Solar Influences on Polar Ozone

by

JUDY ANN ELIZABETH STEPHENSON

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy,
at the
Space Physics Research Institute,
University of Natal, Durban.

Republic of South Africa
February 1994

Preface

The work described in this thesis was carried out at the Space Physics Research Institute (SPRI), Department of Physics, University of Natal, Durban, South Africa, from April 1990 to February 1993, under the supervision of Professor M. W. J. Scourfield and co-supervision of Dr. J. P. S. Rash.

The studies documented in this thesis represent original work by the author and have not been submitted in any form to another university. Where use was made of the work of others it has been duly acknowledged in the text.

Acknowledgements

I gratefully acknowledge all persons who were of assistance during the course of this thesis. In particular:

supervisor Professor M. W. J. Scourfield of the SPRI who initiated this project. His keen interest, kindness and advice will always be appreciated,

Mr G. E. Bodeker for use of computer graphics software and beneficial scientific discussions,

Dr E. Mravlag for valued scientific advice and proof reading,

Ms T. M. Holloway of the SPRI, for computer hardware and software assistance,

Mr J. Doull for image processing and printing,

Ms B. Von Etzdorf for typing parts of the manuscript, and

my family and friends (especially the coffee ladies) for being there.

In addition, I thank P. T. Guimaraes, R. D. McPeters, A. J. Krueger (NASA/GFSC) and members of the TOMS Nimbus Experiment and Ozone Processing teams, and the National Space Data Center-A for Rockets and Satellites for the total column ozone data. Data on solar proton events were supplied by NOAA Space Environment Services Center. Temperature and wind speed profiles at Sanae, Antarctica were courtesy of the South African Weather Bureau.

Finally, I would also like to thank the University of Natal for financial support through Graduate Scholarship Awards during 1989 and 1990, as well as travel funds granted by the University Research Grant Committee toward travel to the Quadrennial Ozone Symposium, Charlottesville, 1992. The Foundation for Research and Development provided financial assistance during 1989 and 1990.

Abstract

Measurements by the TOMS instrument, aboard the Nimbus 7 satellite, of total column ozone over polar regions have been studied to determine the effects of solar induced natural ozone modulation. Two different analysis methods were employed to ascertain short term (days to months) and long term (months to years) solar influences on polar ozone.

Bursts of intense solar activity can result in solar proton events (SPE's). The high energy protons, originating in solar flares, produce secondary electrons which can generate large concentrations of odd nitrogen in the middle atmosphere. These reactive species can catalytically destroy ozone. Three case studies are presented in an attempt to quantify the effect of SPE's on ozone mass over a latitude region 90 to 70°. In order to monitor the ozone response following a SPE over both hemispheres simultaneously, the SPE must occur during the equinox period when both poles are irradiated. Fortuitously, a SPE was recorded in March 1989, the analysis of which forms a case study in this thesis. Ozone depletions of 7.4 $\times 10^9$ kg for the south polar cap and 8.0 $\times 10^9$ kg for the north polar cap indicate the degree of symmetry for this event.

Longer term effects of solar variability are investigated by Fourier techniques. A Fourier transform of eleven years of total ozone mass values, over the region 90 to 70°S, was performed. Inspection of the Fourier spectrum reveals peaks associated with solar cycle, annual and semi-annual oscillations, that may be attributed directly to solar variation. Other peaks, corresponding to QBO and ENSO periodicities, may be ascribed to indirect solar influences i.e. thermally driven dynamics. Finally, a comparison between the phase of the solar cycle peak in this spectrum with that in a spectrum of daily values of solar radio flux, reveals that the austral polar ozone solar cycle periodicity lags solar forcing by 2.8 years.

Portions of chapters have been reported at the 1990 South African Institute of Physics Annual Conference, University of Port Elizabeth, South Africa and as a poster at the 1992 Quadrennial Ozone Symposium, Charlottesville, United States of America, 4-13 June 1992.

In addition, various parts of this work has been submitted for publication, viz:

Stephenson, J. A. E. and M. W. J. Scourfield, Importance of energetic solar protons in ozone depletion, *Nature*, *352*, 137, 1991.

Stephenson, J. A. E. and M. W. J. Scourfield, Ozone depletion over the polar caps caused by solar protons, *Geophys. Res. Lett.*, *19*, 2425, 1992.

Stephenson, J. A. E. and M. W. J. Scourfield, Natural ozone depletion over Antarctica, *S. Afr. J. Antarctic. Res.*, **21**, 222, 1991.

Stephenson, J. A. E. and M. W. J. Scourfield, Solar proton effects on austral ozone during the final months of 1989, *Proceedings of Quadrennial Ozone Symposium*, (in press).

A further publication, involving the Fourier analysis of the eleven year total column ozone data set, is in preparation.

Contents

List of Symbols

B_ν	black body radiance function
C	specific heat capacity
C_p	heat capacity at constant pressure
C_v	heat capacity at constant volume
$C_{p,m}$	specific heat capacity at constant pressure
$C_{v,m}$	specific heat capacity at constant volume
c	speed of light
F	flux density
F_c	coriolis force
F_p	pressure gradient force
G	gravitational constant
$\Delta G^\ominus$	standard Gibb's free energy
g	gravitational acceleration
H	scale height, earth's magnetic field
ΔH	enthalpy of formation
$\Delta H^\ominus$	standard enthalpy of formation
h	Planck's constant
h′	geopotential height
I	intensity
j	differential particle flux
J	photodissociation coefficient
J_ν	radiance source function
K_p	geomagnetic index
k	Boltzmann constant, eddy diffusion coefficient
k	rate constant
k_a	absorption coefficient
k_ν	extinction coefficient
L	scale length, angular momentum
L	radiance
$M_\oplus$	mass of the earth
m_m	molar mass
n_i	number density of species i
P	ion production rate
q	heat
Q	ionization rate

r	gyroradius
R	gas constant; rigidity; distance from centre of the earth
$R_\oplus$	radius of the earth.
R_T	radiancy
$R_T(\nu)$	spectral radiancy
S	entropy, static stability parameter
U	internal energy
u	zonal wind speed
v	meridional wind speed
v_g	geostrophic velocity
w	work
w	vertical wind speed
z	altitude, nuclear charge

Γ	lapse rate
Γ_d	dry adiabatic lapse rate
ζ_g	total relative vorticity
Θ	potential temperature
κ	$\frac{R}{C_{p,m}m_m}$
Λ	geographic longitude
ρ	density
σ	ozone spatial density
σ_a	absorption cross section
τ_a	optical thickness for absorption
τ_ν	total optical thickness at frequency ν
τ_u	zonal dynamical time constant
τ_v	meridional dynamical time constant
τ_w	vertical dynamical time constant
$\tau_{chem}, \tau_{dynam}$	chemical and dynamical time constants
τ_{rad}, τ_{diff}	radiative and diffusion time constants
Φ	geopotential, irradiance
ϕ	geographic latitude, azimuthal direction
χ	solar zenith angle
ω	angular speed, solid angle
$\omega_\oplus$	angular speed of the earth

List of Acronyms

AAOE Airborne Antarctic Ozone Experiment
AASE Airborne Arctic Stratospheric Experiment
ARO Algonquin Radio Observatory
BUV Backscattered Ultraviolet
CPR Chemically Perturbed Region
DFT Discrete Fourier Transform
ENSO El Niño-Southern Oscillation
EPS Energetic Particle Sensor
ERBS Earth Radiation Budget Satellite
FFT Fast Fourier Transform
GMC Geostationary Magnetopause Crossing
GOES Geostationary Operational Environmental Satellites
IGY International Geophysical Year
IMF Interplanetary Magnetic Field
LIMS Limb Infrared Monitor of the Stratosphere
LTE Local Thermodynamic Equilibrium
NASA National Aeronautics Space Administration
NAT Nitric Acid Trihydrate
NOAA National Oceanic and Atmospheric Administration
NSSDC National Space Science Data Centre
PCA Polar Cap Absorption
PJM Pair Justification Method
PSC Polar Stratospheric Cloud
PV Potential vorticity
QBO Quasi Biennial Oscillation
SAGE Stratospheric Aerosol and Gas Experiment
SAM Stratospheric Aerosol Measurement
SAO Semi-Annual Oscillation
SAOZ Système d'Analyse par Observations Zénithales
SBUV Solar Backscattered Ultraviolet
SELDADS Space Environment Laboratory Data Acquisition and Display System
SME Solar Mesospheric Explorer
SMM Solar Maximum Mission
SOI Southern Oscillation Index
SPE Solar Proton Event

SPRI Space Physics Research Institute (at University of Natal)
SRB Schumann Runge Bands
SRC Schumann Runge Continuum
SSBUV Shuttle Solar Backscattered Ultraviolet
SST Sea Surface Temperature
SSW Sudden Stratospheric Warming
TOMS Total Ozone Mapping Spectrometer
TOVS Tiros Operational Vertical Sounder
UARS Upper Atmosphere Research Satellite
USAF United States Air Force
VGA Video Graphics Adapter
WMO World Meteorological Organization
XRS X-ray Sensor

Chemical Families

NO_x: N + NO + NO_2
NO_y: N + NO + NO_2 + NO_3 + HO_2NO_2 + HNO_3 + $2N_2O_5$ + $ClONO_2$
O_x: O + O_3
ClO_x: Cl + ClO + HOCl
HO_x: H + OH + HO_2

Chapter 1

A review of the middle atmosphere

1.1 Introduction

Ozone constitutes a very small fraction (1:1 000 000) of the earth's atmosphere, and yet it is crucial for the sustenance of life. Ozone is the only atmospheric species which effectively absorbs solar ultraviolet radiation around 250-300 nm, protecting plants and animals from harmful exposure. Abundances of ozone maximize at all altitudes near 20-30 km, forming the ozone layer. This layer is particularly vulnerable to external perturbations, be they from natural (e.g. solar variability) or anthropogenic sources.

The purpose of this chapter is to establish the basic chemical (section 1.2), dynamical (section 1.3) and radiative (section 1.4) principles, and their coupling, in the middle atmosphere. Particular emphasis is placed on oxygen species (O_2, O and O_3) to establish their unique contribution.

1.2 Elementary atmospheric chemistry

Almost all chemical constituents present in the middle atmosphere undergo chemical and photochemical changes which affect their distributions. The sun provides photons of the correct wavelength (mostly ultraviolet and visible) to break chemical bonds and ionize species, producing reactive fragments. The absorption of solar energy also plays an important role in determining the thermal budget which, in turn, governs the fundamental dynamics of the atmosphere.

1.2.1 Thermochemistry

Basic principles governing the reactivity of chemical substances lie in the thermodynamic properties of these species. To evaluate the feasibility of chemical processes, the enthalpy of formation (ΔH) of chemical constituents or 'heat content' represents the energy required to make or break the chemical bonds which compose that substance, starting from the elements in their most stable form (e.g. O_2, N_2, S). These substances are conventionally assigned the enthalpy of formation zero. Numerical values for the enthalpies of compounds in this sub-section are from *CRC handbook of chemistry and physics (1984)*. The standard enthalpy of reaction ($\Delta H^{\ominus}$) for a pressure of 1 atm and a temperature of 298 K is

$$\Delta H^{\ominus} = \Sigma \Delta H^{\ominus}(\text{products}) - \Sigma \Delta H^{\ominus}(\text{reactants}).$$

Reactions that have $\Delta H > 0$ are endothermic, and so transfer of heat into the system takes place resulting in cooling of the surroundings. Those that have $\Delta H < 0$ are exothermic and release heat into the surroundings. An exothermic reaction that is central to the thermal structure of the middle atmosphere is ozone formation. The standard enthalpy for this reaction,

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

may be written as

$$\begin{aligned}
\Delta H^{\ominus} &= \Delta H^{\ominus}_{O_3} + \Delta H^{\ominus}_{M} - \Delta H^{\ominus}_{O} - \Delta H^{\ominus}_{O_2} - \Delta H^{\ominus}_{M} \\
\Delta H^{\ominus} &= \Delta H^{\ominus}_{O_3} - \Delta H^{\ominus}_{O} \\
\Delta H^{\ominus} &= 142.7 - 249.2 \ \text{kJ mol}^{-1} \\
\Delta H^{\ominus} &= -106.5 \ \text{kJ mol}^{-1}.
\end{aligned}$$

Endothermic reactions require an external energy source to drive the reaction. Such a source may be a solar photon with wavelength $\lambda = \frac{hc}{E}$ where h is Planck's constant and c is the speed of light. An example of this type of endothermic reaction is the splitting, by solar ultraviolet radiation, of molecular oxygen

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

The standard reaction enthalpy may be calculated,

$$\begin{aligned}
\Delta H^{\ominus} &= 2 \times \Delta H^{\ominus}_{O} - \Delta H^{\ominus}_{O_2} - h\nu \\
\Delta H^{\ominus} &= 2 \times (249.2 \ \text{kJ mol}^{-1}) - 0 - h\nu \\
\Delta H^{\ominus} &= 498.4 \ \text{kJ mol}^{-1}.
\end{aligned}$$

This reaction requires 498.4 kJ mol^{-1} which corresponds to wavelengths $\leq$ 240 nm.

The enthalpy for non-standard temperatures and pressures may be calculated from a knowledge of the heat capacity at constant pressure (C_p) of a species.

$$\Delta H = \Delta H^{\ominus} + \int_{\ominus}^{T} \Delta C_p \, dT. \tag{1.1}$$

To ascertain whether or not a reaction is spontaneous, the Gibb's free energy ($\Delta G^{\ominus}$) must be calculated. This requires consideration of entropy (S) e.g. if a particular reaction results in a more chaotic dispersal of energy, it is spontaneous. At standard temperature and pressure,

$$\Delta G^{\ominus} = \Delta H^{\ominus} + T\Delta S^{\ominus}. \tag{1.2}$$

1.2.2 Chemical kinetics

Chemical kinetics answers the question: how fast can a reaction occur? Some reactions go to completion (they attain thermodynamic equilibrium) in minutes, while other slower reactions may take many hours. The rates at which atmospheric chemical species are created and destroyed determine their concentrations and lifetimes.

The Rate Law states that the measured rate of a reaction is proportional to the concentrations of the reactants raised to some power. The power to which the concentration of a component is raised is the order of the reaction, and the overall order is the sum of the powers of the concentrations. The order of a reaction is derived from experiment and cannot be inferred by inspection of the chemical equation. The chemical reaction

$$A + 2B \rightarrow 3C + D$$

might have a rate equation

$$-\frac{d[A]}{dt} = k[A][B] \tag{1.3}$$

which is first order in A, first order in B and overall second order. The constant of proportionality is the rate constant (k), which is a function of temperature. Square brackets are used throughout this thesis to represent the concentration of the species.

Alternatively, the rate equation could be:

$$-\frac{d[A]}{dt} = k'[A][B]^2, \tag{1.4}$$

which is first order in A, second order in B and overall third order.

In order to determine a species concentration with time we must integrate the rate equation. Consider the first order reaction

$$A \rightarrow C + D.$$

The rate equation is

$$\begin{aligned} -\frac{d[A]}{dt} &= k_1[A] \\ \frac{d[A]}{[A]} &= -k_1\, dt \\ ln[A] &= -k_1 t + C. \end{aligned}$$

Letting $C = [A]_0$ at $t = 0$

$$[A] = [A]_0 e^{-k_1 t}. \tag{1.5}$$

Inspection of equation (1.5) reveals that the time required for [A] to decrease by $\frac{1}{e}$ is $\frac{1}{k_1}$ which is the chemical lifetime of species A. This concept is extremely important in atmospheric relations. For instance, in order to determine the contribution of transport to a species concentration we need to compare the chemical lifetime of the species with the transport lifetime. If $\tau_{chem} < \tau_{dynam}$, transport effects may be neglected to a first approximation. This is discussed more fully in sub-section 1.3.3.

Often, a second order reaction may be reduced to a pseudo first order reaction. In this case, the assessment of chemical lifetimes becomes simpler. The chemical reaction

$$A + B \rightarrow C + D$$

may have a rate equation

$$-\frac{d[A]}{dt} = -\frac{d[B]}{dt} = k_2[A][B].$$

Assuming that [B] is approximately constant over τ_{chem} we may write

$$[A] \approx [A]_0 e^{-k_2[B]t}$$

where $\tau_{chem} = \frac{1}{k_2[B]}$.

In a system like the atmosphere, numerous gases are interacting simultaneously. In order to examine a chemical lifetime and concentration of a species, all production and loss reactions must be considered. For example, let us assume the full set of reactions involving species A is

$$\begin{aligned} A + B &\rightarrow \text{products } (k_1) \\ A + C + M &\rightarrow \text{products } (k_2) \\ A + F &\rightarrow \text{products } (k_3) \\ G + H &\rightarrow A + \text{products } (k_4). \end{aligned}$$

The rate equation for species A is

$$\frac{d[\mathrm{A}]}{dt} = -k_1[\mathrm{A}][\mathrm{B}] - k_2[\mathrm{A}][\mathrm{C}][\mathrm{M}] - k_3[\mathrm{A}][\mathrm{F}] + k_4[\mathrm{G}][\mathrm{H}].$$

and the chemical lifetime is then

$$\tau_{\mathrm{A}} = \frac{1}{k_1[\mathrm{B}] + k_2[\mathrm{C}][\mathrm{M}] + k_3[\mathrm{F}]}. \tag{1.6}$$

Assuming steady state (the concentrations of the other reactions are changing slowly over τ_{A}) we can set $\frac{d[\mathrm{A}]}{dt}$ to zero. So that the equilibrium concentration of A is

$$[\mathrm{A}] = \frac{k_4[\mathrm{G}][\mathrm{H}]}{k_1[\mathrm{B}] + k_2[\mathrm{C}][\mathrm{M}] + k_3[\mathrm{F}]}. \tag{1.7}$$

Unimolecular reactions

Consider an excited species A*. Under what conditions does a molecule decompose into its constituent parts? The theory of unimolecular reactions involves three processes activation, deactivation and decomposition. They are represented by the following set of reactions

$$\begin{aligned} \mathrm{A} + \mathrm{M} &\rightarrow \mathrm{A}^* + \mathrm{M}\ (k_{act}) \\ \mathrm{A}^* + \mathrm{M} &\rightarrow \mathrm{A} + \mathrm{M}\ (k_{deact}) \\ \mathrm{A}^* &\rightarrow \mathrm{B} + \mathrm{C}\ (k_{decomp}). \end{aligned}$$

At high pressures, deactivation occurs more readily than decomposition i.e. $k_{decomp} \gg k_{deact}$. The rate limiting step is therefore decomposition and the reaction is found to be independent of pressure. This results in a first order process

$$\frac{d[A]}{dt} = -k_{decomp}\frac{k_{act}[\mathrm{A}]}{k_{deact}}. \tag{1.8}$$

At low pressures, the rate limiting step will be the formation of the excited species. This results in a second order process

$$\frac{d[\mathrm{A}]}{dt} = -k_{act}[\mathrm{A}][\mathrm{M}]. \tag{1.9}$$

At intermediate pressures, unimolecular processes are neither first nor second order, as is the case when considering the decomposition of N_2O_5 at middle atmospheric pressures.

