A brief history of stratospheric ozone research

ROLF MÜLLER*

ICG-1, Forschungszentrum Jülich, Jülich, Germany

(Manuscript received September 3, 2008; in revised form November 14, 2008; accepted November 14, 2008)

Abstract

Ozone is one of the most important trace species in the atmosphere. Therefore, the history of research on ozone has also received a good deal of attention. Here a short overview of ozone research (with a focus on the stratosphere) is given, starting from the first atmospheric measurements and ending with current developments. It is valuable to study the history of ozone research, because much can be learned for current research from an understanding of how previous discoveries were made. Moreover, since the 1970s, the history of ozone research has also encompassed also the history of the human impact on the ozone layer and thus the history of policy measures taken to protect the ozone layer, notably the Montreal Protocol and its amendments and adjustments. The history of this development is particularly important because it may serve as a prototype for the development of policy measures for the protection of the Earth’s climate.

1 Introduction

In 1839, in a lecture to the “Naturforschende Gesellschaft Basel”, Christian Friedrich Schönbein announced the discovery of a new substance that he would later call “ozone” (SCHÖNBEIN, 1839). In the years after its discovery, the new substance aroused the interest of the scientific community; e.g., Justus von Liebig said “... betrachtete ich die Entdeckung des ozonisirten Sauerstoffs für eine der merkwürdigsten, die je gemacht wur-
den”1 (KAHLBAUM and THON, 1900, p. 18). Interestingly, however, it took many years before the name “ozone”, which we use quite naturally today, was generally accepted for the new substance (see HERRMANN (1982) and RUBIN (2001) for a more extensive report on Schönbein’s work).

Ozone is one of the most important trace species in the atmosphere. The presence of an ozone layer in the lower stratosphere is a key factor determining the vertical temperature profile of the Earth’s atmosphere. Further, UV-B radiation from the sun is absorbed in the ozone layer; if this radiation was not absorbed, it would reach the Earth’s surface to an extent that would be harmful to a variety of life forms (including humans). Excess ozone at the Earth’s surface (“photochemical smog”) occurs because of air pollution and can be harmful to plants, animals, and humans. On the other hand, a certain ozone level in the lower atmosphere is beneficial, because without ozone the chemical cycles that remove pollutants from the atmosphere cannot work.

Owing to the importance of ozone for the Earth’s atmosphere, the history of atmospheric ozone research has received much attention. Over the years, many papers and books have appeared describing various aspects of ozone research and in various levels of detail (DOBSON, 1968; KHURDIAN, 1975; TUCK, 1978; DOTTO and SCHIFF, 1978; HERRMANN, 1982; CRUTZEN, 1988; FARMA, 1989; CRUTZEN, 1996; SOLOMON, 1999; RUBIN, 2001; STOLARSKI, 2001; SOLOMON, 2004; FARRELL, 2005). The history of ozone research in the last several decades is furthermore a good example of the knowledge cycle in geosciences today, where field measurements and laboratory studies are combined to advance the understanding of processes leading to an improvement of numerical models, and the results and forecasts of the numerical models are then confronted again with measurements (BRASSEUR, 2008).

A feature of research on atmospheric issues that becomes obvious when studying the history of ozone research is that the scarcity of observational data has al-

1 Author’s address: Rolf Müller, ICG-1, Forschungszentrum Jülich, 52425 Jülich, Germany, e-mail: ro.mueller@fz-juelich.de

I consider the discovery of ozonised oxygen [i.e., ozone] as one of the most peculiar ever made.
ways been a fundamental problem hindering progress. And in spite of all the technological advances achieved so far, it still is. It is difficult and expensive to obtain measurements several kilometres above the ground and the gases to be measured are only present in the atmosphere in trace amounts (for example the mixing ratios of some chlorofluorocarbons (CFCs) in the stratosphere are orders of magnitude below one particle in $10^{12}$). Measurements in the atmosphere are never repeatable so that long-term, consistent time series are very important. A prominent example is the time series of total column ozone measurements in Halley Bay, Antarctica, that was started in 1956 and which formed the basis for the discovery of the “ozone hole” by Farman et al. (1985). Moreover, the interpretation of atmospheric measurements is always complicated, because advection, mixing, chemistry, and radiation are all important for the composition of the atmosphere.

Arguably, the most eminent example of the impact of human activities on nature today is the ozone hole in the Antarctic (see Section 5.1 below). The discovery of the ozone hole by Farman et al. (1985) was a key factor leading to the “Montreal Protocol on Substances that Deplete the Ozone Layer”, which was signed in 1987. Consequently, the history of research on stratospheric ozone depletion has recently also aroused interest as an example of how the dynamics of scientific learning and scientific assessment processes can be studied (Brasseur, 2008; Crutzen and Oppenheimer, 2008), which is an important issue in the current debate about climate change.

The outline of the paper is as follows. The early observations of atmospheric ozone are introduced in Section 2, the discussion starts with the discovery of ozone and the first tropospheric measurements, covers the discovery of the role of ozone as an absorber of solar radiation, the first measurements of total column ozone, and the findings from the early 1930s that the height of the ozone layer is $\sim 25$ km rather than 40–50 km as had been hitherto thought.

Section 3 presents the development of the understanding of the chemistry of stratospheric ozone, from the early work by Chapman (1930) to the discovery of the importance of catalytic cycles in destroying ozone. The development of concerns about anthropogenic effects on the ozone layer is described in Section 4. The effects discussed are nuclear weapons tests in the stratosphere, the emissions of nitrogen oxides from supersonic aircraft, the use of artificial fertilisers in agriculture, followed by a review of the warning by Molina and Rowland (1974) that the accumulation of anthropogenic CFCs in the atmosphere might ultimately lead to a depletion of the ozone layer.

The discovery of the ozone hole (Farman et al., 1985; Stolarski et al., 1986; WMO, 2007) is described in Section 5, the related issue of polar ozone loss in the Arctic in late winter and spring is introduced, and developments towards an increasing understanding of the phenomenon are described. Finally, in the conclusions (Section 6), the timeline of discoveries in ozone research is compared to milestones in the history of physics.

2 Early observations of atmospheric ozone

Schönbein himself believed that ozone played an important role in the Earth system and suggested in 1853 that long-term ozone measurements in the atmosphere should be performed on an international scale: “Geneigt zu glauben, das atm. Ozon spiele im Haushalte der Erde eine wichtige Rolle, halte ich es f"ur w "unschenswerth, dass m "oglichst zahlreiche, sowohl grosse Zeitr "aume als bedeutende L"anderstrecken umfassende, untereinander vergleichbare Beobachtungen "{u}ber die Ver"anderungen des Ozongehaltes der Atmosph"are angestellt werden...”2 (in a letter to Justus von Liebig, Kahlbaum

2Inclined to believe that atmospheric ozone plays an important role in the balance of the Earth, I consider it desirable that as many observations as pos-
and THON, 1900, p. 10). Indeed numerous measurements were made during the second half of the 19th century using a method that Schönbein had developed (for a detailed discussion of the early measurements of atmospheric ozone see e.g., CRUTZEN, 1988; LONDON and LIU, 1992; SONNEMANN, 1992). Unfortunately, measurements using Schönbein’s method are only semi-quantitative because of poor standardisation and the influence of humidity and wind speed on the values obtained (LINVIL et al., 1980; VOLZ and KLEY, 1988). The only historic measurements of tropospheric ozone that proved reliable enough to allow a reanalysis and a meaningful comparison with modern measurements were taken at the Observatorio de Montsouris, located on the outskirts of Paris (VOLZ and KLEY, 1988). At Montsouris a quantitative wet-chemical method was established in 1876 and used continuously for 34 years.

This early interest in ozone was motivated to a large extent by the fact that ozone, at that time, was considered to be an indicator of clean, healthy air and that ozone-poor air would lead to illness. Measurement series of ozone were performed in spa towns with the intention of demonstrating how beneficial to human health a stay would be (LENDER, 1872). This view persisted for many decades and was held by leading scientists in the field; as late as 1946 Erich Regener, one of the early pioneers in ozone research (Figure 1), stated “Anwesenheit von Ozon ist ein Indikator für gute Luft.”3 (REGENER, 1946). Today, of course, it is known that high concentrations of tropospheric ozone are detrimental to human health and, similarly, harmful to other living systems.

