.

Hurrel, J. W., Kushnir, Y., Ottersen, G., and Visbeck, M.: An overview of the North Atlantic Oscillation, in: The North Atlantic Oscillation: Climate significance and environmental impact, edited by Hurrel, J. W., Kushnir, Y., Ottersen, G., and Visbeck, M., Geophysical Monograph Series, 134, pp. 1–35, American Geophysical Union, 2003.

Jackman, C. H., Fleming, E. L., and Vitt, F. M.: Influence of extremely large solar proton events in a changing atmosphere, J. Geophys. Res., 105, 11 659–11 670, 2000.

Jackman, C. H., McPeters, R. D., Labow, G. L., Fleming, E. L., Praderas, C. J., and Russell, J. M.: Northern hemisphere atmospheric effects due to the July 2000 Solar Proton Event, Geophys. Res. Lett., 28, 2883–2886, 2001.

Jackman, C. H., Deland, M. T., Labow, G. J., Fleming, E. L., Weisenstein, D. K., Ko, M. K. W., Sinnhuber, M., Anderson, J., and Russell, J. M.: The influence of the several very large solar proton events in years 2000 – 2003 on the neutral middle atmosphere, Adv. Space Res., 35, 445–450, doi:10.1016/j.asr.2004.09.006, 2005.

Jackman, C. H., Marsh, D. R., Vitt, F. M., Garcia, R. R., Randall, C. E., Fleming, E. L., and Frith, S. M.: Long-term middle atmospheric influence of very large solar proton events, J. Geophys. Res., 114, D11 304, doi:10.1029/2008JD011415, 2009.

Jazwinski, A. H.: Stochastic Processes and Filtering Theory, Academic Press, 1970.

Jiang, X., Pawson, S., Camp, C. D., Nielsen, J. E., Shia, R.-L., Liao, T., Limpasuvan, V., and Yung, Y. L.: Interannual variability and trends of extratropical ozone, Part I: Northern Hemisphere, J. Atmos. Sci, 65, 3013–3029, 2008a.

Jiang, X., Pawson, S., Camp, C. D., Nielsen, J. E., Shia, R.-L., Liao, T., Limpasuvan, V., and Yung, Y. L.: Interannual variability and trends of extratropical ozone, Part II: Southern Hemisphere, J. Atmos. Sci, 65, 3030–3041, 2008b.

Johnston, H.: Reduction of Stratospheric Ozone by Nitrogen Oxide Catalysts from Supersonic Transport Exhaust, Science, 173, 517–522, doi:10.1126/science.173.3996.517, 1971.

Jones, A. E. and Shanklin, J. D.: Continued decline of total ozone over Halley, Antarctica, since 1985, Nature, 376, 409–411, 1995.

Kalman, R. E.: A new approach to linear filtering and prediction problems, Trans. ASME, Ser. D: J. Basic Eng., 82, 35–45, 1960.

Kawa, S. R., Newman, P. A., Stolarski, R. S., and Bevacquila, R. M.: Fall vortex ozone as a predictor of springtime total ozone at high northern latitudes, Atmos. Chem. Phys., 5, 1655–1663, 2005.

Kazeminejad, S.: The middle atmosphere's response to energetic particle precipitation, Ph.D. thesis, University of Bremen, Bremen, Germany, 2009.

Khattatov, B. V., Gille, J. C., Lyjak, L. V., Brasseur, G. P., Dvortsov, V. L., Roche, A. E., and Waters, J. W.: Assimilation of photochemically active species and a case analysis of UARS data, J. Geophys. Res., 104, 18 715–18 738, doi:10.1029/1999JD900225, 1999.

Khattatov, B. V., Lamarque, J.-F., Lyjak, L. V., Menard, R., Levelt, P., Tie, X., Brasseur, G. P., and Gille, J. C.: Assimilation of satellite observations of long-lived chemical species in global chemistry transport models, J. Geophys. Res., 105, 29 135–29 144, 2000.

Kidson, J. W.: Indices of the Southern Hemisphere zonal wind, J. Climate, 1, 183–194, 1988.

Kiesewetter, G., Sinnhuber, B.-M., Vountas, M., Weber, M., and Burrows, J. P.: A Long-term Stratospheric Ozone Dataset from Assimilation of Satellite Observations: High-Latitude Ozone Anomalies, J. Geophys. Res., 115, D10 307, doi:10.1029/2009JD013362, 2010a.