1.2.3 Term symbols

Atoms and molecules can possess energy in a variety of forms, one of them being the energy associated with a particular electronic configuration of the particle. The energy difference between electronic states is large (~eV) so that photons capable of inducing such transitions (sometimes leading to the splitting of the compound) are found at ultraviolet and visible wavelengths, as illustrated in Figure 1.1. Transitions between modes of vibrational and rotational states involve infrared and far infrared wavelengths, also shown in Figure 1.1. Finally, atoms and molecules also possess kinetic energy which, for our purposes, we assume not to be quantized.

Figure 1.2 depicts typical potential energy profiles with internuclear separation for a diatomic molecule. Curves XY and XY* represent bonded states. The minima of these curves represent the most stable configuration (too close and strong repulsive nuclear forces come into play) of the molecules XY and XY*. An electronic state corresponding to a potential energy curve XY** is repulsive at all distances and so describes an unbonded state. Vibrational states of bonded molecules are represented by fine horizontal lines superimposed on curves XY and XY*. Hyperfine horizontal line structure would correspond to rotational states (not shown in Figure 1.2).

The electronic state of any particle may be denoted, in short hand, by term symbols.

Atomic term symbols

To illustrate the meaning of an atomic term symbol (${}^{M}L_{J}$) we shall use the example of atomic oxygen (O). The Aufbau diagram, which illustrates the orbital placement of electrons, for atomic oxygen is

$$\begin{array}{ll} 2p & \uparrow\downarrow \quad \uparrow \quad \uparrow \\ 2s & \uparrow\downarrow \\ 1s & \uparrow\downarrow \end{array}$$

There are, clearly, two unpaired electrons in the 2p orbitals. Each unpaired electron has an orbital angular momentum quantum number (l) of 1 and a spin quantum number (s) of $\pm\frac{1}{2}$. The total orbital angular momentum quantum number (L in the atomic term symbol) is given by the Clebsch-Gordon series

$$L = l_1 + l_2,\ l_1 + l_2 - 1, \cdots \mid l_1 - l_2 \mid . \tag{1.10}$$

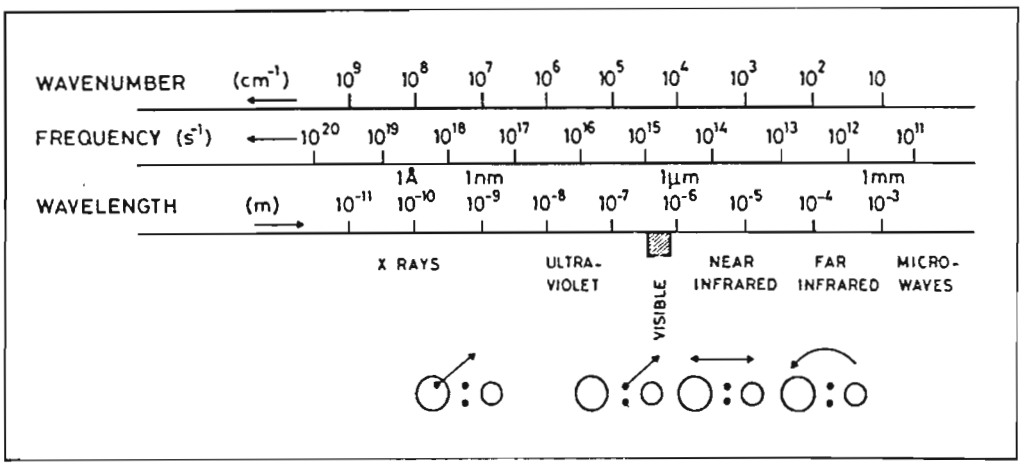

Figure 1.1: The effect of radiation on molecules. Short wavelengths (ultraviolet) may ionize, visible wavelengths may dissociate, long wavelengths (infrared) may alter vibrational and rotational states. From *Brasseur and Solomon (1986).*

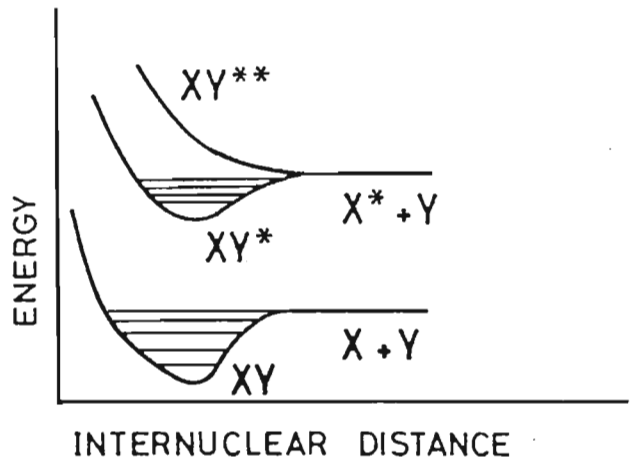

Figure 1.2: The potential energy diagram for a typical diatomic XY. From *Brasseur and Solomon (1986)*.

In this case,

$$L = 2, 1, 0.$$

L is represented by a letter such that

$$\begin{aligned} L &= 0, 1, 2, 3, \cdots \\ L &= S, P, D, F, \cdots \end{aligned}$$

The total spin quantum number (S) is determined by a similar series

$$S = s_1 + s_2,\ s_1 + s_2 - 1, \cdots \mid s_1 - s_2 \mid . \tag{1.11}$$

For atomic oxygen,

$$S = 1, 0.$$

The multiplicity (superscript M in the atomic term symbol) is

$$M = 2S + 1, \tag{1.12}$$

so that

$$M = 3\ (S = 1), 1\ (S = 0).$$

The coupling between the spin and orbital momenta is called the Russell-Saunders coupling (subscript J in the atomic term symbol)

$$J = L + S,\ L + S - 1, \cdots \mid L - S \mid . \tag{1.13}$$

Therefore,

$$\begin{aligned} J &= 3, 2, 1\ (S = 1, L = 2) \\ J &= 2, 1, 0\ (S = 1, L = 1) \\ J &= 1\ (S = 1, L = 0) \\ J &= 2\ (S = 0, L = 1) \\ J &= 1\ (S = 0, L = 1) \\ J &= 0\ (S = 0, L = 0). \end{aligned}$$

All possible electronic states are O($^3D_3$, $^3D_2$, $^3D_1$, $^3P_2$, $^3P_1$, $^3P_0$, $^3S_1$, $^1D_2$, $^1P_1$ and $^1S_0$). The common electronic states in the middle atmosphere, in decreasing energy, are O(^1S, ^1D and ^3P).

Transitions between states obey selection rules based on quantum mechanical principles. Photons have unit angular momentum and, as such, can change the orbital angular momentum by one unit i.e. $\Delta L = \pm 1, 0$. Photons cannot alter the spin of the

electrons i.e. $\Delta S = 0$. This implies $\Delta J = \pm 1, 0$ but $J = 0 \rightarrow J = 0$ is not allowed. If a transition violates these rules it is unlikely to occur and termed 'unfavourable'. If the selection rules are obeyed, transitions are 'allowed' and consequently exhibit short radiative lifetimes.

Some examples of electronic transitions are given in Table 1.1. The transition $O(^1S \rightarrow {}^1D)$ emits a green line ($\lambda = 557.7$ nm). It is not favourable ($\Delta L = 2$) and subsequently has a relatively long radiative lifetime of 0.74 s. The transition $O(^1D \rightarrow {}^3P)$ emits a red line ($\lambda = 630.0$ nm). This transition requires $\Delta S = 2$ which results in a radiative lifetime of 110 s.

Molecular term symbols

To illustrate a molecular term symbol, which has the form $\mathbf{X}^M\Lambda_p$, we shall use the example of the homonuclear diatomic O_2. The molecular orbital energy level diagram for this species is shown in Figure 1.3.

The molecular angular momentum (Λ) is now represented by a capital Greek letter

$$\Lambda = 0, 1, 2, 3, \cdots$$

$$\Lambda = \Sigma, \Pi, \Delta, \Phi, \cdots$$

The **X** is used to denote the ground state of the molecule (X); the first and second excited species are indicated by A, B, or a, b, etc. The superscript M is the multiplicity (as before). The subscript p is either 'g' or 'u' and denotes parity. This indicates whether the bonding orbitals are 'bonding' or 'anti-bonding' and is therefore related to the sign of the amplitude of each of the atomic bonding orbitals. If there is constructive overlap between the σ bonding orbitals (resulting from 'head to head' overlap) then p = g (*gerade*, the German for even) and if there is destructive interference then p = u (from *ungerade*, odd). The opposite is true for π (resulting from broadside overlap of p-orbitals) bonds. The 'g' or 'u' character of each bonding electron must be determined. The overall parity of the molecule is the product (g × g = g, g × u = u, u × u = g) of the electron parities. The electrons forming the chemical bond for ground state molecular oxygen (see Figure 1.3) are g × g = g for the π anti-bonding molecular orbital, and g × g = g for the σ bonding molecular orbital. The overall parity is therefore g. Some of the states of O_2, in decreasing energy, are: $O_2(B^3\Sigma_u^-$, $A^3\Sigma_u^+$, $a^1\Delta_g$ and $X^3\Sigma_g^-)$.

The same selection rules apply to molecules as to atoms. However, there is one additional rule: g → u and u → g are allowed, but u → u and g → g are difficult. Examining Table 1.1, the g → g transition of $O_2(a^1\Delta_g \rightarrow X^3\Sigma_g)$ is unfavourable with a very long

Lower state	Excited state	Radiative lifetime (s)	λ (Å)	Name
$O(^3P)$	$O(^1D)$	110	6300	Red line
$O(^1D)$	$O(^1S)$	0.74	5577	Green line
$O_2(X^3\Sigma_g^-)$	$O_2(a^1\Delta_g)$	2.7(3)	12700+	Infrared atmospheric bands
$O_2(X^3\Sigma_g^-)$	$O_2(b^1\Sigma_g^+)$	12	7619+	Atmospheric bands
$O_2(X^3\Sigma_g^-)$	$O_2(A^3\Sigma_u^+)$	1	2600-3800	Herzberg bands
$OH(X^2\Pi)_{v=0,1,..}$	$OH(X^2\Pi)_{v=9,8,..}$	6(-2)	< 28007	Meinel bands
$N(^4S)$	$N(^2D)$	9.36(4)	5200	
$N(^4S)$	$N(^2P)$	12	3466	
$N_2(X^1\Sigma_g^+)$	$N_2(A^3\Sigma_u^+)$	2	2000-4000	Vegard-Kaplan bands
$NO(X^2\Pi)$	$NO(A^2\Sigma^+)$	2(-7)	2000-3000	γ bands

Table 1.1: Emissions by excited oxygen and nitrogen species in the middle atmosphere. From *Brasseur and Solomon (1986).*

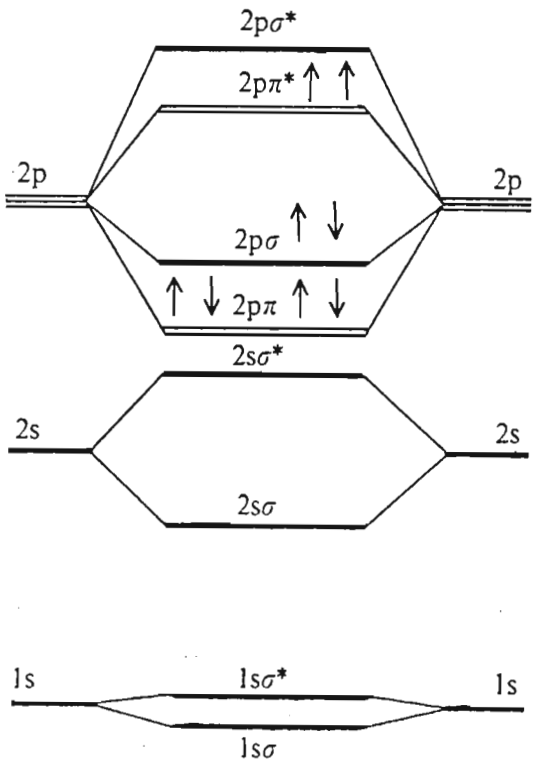

Figure 1.3: The molecular orbital diagram of O_2. Adapted from *Atkins (1982)*.

radiative lifetime of 2.7×10^3 s. It also violates the spin selection rule. On the other hand, the transition $NO(X^2\pi \rightarrow A^2\Sigma)$ is fully allowed and has a lifetime of 2×10^{-7} s.

Oxygen is double bonded (O=O), comprising a single σ bond plus a single π bond. The Pauli principle (electrons of opposite spin pair and form the lowest energy configuration) is violated in the $2p\Pi^*$ orbital. In this case, Hund's rule is applied which states that the lowest energy configuration for two electrons in the highest energy molecular orbital is attained when the spins are aligned. This is the source of oxygen's magnetic properties.

Vibrational and rotational term symbols

Gases in the atmosphere emit radiation as a result of their temperature. The most important radiatively active gases are CO_2, H_2O vapour and O_3. All emitted radiation is at infrared wavelengths and is due to molecules changing vibrational states. Each state corresponds physically to stretching and bending of chemical bonds of the molecule. Figure 1.4 presents a schematic diagram of the possible vibrational modes of diatomic, linear triatomic and bent triatomic molecules. Each mode is labelled ν_1, ν_2 and ν_3. For each molecule an ordered list of quantum numbers e.g. (010) means ground or '0' state for ν_1 and ν_3 modes and a higher energy or '1' state for ν_2. These vibrational states are represented by the fine horizontal lines in Figure 1.2.

Rotational transitions always accompany changes in vibrational modes. This is due to the fact that much less energy is required for such transitions. The result is that spectral lines associated with vibrational transitions have the fine structure of rotational transitions modulated onto them, resulting in band, rather than line, spectra.

1.3 Atmospheric structure and dynamics

The distribution of chemical species in the middle atmosphere depends, in general, on both dynamical and chemical processes. Concentrations of some chemical species are determined by chemistry alone, and others by transport. The criterion determining the relative contribution of these processes will be discussed in sub-section 1.3.3. Transport can be both by prevailing winds (advection) or by turbulent mixing (diffusion). Certain photochemical species, particularly ozone, can influence the radiative budget and indirectly affect temperature and dynamic flow patterns.

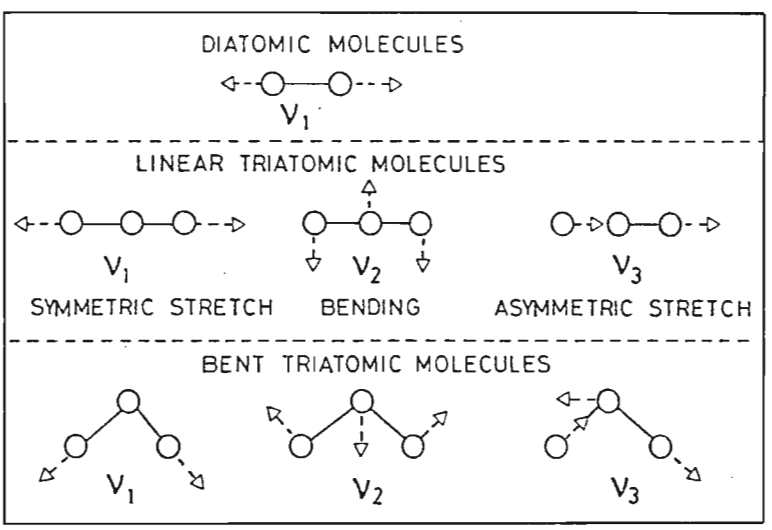

Figure 1.4: The vibrational modes, and their corresponding physical states, of diatomic and triatomic molecules. From *Brasseur and Solomon (1986).*

1.3.1 Vertical structure and some observed dynamical characteristics

The earth's atmosphere is commonly described as a series of layers, each defined by a characteristic temperature gradient which alternates between positive and negative values as illustrated on the left in Figure 1.5. The boundaries between layers (called pauses) are regions of constant temperature. In reality, the pauses are more or less indistinct, with gases mixing upwards and downwards across the boundaries.

The lowest layer, called the troposphere, exhibits decreasing temperatures with altitude and a minimum is attained at the tropopause. The temperature and location of the tropopause varies with latitude and season. At the equator, its mean altitude is located near 18 km and the corresponding temperature is about -80°C. In polar regions its altitude is only 8 km and the temperature -50°C. Above the tropopause, the stratosphere begins and temperatures increase with altitude up to maximum of -10°C at the level of the stratopause, located near 50 km. At still higher altitude, the temperature gradient is again negative, with a minimum at 85 km altitude. This layer is called the mesosphere and its upper boundary is the mesopause. In these layers the mean molecular weight of air varies very little, consisting of roughly 80% N_2 and 20% O_2. The three layers are known collectively as the homosphere.

Above the mesopause the temperature increases rapidly in a layer called the thermosphere. Temperatures exhibit large fluctuations corresponding to changes in solar activity. This layer is largely populated by atomic species (especially O) which have been photolyzed. Above 100 km the mean molecular weight of air varies with altitude and this region is called the heterosphere.