Since the early 1950s, it has been understood that high ozone concentrations in the troposphere (referred to as “photochemical smog”) are caused by pollution; ozone is chemically formed as a result of photochemical reactions involving NOx and non-methane hydrocarbons from automobile exhausts and similar combustion processes (HAAGEN-SMITH, 1952).

Still in the 19th century, another important aspect of ozone was discovered, namely its importance as an absorber of light. In 1879 the French physicist Alfred Cornu recognised that solar radiation with wavelengths below 300 nm does not penetrate to the Earth’s surface (CORNU, 1879a) 4. Only three years later, HARTLEY (1881b) suggested that ozone is responsible for this observation; the ozone absorption bands in this UV region are therefore today referred to as Hartley bands. Already HARTLEY (1881a) believed that “ozone is a normal constituent of the higher atmosphere” (see KHRJIAN (1975) for more details on Hartley’s work). The discovery of the spectroscopic properties of ozone opened up the possibility of making measurements of the total thickness of the atmospheric ozone (i.e., the total atmospheric column of ozone) including the ozone in the stratosphere. The first detailed measurements of the total atmospheric column of ozone were made by FABRY and BUISSON (1921), who firmly established that ozone was responsible for the observed strong cut-off in solar UV spectra at about 300 nm towards shorter wavelengths. FABRY and BUISSON (1921) speculated (correctly as we know today) that ozone was formed by solar UV radiation. They also suggested that the ozone layer was situated at an altitude of about 50 km. And first measurements of the height of the ozone layer indeed confirmed an altitude of the ozone layer of 48–53 km (CABANES and DUFA, 1925; LAMBERT et al., 1926).

In 1924, G. M. B. Dobson designed a spectrograph to measure the total ozone column that was more suited for routine out-of-door use and that was cheaper to build than the FABRY and BUISSON spectrograph. The aim was to develop an instrument that allowed regular measurements to be made over extended time periods. This development was very successful; today, the standard unit for the total atmospheric column of ozone bears Dobson’s name: the Dobson unit (DU). One DU is defined as 2.687 × 1016 ozone molecules per square centimetre. The Dobson unit describes the thickness of a layer of pure ozone, if the total amount of ozone in the atmosphere were brought to standard conditions (15 degrees Celsius and 1013 hPa). For example, an atmospheric ozone column of 300 DU (a typical atmospheric value) brought down to the surface of the Earth would occupy a 3 mm thick layer of pure ozone.

The first observations at Oxford 1924-1925 showed a marked annual variation of ozone. Likewise, the data exhibited a strong day-to-day variability that was closely connected to meteorological conditions (DOBSON and HARRISON, 1926; DOBSON, 1968). Between July 1926 and November 1927 seven instruments were built at Oxford and distributed throughout Europe covering the latitude range 68°N to 47°N and one instrument was operated at 22°S in Chile (DOBSON et al., 1929; DOBSON, 1930). After the 1926-1927 observation period, the instruments were sent to Spitzbergen (via an Italian airship expedition), California, Egypt, India, and New Zealand (see DOBSON, 1968; FARMAN, 1989; BRÖNNMANN et al., 2003, for further details on the history of early total ozone measurements). In the former USSR, the first measurements of total ozone were taken in 1933; later, in 1959, the M-83 ozonometer was developed, which became the backbone of the USSR ozone station network (GUSHCHIN, 1995).

Even today, ground-based total ozone measurements are essential as a reference measurement for satellite data and thus for monitoring the ozone content of the atmosphere. Further, the longest records of continu-
ous reliable measurements are available from stations equipped with Dobson spectrophotometers (e.g., Jones and Shanklin, 1995; Staehelin et al., 1998), whose design has remained essentially unchanged since 1930 (Dobson, 1968). Such long data records are invaluable for time series analysis and for trend detection (e.g., Fioletov et al., 2002; Jánosi and Müller, 2005; WMO, 2007).

In the late twenties, the ozone layer was still assumed to be located in the upper stratosphere. Götz and Dobson (1928), based on their measurements, reported that the “...average height [of the ozone layer] seems to be between 30–40 km above sea-level”. Bjerknes believed in 1929 that “...le fait qui parait aussi être bien établi, que la couche d’ozone se trouve à une altitude de 40–50 kilomètres” (Bjerknes, 1929).

When Götz and Dobson (1928) and Bjerknes (1929) refer to the height of the ozone layer they mean the altitude of the maximum ozone concentration in number density (measured in particles per unit volume or in DU per km). A different, greater altitude results from considering the maximum of the molar mixing ratio (measured in micromole ozone per mole of air, commonly referred to as ppm).

In ground-based measurements in Spitzbergen in 1929, Götz (1931) discovered the so-called “Umkehreffekt”, which allows a low-resolution vertical ozone profile to be deduced (e.g., Paetzold and Regener, 1957; Dutsch and Staehelin, 1992). He reported an average height of the ozone layer for the period 10 July to 28 August 1929 in Spitzbergen of 27.6 ± 8 km. At the time, Götz put forward these new ideas rather carefully “...es fragt sich nun ob die geophysikalisch interessante tiefere Lage der Ozonschicht in der hohen Breite Spitzbergens, vor allem aber auch ihre Schwankung bis herunter zur minimalen Höhe des 26. August [11 km] als genügend gesichert angesehen werden dürften”. However, two years later the Spitzbergen measurements were confirmed by mid-latitude measurements and Götz et al. (1933) stated with much more confidence that “the average height [of the ozone in the atmosphere] at Arosa now appears to be about 20 km, which is much below the former estimates”.

The first attempt to directly measure the vertical profile of the atmospheric ozone concentration (and other atmospheric parameters) was made by James Glaisher and Henry Tracey Coxwell on a manned balloon. On 5 September 1862 they made a balloon ascent from Wolverhampton, England, and measured ozone using Schönbein’s method. Although the balloon reached an altitude of 11 km, the last ozone measurement was only performed at an altitude of 4.8 km because the balloonists lost consciousness at greater altitudes. This balloon flight was described by Lender (1873) and later in the New York Times, 27 June (1909); see also Hoinka (1997) for a report on the flight. Between 1862 and 1866 Glaisher made numerous further balloon ascents in order to measure the temperature and humidity of the atmosphere at greater heights. He was also a founder member of the Meteorological Society in 1850, which later (in 1883) became the Royal Meteorological Society.

The first successful balloon-borne measurements of the vertical distribution of ozone were conducted soon after the invention of the instruments.

---

5. The fact seems also to be well established that the ozone layer is located at an altitude of 40–50 kilometres.

6. The question is now whether the geophysically interesting lower altitude of the ozone layer in the high latitude stations, in particular also its variation down to the minimum height on 26 August [11 km], can be regarded as being sufficiently well established.

### Table 1: The first balloon-borne ozone profile. The table lists the original measurements reported by Regener and Regener (1934) (digitised from their Fig. 4). The ozone column (in cm; 1 cm corresponds to 1000 DU) above the balloon-borne instrument is reported as a function of height.