Kiesewetter, G., Sinnhuber, B.-M., Weber, M., and Burrows, J. P.: Attribution of stratospheric ozone trends to chemistry and dynamics: a modelling study, Atmos. Chem. Phys., 10, 12 073–12 089, doi:10.5194/acp-10-12073-2010, 2010b.

Kodera, K.: Influence of stratospheric sudden warming on the equatorial troposphere, Geophys. Res. Lett., 33, L06804, 2006.

Kühl, S., Dörnbrack, A., Wilms-Grabe, W., Sinnhuber, B.-M., Platt, U., and Wagner, T.: Observational evidence of rapid chlorine activation by mountain waves above northern Scandinavia, J. Geophys. Res., 109, D22 309, doi:10.1029/2004JD004797, 2004.

Lahoz, W. A., Errera, Q., Swinbank, R., and Fonteyn, D.: Data assimilation of stratospheric constituents: a review, Atmos. Chem. Phys., 7, 5745–5773, 2007.

Li, F., Austin, J., and Wilson, J.: The Strength of the Brewer Dobson Circulation in a Changing Climate: Coupled Chemistry Climate Model Simulations, J. Clim., 21, 40–57, doi:10.1175/2007JCLI1663.1, 2008.

Li, F., Stolarski, R. S., and Newman, P. A.: Stratospheric ozone in the post-CFC era, Atmos. Chem. Phys., 9, 2207–2213, 2009.

Lorenz, D. J. and Hartmann, D. L.: Eddy–zonal flow feedback in the Southern Hemisphere, J. Atmos. Sci., 58, 3312–3327, 2001.

Lorenz, D. J. and Hartmann, D. L.: Eddy–zonal flow feedback in the Nouthern Hemisphere winter, J. Climate, 16, 1212–1227, 2003.

McCormick, M. P., Steele, H. M., Hamill, P., Chu, W. P., and Swissler, T. J.: Polar Stratospheric Cloud Sightings by SAM II., J. Atmos. Sci., 39, 1387–1397, doi:10.1175/1520-0469(1982)039<1387:PSCSBS>2.0.CO;2, 1982.

McElroy, M. B., Salawitch, R. J., Wofsy, S. C., and Logan, J. A.: Reductions of Antarctic ozone due to synergistic interactions of chlorine and bromine, Nature, 321, 759–762, doi: 10.1038/321759a0, 1986.

McLinden, C. A., Olsen, S. C., Hannegan, B., Wild, O., Prather, M. J., and Sundet, J.: Stratospheric ozone in 3-D models: A simple chemistry and the cross-tropopause flux, J. Geophys. Res., 105, 14 653–14 665, 2000.

McPeters, R. D., Heath, D. F., and Bhartia, P. K.: Average ozone profiles for 1979 from the Nimbus 7 SBUV instrument, J. Geophys. Res., 89, 5199–5214, doi: 10.1029/JD089iD04p05199, 1984.

McPeters, R. D., Bhartia, P. K., and et al., A. K.: Earth Probe Total Ozone Mapping Spectrometer (TOMS) Data Products User's Guide, NASA Reference Publication, 1998-206895, 1998.

Ménard, R., Cohn, S. E., Chang, L.-P., and Lyster, P. M.: Assimilation of Stratospheric Chemical Tracer Observations Using a Kalman Filter. Part I: Formulation, Mon. Wea. Rev., 128, 2654, 2000.

Miller, A. J., Nagatani, R. M., Flynn, L. E., Kondragunta, S., Beach, E., Stolarski, R., McPeters, R. D., Bhartia, P. K., DeLand, M. T., Jackman, C. H., Wuebbles, D. J., Patten, K. O., and Cebula, R. P.: A cohesive total ozone data set from the SBUV(/2) satellite system, J. Geophys. Res., 107, 4701, doi:10.1029/2001JD000853, 2002.

Molina, M. J. and Rowland, F. S.: Stratospheric sink for chlorofluoromethanes: chlorine atomc-atalysed destruction of ozone, Nature, 249, 810–812, doi:10.1038/249810a0, 1974.

Molina, M. J., Tso, T.-L., Molina, L. T., and Wang, F. C.-Y.: Antarctic Stratospheric Chemistry of Chlorine Nitrate, Hydrogen Chloride, and Ice: Release of Active Chlorine, Science, 238, 1253–1257, doi:10.1126/science.238.4831.1253, 1987.