A typical ozone profile is given on the right side in Figure 1.5 exhibiting a maximum in the stratosphere. Ozone is the major source of heating in the stratosphere due to absorption at ultraviolet wavelengths. Molecular oxygen plays a similar role, at higher altitudes and shorter wavelengths. Radiative cooling occurs through infrared emissions associated with vibrational relaxation of carbon dioxide, water and ozone. Clearly, there is an important relationship between atmospheric chemical composition and the radiation budget and therefore the thermal structure of the atmosphere. However, when only 'chemical' factors are considered when calculating the *net* radiative heating rate (as depicted in Figure 1.6). discrepancies occur with the observed temperature structure (Figure 1.7). In particular, large heating rates are predicted for the summer mesopause (top left hand corner of Figure 1.6) but temperatures in this region are lower than their winter counterpart. In addition, the tropical tropopause is much colder (200 K) than at polar latitudes (240 K), as previously stated, but no radiative heating gradient over these latitudes exists. These facts illustrate the importance of including dynamical

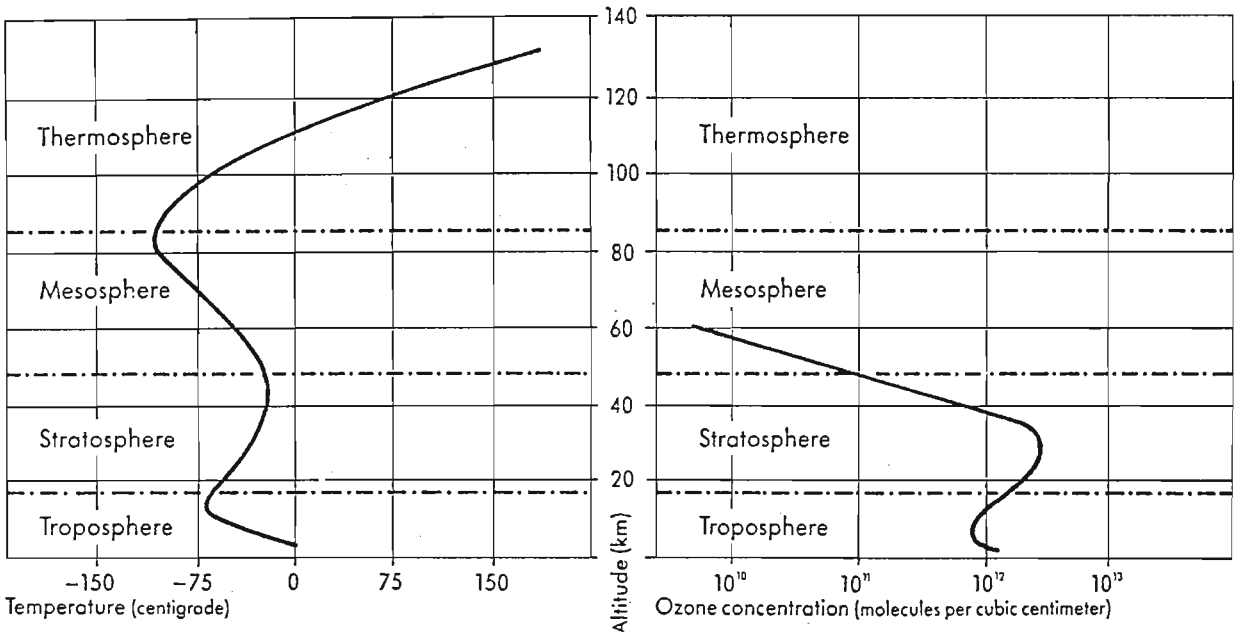

Figure 1.5: The layered structure of the atmosphere is determined by the sign of the temperature gradient, shown on the left side. An ozone profile, showing maximum ozone concentration in the stratosphere is given on the right side. From *Gribbin (1988)*.

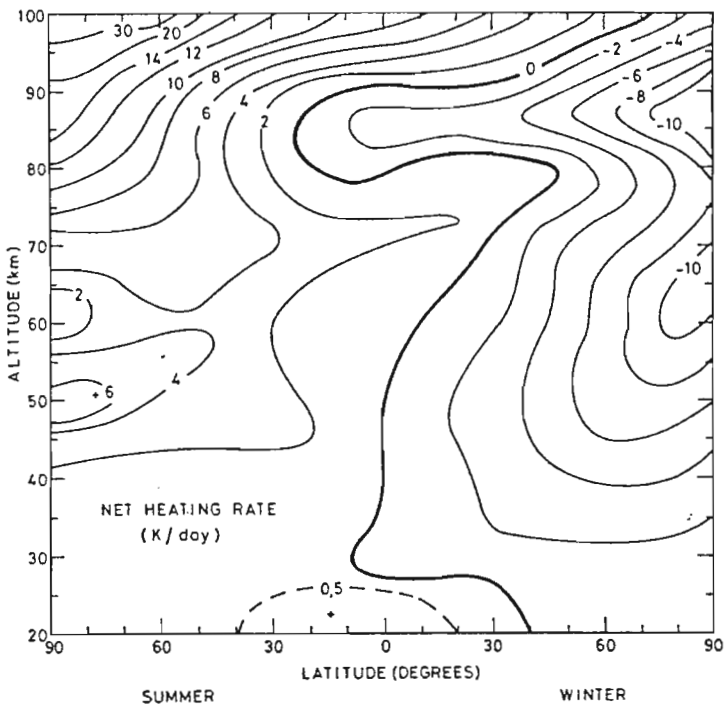

Figure 1.6: Total *net* radiative heating rate (°/day) theoretically determined from absorption of ultraviolet radiation by O_3 and O_2 and emission by CO_2. From *Brasseur and Solomon (1986).*

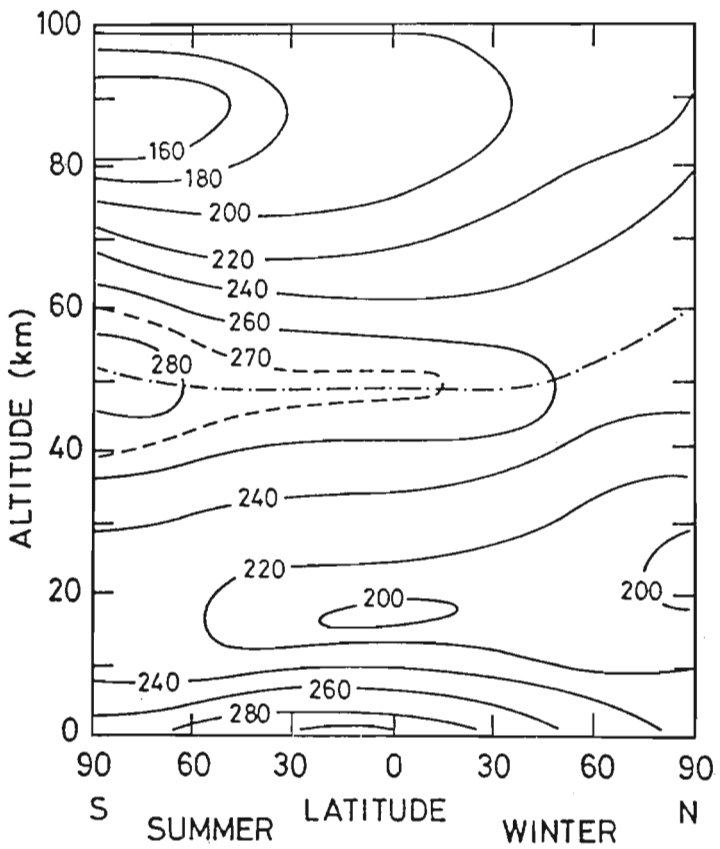

Figure 1.7: Observed atmospheric temperature structure at solstice. From *Brasseur and Solomon (1986)*.

effects. Specifically, an air parcel displaced adiabatically upwards undergoes expansion and net cooling; the opposite is true for adiabatic compression. This principle, derived from thermodynamic considerations, will be described more fully in the next sub-section (1.3.2). Obviously, mean upward motion must be characteristic of the summer mesopause and tropical tropopause. The latter represents one of the ways the lower atmosphere is coupled to the middle atmosphere, which is driven by release of latent heat by cloud formation and precipitation in the tropics.

Other direct dynamical observations are also possible. Zonal (cross-longitude) winds are particularly easy to measure due to their high speeds (many m s^{-1}). In the troposphere they are particularly strong at mid-latitudes. In the mesosphere there is a strong westerly wind in the winter hemisphere while an easterly wind prevails in the summertime. Meridional (cross-latitude) winds can be measured at some altitudes by radar techniques. Most meridional and vertical wind speeds are difficult to measure as they are so small (vertical wind speeds may be as little as $\sim$cm s^{-1}) compared to zonal winds. They can, however, be estimated from theoretical studies.

Throughout this thesis, the magnitude of zonal winds will be represented by u, that of meridional by v, and vertical by w. A mean zonal wind is one for which the wind speed has been averaged over longitude. A zonal eddy is any local departure of zonal wind speed from the zonal mean.

Figure 1.8 represents typical observed temperature and geopotential height fields for summer and winter over the northern hemisphere. Since gravitational acceleration (g) increases with increasing latitude, higher altitudes are required in the polar regions in order to maintain constant gravitational potential energy. This concept is clarified in the following sub-section where the geopotential height is defined. Since air tends to flow along lines of constant geopotential, winds generally do not deviate from geopotentials or lines of latitude. This flow pattern is seen for the summer hemisphere in Figure 1.8 (c) and (d) and is termed zonally symmetric. On the other hand, air flow patterns in winter at high northern latitudes, exhibit regions of relative highs and lows indicative of meridional (cross-latitude) flow. The wave structure in Figures 1.8 (a) and (b) can be described as a wave number one pattern, exhibiting only one major ridge (high) and one trough (low) in a given longitudinal direction. Observations (e.g. *Hare and Boville, 1965*) have established that large scale planetary waves, of wave number < 3 account for most of the wave structure in the stratosphere. The presence of waves plays an important role in transporting ozone rich air from the equator to polar latitudes in spring.

Another possible mechanism for transport is diffusion. This occurs at molecular level due to random motion of atoms and molecules. Diffusion is only important in the thermosphere. At lower altitudes, and of time scales of interest for ozone studies, it

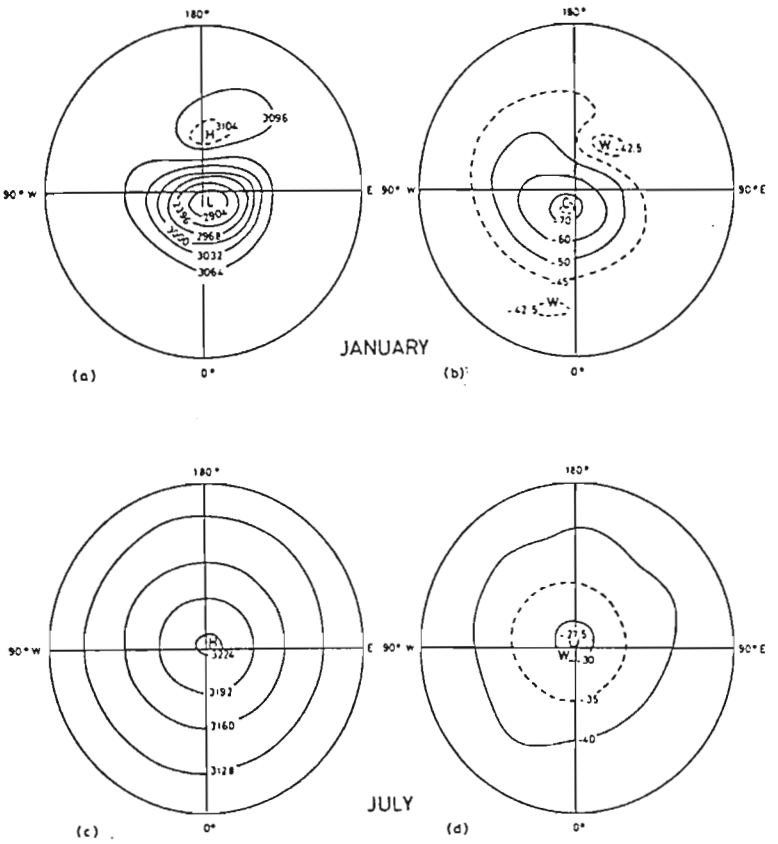

Figure 1.8: Typical geopotential height (dm) and temperature (°C) configurations for winter ((a) and (b)) and summer ((c) and (d)) seasons. From *Brasseur and Solomon (1986)*.

is safe to assume that all transport results from motions of parcels of air, rather than individual particles.

1.3.2 Fundamental atmospheric equations

The hydrostatic equation

In order to derive the hydrostatic equation, we represent the earth's atmosphere by a simple model. This model assumes that the atmosphere is transparent to all radiation, that the temperature of the lower boundary is that of the earth's surface and that the atmosphere is an ideal gas i.e. one in which the molecules move freely, totally without interaction, apart from collisions. The only forces acting on an air parcel can be assumed gravitational. Applying Newton's Second Law we can write

$$g = \frac{GM_{\oplus}}{R^2}, \tag{1.14}$$

where G is the gravitational constant and $M_{\oplus}$ is the mass of the earth. It can be readily seen from this equation that gravitational acceleration is inversely proportional to the square of the distance away from the centre of the earth (R) and so g decreases with altitude. Since the earth spins around an axis its centripetal acceleration (g) varies with latitude, with a minimum at the equator. The definition of a new co-ordinate, to take into account the dependence of g with altitude and latitude, is needed. Lines of constant geopotential (Φ) join heights of equal geopotential energy i.e. they lie parallel to lines of latitude

$$\Phi = \int_0^z g\, dz. \tag{1.15}$$

The geopotential height (h′) is

$$h' = \frac{\Phi}{g_0}, \tag{1.16}$$

where g_0 is the value of g at the earth's surface.

Consider a stationary air parcel (of height dz and area A for upper and lower 'faces') in a gravitational field. The parcel has density ρ, mass m and pressure p. The forces (F) on it are

$$\begin{aligned} F_{bottom} + m\,g &= F_{top} \\ \frac{F_{top} - F_{bottom}}{A} &= \frac{m\,g}{A} \\ p - (p + dp) &= \frac{\rho\, V\, g}{A}. \end{aligned}$$

The pressure difference between the two 'faces' is therefore

$$dp = -g\,\rho\,dz. \tag{1.17}$$

The Ideal Gas Law states

$$\rho = \frac{p\,m_m}{R\,T}, \tag{1.18}$$

where m_m is the molar mass and R is the gas constant. Substituting equation (1.18) into (1.17)

$$\frac{dp}{p} = -\frac{dz}{H}, \tag{1.19}$$

where $H = \frac{R\,T}{m_m\,g}$ is the scale height. The scale height is a function of molar mass and therefore is unique to each chemical species. H is also a function of gravitational acceleration, and therefore latitude and altitude.

We can integrate equation (1.19) so that

$$p = p_0\,e^{-\left(\int_0^z \frac{dz}{H}\right)}, \tag{1.20}$$

where p_0 is standard atmospheric pressure i.e. 1.013×10^5 Pa. By inspection of equation (1.20) it can easily be seen that H represents the increase in altitude (scale height) required to reduce the pressure by $\frac{1}{e}$. The scale height is approximately 6 km in the homosphere. Large variations in scale height occur in the heterosphere where large particle population gradients are prevalent.

An alternative way of structuring the atmosphere is by demarcating a layer each time the pressure is halved. These are Umkehr layers and are generally 5 to 6 km thick.

Adiabatic lapse rate

The simple model applied to the atmosphere in order to derive the hydrostatic equation is assumed once more to obtain the adiabatic lapse rate. In addition, the First Law of Thermodynamics is now applied that relates work (w), heat (q) and internal energy (U) of a system such that

$$dU = dq + dw, \tag{1.21}$$

where dq is the heat supplied to the system, dU is the increase in energy of the system and dw is the work done *on* the system. Work done on or by a gaseous system is frequently work done when a gas expands ($+dV$) or contracts ($-dV$) against an external pressure p. So, $dw = -p\,dV$. An adiabatic expansion ($dq = 0$) implies a decrease in internal energy ($dU = dw$) which results in cooling of the system.

When energy is transferred to a system there is a change of state which may appear as an increase in T. For an infinitesimal transfer of heat the increase in temperature is proportional to the amount of heat supplied viz.

$$dq = C\ dT. \tag{1.22}$$

where C is the heat capacity of a substance. The heat capacity depends on the conditions under which heat transfer is performed. If the system is constrained to a constant volume we denote the heat capacity by C_v. If it is constrained at a constant pressure we denote the heat capacity by C_p

For simplification, we shall assume that the air parcel under consideration has unit mass so that $\rho = \frac{1}{V}$. Specific heat capacities are now denoted by $C_{p,m}$ and $C_{v,m}$. The Ideal Gas Law becomes

$$p\,V = \frac{R\,T}{m_m}. \tag{1.23}$$

Differentiating equation (1.23) and substituting (1.24)

$$C_{p,m} - C_{v,m} = \frac{R}{m_m}, \tag{1.24}$$

we obtain

$$V\,dp + p\,dV = C_{p,m}\,dT - C_{v,m}\,dT. \tag{1.25}$$

A parcel of air exchanges heat with its surroundings much more slowly than the contribution to U by the work of compression or expansion and we may assume any motion of the air parcel to be adiabatic ($dq = 0 \Rightarrow dU = dw$). Equation (1.25) simplifies to

$$C_{v,m}\,dT = -p\,dV. \tag{1.26}$$

Substituting (1.26) into (1.25)

$$V\,dp = C_{p,m}\,dT. \tag{1.27}$$

Recalling equation (1.17)

$$dp = -g\,\rho\,dz$$

since m = 1, the hydrostatic equation takes the form

$$dp = -\frac{g\,dz}{V}. \tag{1.28}$$

Substituting into equation (1.27) we can write

$$-\frac{g\,dz}{V}\,V = C_{p,m}\,dT$$

$$\frac{dT}{dz} = -\frac{g}{C_{p,m}}$$

$$\frac{d\mathrm{T}}{dz} = \Gamma_{\mathrm{d}}. \tag{1.29}$$

The decrease of temperature with increasing altitude is called the lapse rate. Under adiabatic conditions the lapse rate is termed the dry adiabatic lapse rate with a value of $\Gamma_{\mathrm{d}} \sim$ -10 K km^{-1}. Since Γ_{d} is a function of gravitational acceleration it depends on latitude and altitude.

Combining equations (1.27) and (1.23)

$$\frac{\mathrm{R\,T}}{\mathrm{m_m\,p}}\,d\mathrm{p} = \mathrm{C}_{p,m}\,d\mathrm{T}$$

$$\frac{\mathrm{R}}{\mathrm{C}_{p,m}\,\mathrm{m_m}}\,\frac{d\mathrm{p}}{\mathrm{p}} = \frac{d\mathrm{T}}{\mathrm{T}}.$$

This may be integrated and written in the form

$$\mathrm{T} = \mathrm{A\,p}^{\kappa}, \tag{1.30}$$

where $\kappa = \frac{\mathrm{R}}{\mathrm{C}_{p,m}\,\mathrm{m_m}} = 0.286$ for dry air.