<table>
<thead>
<tr>
<th>Height (km)</th>
<th>O₃ column (cm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1</td>
<td>0.266</td>
</tr>
<tr>
<td>5.1</td>
<td>0.254</td>
</tr>
<tr>
<td>8.1</td>
<td>0.250</td>
</tr>
<tr>
<td>10.1</td>
<td>0.240</td>
</tr>
<tr>
<td>12.2</td>
<td>0.230</td>
</tr>
<tr>
<td>14.7</td>
<td>0.218</td>
</tr>
<tr>
<td>16.8</td>
<td>0.205</td>
</tr>
<tr>
<td>18.9</td>
<td>0.189</td>
</tr>
<tr>
<td>20.8</td>
<td>0.172</td>
</tr>
<tr>
<td>22.5</td>
<td>0.152</td>
</tr>
<tr>
<td>24.5</td>
<td>0.126</td>
</tr>
<tr>
<td>26.6</td>
<td>0.107</td>
</tr>
<tr>
<td>28.4</td>
<td>0.087</td>
</tr>
<tr>
<td>31.0</td>
<td>0.071</td>
</tr>
</tbody>
</table>

---

### Figure 2: The first balloon-borne measurement of the vertical distribution of ozone in the atmosphere from Stuttgart (48.8°N) on 31 July 1934 (Regener and Regener, 1934). Here, the original measurements of Regener and Regener are shown reported as column ozone above the balloon. Note that the absolute values reported by Regener and Regener (1934) were deduced from relative measurements assuming a value of 240 DU at 10 km. The altitude was deduced as a log-pressure height from the pressure recorded on the balloon payload.
after the first Umkehr measurements. In a balloon flight on 31 July 1934 from Stuttgart, Regener and Regener (1934) recorded the solar UV spectrum in the stratosphere and deduced an ozone profile from these measurements. These data independently confirmed the first Umkehr observations. The payload for this balloon experiment weighted only 2.7 kg and was launched to an altitude of about 31 km using two standard meteorological balloons. Table 1 lists the original values (ozone column above the balloon) reported by Regener and Regener (1934) and Figure 2 shows these data as they were presented in the original paper. In Figure 3, the vertical profile of the ozone mixing ratio deduced from these early measurements is compared with a corresponding ozone climatology deduced from measurements of the Halogen Occultation Experiment (HALOE) on the Upper Atmosphere Research Satellite (UARS) in the years 1991-2002 (Russell et al., 1993; Grooss and Russell, 2005) demonstrating a rather good agreement of the early measurements with a current climatology.

In the early fifties, Paetzold (1950) developed a method for determining the vertical profile of ozone during lunar eclipses up to altitudes that cannot be reached by balloons (up to about 45 km). In this method, the ozone profile is determined from the ozone absorption in the Chappius band measured at the edge of the Earth’s shadow on the moon. The effectiveness of this method stems from the large effective ozone column (up to 15000 DU) due to the tangential optical path through the Earth’s atmosphere. This effect is visible during lunar eclipses as a small area at the edge of the umbra showing a greenish colour rather than the reddish colour in the core of the umbra. This greenish colour is caused by the extinction of the solar light reaching the moon in the yellow-red spectral region by the Chappius band of ozone (Paetzold, 1951).

Erich Regener was forced to retire from his chair at the University of Stuttgart in 1937 by the Nazi regime. Nonetheless, he continued his scientific work; he first founded a private research institute on 1 January 1938, his “Forschungsstelle für Physik der Stratosphäre”, which was integrated later that year into the “Kaiser Wilhelm Society for the Advancement of Science” (which was refounded in 1948 as the “Max Planck Society”). He developed the first scientific payload for a rocket that allowed it to reach high altitudes (about 150 km); the rocket was known as the A4. He designed an air- and water-tight container containing an assembly of instruments, in particular temperature and pressure gauges and a spectrometer to register solar UV radiation (Ordway III et al., 2007). This payload, referred to as the “Regener-Tonne” (Regener barrel) was never flown, but the first scientific payload (a Geiger counter, which however failed to work) was launched on an A4 rocket from New Mexico as early as 16 April 1946 (Green, 1954). Many more flights followed, including flights of A4 rockets carrying UV spectrometers for ozone measurements (Johnson et al., 1951; Ordway III et al., 2007). The first in situ measurement of an ozone profile above balloon altitudes (Figure 4) was made on 10 October 1946 by a UV spectograph mounted in a tail fin of an A4 rocket (Johnson et al., 1951).

Simpler and less expensive rockets were soon developed for upper atmosphere research (Van Allen et al., 1948). For many years, rocket measurements were the
key experimental tool for exploring ozone chemistry in the upper stratosphere and in the mesosphere. In the early sixties, the first satellite measurements of the vertical distribution of ozone in the stratosphere and mesosphere were made using the same technique as for the lunar eclipses but replacing the moon as a reflector by an artificial satellite (Venkateswaran et al., 1961). Nonetheless, rocket measurements remained an important scientific tool for decades to come (e.g., Miller and Ryder, 1973; Vaughan, 1982, 1984; Gernandt et al., 1989). An important feature of the vertical ozone profile, the secondary ozone maximum at about 90 km, was discovered both in ozone measurements from rocket payloads and in ozone concentrations in the upper atmosphere deduced from stellar occultation measurements made from a satellite (Miller and Ryder, 1973; Hays and Roble, 1973).

3 The chemistry of stratospheric ozone

When Dobson and Harrison first published their measurements of total ozone in 1926, the formation mechanism of stratospheric ozone was still unclear. They discussed “whether the ozone is formed in the extreme upper atmosphere by ultra-violet radiation from the sun, or by electrical discharges in aurorae...”. Four years later, Chapman (1930) proposed the first photochemical theory of the formation of ozone; four reactions that involved only allotropes of oxygen and that are today known as the Chapman reactions. He suggested that stratospheric ozone is produced by the photolysis of molecular oxygen (O$_2$) at ultraviolet wavelengths below 242 nm,

$$R1: \ O_2 + h\nu \rightarrow 2O(^3P),$$

where $h\nu$ denotes an ultraviolet photon.

The atomic oxygen (O($^3P$), hereafter denoted by O) produced in reaction R1 reacts rapidly with molecular oxygen to form ozone (O$_3$)

$$R2: \ O + O_2 + M \rightarrow O_3 + M,$$

where M denotes a collision partner (N$_2$ or O$_2$) that is not affected by the reaction. Ozone is photolysed rapidly

$$R3: \ O_3 + h\nu \rightarrow O + O_2.$$

Through the reaction

$$R4: \ O + O_3 \rightarrow 2O_2,$$

an O atom and an O$_3$ molecule are lost. Because O and O$_3$ are in rapid photochemical equilibrium, the loss of one oxygen atom effectively implies the loss of an ozone molecule, so that R4 destroys two molecules of odd oxygen.

The Chapman reactions R1-R4 were accepted for decades as sufficient to describe the observed vertical distribution of ozone. However, new measurements of the rate constants for the Chapman reactions (Benson and Axworthy, 1957; Jones and Davidson, 1962) demonstrated that a quantitative understanding of ozone chemistry was lacking. It became clear that the destruction of ozone by reaction R4 alone cannot explain the observed ozone abundances in the stratosphere (e.g., Hunt, 1966b; Schiff, 1969); as Hunt (1966a) put it “…the photochemical reaction scheme normally considered is no longer adequate, and there must be reactions occurring in the atmosphere which destroy ozone but which have been neglected in the past”.

Bates and Nicolet (1950) had originally suggested that reactions involving OH and HO$_2$ radicals would lead to a significant catalytic ozone loss. Based on this work, HO$_2$-catalysed reactions were suggested as an ozone destruction mechanism by Hampson (1964, 1966) and, using laboratory measurements by Norrish and Wayne (1965), implemented in a model calculation by Hunt (1966b). Crutzen (1969) found that the mechanism involving OH and HO$_2$ could not explain ozone observations in the middle stratosphere between 30–35 km and, a year later, proposed that NO- and NO$_2$-catalysed reactions control the ozone concentrations in the middle stratosphere (Crutzen, 1970). At that time, the very first measurements of nitrogen compounds in the stratosphere had just become available (Murcray et al., 1968; Rhine et al., 1969). The possibility that nitrous oxide (N$_2$O) is decomposed in the stratosphere via the reaction

$$R5: \ N_2O + O(^1D) \rightarrow 2NO,$$

and thus constitutes a possible source for stratospheric NO and NO$_2$ had already been suggested by Hampson (1964)$^5$. At that time he concluded that “it is unlikely that the nitrogen oxide reactions can affect total

---

$^5$The author learned about the CARDE (Canadian Armament Research and Development Establishment) Report No. 1627 (Hampson, 1964) and obtained a copy of the report from Adrian Tuck who recalls the following story: “At the second joint Anglo-French meeting at Jesus College, Oxford in September 1974, on the effects of Concorde on the ozone layer, Hampson was there and referred to CARDE Report No. 1627 in his talk. I asked the Met Office library to get me a copy of Report 1627 from CARDE in 1975. The CARDE library replied, saying they had kept a copy until a month ago, when they had a request for it from another scientist, and because it was an old report they had let him keep this last copy. Fast forward to 1997, when I was sitting in Carl Howard’s office at the Aeronomy Lab and we were reminiscing about the history of the subject. I told him about the above account, he grinned and picked that last copy from CARDE off his shelf. I made a copy, which I still have.”