Newchurch, M. J., Yang, E.-S., Cunnold, D. M., Reinsel, G. C., Zawodny, J. M., and Russell III, J. M.: Evidence for slowdown in stratospheric ozone loss: First stage of ozone recovery, J. Geophys. Res., 108, 4507, doi:10.1029/2003JD003471, 2003.

Newman, P. A., Gleason, J. F., McPeters, R. D., and Stolarski, R. S.: Anomalously low ozone over the Arctic, Geophys. Res. Lett., 24, 2689–2692, doi:10.1029/97GL52831, 1997.

Newman, P. A., Nash, E. R., and Rosenfield, J. E.: What controls the temperature of the Arctic stratosphere during the spring?, J. Geophys. Res., 106, 19 999–20 010, doi: 10.1029/2000JD000061, 2001.

Newman, P. A., Nash, E. R., Kawa, S. R., Montzka, S. A., and Schauffler, S. M.: When will the Antarctic ozone hole recover?, Geophys. Res. Lett., 33, L12814, doi: 10.1029/2005GL025232, 2006.

Newman, P. A., Daniel, J. S., Waugh, D. W., and Nash, E. R.: A new formulation of equivalent effective stratospheric chlorine (EESC), Atmos. Chem. Phys., 7, 4537–4552, 2007.

NOAA: U.S. Standard Atmosphere 1976, Tech. Rep. NOAA-S/T 76-1562, U.S. Government Printing Office, Washington, D.C., 1976.

Oman, L., Waugh, D. W., Pawson, S., Stolarski, R. S., and Newman, P. A.: On the influence of anthropogenic forcings on changes in the stratospheric mean age, J. Geophys. Res., 114, D03 105, doi:10.1029/2008JD010378, 2009.

Orsolini, Y. J. and Doblas-Reyes, F. J.: Ozone signatures of climate patterns over the Euro-Atlantic sector in the spring, Quart. J. Roy. Meteor. Soc., 129, 3251 – 3263, 2003.

Prather, M.: Catastrophic Loss of Stratospheric Ozone in Dense Volcanic Clouds, J. Geophys. Res., 97, 10 187–10 191, 1992.

Prather, M. J.: Numerical advection by conservation of second-order moments, J. Geophys. Res., 91, 6671–6681, doi:10.1029/JD091iD06p06671, 1986.

Price, J. C.: Timing of NOAA afternoon passes, Int. J. Remote Sensing, 12, 193–198, doi: 10.1080/01431169108929644, 1991.

Ramanathan, V. and Collins, W.: Thermodynamic regulation of ocean warming by cirrus clouds deduced from observations of the 1987 El Nino, Nature, 351, 27–32, doi: 10.1038/351027a0, 1991.

Randel, W. and Wu, F.: Cooling of the Arctic and Antarctic Polar Stratospheres due to Ozone Depletion, J. Clim., 12, 1467–1469, 1999.

Randel, W. J. and Wu, F.: A stratospheric ozone profile data set for 1979–2005: Variability, trends, and comparisons with column ozone data, J. Geophys. Res., 112, D06313, doi: 10.1029/2006JD007339, 2007.

Randel, W. J., Wu, F., Russell III, J. M., Waters, J. W., and Froidevaux, L.: Ozone and temperature changes in the stratosphere following the eruption of Mt. Pinatubo, J. Geophys. Res., 100, 16 753–16 764, 1995.

Randel, W. J., Wu, F., and Stolarski, R.: Changes in column ozone correlated with the stratospheric EP flux, J. Met. Soc. Japan, 80, 849–862, 2002.

Reinsel, G. C., Weatherhead, E. C., Tiao, G. C., Miller, A. J., Nagatani, R. M., Wuebbles, D. J., and Flynn, L. E.: On detection of turnaround and recovery in trend for ozone, J. Geophys. Res., 107, 4078, doi:10.1029/2001JD000500, 2002.

Reinsel, G. C., Miller, A. J., Weatherhead, E. J., Flynn, L. E., Nagatani, R. M., Tiao, G. C., and Wuebbles, D. J.: Trend analysis of total ozone data for turnaround and dynamical contributions, J. Geophys. Res., 110, 16 306, doi:10.1029/2004JD004662, 2005.