Equation (1.30) may be rewritten adopting a reference pressure of 1000 mb (or 1000 hPa) so that

$$\frac{\mathrm{T}}{\Theta} = \left[\frac{\mathrm{p}}{1000}\right]^{\kappa}, \tag{1.31}$$

where Θ is the potential temperature. This is the temperature an air parcel would attain of it were adiabatically (dq = 0) compressed or expanded starting from a temperature T and a pressure of 1000 hPa. Potential temperature is a conservative property in any adiabatic displacement, and can be used to evaluate air parcel trajectories. When Θ does vary with height, the actual lapse rate (Γ) will differ from the dry adiabatic lapse rate.

This departure from the dry adiabatic rate indicates the presence of non-adiabatic processes, such as the absorption of ultraviolet radiation by ozone.

The difference between the actual and the dry adiabatic lapse rates is also related to the tendency of a displaced air parcel to return to its original position. This quantity is called the static stability parameter (S) and is related to Γ - Γ_{d}. If $\Gamma = \Gamma_{\mathrm{d}}$ then an air parcel adiabatically displaced from its position will tend to remain at its new location, since its temperature will be the same as the ambient temperature. On the other hand, if $\Gamma < \Gamma_{\mathrm{d}}$ then an air parcel lifted (lowered) from its equilibrium position will tend to sink (rise) back to its original position. Such oscillatory motion is the source of gravity waves. The atmosphere is then stably stratified.

Quasi-geostrophic potential vorticity equation

Any large air mass moving at speed will be deflected sideways; to the left in the southern hemisphere and to the right in the northern. So, in the southern hemisphere, a west wind will be deflected equatorward and a south wind will be deflected to the west. This is due to the fact that the earth is a non-inertial reference frame as it is spinning around its north-south axis with an angular speed of $\omega_{\oplus} = 7.3 \times 10^{-5}$ rad s^{-1}. In order to retain Newton's Laws of Motion certain 'fictitious forces' must be introduced. Consider placing a mass, m, in a reference frame of a rotating disc. The angular momentum (L) is then

$$L = m\omega r^2, \tag{1.32}$$

where ω is the speed of rotation of the disc and r is the radial distance from the centre. So, when the mass is close to the centre, it has relatively little angular momentum, but if it is moved to a new position further out, m has more angular momentum, so a torque must be exerted in order to move it along the radius

$$\tau = F_c r = \frac{dL}{dt} = 2 m \omega r \frac{dr}{dt}. \tag{1.33}$$

The Coriolis force (F_c) may be expressed as

$$F_c = 2 m \omega v. \tag{1.34}$$

If we now apply equation (1.34) to a surface of a sphere, instead of a disc, we must take into account its curvature

$$F_c = 2 m \omega_{\oplus} v \sin(\phi), \tag{1.35}$$

where ϕ is geographic latitude. There is no Coriolis force at the equator, and a maximum value exists at the poles of $2 m \omega_{\oplus} v$.

For large air masses (scale length L $\sim$ 1000 km) and altitudes $>$ 1 km, where frictional forces may be neglected, we need only consider horizontal pressure gradient and coriolis (ϕ non-zero) forces to be acting. The horizontal pressure gradient force (F_p), over a distance l, is directed from regions of high pressure to low and is given by the formula

$$F_p = -\rho \frac{dp}{dl}. \tag{1.36}$$

A geostrophic wind is achieved when the coriolis force is almost equal, but in an opposite direction, to the horizontal pressure gradient force. The wind therefore has a constant velocity, the geostrophic velocity (v_g), given by

$$v_g = \frac{1}{2 \omega_{\oplus} \sin(\phi) \rho} \frac{dp}{dl} \tag{1.37}$$

which is directed along isobars and thus contours of the geopotential field. These winds are the source of polar vortices every winter.

Geostrophy is an excellent approximation however, equation (1.37) does not have any derivatives with time i.e. it is only a diagnostic and cannot be used for prediction purposes. This necessitates the quasi-geostrophic potential vorticity equation which has geostrophic balance as its lowest order of approximation.

Large air mass dynamics are very similar to fluid flow. The most basic equation to be obeyed is the continuity equation

$$\frac{1}{R_\oplus \cos(\phi)} \frac{\partial u}{\partial \Lambda} + \frac{1}{R_\oplus} \frac{\partial v}{\partial \phi} + \frac{\partial w}{\partial z} = 0, \tag{1.38}$$

where u, v, w are zonal, meridional and vertical wind velocity components and Λ is geographic longitude. This implies that if an air parcel is squeezed horizontally, it expands vertically and *vice versa*.

The occurrence of closed-vortex perturbations in the circulation of the atmosphere is common. The vorticity (ζ_g) or spin of a large scale vortex is given by the curl of the geostrophic velocity

$$\zeta_g = \nabla \times v_g = \frac{1}{R_\oplus \cos(\phi)} \frac{\partial v_g}{\partial \Lambda} - \frac{1}{R_\oplus} \frac{\partial u_g}{\partial \phi} \tag{1.39}$$

ζ_g is also termed the total relative vorticity.

Other vorticity factors must also be considered. One, already mentioned, is of planetary origin. The variation of coriolis force with latitude, if air is moving equatorward the coriolis force decreases and the wind then experiences a net torque due to the horizontal pressure gradient and its vorticity increases. The other vorticity component to be considered is due to stratification of the atmosphere. Total vorticity (sum of all three vorticity factors mentioned above) is conserved with time under adiabatic conditions (q = 0) viz.

$$\frac{d_g}{dt} [\nabla \times v_g \text{ term} + \text{coriolis term} + \text{fluid stratification term}] = \text{term in q.} \tag{1.40}$$

This is the general form of the quasi-geostrophic potential vorticity equation. The details of the equation depend, among other things, on the horizontal scale of the motions of interest.

Variations in the planetary vorticity factor with latitude leads to horizontal wave structure since air that accelerates (decelerates), diverges (converges). Some waves have very long wavelengths (~10 000 km) and are usually stationary, these are called planetary waves, depicted in Figure 1.8 (b).

1.3.3 Dynamical effects on chemical species

This sub-section establishes a rough guide of when and why atmospheric transport processes influence chemical species. The particular effects on ozone will be discussed in following sections of this chapter. It has already been shown (section 1.2.2) that the time constant appropriate to photochemical processes (τ_{chem}) can be evaluated from a knowledge of the rate of loss of chemical species. Dynamical time constants are somewhat harder to evaluate.

If n_i is the number density for a vertical distribution of species i, and if one assumes the vertical distribution of species i to be described by a scale height H_i, then

$$n_i = n_{i,0}\, e^{-(\frac{z}{H_i})}, \tag{1.41}$$

where z is altitude and $n_{i,0}$ is the density of chemical species i at a reference altitude. If w (vertical speed) is constant over spatial distance H, then z = -wt and

$$n_{i,t} = n_{i,t_0}\, e^{(\frac{w1}{H_i})} \tag{1.42}$$

so that the time required for $n_{i,t}$ to change by $\frac{1}{e}$ relative to its initial value n_{i,t_0} is $\frac{H_i}{w}$. This defines a time constant for transport by vertical winds. Assuming H $\sim$ 5 km for zonally averaged mean vertical winds, $\tau_w \sim$ months in stratosphere and days in upper mesosphere.

Time constants for meridional and zonal winds are derived in a similar fashion, but it is far more difficult to characterize their gradients. Assuming a scale length (L) of $\sim$1000 km in the zonal and meridional directions, the time constant for zonal transport (τ_u) is of the order of days throughout the middle atmosphere. The time constant for mean meridional transport (τ_v) is approximately months in the stratosphere and days in the upper mesosphere.

Three different cases can be identified to highlight the relative importance of dynamics and chemistry:

$\tau_{chem} \ll \tau_{dynam}$: In this case the species is in photochemical equilibrium and the effect of dynamics is not directly important. Although dynamics may be important indirectly e.g. through temperature or coupling with a chemically longer lived species.

$\tau_{chem} \gg \tau_{dynam}$: In this case the dynamics will mix chemical constituents well to eliminate gradients e.g. τ_{chem} for N_2O is of the order of years in the lower stratosphere while τ_u is days, resulting in a uniform mixing ratio of N_2O with lines of longitude. Species such as N_2 and O_2 are so long lived that they are thoroughly mixed in all directions in the middle atmosphere.

$\tau_{chem} \sim \tau_{dynam}$: In this case both dynamics and chemistry play vital roles. For example N_2O meridional and vertical dynamical lifetimes are similar to its chemical lifetime so that meridional transport is important in determining N_2O density. The case of zonal asymmetries first mentioned in sub-section 1.3.1 should be again focused on here. In summer the vector wind is almost exclusively zonal, but in winter the influence of planetary waves induces cross-latitude motion. Under these conditions $\tau_v \sim \tau_u$, and so, any species exhibiting a zonal gradient will then exhibit a meridional one. This is best illustrated by planetary wave activity whose influence on the distribution of ozone has long been recognized (*Berggren and Labitzke, 1968*).

1.3.4 Dynamics in 2-dimensions

Three-dimensional models for global atmosphere dynamics present computational difficulties and a popular solution is zonally averaged (latitude versus altitude) plots. An example of a 2-dimensional plot is shown in Figure 1.6. Unfortunately these plots are often, by their nature, inadequate.

Although the earth-atmosphere climatic system is in global energy balance, local radiative imbalances do occur leading to atmospheric transport processes. Figure 1.9 shows the solar energy absorption (dashed line) is strong at tropical latitudes whereas, reflection becomes dominant over polar regions. On the other hand, the energy emitted (solid line) depends on temperature making its latitude variation small. To establish an equilibrium, a net transport of energy from the equator to the pole must occur. This circulation, shown in Figure 1.10 (attributed to *Murgatroyd and Singleton, 1961*), was derived by tracer studies of O_3 and H_2O. *Brewer (1949)* suggested a circulation exhibiting rising motion in the tropics, and descending motion at extra tropical latitudes to explain the dryness of the stratosphere. Heating rates were theoretically calculated for radiative gases such as O_3, O_2 and CO_2. Such a circulation pattern is referred to as a Hadley cell.

The earth-atmosphere system must have constant angular momentum unless an external torque is applied to it. Hadley cell circulation which is equatorward at low altitudes imparts easterly momentum to the earth's surface (due to coriolis force). Since, on average, there can be no exchange of momentum between the atmosphere and the surface the transfer of easterly momentum in Hadley circulation must be balanced by a transfer of westerly momentum at higher latitudes. This circulation represents a Ferrell cell. Using this additional consideration, model mean meridional circulations (Figure 1.11, attributed to *Cunnold et al. (1975)*) were devised.

Which circulation is correct? The answer is both. Transport may be traced to three

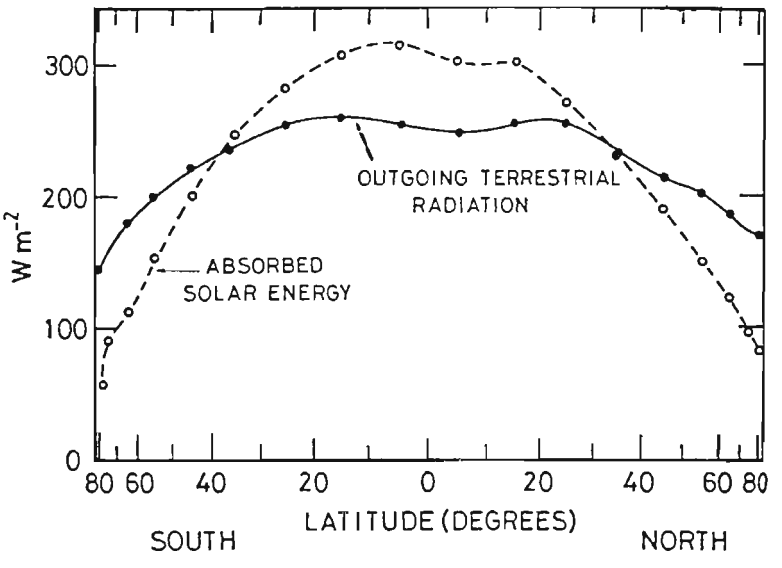

Figure 1.9: Solar energy (W m^{-2}) absorbed by the earth-atmosphere system (dashed line) and the emitted terrestrial radiative energy (solid line) as a function of latitude. From *Brasseur and Solomon (1986).*

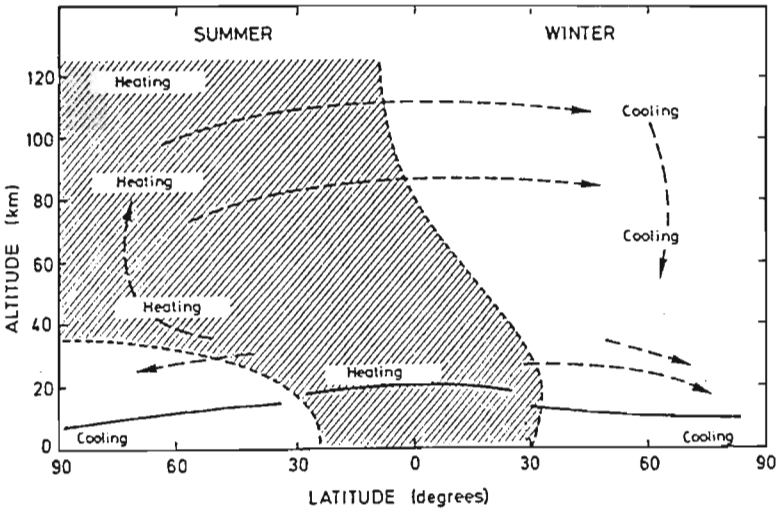

Figure 1.10: Meridional circulation at solstice based on a study by *Murgatroyd and Singleton (1961)*. From *Brasseur and Solomon (1986)*.

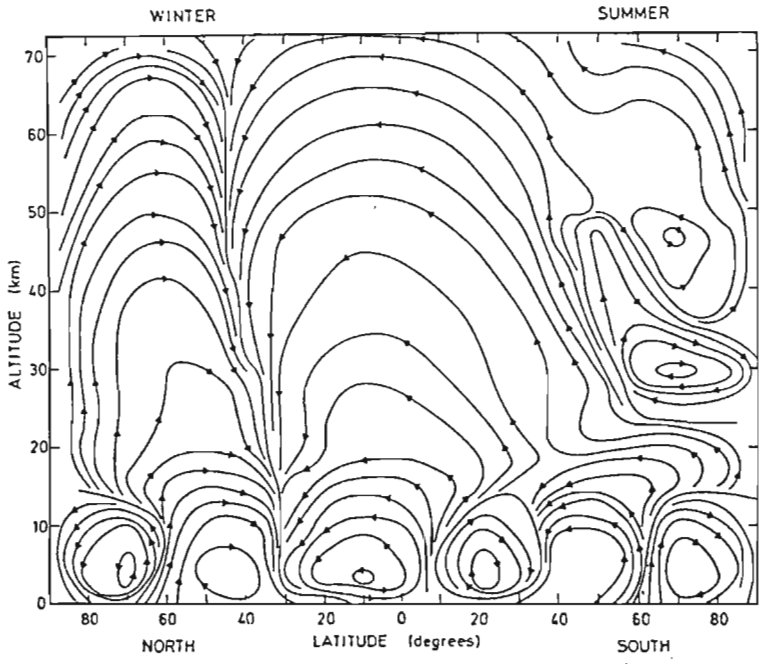

Figure 1.11: Meridional circulation at solstice deduced from a theoretical study by *Cunnold et al. (1975)*. From *Brasseur and Solomon (1986)*.

driving forces: (a) the global energy imbalance; (b) mean meridional circulation (classical Eulerian) derived from momentum considerations; (c) eddy transport due to gradients along lines of latitude. Figure 1.11 represents a circulation whose source is (b). If we assume zonal wave motion is steady and conservative (we can safely average zonal heating rates) then eddy transport will approximately cancel mean meridional circulation. The *net* transport is a small difference between the two (*Mahlman et al., 1980*) which is represented in Figure 1.10. This implies that mean and eddy transports are intimately coupled. Since the large scale planetary waves (wave number ≤ 3) which dominate the wave structure of the stratosphere are generally conservative, zonally averaging is acceptable. However, the eddy mean flow cancellation can represent a difficulty for 2-dimensional descriptions.

Of course, one could calculate the net mean circulation from the first principles of diabatic heating rate. For this, temperatures have to be known *à priori*. *Dunkerton (1978)* has performed these calculations and the resulting 2-dimensional plot corresponds exactly to that of Murgatroyd and Singleton. So, by considering only thermodynamic and continuity (omitting momentum) equations net transport can be derived, provided that eddy transience is ignored. It is also called diabatic circulation or residual Eulerian circulation.

In the absence of chemical sources or sinks, a chemical tracer has constant potential temperature and the transport of energy may be applied to the transport of chemical species. If $\tau_{chem} \sim \tau_u$ then the effects of production and loss along wave trajectories cannot be ignored. Under these special circumstances chemical eddy transport becomes important. One species affected by eddy transport is stratospheric ozone at high latitudes in winter.

Wave transience and dissipation

The preceding section shows that the meridional circulation derived from transformed thermodynamic equations by neglecting wave transience (growth or decay of wave amplitude) and dissipation (thermal damping, turbulent diffusion) is quite useful, since it corresponds approximately to the *net* atmospheric transport. The wave structure is not always conservative (e.g. during deceleration of zonal winds) and the eddy and mean meridional processes do not cancel. For example, large variations of the zonal wind are observed in the tropical lower stratosphere with a mean period of 26 months. This phenomenon is the quasi-biennial oscillation (QBO). Another example of non-conservative wave activity is the phenomenon of sudden stratospheric warmings (SSW). *Matsuno and Nakamura (1979)* have shown that sudden warmings are probably related to strong planetary wave drag due to wave transience. This induces a mean meridional

circulation with upward motion in the tropics and strong downward motion at high latitudes resulting in warming of the middle atmosphere by adiabatic compression. Consequently, the meridional temperature gradient is decreased and the polar jet begins to break up, finally reversing. SSW's have significant effects in the annual cycle of total ozone.