In summer 2008 the author wrote to CARDE asking again whether copies of the report were still available. Months later, together with the reviews for this paper, a letter from CARDE arrived. The CARDE library, this time, provided a photocopy of the report. However, even in 2008, release of the document to a scientist had to be authorized first; the document had to be downgraded from “defense departments only” although it was always unclassified.
ozone but it is possible that observation of them, knowing the rates of the reactions involved, can assist our understanding of the distribution and behaviour of ozone” (HAMPSON, 1964, p. 267).

Again a few years later, the possibility of chlorine catalysed ozone loss was put forward by STOLARSKI and CICERONE (1974). Their original motivation for investigating the impact of chlorine on ozone was studying the possible effects on the stratosphere of HCl originating from space shuttle exhaust, but soon the focus shifted to chlorine injected into the stratosphere by volcanoes (STOLARSKI, 2001). At that time, laboratory studies of chlorine reactions had been begun employing the latest technique. STOLARSKI (2001) recalls that as early as 1972 he heard laboratory chemist Don Stedman say in a conversation: “Chlorine destroys ozone. Everybody knows that!”.

The different catalytic ozone loss cycles can be summarised in the form

\[
\begin{align*}
R6: & \quad \text{XO} + \text{O} & \rightarrow & X + \text{O}_2 \\
R7: & \quad 	ext{X} + \text{O}_3 & \rightarrow & \text{XO} + \text{O}_2
\end{align*}
\]

where the net reaction is identical to reaction R4. For the cycle catalysed by hydrogen radicals, \( X = \text{H}, \text{OH} \), for the nitrogen radical cycle, \( X = \text{NO} \), and for the chlorine radical cycle, \( X = \text{Cl} \).

Today the relative importance of the various ozone loss cycles is well understood. Because of the large increase of atomic oxygen (O) with altitude, the rates of all cycles increase strongly between 25 and 40 km and the same is true of the rate of ozone production through reaction R1 (e.g., CRUTZEN et al., 1995; GROOSS et al., 1999). Further, the relative importance of the cycles for ozone loss varies considerably with altitude. Between 25–40 km the NOx cycle is the dominant ozone loss process, whereas above 45 km HOx catalysed ozone loss dominates (see Figure 5). The HOx cycle is also the strongest loss cycle below about 25 km. The loss through the ClOx cycle (which also depends on the stratospheric chlorine loading) peaks at 40 km.

Furthermore, interestingly, the importance of heterogeneous reactions (albeit no heterogeneous reactions involving chlorine species) in the stratosphere was already suggested in the mid-seventies (CADLE et al., 1975). Based on laboratory experiments, heterogeneous reactions were, however, for a long time thought to be unimportant. This situation changed only in 1986 when SOLOMON et al. suggested that heterogeneous chemistry could greatly enhance the ability of chlorine to destroy ozone in polar regions (see Section 5 below). The paper by SOLOMON et al. prompted numerous laboratory studies and today, a large number of heterogeneous reactions are known that are important for stratospheric chemistry (Table 2, see also PETER, 1997; SOLOMON, 1999).

4 Anthropogenic influence on the ozone layer

4.1 The effect of nitrogen compounds on stratospheric ozone

The first concern about the human influence on stratospheric ozone was formulated in January 1958, when, in a discussion9 in Washington, DC, the question was raised as to whether or not the ultraviolet emissions from planned nuclear tests in the stratosphere (the so-called Teak and Orange events) would “burn a hole” in the natural ozone layer (HOERLIN, 1976). It was argued that, on the one hand, UV radiation originating from a nuclear test would be absorbed by ozone thus leading to dissociation, but that, on the other hand, shorter wavelength UV radiation would lead to dissociation of O2 and thus to ozone production – therefore these pre-event discussions were inconclusive (HOERLIN, 1976).

Nonetheless, there was a period of large-scale testing of nuclear weapons in the stratosphere (1952-1962). Much later, in 1973, it was recognised that large amounts of NOx were produced in air heated by the stratospheric nuclear explosions. However, no substantial impact of these NOx emissions on stratospheric ozone amounts was detected (FOLEY and RUDEMAN, 1973; JOHNSTON et al., 1973; GOLDSMITH et al., 1973).

9The participants in this discussion were J. Chamberlain (Yerkes Observatory), F. Gilmore (Rand Cooperation), H. Steward (NRL), and H. Hoerlin (LASL).
Another anthropogenic source of stratospheric NOx was suggested by Crutzen (1976), Crutzen and Ehhalt (1977), and by McElroy et al. (1977). In these studies it was argued that enhanced atmospheric levels of nitrous oxide (N2O) caused by increasing use of artificial fertilisers in agriculture could lead to enhanced NOx in the stratosphere and, in turn, to enhanced NOx-catalysed ozone destruction.

In the early seventies, attention was focused on another possible effect of human activities on stratospheric ozone: supersonic aircraft, since at that time it was thought they would cruise in the stratosphere by the turn of the century10. First it was suggested in 1970 that water vapour emitted by the planned large fleets of supersonic aircraft could result in HOx-catalysed ozone loss. It was soon clarified by Johnston (1971) that the emission of nitrogen oxides from the jet engines of these aircraft would pose a much greater threat to the ozone layer than the increase in water vapour. The possible relevance of this problem was also pointed out by Crutzen (1971).

For economic reasons, the planned fleets of supersonic transport aircraft were never built; nonetheless, knowledge about the chemistry of stratospheric ozone was greatly improved through research programmes directed at assessing the impact of supersonic transport on the stratosphere. This improved knowledge paved the way for further research on the impact of anthropogenic stratospheric chlorine on the ozone layer (see below).

4.2 The effect of chlorofluorocarbons on stratospheric ozone

At a meeting of the American Chemical Society in 1930, the engineer Thomas Midgley reported the invention of a new chemical compound, called Freon, that he proposed as a replacement for a variety of compounds in common use as refrigerants at the time. This was the starting point for the successful application of a whole family of related compounds, the chlorofluorocarbons (CFCs). Many years later CFCs were used in essentially all refrigeration and air conditioning systems, and halons, compounds that also contained bromine atoms, were frequently used as fire extinguishing agents. More than forty years after the invention of Freon, Lovelock et al. (1973) discovered that, because of their extremely long atmospheric lifetime, CFCs accumulate in the atmosphere and are globally distributed; the first indication of a potential environmental impact (although Lovelock et al. concluded at the time that “the presence of these compounds constitutes no conceivable hazard….”).

Atmospheric concentrations of CFCs and halons continued to increase until the past ten years when the rise first slowed down and was then reversed (WMO, 2007). Today, although production of practically all such substances has ceased because of the provisions of the Montreal Protocol and its amendments and adjustments, halogen source gases are still present in existing equipment, chemical stockpiles, foams etc., so that emissions continue.

Methyl chloride is the most important chlorine source gas, whose emissions are dominated by natural production. Methyl chloride has a global mean mixing ratio in the troposphere of about 550 ppt and its trend is presumed to be zero (e.g., Kopmann et al., 1993; WMO, 2007). At the end of the 21st century, when the abundance of anthropogenic chlorine source gases (e.g. CFCs) will be greatly reduced as a consequence of the Montreal Protocol, methyl chloride is expected to account for a large fraction of the remaining stratospheric chlorine.