Rex, M., Salawitch, R. J., Harris, N. R. P., von der Gathen, P., Braathen, G. O., Schulz, A., Deckelmann, H., Chipperfield, M., Sinnhuber, B.-M., Reimer, E., Alfier, R., Bevilacqua, R., Hoppel, K., Fromm, M., Lumpe, J., Kullmann, H., Kleinbohl, A., Bremer, H., von Konig, M., Kunzi, K., Toohey, D., Vomel, H., Richard, E., Aikin, K., Jost, H., Greenblatt, J. B., Loewenstein, M., Podolske, J. R., Webster, C. R., Flesch, G. J., Scott, D. C., Herman, R. L., Elkins, J. W., Ray, E. A., Moore, F. L., Hurst, D. F., Romashkin, P.,

Toon, G. C., Sen, B., Margitan, J. J., Wennberg, P., Neuber, R., Allart, M., Bojkov, B. R., Claude, H., Davies, J., Davies, W., Backer, H. D., Dier, H., Dorokhov, V., Fast, H., Kondo, Y., Kyro, E., Litynska, Z., Mikkelsen, I. S., Molyneux, M. J., Moran, E., Nagai, T., Nakane, H., Parrondo, C., Ravegnani, F., Skrivankova, P., Viatte, P., and Yushkov, V.: Chemical depletion of Arctic ozone in winter 1999/2000, J. Geophys. Res., 107, 8276, doi:10.1029/2001JD000533, 2002.

Rex, M., Salawitch, R. J., von der Gathen, P., Harris, N. R. P., Chipperfield, M. P., and Naujokat, B.: Arctic ozone loss and climate change, Geophys. Res. Lett., 31, L04116, doi:10.1029/2003GL018844, 2004.

Rodgers, C. D.: Inverse methods for atmospheric sonding: Theory and practice, Series on Atmospheric, Oceanic and Planetary Physics, World Scientific, 2000.

Rosenfield, J. E., Considine, D. B., Meade, P. E., Bacmeister, J. T., Jackman, C. H., and Schoeberl, M. R.: Stratospheric effects of Mount Pinatubo aerosol studied with a coupled two-dimensional model, J. Geophys. Res., 102, 3649–3670, doi:10.1029/96JD03820, 1997.

Rossby, C. G.: Relation between variations in the intensity of the zonal circulation of the atmosphere and the displacements of the semipermanent centers of action, J. Marine Res., 2, 38–55, 1939.

Rozanov, A., Eichmann, K.-U., von Savigny, C., Bovensmann, H., Burrows, J. P., von Bargen, A., Doicu, A., Hilgers, S., Godin-Beekmann, S., Leblanc, T., and McDermid, I. S.: Comparison of the inversion algorithms applied to the ozone vertical profile retrieval from SCIAMACHY limb measurements, Atmos. Chem. Phys., 7, 4763–4779, 2007.

Sander, S. P., Friedl, R. R., Ravishankara, A. R., Golden, D. M., Kolb, C. E., Kurylo, M. J., Molina, M. J., Moortgat, G. K., Finlayson-Pitts, B. J., Wine, P. H., Huie, R. E., and Orkin, V. L.: Chemical Kinetics and Photochemical Data for Use in Atmospheric Studies, Evaluation Number 15, NASA Jet Propusion Laboratory, Pasadena, California, 2006.

Santee, M. L., Manney, G. L., Froidevaux, L., Zurek, R. W., and Waters, J. W.: MLS observations of ClO and HNO_3 in the 1996-97 Arctic polar vortex, Geophys. Res. Lett., 24, 2713–2716, 1997.

Searle, K. R., Chipperfield, M. P., Bekki, S., and Pyle, J. A. .: The impact of spatial averaging on calculated polar ozone loss. 2. Theoretical analysis, J. Geophys. Res., 103, 25 409–25 416, 1998.

Shepherd, T. G. and Jonsson, A. I.: On the attribution of stratospheric ozone and temperature changes to changes in ozone-depleting substances and well-mixed greenhouse gases, Atmos. Chem. Phys., 8, 1435–1444, 2008.

Shine, K. P.: The middle atmosphere in the absence of dynamical heat fluxes, Q. J. R. Meteorol. Soc., 113, 603–633, 1987.

Sinnhuber, B.-M., Chipperfield, M. P., Davies, S., Burrows, J. P., Eichmann, K.-U., Weber, M., von der Gathen, P., Guirlet, M., Cahill, G. A., Lee, A. M., and Pyle, J. A.: Large loss of total ozone during the Arctic winter of 1999/2000, Geophys. Res. Lett., 27, 3473–3476, doi:10.1029/2000GL011772, 2000.