1.4 Radiation

1.4.1 Introduction

In section 1.3 we considered the atmosphere to be heated by the earth's surface and transparent to all radiation. This of course is not true. The absorption or emission of photons (at both ultraviolet and infrared wavelengths) can alter a particle's energetic state. The opacity of the atmosphere is important in determining local temperature, which in turn affects dynamics, chemical kinetics, and the production of important reactive ions and radicals.

Solar energy is absorbed by ozone (stratosphere and mesosphere) and molecular oxygen (upper mesosphere and lower thermosphere). It is rapidly converted to thermal energy through chemical reactions involving the recombination of these species. To maintain an equilibrium cooling must take place. Thermal emission due to vibrational relaxation of CO_2, O_3 and H_2O takes place in the infrared part of spectrum. In addition, the production of latent heat in the troposphere and dissipation of atmospheric waves contribute to heat loss. An analysis of the earth's energy balance is outlined in Figure 1.12. Short wave radiation from the sun (100%) is absorbed by the atmosphere (19%) and is reflected back into space (30%). The rest of incoming energy (51%) reaches the surface, where it heats the earth-atmosphere system and subsequently radiates in the infrared region. Infrared emitted by the surface is 21%, by atmospheric gases 38% and by clouds 26%. The surface also emits radiation by sensitive heat and latent heat. Of the total received, 30% are reflected and 70% are emitted as long wave radiation.

Both the sun's and earth's emitted radiation approximate a blackbody spectrum (described in sub-section 1.4.2.2) of 5900 K and 300 K respectively. Radiation of wavelengths < 4 μm is of solar origin and wavelengths > 4 μm can be assumed to be terrestrial and atmospheric in origin, depicted in Figure 1.13.

The effect of absorbers on the solar blackbody spectrum is depicted in Figure 1.14. For $\lambda < 100$ nm radiation is almost completely absorbed above 100 km by O, O_2 and N_2. Some X-rays (< 1 nm) can penetrate to the middle atmosphere causing

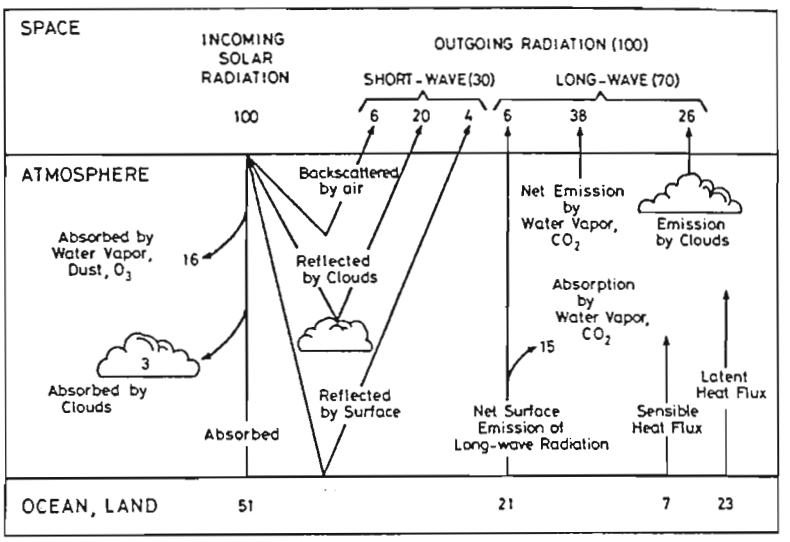

Figure 1.12: The radiation budget for the earth-atmosphere system. From *Brasseur and Solomon (1986)*.

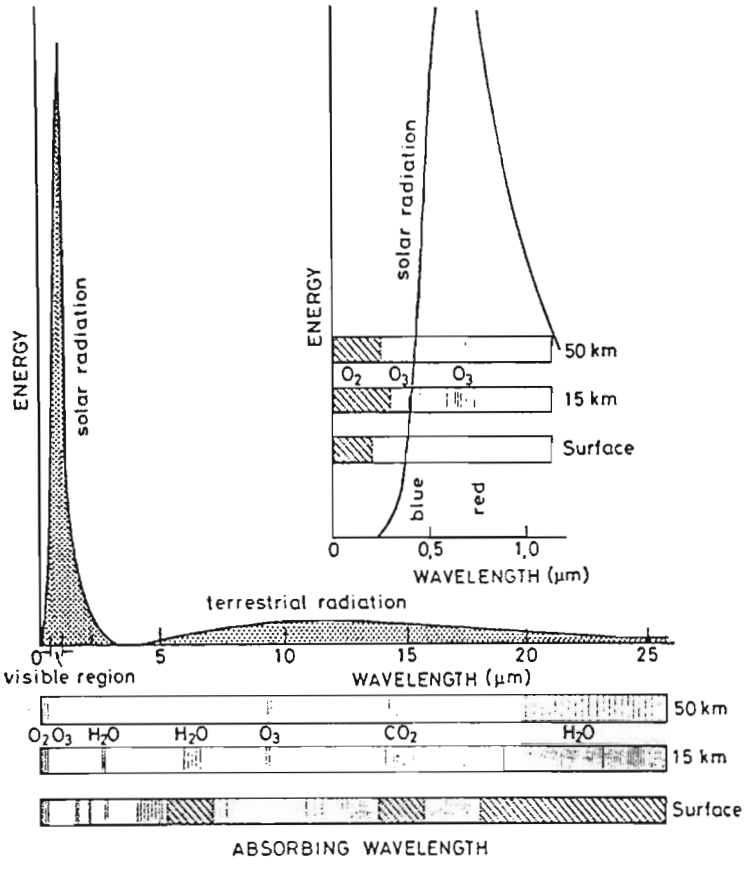

Figure 1.13: The spectra of the electromagnetic radiation emitted by the sun and the earth. Wavelengths less than 4 μm represent radiation of solar origin, while wavelengths longer than 4 μm represent terrestrial and atmospheric radiation. The spectral bands, at which atmospheric absorbers are active, are given at 50 km, 15 km and the earth's surface for the terrestrial spectrum (main graph) and solar spectrum (inset). From *Brasseur and Solomon (1986)*.

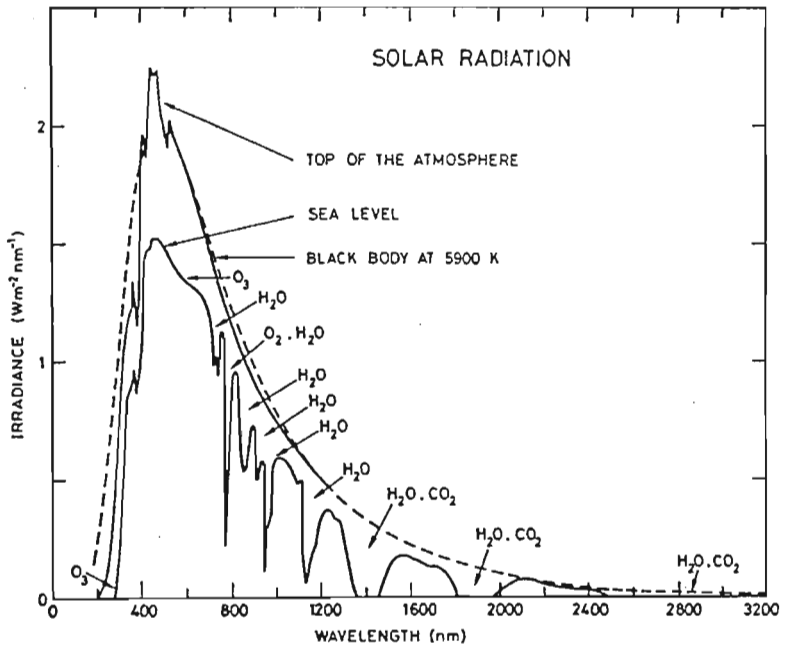

Figure 1.14: The spectrum of solar radiation as approximated by a black-body spectrum at 5900 K (dashed curve), outside the earth's atmosphere and at sea level. The spectrum at sea level clearly illustrates the effect of atmospheric absorbers. From *Brasseur and Solomon (1986).*

sporadic ionospheric disturbances. The Lyman α (n=2 → n=1) line (λ = 121.6 nm) is very intense and it is weakly absorbed in the upper atmosphere and penetrates to the middle atmosphere where it dissociates water vapour, CO_2 and CH_4. Furthermore, it photoionizes NO, a principle source of the D-region. At longer ultraviolet wavelengths the solar spectrum is divided into regions of absorption by two species: O_2 and O_3. O_2 absorbs at wavelengths less than 240 nm. Ozone, abundant in the stratosphere, absorbs between 2 and 300 nm but also to some extent in the visible and infrared. For $\lambda > 310$ nm, a large portion of solar photons reach the troposphere and so molecular scattering, cloud and surface albedo must be considered. At even longer wavelengths, absorption by H_2O and CO_2 becomes dominant.

1.4.2 Solar irradiation

The physical structure of the sun

In order to understand the nature and variations of the solar radiation incident on the earth we need to outline aspects of the sun's atmosphere.

The sun is still a relatively young star as it burns hydrogen to form helium as its energy source. These fusion reactions generate temperatures of $\sim 2 \times 10^7$ K in the interior of the sun.

Most energy reaching the earth's atmosphere originates from a thin layer (1000 km) called the photosphere. This layer defines the visible 'surface' of the sun. It is a relatively cooler layer with an effective temperature of ~5900 K. Observations indicate that the brightness of the photosphere is not continuous, but is characterized by granules which are associated with strong convective processes. Transient phenomena such as sunspots appear in the photosphere and influence the variability in the solar emission at short wavelengths. Sunspots are relatively cool, dark regions (T $\sim$ 3000 K) with a typical diameter of less than 5 $\times 10^4$ km and intense magnetic fields ($\sim$ 1 T). Most sunspots are bipolar and appear in a distribution approximately symmetrical with respect to the equator. The latitude at which they appear varies with the solar cycle according to Maunder 'butterfly' diagram. The solar cycle is an eleven year cycle, and is traditionally observed in the number of sunspots (Wolf number). Larger numbers of sunspots are indicative of an active sun. Sunspots form in regions where there is bunching (or kinking) of solar magnetic field lines. The sun's magnetic field is not static but is 'frozen in' to its plasma (due to high electrical conductivity) and so differentially rotates. It has differential rotation, as it is a non-solid body, 26 days at equator and 37 days at pole. This results in its magnetic poles reversing every $\sim$ 11 years. The lifetime of a sunspot is variable (days - months) but is often long enough to 'be seen again'

by the an observer on the earth as the sun's mean rotation is 27 days. Light emitted lower down in the photosphere (a continuous spectrum) is absorbed by cooler gases higher up resulting in dark, absorption lines ('Fraunhofer lines'). The chromosphere is sandwiched between the cool photosphere and the very hot corona. The corona is the halo around the photosphere, visible at times of total eclipse. It is very hot $\sim 10^6$ K and tenuous and, at these temperatures, is a fully ionized plasma. The corona emits X-rays. The plasma flows outwards along the field lines to produce the solar wind.

The quiet sun

The radiation emitted by a body as a result of its temperature is called thermal radiation. A blackbody is one that has a surface that absorbs all thermal radiation incident on it, so that, if cool enough we cannot see the object in reflected light. The sun is therefore a good example of a blackbody, whereas the earth is less so, since it has an albedo of ~0.3. The spectral distribution of a blackbody is a function only of its surface temperature and independent of composition of the body. The spectral distribution is specified by the quantity $R_T(\nu)$ called the spectral radiancy which is defined so that $R_T(\nu)d\nu$ is equal to the energy emitted per unit time in the radiation of frequency in the interval ν to $\nu + d\nu$ from a unit area of the surface at absolute temperature T. This quantity is plotted for different surface temperatures 1000 K, 1500 K and 2000 K in Figure 1.15. Planck's formula describes these curves

$$R_T(\nu)\,d\nu = \frac{2\pi h \nu^3}{c^2}\frac{1}{e^{\frac{h\nu}{kT}} - 1}\,d\nu, \tag{1.43}$$

where k is the Boltzmann constant. It is seen by inspection of Figure 1.15 that the peak of the spectrum shifts toward higher frequencies as temperature increases. This is expressed by the Wien's displacement law

$$\nu_{max} \propto T, \tag{1.44}$$

where ν_{max} is the frequency at which $R_T(\nu)$ has a maximum value for a particular T.

The integral of the spectral radiancy $R_T(\nu)$ over all ν is the total energy emitted per unit time per unit area from a blackbody at temperature T. The radiancy, R_T, is therefore

$$R_T = \int_0^\infty R_T(\nu)\,d\nu. \tag{1.45}$$

In Figure 1.15 we see that R_T increases rapidly with increasing temperature, as expressed in Stefan's Law

$$R_T = \sigma\, T^4, \tag{1.46}$$

where σ is the Stefan-Boltzmann constant.

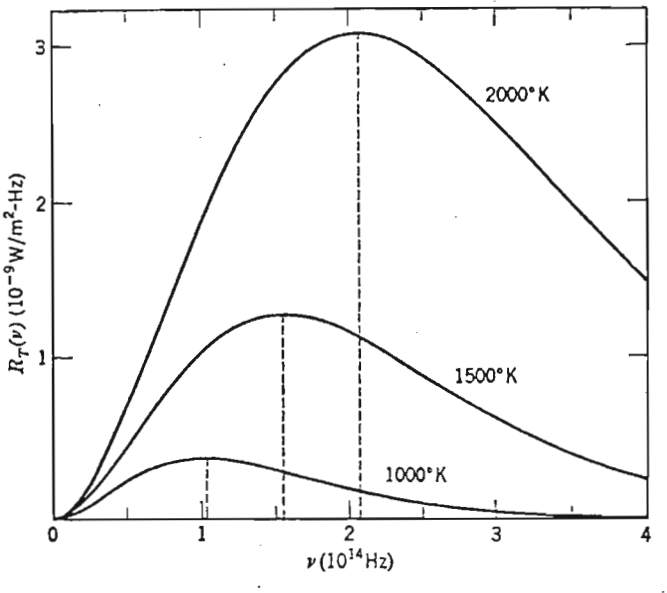

Figure 1.15: The spectral radiancy of a blackbody at 2000 K, 1500 K and 1000 K plotted against frequency. From *Eisberg and Resnick (1985).*

The solar constant, which is defined as the total radiative energy flux outside the earth's atmosphere i.e. at 1 AU is 1367 W m^{-2}. This would be R_T scaled by $\frac{1}{[1AU]^2}$ for flux at earth's surface.

Besides electromagnetic radiation the quiet sun also emits streams of particles, called the solar wind, discussed in more detail in chapter 2.

The active sun: solar flares

Solar flares produce increases in radiation intensity over a broad range of wavelengths on time scales of minutes to hours. Optically they are observed in the H_α (n=3 $\rightarrow$ n=2) line, but they show up in shorter (X-rays) and longer (radio) wavelengths. They are classified in level of importance both visually and in X-rays, *Solar Geophysical Data (1987)*. The optical importance has a five level scale (1-5) for increasing area of plague followed by a letter F, N, B standing for faint, normal, bright for brilliance. X-rays are quoted in W m^{-2} for peak flux in the wavelength interval 1-8 Å. A digit multiplier succeeds letters C (corresponding to 10^{-5}) or M (corresponding to 10^{-4}) or X (corresponding to $> 10^{-4}$). The large solar flare on 9 March 1989 was classified as 4B/X4.0 (*Solar Geophysical Data, 1989*).

The basic mechanism of solar flares is thought to be the conversion of magnetic energy to particle energy. This magnetic energy comes from 'annihilation' of the magnetic field over solar active regions. The energy transferred to the particles is then manifested as high energy protons and electrons, Bremstrahlung X-rays, Cerenkev radiation, Synchrotron radiation at radio wavelengths, etc. The radio emissions are monitored at 10.7 cm (28 GHz) and their intensities are closely correlated with the solar cycle.

Solar flares have perhaps their most important manifestation in proton fluxes, the flux can increase more than 100% over the galactic cosmic ray background level. Energies of cosmic ray (including solar) particles are commonly specified in terms of rigidity (R) or momentum per unit charge.

$$R = \frac{p\,c}{z\,e} \tag{1.47}$$

where z is the nuclear charge. Rigidity is measured in GV.

All particles with the same rigidity follow the same trajectory in a magnetic field. For observation on earth, assuming a dipole field, particles can only reach dipole latitude ϕ_c if $R > R_c$ the cut-off rigidity where $R_c = 14.9\ \cos^4(\phi_c)$ (*Hargreaves, 1979*).

A solar proton event by definition is said to have occurred when the flux of protons of energy > 10 MeV exceeds 10 particles $cm^{-2}\ s^{-1}\ ster^{-1}$.

Energetic solar protons produce secondary electrons which, in turn, lead to the production of odd nitrogen species in the middle atmosphere. Nitric oxide (NO) catalytically destroys ozone molecules. This forms the basis of work described in chapter 4. Theoretical aspects of nitric oxide formation are described fully in chapter 2.

1.4.3 Attenuation of solar radiation in the earth's atmosphere

As solar photons penetrate into the earth's atmosphere they undergo collisions with atmospheric molecules and are progressively absorbed and scattered.

Molecular absorption

The probability of absorption by a molecule depends on the nature of the molecule and the wavelength of the incoming photon. An effective absorption cross-section $\sigma_a(\lambda)$ can be defined for each photochemical species. This quantity is expressed in cm^2 and is independent of the concentration of the species under consideration.

The Beer-Lambert Law describes the absorption (dI) of a ray of incident intensity $I_0(\lambda)$ and wavelength λ passing through an infinitesimally thin layer ds. The variation in intensity is given by

$$dI = -k_a(\lambda)\, I_0(\lambda)\, ds, \tag{1.48}$$

where k_a is the absorption coefficient (units of cm^{-1}). This coefficient is proportional to the concentration (cm^{-3}) of the absorbing particles (n) and is related to σ_a by the expression

$$k_a(\lambda) = \sigma_a(\lambda)\, n. \tag{1.49}$$

Integrating equation (1.48) and substituting equation (1.49)

$$I(\lambda) = I_0(\lambda)\, e^{-\left(\int \sigma_a(\lambda)\, n(s)\, ds\right)}. \tag{1.50}$$

The optical thickness over length s is defined as

$$\tau_a = \int_s \sigma_a(\lambda)\, n(s)\, ds. \tag{1.51}$$

Solar radiation penetrates the atmosphere at an angle of incidence (the solar zenith angle χ) which depends on local time, season and latitude. Neglecting the curvature of the earth, we can write

$$ds = \sec(\chi)\, dz. \tag{1.52}$$

Using equation (1.42)

$$n(z) = n_0 e^{-(\frac{z}{H})} \tag{1.53}$$

we can now express the variation of monochromatic solar radiation with altitude (assuming only one absorbing gas)

$$I(z) = I_\infty e^{\left(-\sec(\chi)\int_z^\infty \sigma_a n_0 e^{-(\frac{z}{H}} dz'\right)}$$

$$I(z) = I_\infty e^{\left(-\sec(\chi)\sigma_a n_0 H e^{-\frac{z}{H}}\right)}, \tag{1.54}$$

where I_∞ represents solar irradiation at the top of the atmosphere.