Molina and Rowland (1974) related the possibility of chlorine-catalysed ozone loss to the accumulation of anthropogenic CFCs in the atmosphere. Early model studies (Crutzen, 1974; Rowland and Molina, 1975) predicted that enhanced levels of chlorine in the stratosphere would lead to a depletion of upper stratospheric ozone via cycle C1 with $X = \text{Cl}$. This notion was subsequently confirmed by a variety of model studies (WMO, 1986). Based on these early warnings (ozone reductions by 15 or 20% were predicted for continuing CFC emissions, Rowland and Molina, 1975), the first policy measures were taken. The United States phased out the use of CFCs as a propellant in aerosol sprays as of 1 January 1979. This action was soon followed by similar bans in Canada, Sweden, and Norway. However, it took many years, until 1987, for the first international agreement to be signed, the Montreal Protocol, regulating the use of CFCs and halons.

The predictions of the early studies on the future of upper stratospheric ozone were remarkably farsighted. It was understood that CFCs accumulate in the lower atmosphere, with photolysis in the middle and upper stratosphere being the major sink. It was understood already in 1974 that this constitutes a long-term problem, on a timescale of many decades. Most importantly, it was predicted that enhanced levels of stratospheric chlorine would lead to a decline of ozone in the upper stratosphere by a catalytic ozone loss cycle (cycle C1, with $X = \text{Cl}$, Molina and Rowland, 1974; Rowland and Molina, 1975) with peak impact at about 40 km altitude.

Today, more than thirty years after the first scientific studies were published linking the danger of future stratospheric ozone loss to the accumulation of anthropogenic CFCs in the atmosphere (Molina and Rowland, 1974; Crutzen, 1974), ozone decline has been unequivocally detected in the altitude region between 30 and 50 km by a variety of ground-based and space-borne instruments (WMO, 1995; Harris et al., 1998; WMO, 1998).

---

10The predictions in the mid-seventies were that in the year 2000 the supersonic fleet would comprise 615–1400 aircraft (Dotto and Schiff, 1978, pp. 78–79).
A significant reduction in the concentration of ozone was observed during the 1980s and 1990s, with the largest losses reaching $7.4 \pm 1\%$/decade at 40 km altitude (HARRIS et al., 1998, Figure 6, panel b). The observed altitude variation of loss of upper stratospheric ozone – peak percentage losses around 40 km – was correctly predicted by the first model studies (CRUTZEN, 1974, see also Figure 6, panel a). These calculations were, however, based on a projected increase of Cl, to a level of 5.3 ppb, a value that was fortunately never reached in the contemporary stratosphere. During the past decade, the decline of upper stratospheric ozone, observed for 1980-1996, has substantially slowed and there is evidence that this decline can be attributed to the decline of halogen source gases in the atmosphere (NEWCHURCH et al., 2003; STEINBRECHT et al., 2006; WMO, 2007).

5 Polar stratospheric ozone loss

5.1 The Antarctic ozone hole

In 1985, Farman, Gardiner, and Shanklin reported that in the Antarctic spring strongly reduced total ozone values occurred at the British Antarctic Survey research station at Halley (FARMAN et al., 1985; JONES and SHANKLIN, 1995). The data from the original publication are shown in Figure 7 (black symbols) together with measurements up to 2008 (grey symbols). This phenomenon that was soon referred to as the Antarctic ozone hole is one of the most striking examples of the direct impact of human activities on the atmosphere. The discovery by FARMAN et al., which was based on a series of measurements by classical Dobson instruments, started by Dobson himself in 1956 (DOBSON, 1968) was soon confirmed by satellite measurements\(^\text{11}\) showing that the ozone depletion extended over roughly the entire Antarctic continent (STOLARSKI et al., 1986). Independent methods for measuring total column ozone provided further support for the depletion of springtime ozone in Antarctica (e.g., SOLOMON, 1999).

The term ozone hole for the phenomenon of low total ozone values in the Antarctic spring was first used by STOLARSKI et al. (1986): “The deep minimum, or hole...”. But, of course, the Antarctic ozone hole is not a real hole. Some column ozone always remains; e.g., the October mean at Halley never dropped below 100 DU (Figure 7). Interestingly, however, CHAPMAN (1934) already asked the question “Can a hole be made in the ozone layer?”. The answer he suggested was the artificial removal of most of the ozone in a column of air to allow astronomers better observations in the ultraviolet (see also TUCK, 1978; WELLS, 1997). Likewise, REGENER (1946) used the term “ein Loch in der Ozonschicht” (“a hole in the ozone layer”) and pointed out that such a hole would be welcomed by astronomers. Moreover, the question whether nuclear explosions in

\(^{11}\)Of course the question arises as to why the ozone hole was not discovered in the satellite measurements that were available since 1979. There is some debate on this question (see e.g., PUARCE, 2008) but it is clear that the low ozone values were “flagged” by the processing software as unreliable because they were so low.
the stratosphere would “burn a hole” into the ozone layer was raised in discussions in early 1958 about the possible effects of nuclear weapons testing in the stratosphere (HOERLIN, 1976).

The discovery of the ozone hole came as a complete surprise. Ozone was considered to be almost chemically inert in the polar winter and early spring when the sun was low. In the results of model calculations there was no indication that such a phenomenon could occur. FARMAN et al. stated that “Recent attempts to consolidate assessments of the effect of human activities on stratospheric ozone (O₃) using one-dimensional models for 30°N have suggested that perturbations of total O₃ will remain small for at least the next decade. Results from such models are often accepted by default as global estimates. The inadequacy of this approach is here made evident by observations that the spring values of total O₃ in Antarctica have now fallen considerably”. Indeed it turned out in the following years that the fundamental chemical mechanisms responsible for the formation of the ozone hole were unknown in 1985 (see also Section 5.3 below).

The first measurement of a vertical ozone profile under ozone hole conditions was presented by CHUBACHI (1984). Measurements describing the development of the vertical structure of ozone in the Antarctic spring rapidly followed. A time series of 66 sondes flown between May and December 1985 from the Georg Forster Station (71°S) of the then German Democratic Republic was reported by GERNANDT et al. (1987) and GERNANDT (1987). IWASAKA and KONDOH (1987) showed the 1981-1984 ozone sonde measurements from Syowa Station (69°S) in detail, and HOFMANN et al. (1987) described measurements during the period from 25 August to 6 November 1986 from McMurdo Station (78°S). GERNANDT et al. (1989) compared ozone sonde measurements for the years 1985, 1986, and 1987 from the Georg Forster, Syowa, Halley, and South Pole Stations. When ozone measurements in mid-winter and spring are compared (Figure 8), the signature of the ozone hole becomes obvious through substantially lower ozone partial pressures in October in the altitude range (~200–30 hPa) where the ozone maximum occurs during mid-winter. In recent years, satellite measurements (e.g., HOPPEL et al., 2003; TILMES et al., 2006b) provided additional, detailed information on the vertical ozone distribution and on chemical ozone loss in the ozone hole. Nonetheless, ozonesonde observations remain important as they provide unique information. Measurements of extremely low ozone mixing ratios (below 0.01 ppm) are not possible with satellite instruments; such observations could potentially provide insights into future ozone layer recovery (SOLOMON et al., 2005, 2007).

5.2 Arctic ozone loss

The Arctic wintertime stratosphere resembles its Antarctic counterpart in many respects; it exhibits a cold polar vortex separating the air enclosed in it from mid-latitude air. Strong diabatic descent throughout the winter transports air from the upper stratosphere and partly from the mesosphere to the lower stratosphere (e.g., TUCK, 1989). Indeed, signs of a perturbed chlorine chemistry showing that the Arctic was primed for chemical ozone depletion were reported in the late eighties (e.g., SOLOMON et al., 1988; SCHILLER et al., 1990; BRUNE et al., 1990).

However, the Arctic polar vortex is warmer and much more variable than the Antarctic vortex leading to a stronger interannual variability in both chemical loss of ozone and in dynamical supply of ozone-rich air to high latitudes. Therefore, compared to the Antarctic, Arctic ozone abundances in the winter and spring are much
5.3 Chemical mechanisms of polar ozone destruction

5.3.1 Observations of a perturbed polar chlorine chemistry

Farman et al. (1985) already tried to link their observations of Antarctic ozone loss to the increase of anthropogenic CFCs in the atmosphere. Alternative explanations were put forward (Tung et al., 1986; Mahlman et al., 1986; Callis and Natarajan, 1986), but these were eventually discarded in favour of a linkage with CFCs, albeit via completely different chemistry than that proposed by Farman et al. (Solomon, 1999).