Sinnhuber, B.-M., Weber, M., Amankwah, A., and Burrows, J. P.: Total ozone during the unusual Antarctic winter of 2002, Geophys. Res. Lett., 30, 1580, doi:10.1029/2002GL016798, 2003.

Sinnhuber, B.-M., von der Gathen, P., Sinnhuber, M., Rex, M., König-Langlo, G., and Oltmans, S. J.: Large decadal scale changes of polar ozone suggest solar influence, Atmos. Chem. Phys., 6, 1835–1841, 2006.

Sinnhuber, B.-M., Sheode, N., Sinnhuber, M., Chipperfield, M., and Feng, W.: The contribution of anthropogenic bromine emissions to past stratospheric ozone trends: a modelling study, Atmos. Chem. Phys., 9, 2863–2871, 2009.

Sinnhuber, M., Kazeminejad, S., and Wissing, J. M.: Interannual variation of NOx from the lower thermosphere to the upper stratosphere in the years 1991-2005, J. Geophys. Res., 116, A02 312, 2010.

Solomon, S.: Antarctic ozone: Progress towards a quantitative understanding, Nature, 347, 347–354, 1990.

Solomon, S.: Stratospheric ozone depletion: A review of concepts and history, Rev. Geophys., 37, 275–316, 1999.

Solomon, S., Garcia, R. R., Rowland, F. S., and Wuebbles, D. J.: On the depletion of Antarctic ozone, Nature, 321, 755–758, doi:10.1038/321755a0, 1986.

Solomon, S., Portmann, R. W., Garcia, R., Thomason, L. W., Poole, L. R., and McCormick, M.: The role of aerosol variations in anthropogenic ozone depletion at northern midlatitudes, J. Geophys. Res., 101, 6713–6727, 1996.

Soukharev, B. E. and Hood, L. L.: Solar cycle variation of stratospheric ozone: Multiple regression analysis of long-term satellite data sets and comparisons with models, J. Geophys. Res., 111, D20 314, doi:10.1029/2006JD007107, 2006.

Steinbrecht, W., Claude, H., Köhler, U., and Hoinka, K. P.: Correlations between tropopause height and total ozone: Implications for long-term changes, J. Geophys. Res., 103, 19 183–19 192, 1998.

Stolarski, R. S. and Cicerone, R. J.: Stratospheric chlorine: A possible sink for ozone, Canad. J. Chem., 52, 1610, 1974.

Stolarski, R. S. and Frith, S. M.: Search for evidence of trend slow-down in the long-term TOMS/SBUV total ozone data record: the importance of instrument drift uncertainty, Atmos. Chem. Phys., 6, 4057–4065, 2006.

Stolarski, R. S., Krueger, A. J., Schoeberl, M. R., McPeters, R. D., and Herman, J. R.: Nimbus 7 measurements of the springtime Antarctic ozone decrease, Nature, 322, 808–811, 1986.

Stolarski, R. S., Bloomfield, P., McPeters, R. D., and Herman, J. D.: Total ozone trends deduced from Nimbus 7 TOMS data, Geophys. Res. Lett., 18, 1015–1018, 1991.

Swider, W. and Keneshea, T. J.: Decrease of ozone and atomic oxygen in the lower mesosphere during a PCA event, Planetary and Space Sci., 21, 1969–1973, doi:10.1016/0032-0633(73)90126-8, 1973.

Tegtmeier, S., Fioletov, V. E., and Shepherd, T. G.: Seasonal persistence of northern low- and middle latitude anomalies of ozone and other trace gases in the upper stratosphere, J. Geophys. Res., 113, D21308, 2008.

Thomason, L. W. and Peter, T., eds.: Assessment of Stratospheric Aerosol Properties (ASAP). SPARC Rep. 4, 2006.

Thomason, L. W., Poole, L. R., and Deshler, T.: A global climatology of stratospheric aerosol surface area density deduced from Stratospheric Aerosol and Gas Experiment II measurements: 1984 – 1994, J. Geophys. Res., 102, 8967–8976, 1997.

Thompson, D. J. and Wallace, J. M.: Annular Modes in the Extratropical Circulation. Part I: Month-to-Month Variability, J. Clim., 13, 1000, 2000.

Thompson, D. W. J. and Solomon, S.: Interpretation of Recent Southern Hemisphere Climate Change, Science, 296, 895–899, doi:10.1126/science.1069270, 2002.

Thompson, D. W. J. and Wallace, J. M.: Regional climate impacts of the Northern Hemisphere Annular Mode, Science, 293, 85, 2001.