Scattering by molecules and aerosol particles

Scattering is a physical process by which a particle in the path of an electromagnetic wave abstracts (not true absorption) energy from this incident wave and re-radiates the energy in all directions.

The atmospheric particles responsible for scattering include gas molecules, dust, raindrops and hail. The relative intensity of the scattered light depends strongly on the ratio of particle size to wavelength of the incoming radiation. When this ratio is small (for small particles) the light is scattered isotropically. This is called Rayleigh scattering which is $\propto \lambda^{-4}$. When particles are larger e.g. aerosols, an increasing portion is preferentially scattered in the forward direction. This is called Mie scattering. This form of scattering is relatively rare in the middle atmosphere compared with Rayleigh scattering.

1.4.4 Radiative Transfer

General laws and equations which govern the interaction of radiation with the atmosphere are developed in this section. We need to define a radiation field for each point in space and for each wavelength which plays a part in this process.

The radiance L represents the amount of energy (dE) traversing a unit of surface (dS) per unit time (dt) in a cone of solid angle $d\omega$ whose axis is at an angle θ relative to the normal to the surface i.e.

$$L = \frac{dE}{dt\, dS\, d\omega \cos(\theta)}. \tag{1.55}$$

L is measured in W m^{-2} ster^{-1}. L is integrated over all frequencies i.e.

$$L = \int L_\nu \, d\nu. \tag{1.56}$$

In order to evaluate the rate of radiative heating due to absorption of energy per unit volume at any point in space located by vector **r**. the net flux density traversing the surface must be determined. The flux density (F) across the surface at angle ω due to all the cone of direction ω' is

$$F(\mathbf{r},\omega) = \int_{4\pi} L(\mathbf{r},\omega')\cos(\omega,\omega')\,d\omega'. \tag{1.57}$$

The atmosphere can be assumed to be divided into horizontal layers,

$$F(\mathrm{z}) = \int_0^{2\pi} d\phi \int_{-1}^{1} \mu\, L(\mu,\phi)\,d\mu, \tag{1.58}$$

where $\mu = \cos(\omega,\omega')$ and ϕ is azimuthal direction. In the case of a plane parallel atmosphere we may separate the flux into upward propagating flux, and downward propagating flux viz.

$$F\uparrow(\mathrm{z}) = \int_0^{2\pi} d\phi \int_0^{1} \mu\, L(\mathrm{z},\mu,\phi)\,d\mu \quad \mu > 0 \tag{1.59}$$

and

$$F\downarrow(\mathrm{z}) = \int_0^{2\pi} d\phi \int_0^{-1} \mu\, L(\mathrm{z},\mu,\phi)\,d\mu \quad \mu < 0. \tag{1.60}$$

The net flux density is therefore

$$F(\mathrm{z}) = F\downarrow(\mathrm{z}) - F\uparrow(\mathrm{z}). \tag{1.61}$$

To determine heating rate over an altitude dz we use the Divergence Theorem which states that the energy absorbed per unit volume = net flux divergence as $dz \to 0$. Therefore,

$$\frac{d\mathrm{T}}{dt} = -\frac{1}{\rho \mathrm{C}_p}\frac{dF}{dz}. \tag{1.62}$$

Assuming the radiation to be isotropic, the amount of energy entering unit volume at any point in space is called the irradiance

$$\Phi(\mathbf{r}) = \int_{4\pi} L(\mathbf{r},\omega)\,d\omega. \tag{1.63}$$

For a plane stratified atmosphere we can express the irradiance as

$$\Phi(\mathrm{z}) = \int_0^{2\pi} d\acute{o} \int_{-1}^{1} L\,d\mu. \tag{1.64}$$

The equation of radiative transfer is an expression of the energy balance in each unit volume of the atmosphere, including absorption, scattering and emission. In the general

case of a horizontally stratified medium, the following expression can be used to describe radiative transfer in a layer bounded by two infinite parallel planes *(Lenoble, 1977)*

$$\mu \frac{dL_\nu}{dz} = -k_\nu (L_\nu - J_\nu), \tag{1.65}$$

where k_ν is the extinction coefficient and J_ν is the source function. J_ν expresses the incoming radiation due to scattering from all other directions, solar radiation, or thermal emission. J_ν can be neglected in the case of molecular absorption. The source function depends on the spectral domain under consideration. If we let $d\tau_\nu = -k_\nu\, dz$ then

$$\mu \frac{dL_\nu}{d\tau_\nu} = L_\nu - J_\nu. \tag{1.66}$$

Radiative transfer: $\lambda < 4\ \mu$m

Wavelengths shorter than 4 μm may be assumed to be of solar origin. When direct solar flux penetrates into the atmosphere, it is progressively attenuated by absorption and scattering. In order to estimate the total flux density entering a parcel of air (Φ in sub-section 1.4.4) we have to consider both the direct beam flux and diffuse flux (reflected, scattered radiation). The total flux density is also called the actinic flux density. The combination of the effects of multiple scattering (mainly Rayleigh in the denser, lower atmosphere) and albedo produces a substantial increase in the availability of solar radiation for photochemical processes.

Figure 1.16 (a) shows the results of a computation of actinic flux density (by *Meier et al., 1982*) for overhead sun ($\chi = 0^o$) which ignores the effects of multiple scattering and albedo on the solar photon flux density and considers only the direct attenuation effects in terms of optical depth. The stratospheric enhancement factor $\frac{\Phi}{\Phi_\infty}$ represents the factor by which the solar photon flux density at the top of the atmosphere must be multiplied in order to obtain the flux density at a given altitude for the indicated wavelengths. Figure 1.16 (b) shows the same factor when there is an albedo of 0.5 and the effects of multiple scattering in air are incorporated. In this figure the source function, J_ν, incorporates both diffuse scattering and scattering from direct sunlight so that the product of the value of $\frac{\Phi}{\Phi_\infty}$ and the extraterrestrial solar photon flux density is the actinic flux density for the altitude and wavelength concerned. Comparison of Figures 1.16 (a) and (b) emphasizes the importance of albedo and scattering (particularly Rayleigh) on flux densities of a parcel of air. It should also be noted that wavelengths > 360 nm (J-N in Figures 1.16(a) and (b)) are particularly sensitive to scattering and albedo, such that their enhancement factors > 2.

The flux density change with solar zenith angle is slight for $\chi \leq 30^o$ but is quite large when the sun is near the horizon. During early morning and late afternoon

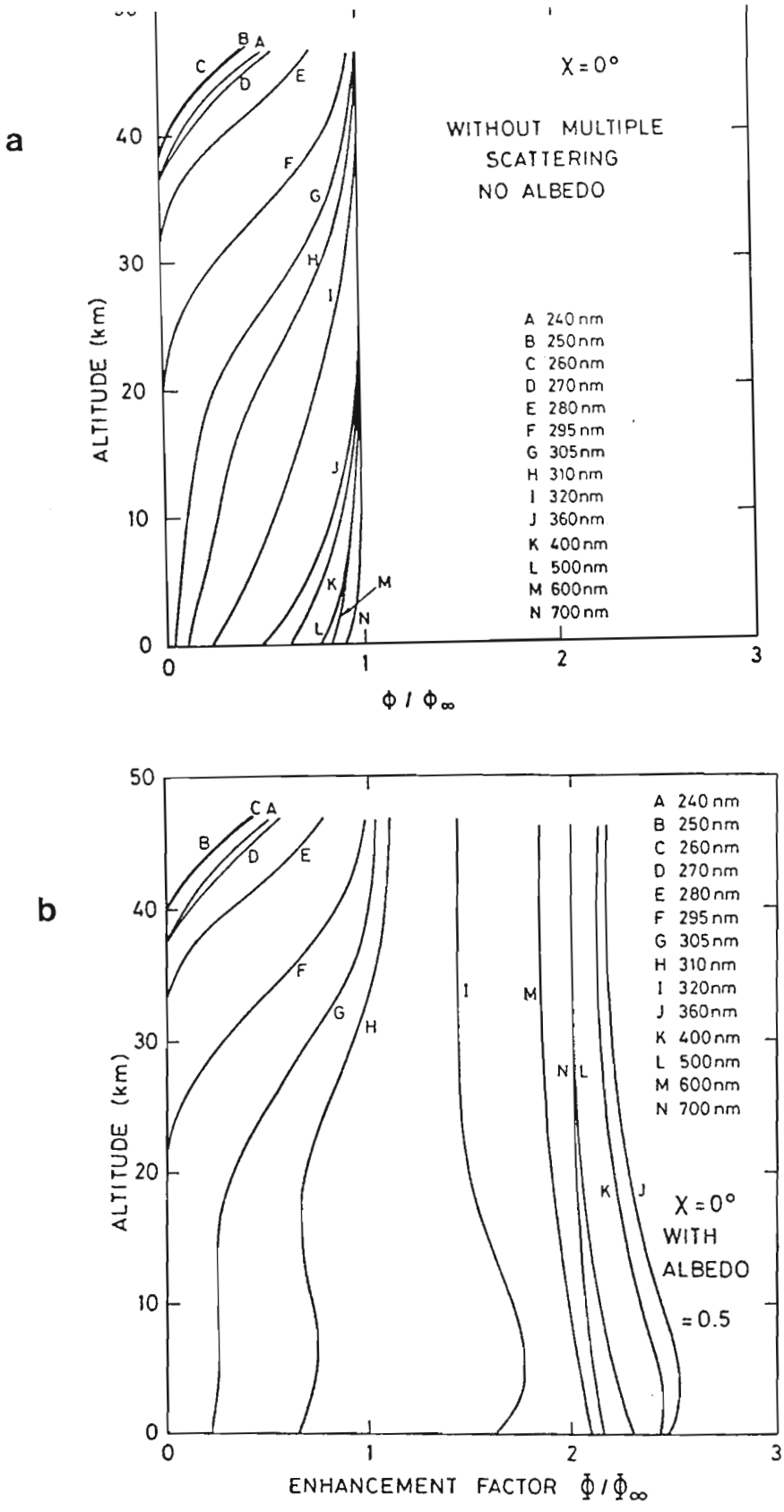

Figure 1.16: (a) Stratospheric enhancement factor as a function of altitude for solar zenith angle of 0°. Only absorption is considered. (b) Same as (a). However, all three processes of absorption, multiple scattering and albedo are included. From *Brasseur and Solomon (1986).*

therefore, the actinic flux density available for inducing photochemical reactions varies very rapidly as compared to near midday.

Radiative transfer: $\lambda > 4$ μm

Radiances at wavelengths > 4 μm are largely of terrestrial origin and clearly dominate solar values (marked 'terrestrial radiation' in Figure 1.13). Thermal emission and absorption by atmospheric gases and aerosols is effective in this spectral region but, scattering can be neglected. At these longer wavelengths the source function, J_ν, represents incoming radiation from thermal emission by atmospheric molecules.

In regions of the atmosphere where local thermodynamic equilibrium (LTE) applies the energy levels are populated according to a Boltzmann distribution. Thus, local emission corresponds to a blackbody radiance (i.e. $J_\nu = B_\nu$) at the temperature of the parcel of air under consideration. The equation of radiative transfer now becomes

$$\mu \frac{\partial L_\nu}{\partial \tau_\nu} = L_\nu - B_\nu. \tag{1.67}$$

The radiance over most of the middle atmosphere can be derived from the transfer equation above. This particular form of the radiative transfer equation is often referred to as the Schwarzchild's equation.

Assuming azimuthal symmetry, the upward and downward components $L_\nu \uparrow$ and $L_\nu \downarrow$ may be evaluated.

The radiatively active trace gases which are the most important from a thermal point of view are: CO_2, O_3 and H_2O vapour. Other gases, whose sources are significantly related to anthropogenic activity such as CH_4, N_2 and chlorofluorocarbons can, in part, contribute to the radiation budget. The characteristic absorption of some gases in the atmosphere is shown in Figure 1.17.

CO_2 is a linear molecule which has a relatively simple absorption spectrum (fifth panel in Figure 1.17). The strongest band is the ν_2 (bending) fundamental at 15 μm, which contributes significantly to the energy budget of the atmosphere because it is located in the spectral region where the emission of the terrestrial environment is very intense (see CO_2 absorption lines in Figure 1.13). The ν_3 (asymmetric stretch) band at 4.3 μm absorbs strongly but has marginal influence since it is located at a wavelength where both solar and terrestrial emissions are weak.

Ozone is a non-linear molecule whose spectrum (fourth panel in Figure 1.17) exhibits a strong rotational structure as well as 3 fundamental vibrational bands ν_1, ν_2 and ν_3 at 9.066, 14.27 and 9.597 μm respectively. The ν_2 band is masked by a CO_2 band but

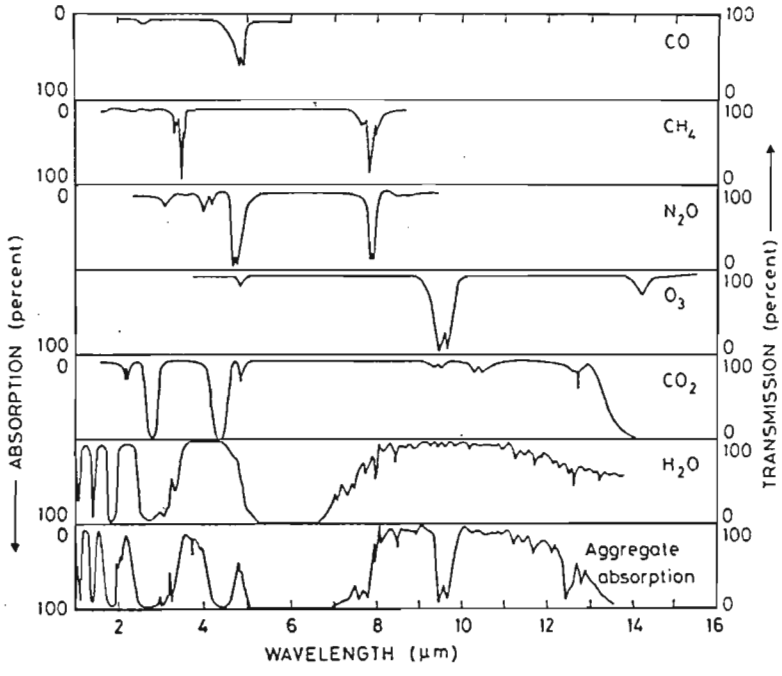

Figure 1.17: Spectral distribution of the absorption (left axis) and transmission (right axis) of several radiatively active gases plotted from the earth's surface to infinity. The aggregate spectrum, due to all absorbers, is in the bottom panel. From *Brasseur and Solomon (1986)*.

the ν_1 and ν_3 both absorb and combine to form the important 9.6 μm band. Another strongly absorbing band at 4.7 μm is located in a region with weak solar and terrestrial intensity.

Water vapour, like ozone, is a non-linear molecule with a complex vibrational-rotational spectrum (sixth panel in Figure 1.17). The ν_2 fundamental band centred at 6.25 μm is overlaid by a series of rotational transitions to make a broad, intense band centred at 6.3 μm. A wide pure rotation band extends from 18 to beyond 100 μm of varying intensity. The ν_1 and ν_3 fundamentals at 2.74 and 2.66 μm absorb significant amounts of solar energy in the troposphere where H_2O vapour is abundant.

1.4.5 Photochemical effects of radiation

Absorption of ultraviolet and visible photons by atmospheric molecules can induce transitions into electronically excited states. Some of the higher energy states are unstable and result in photodissociation of the molecule.

A quantitative value of the photodissociation rate of a molecule A may be derived using the Rate Law

$$\frac{d[\mathrm{A}]}{d\mathrm{t}} = -J(\mathrm{A})\,[\mathrm{A}], \tag{1.68}$$

where $J(\mathrm{A})$ is the photodissociation coefficient with units s^{-1}. The inverse of J represents the lifetime of the molecule against photolysis. The photodissociation coefficient is determined experimentally. For a wavelength interval $\Delta\lambda$ the contributions of photon flux, quantum efficiency (usually unity) and absorption cross-section (σ_a) must be evaluated. The latter two are temperature dependent.

Molecular oxygen (O_2)

The photodissociation of molecular oxygen produces atoms which are crucial to formation of many species, especially O_3. A potential energy diagram for the principal electronic states of the O_2 molecule is presented in Figure 1.18. The transition $X^3\Sigma_g^- \rightarrow A^3\Sigma_u^+$ requires $185 < \lambda < 242$ nm and this wavelength interval constitutes the Hertzberg system. This is an unfavourable transition and therefore has a correspondingly small absorption cross-section. The excited oxygen molecule dissociates, resulting in two ground state (^{3}P) oxygen atoms. The $X^3\Sigma_g^- \rightarrow B^3\Sigma_u^-$ transition constitutes the Schumann-Runge System. This wavelength interval is characterized by a banded structure from 175-200 nm and a continuum from 137-175 nm. The continuum wavelengths result in the formation of one O(^{3}P) atom and one O(^{1}D). Wavelengths

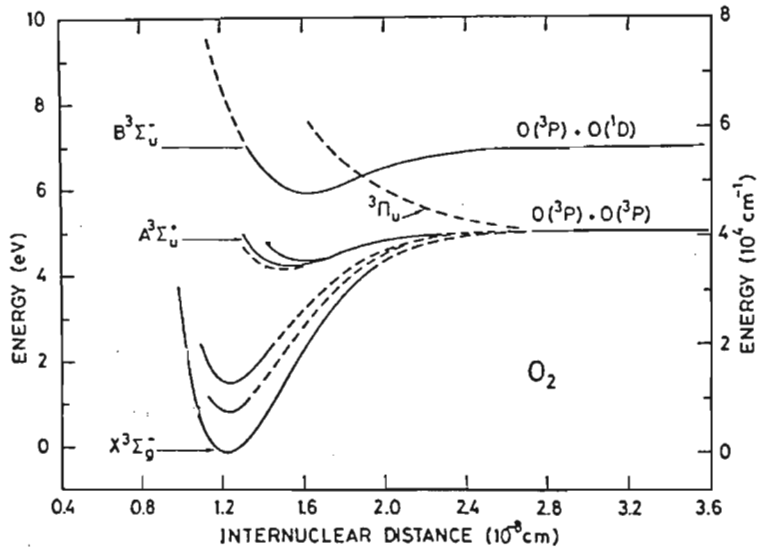

Figure 1.18: Potential energy diagram illustrating the principal energetic states of molecular oxygen. Adapted from *Brasseur and Solomon (1986)*.

in the Schumann-Runge Band result in the formation of two $O(^3P)$ atoms, from a predissociation state ($^3\Pi_u^-$). For $\lambda < 137$ nm some diffuse bands occur, interspersed with windows which allow some wavelengths (including Lyman α) to penetrate deep into the atmosphere.