First signs of a strong perturbation of the Antarctic chlorine chemistry were already seen during the National Ozone Expedition (NOZE) in 1986 through ground-based observations of strongly reduced column values of HCl (Farman et al., 1987), strongly enhanced column values of OClO (Solomon et al., 1987), and high concentrations of ClO below 20 km (De Zafra et al., 1987). In 1987, within the framework of the Airborne Antarctic Ozone Experiment (AAOE, Tuck et al., 1989), aircraft measurements in the stratosphere provided even clearer evidence that the Antarctic ozone hole is indeed caused by a strongly perturbed chlorine chemistry. These first measurements were subsequently confirmed by a great number of further measurements in the Antarctic from a variety of field campaigns and remote sensing instruments. Likewise, a perturbed chlorine chemistry was demonstrated by data obtained in the Arctic stratosphere in a number of measurement campaigns. In both polar regions, in winter, strongly enhanced ClO mixing ratios were detected from balloon, aircraft, and remote sensing experiments (e.g., Anderson et al., 1991; Waters et al., 1993; Woyke et al., 1999; Vogel et al., 2003; Von Hobe et al., 2006), enhanced OClO, (e.g., Schiller et al., 1990; Brandt-jen et al., 1994; Kreher et al., 1996; Wagner et al., 2001), and strongly depleted HCl and ClONO$_2$, (e.g., Webster et al., 1993; Oelhaf et al., 1994; Müller et al., 1996; Tilmes et al., 2004; Santee et al., 2008).

5.3.2 Heterogeneous reactions and polar stratospheric clouds

Solomon et al. (1986) first proposed that the reaction of HCl and ClONO$_2$ on the surfaces of polar strato-
spheric clouds (PSCs) plays an important role in bringing about the perturbed polar chlorine chemistry. In the same year, TOON et al. (1986) and CRUTZEN and ARNOLD (1986) suggested that PSCs could consist of crystalline particles consisting of nitric acid trihydrate (NAT) rather than only of ice as was previously thought. In both studies, the analysis had to be based on extrapolation over more than 50 K down to stratospheric temperatures due to the lack of experimental data. Based on this information, TOON et al. (1986) and CRUTZEN and ARNOLD (1986) suggested that NAT-containing PSC should be formed well above the condensation temperature of pure water vapour. Two years later, HANSON and MAUERSBERGER (1988) reported the results of the first modern laboratory measurements of nitric acid hydrates. Interestingly, the last laboratory study on NAT at low temperatures prior to their work dates from 1904 and is described in a paper in German (KÜSTER and KREMMANN, 1904). The freezing temperatures of H₂O/HNO₃ mixtures measured by KÜSTER and KREMMANN (1904) and the experimental set-up they used is shown in Fig. 11. It was as late as 2000 before NAT particles were detected in situ by a balloon-borne mass spectrometer in the Arctic (VOIGT et al., 2000).

The characteristic radii of type I PSCs were originally assumed to be 0.5–3.0 μm (e.g., HOFMANN et al., 1989; WMO, 1990). However, from January to March 2000 large NAT particles with radii of 20–40 μm (also referred to as “NAT rocks”) were detected by in situ measurements on a high-altitude research aircraft (FAHEY et al., 2001). This observation demonstrated that NAT particles can reach sedimentation velocities large enough to allow HNO₃ to be removed from the polar stratosphere thereby causing “denitrification” of the polar stratosphere. There are also liquid PSC particles, super-cooled ternary solutions (STS) consisting of liquid H₂O/HNO₃/H₂SO₄ aerosol particles, and, when temperatures drop several degrees Kelvin below the frost point, ice crystals may form (e.g., CARSLAW et al., 1997; PETER, 1997; SOLOMON, 1999).

The heterogeneous reaction of HCl and ClONO₂ (SOLOMON et al., 1986)

\[
R8: \quad \text{HCl} + \text{ClONO}_2 \rightarrow \text{Cl}_2 + \text{HNO}_3
\]

converts the relatively stable chlorine reservoir species HCl and ClONO₂ into photolabile Cl₂, thereby greatly perturbing gas phase chlorine partitioning. It is remarkable that only one year before the suggestion by SOLOMON et al. (1996), it was stated in the Journal of Physical Chemistry that “...it is not likely for such a heterogeneous process [reaction R8] to contribute significantly to the release of Cl from HCl” (MOLINA et al., 1985). This view was, however, corrected soon (MOLINA et al., 1987). Later PRATHER (1992) and CRUTZEN et al. (1992) argued that the heterogeneous reaction

\[
R9: \quad \text{HCl} + \text{HOCI} \rightarrow \text{H}_2\text{O} + \text{Cl}_2
\]

is an important reaction channel for the activation of HCl and thus for a complete activation of the stratospheric chlorine reservoir.

Today, a variety of heterogeneous reactions of importance to stratospheric chemistry are known (e.g.,

<table>
<thead>
<tr>
<th>Table 2: Key Heterogeneous Reactions</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \text{HCl} + \text{ClONO}_2 \rightarrow \text{HNO}_3 + \text{Cl}_2 )</td>
</tr>
<tr>
<td>( \text{N}_2\text{O}_5 + \text{H}_2\text{O} \rightarrow 2 \times \text{HNO}_3 )</td>
</tr>
<tr>
<td>( \text{ClONO}_2 + \text{H}_2\text{O} \rightarrow \text{HNO}_3 + \text{HOCl} )</td>
</tr>
<tr>
<td>( \text{HCl} + \text{HOCI} \rightarrow \text{H}_2\text{O} + \text{Cl}_2 )</td>
</tr>
<tr>
<td>( \text{BrONO}_2 + \text{H}_2\text{O} \rightarrow \text{HNO}_3 + \text{HOBr} )</td>
</tr>
<tr>
<td>( \text{HCl} + \text{BrONO}_2 \rightarrow \text{HNO}_3 + \text{BrCl} )</td>
</tr>
<tr>
<td>( \text{HCl} + \text{HOBr} \rightarrow \text{H}_2\text{O} + \text{BrCl} )</td>
</tr>
</tbody>
</table>

Figure 11: The measurements of nitric acid hydrates by KÜSTER and KREMMANN (1904). Top panel shows the reported freezing temperatures of H₂O/HNO₃ mixtures, bottom panel the apparatus used by KÜSTER and KREMMANN. Note that in the top panel KÜSTER and KREMMANN’s measurements are shown as plus signs together with earlier measurements (shown as asterisks and referred to as “von Erdmann angegebene Schmelzpunkte”) that they refute as unreliable in their paper. (Copyright Wiley-VCH Verlag GmbH & Co. KGaA, reproduced with permission).
SANDER et al., 2006), the most important are listed in Table 2. Originally, it was thought that stratospheric heterogeneous reactions in the polar regions occur only on solid surfaces (e.g., PETER, 1997). Heterogeneous reactions on the ubiquitous stratospheric sulphate aerosol, which is noncrystalline, were known to be important for mid-latitude chemistry (e.g., HOFMANN and SOLOMON, 1989; BRASSEUR et al., 1990), but were not thought to be of great relevance for polar chlorine activation. WOLFF and MULVANEY (1991) and COX et al. (1994) first pointed out the potential importance of heterogeneous reactions on noncrystalline aerosol particles. Today, there is information from laboratory studies on the reaction probabilities of stratospheric species of a variety of solid and liquid aerosol particles. These reaction probabilities are frequently strongly temperature dependent (being relevant only at low temperatures), with the important exception of the reaction of N$_2$O$_5$ with H$_2$O (Table 2), which is important at all temperatures occurring in the stratosphere.