Toohey, D. W.: A critical review of stratospheric chemistry research in the U.S.: 1991–1994, Rev. Geophys., 33, 759–773, 1995.

Trenberth, K. E. and Guillemot, C. J.: The total mass of the atmosphere, J. Geophys. Res., 99, 23 079–23 088, doi:10.1029/94JD02043, 1994.

Uppala, S. M., KÅllberg, P. W., Simmons, A. J., Andrae, U., Bechtold, V. D. C., Fiorino, M., Gibson, J. K., Haseler, J., Hernandez, A., Kelly, G. A., Li, X., Onogi, K., Saarinen, S., Sokka, N., Allan, R. P., Andersson, E., Arpe, K., Balmaseda, M. A., Beljaars, A. C. M., Berg, L. V. D., Bidlot, J., Bormann, N., Caires, S., Chevallier, F., Dethof, A., Dragosavac, M., Fisher, M., Fuentes, M., Hagemann, S., Hólm, E., Hoskins, B. J., Isaksen, L., Janssen, P. A. E. M., Jenne, R., McNally, A. P., Mahfouf, J.-F., Morcrette, J.-J., Rayner, N. A., Saunders, R. W., Simon, P., Sterl, A., Trenberth, K. E., Untch, A., Vasiljevic, D., Viterbo, P., and Woollen, J.: The ERA-40 re-analysis, Quart. J. Roy. Meteor. Soc., 131, 2961–3012, doi:10.1256/qj.04.176, 2005.

von Clarmann, T., Höpfner, M., Kellmann, S., Linden, A., Chauhan, S., Funke, B., Grabowski, U., Glatthor, N., Kiefer, M., Schieferdecker, T., Stiller, G. P., and Versick, S.: Retrieval of temperature, H_2O, O_3, HNO_3, CH_4, N_2O, $ClONO_2$ and ClO from MIPAS reduced resolution nominal mode limb emission measurements, Atmos. Meas. Tech., 2, 159–175, 2009.

Waibel, A. E., Peter, T., Carslaw, K. S., Oelhaf, H., Wetzel, G., Crutzen, P. J., Pöschl, U., Tsias, A., Reimer, E., and Fischer, H.: Arctic Ozone Loss Due to Denitrification, Science, 283, 2064–2069, doi:10.1126/science.283.5410.2064, 1999.

Wallace, J. M. and Hobbs, P. V.: Atmospheric science: an introductory survey, International geophyscs series, volume 92, Elsevier Academic Press, Burlington, San Diego, London, 2nd edn., ISBN: 978-0-12-732951-2, 2006.

Waugh, D. and Hall, T.: Age of Stratospheric Air: Theory, Observations, and Models, Reviews of Geophysics, 40, 1010, doi:10.1029/2000RG000101, 2002.

Waugh, D. W., Considine, D. B., and Fleming, E. L.: Is upper stratospheric chlorine decreasing as expected?, Geophys. Res. Lett., 28, 1187–1190, doi:10.1029/2000GL011745, 2001.

Waugh, D. W., Oman, L., Kawa, S. R., Stolarski, R. S., Pawson, S., Douglass, A. R., Newman, P. A., and Nielsen, J. E.: Impacts of climate change on stratospheric ozone recovery, Geophys. Res. Lett., 36, L03805, doi:10.1029/2008GL036223, 2009.

Weber, M., Dohmse, S., Wittrock, F., Richter, A., Sinnhuber, B.-M., and Burrows, J. P.: Dynamical control of NH and SH winter/spring total ozone from GOME observations 1995–2002, Geophys. Res. Lett., 30, 1583, doi:10.1029/2002GL016799, 2003.

Weber, M., Lamsal, L. N., Coldewey-Egbers, M., Bramstedt, K., and Burrows, J. P.: Pole-to-pole validation of GOME WFDOAS total ozone with groundbased data, Atmos. Chem. Phys., 5, 1341–1355, 2005.

Weber, M., Lamsal, L. N., and Burrows, J. P.: Improved SCIAMACHY WFDOAS total ozone retrieval: Steps towards homogenising long-term total ozone datasets from GOME, SCIAMACHY, and GOME2, in: Proc. 'Envisat Symposium 2007', Montreux, Switzerland, 23-27 April 2007, ESA SP-636, 2007.

Wofsy, S. C., McElroy, M. B., and Yung, Y. L.: The chemistry of atmospheric bromine, Geophys. Res. Lett., 2, 215–218, doi:10.1029/GL002i006p00215, 1975.