Despite the weakness of the Hertzberg transition it is dominant in the stratosphere with a photodissociation coefficient of 10^{-9} s^{-1} for zero optical depth. The rate of ozone production in the stratosphere depends critically on the value of absorption cross-sections in this region. Schumann-Runge band are also effective at stratospheric altitudes although the absorption cross-section varies ~ 5 orders of magnitude between 175 and 205 nm, increasing at shorter wavelengths.

Since the intensity of these wavelengths varies with solar activity (section 2.8), the photodissociation coefficient, J, changes accordingly.

Ozone (O_3)

The primary absorption by ozone occurs in the Hartley band, which is located in the wavelength interval 200 to 310 nm. The absorption cross-section maximizes at about 250-260 nm where $\sigma_a \approx 10^{-17}$ cm^2, depicted in Figure 1.19. These cross-sections show only a weak dependence on temperature.

Around 300 nm, the Hartley band becomes weak, and from 310 to 350 nm it blends with the temperature dependent Huggins bands. Figure 1.20 gives absorption cross-section values determined by *Simons et al. (1973)* from 310 to 340 nm and demonstrates the importance of temperature sensitivity, especially at long wavelengths. Discrepancies in absorption coefficients due to temperature dependence are a source of uncertainty when measuring total column ozone by Dobson spectrophotometers.

Ozone's Chappuis bands are where absorption takes place in the visible region. The absorption coefficients of this band are shown in Figure 1.21. This spectral interval (400 to 850 nm) contributes significantly to the photodissociation of ozone and plays a dominant role in the lower stratosphere and troposphere.

Most radiation is absorbed by O_2 at $\lambda < 200$ nm in the Hertzberg continuum at higher altitudes, although Hartley bands do contribute at lower altitudes and are superimposed on the continuum.

Photodissociation of ozone is energetically possible for all wavelengths less than 1.14 μm (near infrared). The products O and O_2 can be found in different electronic states depending on the energy of the incident photon. Listed from lowest energy, atomic

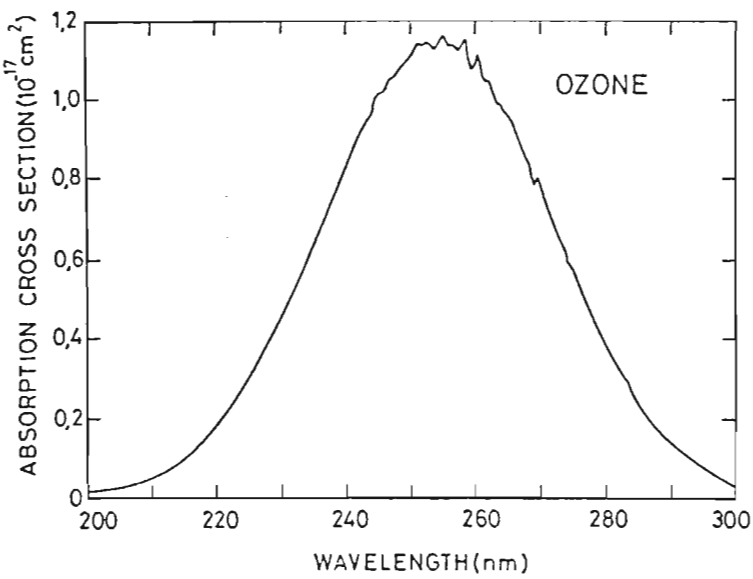

Figure 1.19: The Hartley absorption band of ozone. Absorption cross-sections maximize at ~250 nm. From *Brasseur and Solomon (1986)*.

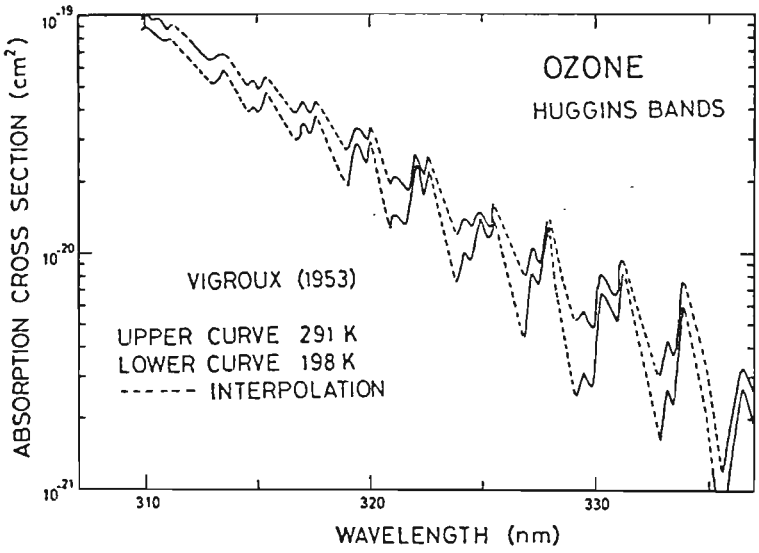

Figure 1.20: The highly temperature dependent Huggins band of ozone, illustrated by absorption cross-section from 310 to 370 nm. From *Brasseur and Solomon (1986).*

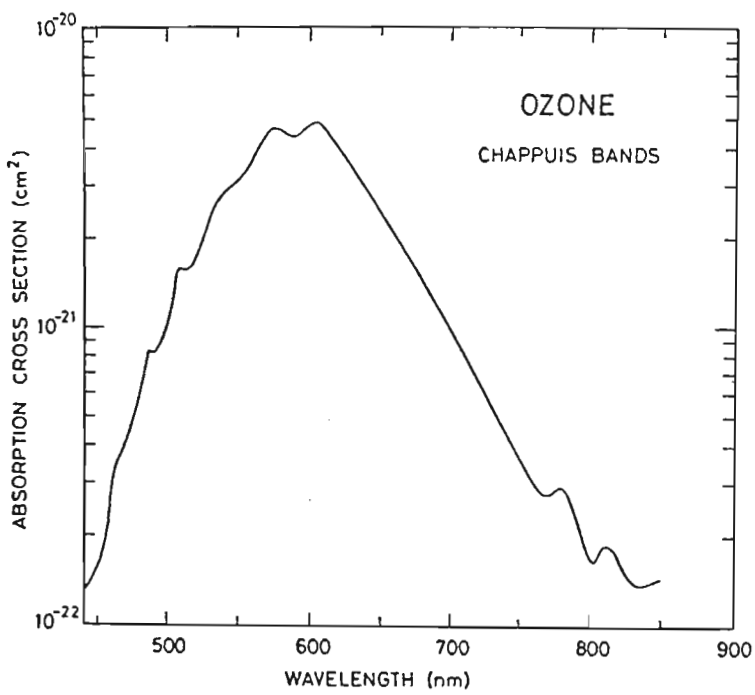

Figure 1.21: The Chappuis absorption bands of ozone. Absorption cross-sections from 450-850 nm are plotted. From *Brasseur and Solomon (1986)*.

oxygen can be $O(^3P)$, $O(^1D)$ and $O(^1S)$. Common states for molecular oxygen are $O_2(^3\Sigma_g)$, $O_2(^1\Delta_g)$, $O_2(^1\Sigma_g^+)$, $O_2(^3\Sigma_u^+)$, $O_2(^3\Sigma_u^-)$. Some transitions are less probable than others and the most common products are

$$O_3 + h\nu \rightarrow O(^3P) + O_2(^3\Sigma_g^-) \quad J_{O_3} \quad \lambda > 310\,\text{nm} \tag{1.69}$$

$$O_3 + h\nu \rightarrow O(^1D) + O_2(^1\Delta_g) \quad J_{O_3}' \quad \lambda < 310\,\text{nm} \tag{1.70}$$

$O(^1S)$ can be formed for $\lambda < 196$ nm.

Figure 1.22 illustrates values of photodissociation coefficients (J) for different spectral bands. Chappuis and Huggins bands (marked 1 and 2) are clearly dominant at low altitudes, whereas wavelengths in the Hartley band (marked 3) region are absorbed at higher altitudes.

1.4.6 Thermal effects of radiation

Heating due to absorption of radiation

The absorption of ultraviolet radiation by ozone in the Huggins and Hartley bands constitutes the principal source of heat in the stratosphere and mesosphere, as depicted by the solid curve in the right panel of Figure 1.23. The heating rate reaches as much as 12 K/day near the stratopause (45 km), with a maximum of approximately 18 K/day near the summer pole (not shown in Figure 1.23). The effect of the Huggins bands in the visible region becomes important in the lower stratosphere, resulting in heating of $\sim$ 1 K/day. These heating rates are obviously related to the amount of ozone present, and an increase (decrease) in ozone concentration would lead to an increase (decrease) in stratospheric and mesospheric temperatures. This could in turn result in possible changes in locations of the stratopause and mesopause.

Above 75 km absorption by the SRC of O_2, dashed curve in the right panel of Figure 1.23, contributes to the major portion of heating of the atmosphere. This absorption results in a mean heating rate of 10 K/day, with large variations over latitude and season. Maximum heating may be as high as 100 K/day. At high altitudes, $\tau_O >$ 1 day, and since its concentration is dependent on transport, thermal energy may be released at a different location. This process, along with adiabatic heating as a result of downward velocities, leads to the warm mesopause observed in polar winters.

The meridional distribution of heating rate due to O_2 and O_3 can be determined by

$$\frac{dT}{dt} = \frac{\cos(\chi)}{\rho\, C_p} \int_\nu \frac{dF}{dz}\, d\nu, \tag{1.71}$$

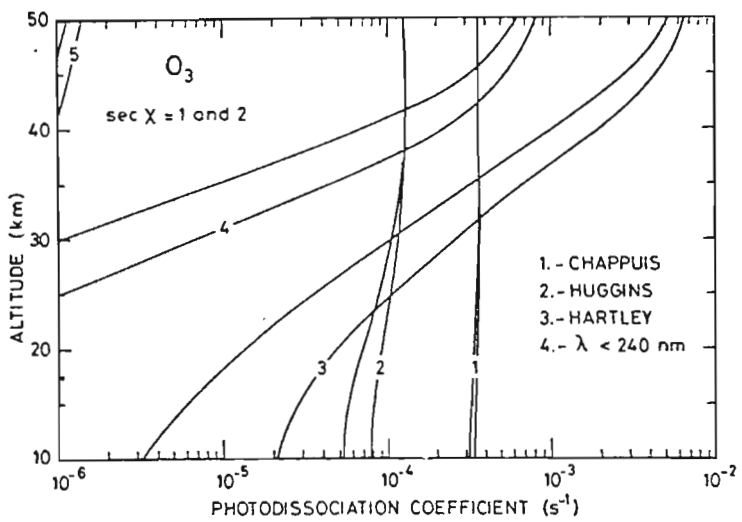

Figure 1.22: The photodissociation coefficient of ozone between 50 and 10 km altitude. Values are plotted for each spectral absorption band of ozone for solar zenith angles of 0° and 60°. From *Brasseur and Solomon (1986)*.

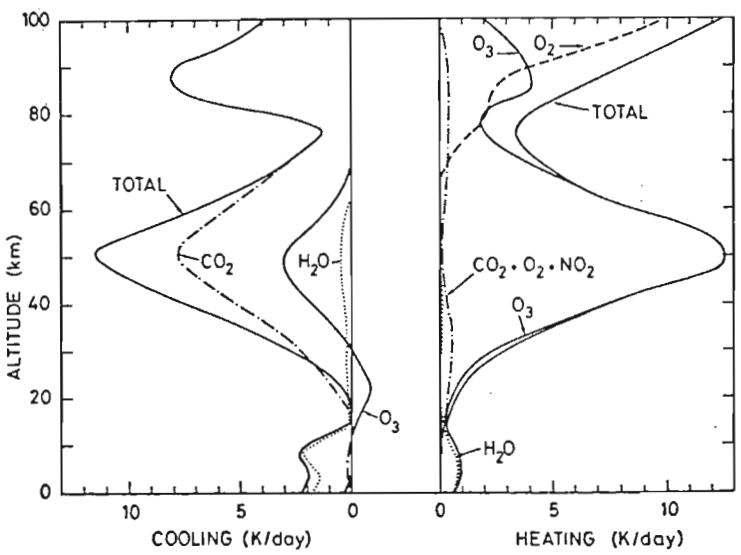

Figure 1.23: Vertical profile of heating rates by O_3, O_2, NO_2, H_2O and CO_2 (right side panel) and cooling rates by CO_2, O_3 and H_2O (left side panel). Note that heating occurs over shorter wavelengths due to absorption of solar radiation, while cooling is achieved by thermal radiative emission by the atmosphere at infrared wavelengths. From *Brasseur and Solomon (1986)*.

where F is the incident solar flux. The heating rate may be calculated from the attenuated solar flux by the expression

$$\frac{d\mathrm{T}}{dt} = \frac{1}{\rho \mathrm{C}_p} \left([\mathrm{O}_3] \int_\nu \sigma(\mathrm{O}_3)\, F\, d\nu + [\mathrm{O}_2] \int_\nu \sigma(\mathrm{O}_2)\, F\, d\nu \right). \tag{1.72}$$

Cooling by radiative emission

The cooling produced by radiative emission in the stratosphere and lower mesosphere is principally due to 15 μm band of CO_2. This is represented by the dot-dash-dot curve in the left panel of Figure 1.23. This band results in a cooling rate of up to 7 K/day around the stratopause. The maximum cooling is found at the winter mesopause, where the temperature is relatively warm.

A second contribution to infrared cooling in the middle atmosphere results from infrared emission by O_3 at 9.6 μm, corresponding to the solid line in Figure 1.23. This contribution is important at a layer near the stratopause, and extends over 10 km of altitude. In the lower stratosphere, the radiative transfer at 9.6 μm leads to a net heating as ozone absorbs some of the terrestrial emission at this frequency at altitudes below the ozone layer.

The effect of water vapour (dotted line in the left panel of Figure 1.23) is manifested in the middle atmosphere at 80 μm. This contribution is relatively weak. In the troposphere where the temperatures are warmer than in the stratosphere the 6.3 μm band of H_2O determines the net cooling rate.

The cooling rate ($\frac{dT}{dt}$) due to infrared emission may be assumed to be proportional to $\frac{dF\uparrow}{dz}$ if the surface emission is neglected (the 'cool to space' approximation) i.e.

$$\frac{d\mathrm{T}}{dt} = -C_1\, B_\nu(\mathrm{T}), \tag{1.73}$$

where C_1 is a constant determined for each radiatively active gas. If the temperature varies little about a reference value T_0 the above equation can be linearized such that

$$\frac{d\mathrm{T}}{dt} = (\frac{d\mathrm{T}}{dt})_{\mathrm{T}_0} + \alpha(\mathrm{T} - \mathrm{T}_0). \tag{1.74}$$

The coefficient α defines the time constant for thermal radiation i.e $\tau_{rad} = \frac{1}{\alpha}$. This time is $\sim$ 15-20 days in the troposphere and 3-5 days in the stratosphere. Transport of heat can only occur if τ_{dynam} is less than or comparable with τ_{rad}. In the troposphere where $\tau_{dynam} < \tau_{rad}$ the temperature field is dependent on transport. However, in the middle atmosphere where $\tau_{rad} \ll \tau_{dynam}$ the thermal structure results from an equilibrium between absorption of ultraviolet radiation and emission in the infrared. Since $\tau_{rad} \sim$ days only small diurnal variations are detected.

1.5 Ozone dynamics and chemistry

1.5.1 The odd oxygen family

In this section all important atmospheric oxygen reactions are listed. Where electronic states of particles are unspecified they may be assumed to be in their ground state i.e. O_2 and O represent $O_2(^3\Sigma_g^-)$ and $O(^3P)$ respectively.

The photodissociation of molecular oxygen by ultraviolet at $\lambda < 242$ nm produces atomic oxygen

$$O_2 + h\nu \rightarrow O + O \quad (J_2). \tag{1.75}$$

These atoms may recombine directly in a three body process

$$O + O + M \rightarrow O_2 + M \quad (k_1), \tag{1.76}$$

where M is usually N_2. Alternatively, atomic oxygen may react with molecular oxygen to produce ozone

$$O + O_2 + M \rightarrow O_3 + M \quad (k_2). \tag{1.77}$$

Ozone may then recombine with atomic oxygen

$$O + O_3 \rightarrow 2\,O_2 \quad (k_3). \tag{1.78}$$

The photodissociation of ozone leads to products in either their ground state

$$O_3 + h\nu(\lambda > 310\,\text{nm}) \rightarrow O_2 + O \quad (J_3), \tag{1.79}$$

or their excited state

$$O_3 + h\nu(\lambda < 310\,\text{nm}) \rightarrow O_2(^1\Delta_g) + O(^1D) \quad (J_3^*). \tag{1.80}$$

In the upper middle atmosphere some $O(^1D)$ atoms are produced by molecular oxygen photolysis

$$O_2 + h\nu(\lambda < 175.9\,\text{nm}) \rightarrow O + O(^1D) \quad (J_2^*). \tag{1.81}$$

Molecular oxygen in an excited state may be quenched by collision with a ground state molecular oxygen viz.

$$O_2(^1\Delta_g) + O_2 \rightarrow 2\,O_2 \quad (k_6), \tag{1.82}$$

or it may relax radiatively by emitting a photon of $\lambda = 1.27$ μm

$$O_2(^1\Delta_g) \rightarrow O_2 + h\nu \quad (A_{1.27}). \tag{1.83}$$

It is possible, by monitoring this emission, to determine the concentrations of $O_2(^1\Delta_g)$ and O_3. $O(^1D)$ may be quenched by collision with either N_2 or O_2 viz.

$$O(^1D) + N_2 \rightarrow O + N_2 \quad (k_{4a}) \tag{1.84}$$

$$O(^1D) + O_2 \rightarrow O + O_2 \quad (k_{4b}). \tag{1.85}$$

The chemical reactions listed above are the basis of a simple theory of stratospheric ozone for a pure oxygen atmosphere, first proposed by Chapman in 1930. Subsequently, substantial modifications have been made to this theory involving hydrogen, nitrogen and chlorine species (details in chapter 2). However, this system provides a framework to which other processes may be added.