As recently as 2001, a completely new PSC mechanism was proposed (LU and SANCHE, 2001); it was suggested that the heterogeneous decomposition of CFCs on the surface of PSC particles by dissociative electron attachment (DEA) would constitute “a new pathway contributing to the formation of the ozone hole”. This issue is still debated in the literature today. For example WANG et al. (2008) assert that DEA-induced heterogeneous decomposition of CFCs “has been proposed to be the key mechanism for the formation of the Earth’s ozone hole”. However, there is now a number of studies that provide strong arguments, based both on model simulations and on observations, that heterogeneous reactions of CFC on PSC cannot be of substantial importance to polar ozone chemistry (HARRIS et al., 2002; PATRA and SANTHANAM, 2002; MÜLLER, 2003, 2008).

Recently, in presentations at the AGU fall meeting in San Francisco in 2005 and at the EGU general assembly in Vienna in 2006, Katja Drdla put forward the concept that PSC are not necessary for polar chlorine activation, but that instead liquid binary H$_2$SO$_4$/H$_2$O sulphate aerosol particles dominate activation. This is an issue of current scientific research (see also TILMES et al., 2007, 2008a; FECK et al., 2008).

5.3.3 Catalytic ozone loss cycles for polar winter conditions

A problem with the chemical theories put forward as an explanation of the Antarctic ozone hole soon after its discovery was that in the polar regions the conventional ozone loss cycles (C1) are not effective in winter and early spring because of the low sun in winter high latitudes and the resulting lack of atomic oxygen (O). For example, HAYMAN et al. (1986) stated: “This restricts photochemical reactions to those molecules which absorb in the near U.V. and in the visible regions of the solar spectrum giving rather low concentrations of O$^3P$), a key species in most of the important catalytic cycles presently included in the chemical models.” Therefore, although the activation of chlorine through heterogeneous reactions could explain observations of a perturbed stratospheric chlorine chemistry, new ozone loss cycles were required for a complete understanding of chemical ozone loss.

Today it is known that in the polar stratosphere two different catalytic cycles are responsible for the major fraction of chemical ozone loss; the efficiency of both cycles depends on the concentration of ClO. The most important cycle was originally proposed by MOLINA and MOLINA (1987):

$$\begin{align*}
R10: & \quad \text{ClO} + \text{ClO} + \text{M} \rightarrow \text{Cl}_2\text{O}_2 + \text{M} \\
R11: & \quad \text{Cl}_2\text{O}_2 + \text{hv} \rightarrow \text{Cl} + \text{ClO} \quad (\lambda \leq 400\text{nm}) \\
R12: & \quad \text{ClO} + \text{O}_3 \rightarrow \text{Cl} + \text{O}_3 + \text{O}_2 \\
R13: & \quad 2 \times (\text{Cl} + \text{O}_3) \rightarrow \text{ClO} + \text{O}_2 \\
\end{align*}$$

C2: Net: 2O$_3$ → 3O$_2$

where hv denotes a photon and M a collision partner (N$_2$ or O$_2$). The second cycle depends, in addition, on BrO concentrations and thus on the stratospheric bromine loading (TUNG et al., 1986; McELROY et al., 1986):

$$\begin{align*}
R14: & \quad \text{ClO} + \text{BrO} \rightarrow \text{Cl} + \text{Br} + \text{O}_2 \\
R13: & \quad \text{Cl} + \text{O}_3 \rightarrow \text{ClO} + \text{O}_2 \\
R15: & \quad \text{Br} + \text{O}_3 \rightarrow \text{BrO} + \text{O}_2 \\
\end{align*}$$

C3: Net: 2O$_3$ → 3O$_2$

A further contribution to polar ozone destruction comes from a cycle involving HOCl; this cycle was first proposed by SOLOMON et al. (1986):

$$\begin{align*}
R16: & \quad \text{HO}_2 + \text{ClO} \rightarrow \text{HOCl} + \text{O}_2 \\
R17: & \quad \text{HOCl} + \text{hv} \rightarrow \text{OH} + \text{Cl} \\
R13: & \quad \text{Cl} + \text{O}_3 \rightarrow \text{ClO} + \text{O}_2 \\
R18: & \quad \text{OH} + \text{O}_3 \rightarrow \text{HO}_2 + \text{O}_2 \\
\end{align*}$$

C4: Net: 2O$_3$ → 3O$_2$

The net result of cycles 2, 3, and 4 is the destruction of two ozone molecules. Like cycle C1, cycles C2, C3, and C4 are catalytic, and chlorine (Cl) and bromine (Br) are not lost in the reaction cycle. In the stratosphere, chlorine is much more abundant than bromine (~160 times). Nonetheless, cycle C3 is important as bromine atoms are about 60 times more efficient than chlorine atoms in destroying ozone chemically (CHIPPERFIELD and PYLE, 1998; WMO, 2007). Clearly, sunlight is necessary to
maintain a large ClO abundance through the photolysis of Cl₂O₂. (And also to photolyse the Cl₂ formed in the heterogeneous reactions R8 and R9 to form Cl). However, in contrast to the shortwave radiation (wavelengths less than 242 nm) required to produce atomic oxygen (the species that is essential for cycle C1), the photolysis of Cl₂O₂ proceeds at rather long wavelengths. Radiation at long wavelengths prevails under conditions of low sun at the poles in spring so that the Cl₂O₂ is photolysed.

The kinetic parameters of the reactions constituting cycle C2 still remain a matter of scientific debate (e.g., POPE et al., 2007; VON HOBÉ et al., 2007; VON HOBÉ, 2007), so that it cannot be claimed that a full quantitative understanding of polar ozone loss has been reached. Nonetheless, there is consensus that cycles C2 and C3 account for the majority of the ozone loss observed in late winter-spring in the polar stratosphere (e.g., IPCC/TEAP, 2005; WMO, 2007). Under cold polar vortex conditions, with high ClO abundances, the rate of ozone destruction is reported to reach substantial values of up to 2–3% per day (e.g., MCKENNA et al., 1990; SALAWITCH et al., 1990; REX et al., 1997; BECKER et al., 1998, 2000). Outside of the polar regions, both cycles are of only minor importance; C2 is negligible because it is only effective at the low temperatures in polar winter and spring (because Cl₂O₂ is thermally unstable at typical stratospheric temperatures) and C3 because of the much lower ClO concentrations in extrapolar regions.

Furthermore, an important aspect of all early theories of ozone hole chemistry was that they required low concentrations of NO and NO₂ (CRUTZEN and ARNOLD, 1986; McELROY et al., 1986; HAYMAN et al., 1986; SOLOMON et al., 1986). Today it is known that the heterogeneous reactions R8, R9, and related reactions (Table 2) suppress the concentration of NO₂ by forming HNO₃. If NO₂ concentrations were not suppressed, the released ClO would readily reform the ClONO₂ reservoir (e.g., MÜLLER et al., 1994; DOUGLASS et al., 1995).

5.4 The future of the stratospheric ozone layer

As the halogen source gases that cause ozone depletion decrease in the coming decades under the provisions of the Montreal Protocol and its adjustments and amendments, a recovery of the stratospheric ozone layer is expected. However, the atmosphere will not return to the same conditions that prevailed before the onset of significant anthropogenic ozone loss; global change, in principle, can either accelerate or delay ozone recovery (IPCC/TEAP, 2005; WMO, 2007).

Increases in carbon dioxide (CO₂) in the atmosphere are expected to cool the stratosphere. Because with lower temperature, the rates of the gas-phase ozone destruction cycles decrease, a cooling of the stratosphere leads to an increase of ozone concentrations above about 25 km (IPCC/TEAP, 2005). The fact that increasing CO₂ in the stratosphere leads to an ozone enhancement in the upper stratosphere was first pointed out by GROVES et al. (1978) and GROVES and TUCK (1979). In the polar regions, a cooling of the stratosphere, if it were to occur because of increasing CO₂, would enhance the formation of polar stratospheric clouds and thus would lead to stronger polar ozone depletion. This effect is expected to be most pronounced in the Arctic (e.g., AUSTIN et al., 1992; SHINDELL et al., 1998; IPCC/TEAP, 2005).