Wohltmann, I., Lehmann, R., Rex, M., Brunner, D., and Mader, J.: A process-oriented regression model for column ozone, J. Geophys. Res., 112, D12 304, doi:10.1029/2006JD007573, 2007.

World Meteorological Organization: Scientific Assessment of Ozone Depletion: 2002. Global Ozone Research and Monitoring Project – Report No. 47, 2003.

World Meteorological Organization: Scientific Assessment of Ozone Depletion: 2006. Global Ozone Research and Monitoring Project – Report No. 50, 2007.

World Meteorological Organization: Scientific Assessment of Ozone Depletion: 2010. Global Ozone Research and Monitoring Project – Report No. 52, 2011.

Yang, E.-S., Cunnold, D. M., Newchurch, M. J., Salawitch, R. J., McCormick, M. P., Russell III, J. M., Zawodny, J. M., and Oltmans, S. J.: First stage of Antarctic ozone recovery, J. Geophys. Res., 113, D20308, doi:10.1029/2007JD009675, 2008.

Danksagung

Wissenschaft ist stets Teamwork, und so haben viele Menschen zum Erfolg meiner Arbeit beigetragen. Vor allem möchte ich mich bei Prof. John Burrows bedanken – für die Möglichkeit, in seiner Gruppe am Institut für Umweltphysik (IUP) Doktorand zu sein, für das Vertrauen, das er mir entgegenbrachte, dafür, dass er mich stets gleichermaßen förderte und forderte. John hatte immer ein offenes Ohr und war stets eine Quelle für sowohl wissenschaftlichen als auch persönlichen Rat ("What you can't finish between 7 a.m. and 7 p.m., you shouldn't try to finish between 7 p.m. and 7 a.m."). Auch wenn ich nicht immer geschafft habe, alle (wissenschaftlichen wie persönlichen, z.B. den oben genannten) Ratschläge in die Tat umzusetzen, bin ich doch stolz darauf, behaupten zu können, vieles von John gelernt zu haben.
Die Hauptlast meiner Betreuung als Doktorand lag bei Dr. Björn-Martin Sinnhuber. Ich stehe tief in Björn-Martins Schuld für die kompetente und umfassende Betreuung, die er mir entgegenbrachte, für die ungezählten fruchtbaren Diskussionen, für all die Ideen, die er hatte, und die Unterstützung bei der Umsetzung. Auch nach seinem Wechsel ans Karlsruhe Institute of Technology (KIT) im Sommer 2010 war Björn-Martin weiterhin eine Stütze für meine Arbeit. Björn-Martin ist es maßgeblich zu verdanken, dass ich den Umstieg von der experimentellen Quantenphysik, meinem Diplomarbeitsgebiet, in die Modellierung der Atmosphäre einigermaßen rasch und problemlos schaffte.
Weiters möchte ich mich bei meinen Kollegen am IUP bedanken, allen voran meinen Zimmerkollegen Dr. Shahin Kazeminejad, vormals Mag. Shahin Kazeminejad (do wor ma r obar aa no jünga!), sowie Jan Aschmann und Martin Langowski. Sie alle haben durch ihren Humor, ihre Herzlichkeit und ihre wissenschaftliche Kompetenz unser Büro zu einem angenehmen Ort gemacht. Jan und Martin haben bewiesen, dass Integration wahrlich keine Einbahnstraße ist, und viel Geduld aufgewendet, um eine solide Basis der Wiener Umgangssprache zu erlernen. Vor allem bei Jan möchte ich mich für die perfekte Zusammenarbeit bedanken; sei es MatLab, Fortran oder LaTeX, Jan hatte als Informatiker fast immer den richtigen Tip auf Lager, um anstehende Probleme zu lösen. Außerdem möchte ich Jan für die vielen hilfreichen Anmerkungen zu dieser Dissertation danken.
Besonderer Dank geht auch an Dr. Mark Weber und Dr. Marco Vountas, die mir mit fachlichem Rat und hilfreichen Anmerkungen bei meinen Veröffentlichungen zur Seite standen. Ferner möchte ich auch meinem Zweitgutachter, Prof. Justus Notholt, danken, der sich kurzfristig bereiterklärt hat, meine Doktorarbeit zu begutachten. Meinen Kollegen Ali, Conny, Luca, Janina möchte ich sehr herzlich für zahllose Zwischendurch-Plaudereien, Passivrauchpausen und Mensabesuche danken, sowie auch allen anderen Arbeitskollegen für die kollegiale Atmosphäre am Institut.
Viele Kollegen und Institutionen in aller Welt haben zu dieser Arbeit beigetragen. Prof. Martyn Chipperfield (University of Leeds, UK) hat freundlicherweise die Assimilationsroutine zur Verfügung gestellt, die hier in leicht abgewandelter Form zum Einsatz kam. Dr. Michael Prather und Dr. Juno Hso (University of California at Irvine, USA) haben die hier verwendeten Linoz-Chemietabellen für mich angefertigt. Die in der Assimilation verwendeten