Numerical models are used to assess past changes in the global ozone distribution and to project future changes (e.g., EYRING et al., 2007). Figure 12 shows total ozone averaged between latitude 60°N and 60°S between 1980 and 2050 based on model projections and, for the period up to 2005, on observations. The model results indicate that for 60°N–60°S total ozone the first two stages of ozone recovery (a slowing of the decline and a turnaround, WMO, 2007) will be reached before 2020. Full recovery, with ozone reaching or exceeding pre-1980 values, is expected to occur by the middle of the 21st century.

Over the Antarctic, models predict that column ozone will increase in spring between 2000 and 2020 by about 5 to 10% (WMO, 2007). In the Antarctic, ozone loss is saturated, which means all the ozone in the lower stratosphere is chemically destroyed (e.g., WMO, 2007; TILMES et al., 2006a). For such conditions, ozone loss is expected to remain constant for some time even when
stratospheric halogen levels decline. Moreover, the decline in halogen levels will occur later over the Antarctic than at lower latitudes because air in the Antarctic stratosphere is older than air found at lower latitudes. As a result, reductions in halogen loading to pre-1980 values will occur 10-15 years later in the Antarctic stratosphere than in the midlatitude stratosphere (NEWMAN et al., 2007).

For the Arctic, most models currently predict that springtime column ozone in 2020 will be 0–10% above 2000 levels and that ozone turnaround in the Arctic will occur before 2020 (WMO, 2007). For Arctic conditions, there is a large interannual variability in the model projections that does not allow a year to be identified when the ozone turnaround due to a decreasing halogen burden will occur. Model predictions show Arctic ozone increasing to pre-1980 values before 2050. Further, in these predictions, Arctic ozone increases to 1980 values before Antarctic ozone does. However, the strong natural year-to-year variability in the Arctic makes it difficult to obtain accurate model simulations for this region, in particular accurate simulations of polar temperatures and of transport barriers (SANKEY and SHEPHERD, 2003; TILMES et al., 2007). A further delay in the recovery of polar ozone would occur if the polar lower stratosphere in winter and spring was to cool as a consequence of climate change.

6 Discussion

Ozone is one of the most important atmospheric trace substances and changes in atmospheric ozone have important consequences for plants, animals, and humans. Nonetheless, the scientific understanding of atmospheric ozone was historically often rather limited and did not allow reliable assessments of its impact on nature and mankind to be made. For example, it was only understood as late as the early 1950s that high ozone concentrations in the troposphere (today referred to as "photochemical smog") are caused by pollution and are detrimental to human health, while, for instance, in 1946 the opinion was still put forward in scientific papers (REGENER, 1946) that the presence of ozone was an indication of good air.

Further, the fact that the discovery of the Antarctic ozone hole in 1985 came as a complete surprise to the scientific community speaks volumes. TUCK (1986) stated: “These experimental findings are a major surprise, and are taxing the theoretical ingenuity of stratospheric researchers”. Today, it is well established that the ozone hole is caused by anthropogenic emissions of chlorofluorocarbons; it is one of the most eminent examples of the effect of humans on the global environment. The growing effect of humans on the environment has led CRUTZEN (2002) to suggest assigning the term “Anthropocene” to the present geological epoch; a suggestion that is becoming increasingly accepted in the geological literature (e.g., ZALASIEWICZ et al., 2008).

It is interesting to compare the time of important milestones in stratospheric ozone research with that of fundamental discoveries in physics (Figure 13). It becomes obvious that before the role of ozone as an UV absorber was understood, the Maxwell equations and Boltzmann’s concept of entropy were established in physics. The first column measurements of ozone in the atmosphere were made at a time when the electron and the proton had been discovered and Rutherford’s model of the atom had been developed. At a time, in 1929, when the height of the ozone layer was grossly underestimated, the wave-particle duality of quantum mechanics had already been established. The prediction of the meson preceeded...
the understanding of the importance of HOx-catalysed ozone loss in the stratosphere and the formulation of the “standard model”, unifying the strong, weak, and electromagnetic force, proceeded the discovery that NOx- and ClOx-catalysed ozone loss cycles would play an important role in stratospheric chemical ozone loss cycles.

Further, in the mid-eighties, when the first chemical theories emerged that proved successful in explaining the observation of an Antarctic ozone hole, the theory of black hole radiation had been formulated and the W and Z bosons had already been detected. Giant magnetoresistance was discovered in the year when the first modern laboratory investigations of nitric acid trihydrate (NAT) were conducted, more than eighty years after the last scientific study on NAT at low temperatures has been published (Küster and Kremann, 1904).

In general, in the history of ozone research there are many examples where progress was hindered by the lack of sufficient and sufficiently accurate data. This is true of the ozone hole (Section 5.1), the understanding of the chemistry of stratospheric ozone (Section 3), and the vertical ozone profile (Section 2). In the case of the ozone hole, even fundamental processes were unknown at the time of its discovery, namely the microphysics of polar stratospheric clouds and the heterogeneous and gas-phase chemistry at low temperature. Moreover, false views persisted in ozone research for years before they were eventually discarded because of new observations; examples are the altitude of the ozone layer, the notion that the presence of ozone was an indication of good, healthy air, and the firm belief that heterogeneous reactions were unimportant in the stratosphere.

Throughout the history of ozone research, progress was made by combining new experimental results, both from laboratory and from atmospheric measurements, to build new theories and new numerical models of the processes occurring in the atmosphere. In recent decades, the complexity of these models has greatly increased with time. Both the development of these models and scientific work based on model results means that the results of the numerical calculations have to be continuously compared with atmospheric observations.

Finally, part of the history of ozone research since the 1970s is the history of research into the human impact on the ozone layer. The dynamics of the development from scientific insight, via the scientific assessment process, to the impact on public policy that is evident from recent scientific work based on model results means that the processes occurring in the atmosphere. In recent decades, the complexity of these models has greatly increased with time. Both the development of these models and scientific work based on model results means that the results of the numerical calculations have to be continuously compared with atmospheric observations.

Future scientific work on the stratospheric ozone layer will address questions including the way in which global change will affect ozone depletion, when a substantial improvement of Antarctic ozone depletion can be expected, whether there could be strong Arctic ozone losses and how possible changes in the stratospheric concentrations of methane, water vapour, and nitrous oxide might affect stratospheric ozone (Solomon, 2004). And, undoubtedly, further questions will arise which we do not foresee today. Clearly, there is a rich field of research in order to answer these questions.

Acknowledgments

I am grateful to G. Brasseur, P. Crutzen, D. Ehnhalt, H. Gernandt, J. Shanklin, S. Solomon, J. Stähelin, and A. Tuck for providing very helpful comments on earlier versions of this manuscript. I am particularly indebted to A. Tuck for pointing out a considerable number of important references that I was not aware of. Thanks also goes to two anonymous reviewers for a very helpful critique of the original version of the paper. I thank J. Shanklin for the total ozone measurements from Halley Station, Antarctica, and H. Gernandt for the ozone sonde measurements from Georg Forster Station, Antarctica, in 1985. The photograph of E. Regener (Figure 1) was provided by the Max Planck Society. For this paper it was necessary to retrieve a lot of historic papers that are difficult to obtain; the help of the Central Library of the Forschungszentrum Jülich was essential for this task. I thank J. Carter-Sigglov for revising the grammar and style of the manuscript.

References


Brasseur, G.P., 2008: Creating knowledge from the confrontation of observations and models: The case of stratospheric ozone. – In: Climate Variability and Extremes During the Past 100 Years, Vol. 33 of Advances in Global Change Research, Brönnimann, S. et al. (Eds.), 303–316.


CRUTZEN, P.J., 1969: Determination of parameters appearing in dry and wet photochemical theories for ozone in stratosphere. – Tellus 21, 368–388.


HARRIS, N., R. HUDSON, C. PHILLIPS (Eds.), 1998: Assessment of Trends in the Vertical Distribution of Ozone. – SPARC Report 1, WMO.


PRAITHER, M.J., 1992: More rapid ozone depletion through the reaction of HOCI with HCl on polar stratospheric clouds. – Nature 355, 534–537.