SBUV-Satellitendaten wurden mir von der National Aeronautics and Space Administration (NASA) zur Verfügung gestellt, ebenso wie die TOMS/SBUV sowie SBUV merged Ozon-Datensätze. Die im SBUV MOD angewendeten Korrekturterme wurden mir von Dr. Stacey Frith (NASA Goddard Space Flight Center, Maryland, USA) persönlich zur Verfügung gestellt. Meteorologische Reanalysen, mit denen das CTM angetrieben wird, wurden unter dem Special Project DECDIO vom European Centre for Medium Range Weather Forecasts (ECMWF) bereitgestellt. Ozon-Sondendaten stammen vom Network for the Detection of Atmospheric Composition Change (NDACC), www.ndacc.org (Ny-Ålesund, Neumayer, South Pole), sowie von Dr. Rigel Kivi (Finnish Meteorological Institute) für Sodankylä. Der GOME/SCIAMACHY/GOME2 Totalozon-Datensatz wurde von Dr. Mark Weber zur Verfügung gestellt; SCIAMACHY-Profilozondaten stammen von Dr. Alexei Rozanov (IUP). MIPAS-Daten wurden freundlicherweise von Dr. Gabi Stiller (Karlsruhe Institute of Technology) bereitgestellt. Der in der Analyse verwendete Northern Hemisphere Annular Mode (NAM)-Index wurde von Dr. Mark Baldwin (Northwest Research Associates, Seattle, USA) berechnet und von Dr. Baldwins Annular Modes Homepage heruntergeladen (http://www.nwra.com/resumes/baldwin/).
Meine Doktorarbeit entstand im Rahmen meiner Mitarbeit an dem Projekt "DACCS – Data Assimilation and Model Calculations to Study Chemistry Climate Interactions and Solar Influence in the Polar Stratosphere", das von der Deutschen Forschungsgemeinschaft (DFG) unter dem Schwerpunktprogramm "CAWSES – Climate And Weather of the Sun Earth System" gefördert wurde. Ohne diese Finanzierung wäre meine Arbeit nicht möglich gewesen. Auch unserer Projektpartnerin an der Freien Universität Berlin, Prof. Ulrike Langematz, möchte ich sehr herzlich für die gute Zusammenarbeit danken, sowie für die Gelegenheit, meine Forschungsergebnisse im Rahmen eines Institutskolloquiums an der FUB zu präsentieren.
Schließlich möchte ich auch noch allen Menschen in meinem Umfeld danken, die dazu beigetragen haben, meine Zeit in Bremen angenehm zu gestalten, und Freude wie Frust während der Doktorarbeit teilten. Ich habe hier wunderbare Menschen kennengelernt, die es hervorragend verstanden haben, mich stets aufzumuntern, abzulenken und zu unterstützen. Hier sind vor allem meine Freunde, Shahin, Bobby, Matze, Farid, Marco, Lore zu nennen, die sich in vielen Abenden das Leid des Doktoranden anhörten, sowie meine Mitbewohner, Clara, Nathalie, Susi, Christoph, Lena, die geduldig über manche Verzögerungen des Putzplanes zugunsten meiner wissenschaftlichen Arbeit hinwegsahen, und mit denen ich nicht missen möchte, zusammen gelebt zu haben. Besonders bedanken möchte ich mich bei meiner Freundin Anna, die die Entbehrungen einer Fernbeziehung auf sich genommen und dabei das Kunststück geschafft hat, mir auch über eine Distanz von 1000 km viel Liebe, Geborgenheit und Unterstützung zu geben. Nicht zuletzt möchte ich mich bei meinen Eltern Roland und Erika bedanken, die mich auf meinem Weg immer unterstützt und mir mein Studium ermöglicht haben. Durch sie wurde meine Begeisterung für die Wissenschaft geweckt, und ohne ihren Rückhalt hätte ich es nicht so weit geschafft.