Neglecting transport, we can write the following continuity equations for odd oxygen

$$\frac{d[O_3]}{dt} + (J_3 + J_3^*)[O_3] + k_3[O][O_3] = k_2[M][O_2][O] \tag{1.86}$$

$$\begin{aligned} \frac{d[O]}{dt} + 2\,k_1[M][O]^2 + k_2[M][O_2][O] + k_3[O_3][O] &= 2\,J_2[O_2] + J_2^*[O_2] + J_3[O_3] \\ &\quad + k_{4a}[N_2][O(^1D)] \\ &\quad + k_{4b}[O_2][O(^1D)] \end{aligned} \tag{1.87}$$

$$\frac{d[O(^1D)]}{dt} + (k_{4a}[N_2] + k_{4b}[O_2])[O(^1D)] = J_3^*[O_3] + J_2^*[O_2] \tag{1.88}$$

$$\frac{d[O_2(^1\Delta_g)]}{dt} + A_{1.27}[O_2(^1\Delta_g)] + k_6[O_2][O_2(^1\Delta_g)] = J_3^*[O_3]. \tag{1.89}$$

The continuity equation for O_2 is omitted since it is not independent, as the total amount of oxygen is always conserved viz.

$$3\frac{d[O_3]}{dt} + 2\frac{d[O_2]}{dt} + 2\frac{d[O_2(^1\Delta_g)]}{dt} + \frac{d[O]}{dt} + \frac{d[O(^1D)]}{dt} = 0. \tag{1.90}$$

Chemical lifetime equations for the different oxygen species can be easily determined. They are evaluated below, for a stratospheric altitude of 30 km

$$\tau_{O_3} = \frac{1}{J_3 + J_3^* + k_3[O]} \sim 2000 \text{ s} \tag{1.91}$$

$$\tau_O = \frac{1}{k_2[O_2][M] + k_3[O_3] + 2\,k_1[M][O]} \sim 0.04 \text{ s} \tag{1.92}$$

$$\tau_{O(^1D)} = \frac{1}{k_{4a}[N_2] + k_{4b}[O_2]} \sim 10^{-8} \text{ s} \tag{1.93}$$

$$\tau_{O_2(^1\Delta_g)} = \frac{1}{k_6[O_2] + A_{1.27}} \sim 1 \text{ s} \tag{1.94}$$

The chemical loss of $O_2(^1\Delta_g)$ always results in $O_2(^3\Sigma_g^-)$. Since O_2 has an almost constant mixing ratio throughout the middle atmosphere, the loss of $O_2(^1\Delta_g)$ is not photochemically important. On the other hand, O_3, O and $O(^1D)$ are a coupled set of partial differential equations whose time constants differ by several orders of magnitude. To solve these equations numerically a time step of $\ll 1$ s would have to be used, making such calculations prohibitively expensive. In view of this, it is very useful to define chemical families for computational purposes, whose lifetimes can be very much longer than those of the constituent members. We therefore define the odd oxygen family as

$$[O_x] = [O(^3P)] + [O(^1D)] + [O_3], \tag{1.95}$$

so that the chemical lifetime (evaluated at 30 km) for odd oxygen is

$$\tau_{O_x} = \frac{[O_x]\ .}{2\,k_1[M][O]^2 + 2\,k_3[O][O_3] + \frac{\text{other terms}}{\text{in N,H,Cl}}} \sim \text{weeks.} \tag{1.96}$$

Figure 1.24 represents an altitude profile of the photochemical lifetimes of the O_x family (as given by equation (1.96)), O_3 and O, as well as approximate time constants associated with winds and vertical diffusion (τ_{diff}). Zonal (τ_u), meridional (τ_v) and vertical (τ_w) are determined for mid-latitudes using characteristic scale lengths of 1000 km, 100 km, and 5 km respectively. The very fast reactions (equations (1.77), (1.80), (1.85)) produce only an exchange, or partitioning, among members of the family and do not appear as production or loss of $[O_x]$. The use of chemical families allow a clearer distinction to be drawn between reactions that result in *net* and *gross* production and loss. For example, the photolysis of O_3 below 80 km will not result in a net loss of ozone over time scales > seconds since recombination would result in more ozone formation. On the other hand, when ozone reacts with atomic O and forms O_2 there is a net loss of ozone. Further, since $\tau_v \sim$ months (see Figure 1.24) which is $\sim \tau_{O_x}$ it is easy to see how the odd oxygen family is affected by meridional transport. Time constants associated with transport are short or similar to τ_{O_x} near the stratopause so odd oxygen is dynamically controlled in that region (sub-section 1.5.2). From the upper stratosphere into the mesosphere $[O] \sim [O_3]$ and ozone becomes critically dependent on atomic oxygen, so much so that diurnal variations (night-time ozone increase) are significant in this altitude range.

1.5.2 Ozone: Historical aspect

The first quantitative analysis of atmospheric ozone was made by Fabry and Buisson at the University of Marseilles in 1912, but their work was interrupted by the First World War. They resumed in 1919 and their work was subsequently published (*Fabry*

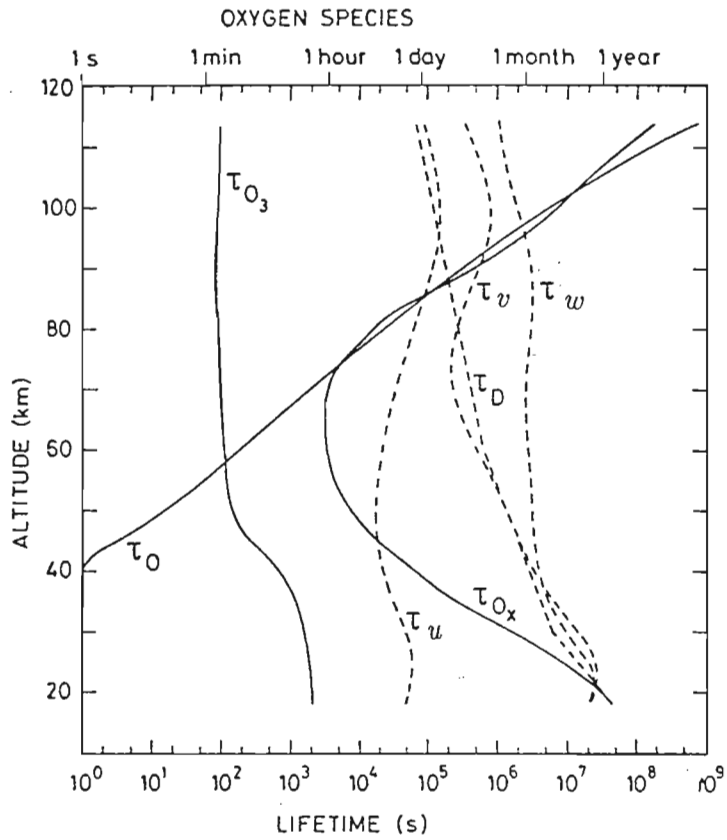

Figure 1.24: Photochemical lifetimes of O_x, O_3 and O together with characteristic transport lifetimes in the zonal (u), meridional (v) and vertical (w) directions. From *Brasseur and Solomon (1986)*.

and Buisson, 1921). Dobson refined their instrument and set up a chain of spectrophotometers at Oxford (England), Valentia (Ireland), Lerwick (Shetland), Abisko (Sweden), Lindenberg (Germany), Arosa (Switzerland) and Montezuma (Chile) in 1926. Ground-based, balloon-borne, rocket and satellite observations (see chapter 3) made since that time have substantially extended the data base for describing the essential features of three-dimensional ozone distribution together with time variations.

It is well known that ozone is characterized by latitudinal and seasonal variations. Figure 1.25 shows total column ozone versus month for a zonally averaged strip in the northern hemisphere. The greatest production of ozone occurs in the tropics (due to greater irradiance of solar ultraviolet wavelengths) and yet, maximum ozone abundance occurs during winter-spring at high latitudes. This maximum in total ozone is due to downward and poleward net transport associated with mean motion in the transformed Eulerian or diabatic framework, as discussed in sub-section 1.3.4. Decrease in summer is attributed to chemical destruction by reactive fragments. Smaller, semi-annual variations (not shown in Figure 1.25) may be attributed to semi-annual zenith crossings of the sun in the equatorial zone.

The global distribution of total ozone versus latitude as derived from a network of ground-based stations at different longitudes is given in Figure 1.26. These values are expressed in Dobson units, which correspond to the height (matm cm) which the ozone column would have if all the gas were at 1 atm and 273.15 K, viz. 300 DU would correspond to 3 mm of ozone. Annual variations for each hemisphere, showing a 180^o shift with respect to each other, are evident in this figure. The maximum ozone observed in the southern hemisphere is much smaller and at lower latitudes. This may be ascribed to differences in stratospheric dynamics of the two hemispheres. The ozone hole, which covers a large portion of the southern polar cap every austral spring is not shown in Figure 1.26. Details of the dynamics and chemistry of polar ozone will be discussed in more detail in chapter 2.

Identification and examination of the altitude profile of ozone (shown in Figure 1.5) was pioneered by Götz and his co-workers with their Umkehr technique. Using this technique, primary changes in ozone content were found to occur between 10 and 20 km, a fact of meteorological significance *(Dobson et al., 1929)*. Dobson subsequently determined that high total ozone abundances were highly correlated with low pressure systems and thus lower tropopause height.

Longer term variations in ozone are also detectable. The QBO (described in sub-section 1.3.4) affects ozone concentrations as the O_x family is dynamically sensitive at stratospheric altitudes. Solar cycle influences are detectable and must be removed in order to perform any trend analysis such as, the determination of global ozone depletion by the burden of chlorofluorocarbons in the atmosphere *(Stolarski et al., 1991)*.

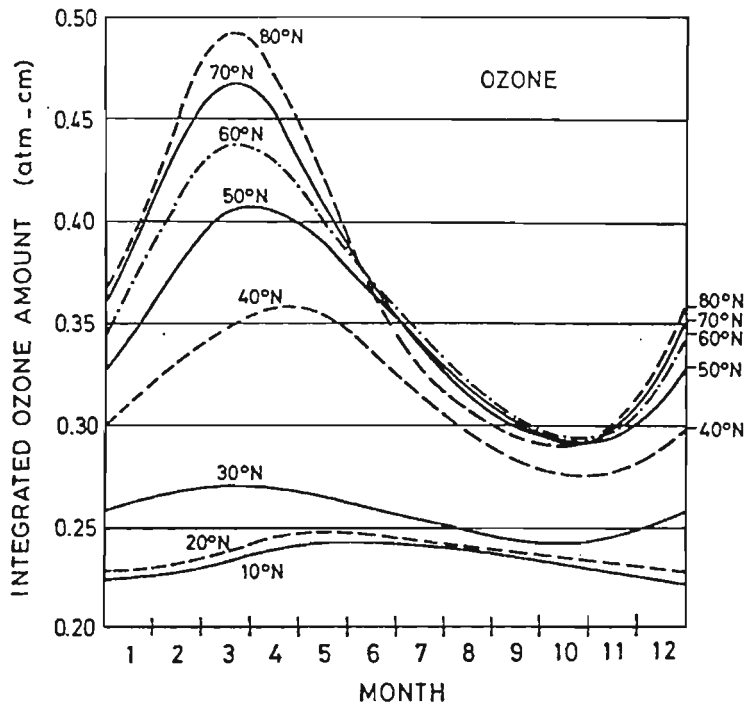

Figure 1.25: The mean annual variation in total ozone at selected northern hemisphere latitudes. From *Brasseur and Solomon (1986)*.

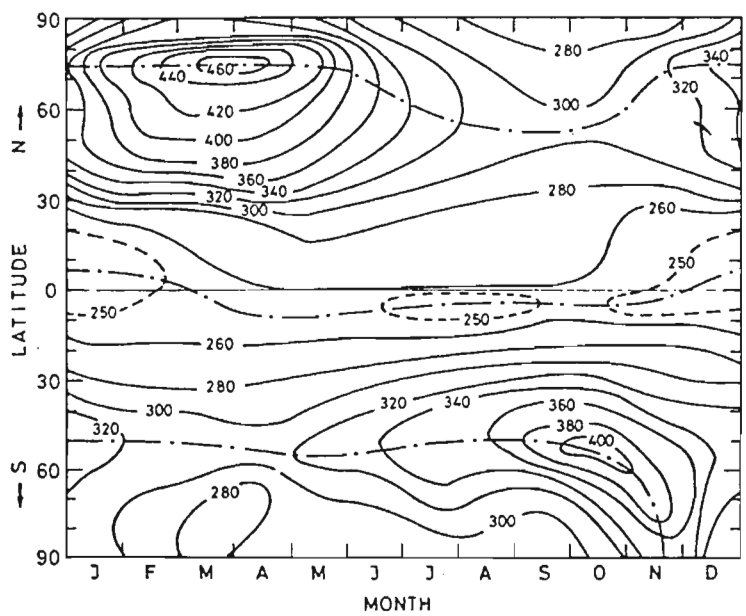

Figure 1.26: Variation of total ozone (DU) over the earth's surface during a typical year. From *Brasseur and Solomon (1986).*

1.6 The scope of this thesis

The major portion of the work in this thesis is concerned with the effects of solar variability on total column ozone in the polar regions. Daily global ozone measurements by the Total Ozone Mapping Spectrometer (TOMS) instrument aboard the Nimbus 7 satellite provide the prime data base. Bursts of solar flare activity result in solar proton events (SPE's) whereby high energy solar protons produce, in the middle atmosphere, enhanced concentrations of odd nitrogen. These reactive species give rise to natural ozone depletion. The odd nitrogen plays a similar role to chlorine, derived from chlorofluorocarbons, in anthropogenic ozone depletion.

In chapter 1 important features of the physics and chemistry of the atmosphere have been presented, those of prime importance for a sound understanding of ozone depletion. Chapter 2 is an in-depth review of polar ozone chemistry and dynamics. Emphasis is placed on the possibility of solar influences on the levels of ozone at polar latitudes. This will enable the reader to assess the results of the present work, with respect to that of earlier studies. Satellite data acquisition of ozone, solar particle and radio flux data are discussed in chapter 3. A detailed description of the analysis techniques employed by the author appears in chapter 4, together with three case studies involving the quantitative short term (days to months) influences of SPE's on ozone amounts in the latitude region 90° to 70°. Longer term (months to years) variation in Antarctic ozone, together with time lags, are discussed in chapter 5, based upon a Fourier transform analysis of eleven years of total column ozone measurements.

Chapter 2

Aspects of polar ozone distribution

2.1 Introduction

Polar ozone has become a focal point of atmospheric research since *Farman et al. (1985)* reported a remarkable 50% column change of ozone over Halley, Antarctica. Furthermore, Farman and his co-workers proposed that inorganic chlorine may be responsible for the ozone depletion. These measurements stood in stark contrast to model predictions of chlorine induced ozone loss near 40 km at mid-latitudes via the catalytic gas phase reactions

$$Cl + O_3 \rightarrow ClO + O_2 \tag{2.1}$$

$$ClO + O \rightarrow Cl + O_2 \tag{2.2}$$

first proposed by *Molina and Rowland (1974)*. Indeed, it was thought that because the chemical lifetime of the ozone layer over Antarctica during polar night was of the order of many years, that it would show little sensitivity to anthropogenic chlorine increases. The evidence presented in section 2.5 confirms that a new chemical mechanism, requiring ice surfaces, is responsible for the chlorine catalysed destruction of ozone over Antarctica.

Chapter 1 of this thesis describes some of the fundamental processes in the middle atmosphere. Chemical, dynamical and radiative processes are all important in determining ozone abundances at any fixed point in space. These processes can, in general, be decoupled with a knowledge of their respective time constants. This chapter focuses on the influences of certain phenomena on the polar ozone distribution which are important for the work described in this thesis. These may originate in solar variations (annual cycle, solar cycle and solar proton events) or in dynamics (QBO, polar vor-

tex and planetary waves) while others may be attributed to increased anthropogenic emissions of chlorine, bromine, odd nitrogen and carbon dioxide.

2.2 Satellite observations of polar total column ozone

The latitudinal and seasonal variations of global total ozone, measured by the TOMS (Total Ozone Mapping Spectrometer) instrument aboard the Nimbus 7 satellite, are illustrated by the four colour panels in Figure 2.1. The maps represent the ozone distribution for the solstices and equinoxes of 1985. The ozone data have a resolution of 25 DU, black and red colours represent high column amounts, whereas blues and purples indicate the areas of lowest total ozone. TOMS determines ozone amounts by the solar backscatter technique implying that TOMS coverage can only extend over latitudes which are irradiated: less than 65°N on December 21 and 65°S on June 21. Immediately apparent are the low, and nearly constant, levels of ozone in the tropics. The broad scale structure in this region is associated with Hadley circulation. The annual cycle becomes more evident with increasing latitude. High latitude ozone observations are discussed below.

2.2.1 Northern polar cap

The highest global total ozone values are found in winter and spring at high latitudes, the region of greatest temporal and spatial variability. These features are evident in the TOMS maps for March 21 and December 21 in Figure 2.1. The black areas poleward of 50°N on March 21 across Siberia, Alaska and the Arctic ocean have greater than 475 DU of total ozone. These are embedded in wave-like regions with rather steep boundaries on the equatorward side. A complex region of decreased ozone is present over the North Atlantic and Europe on March 21 forming gradients of 150 DU in less than 1000 km. The motion of such features is responsible for the large daily changes observed at mid-latitudes.

The pattern of ozone highs on December 21 is somewhat simpler with minima over western North America and Eurasia, and maxima over the North Atlantic and eastern Siberia. The small width of the maximum over eastern North America is particularly striking.

During the summer months the amplitude of the wave-like ozone distribution diminishes, the gradients decrease, and the higher mid-latitude ozone levels expand into the

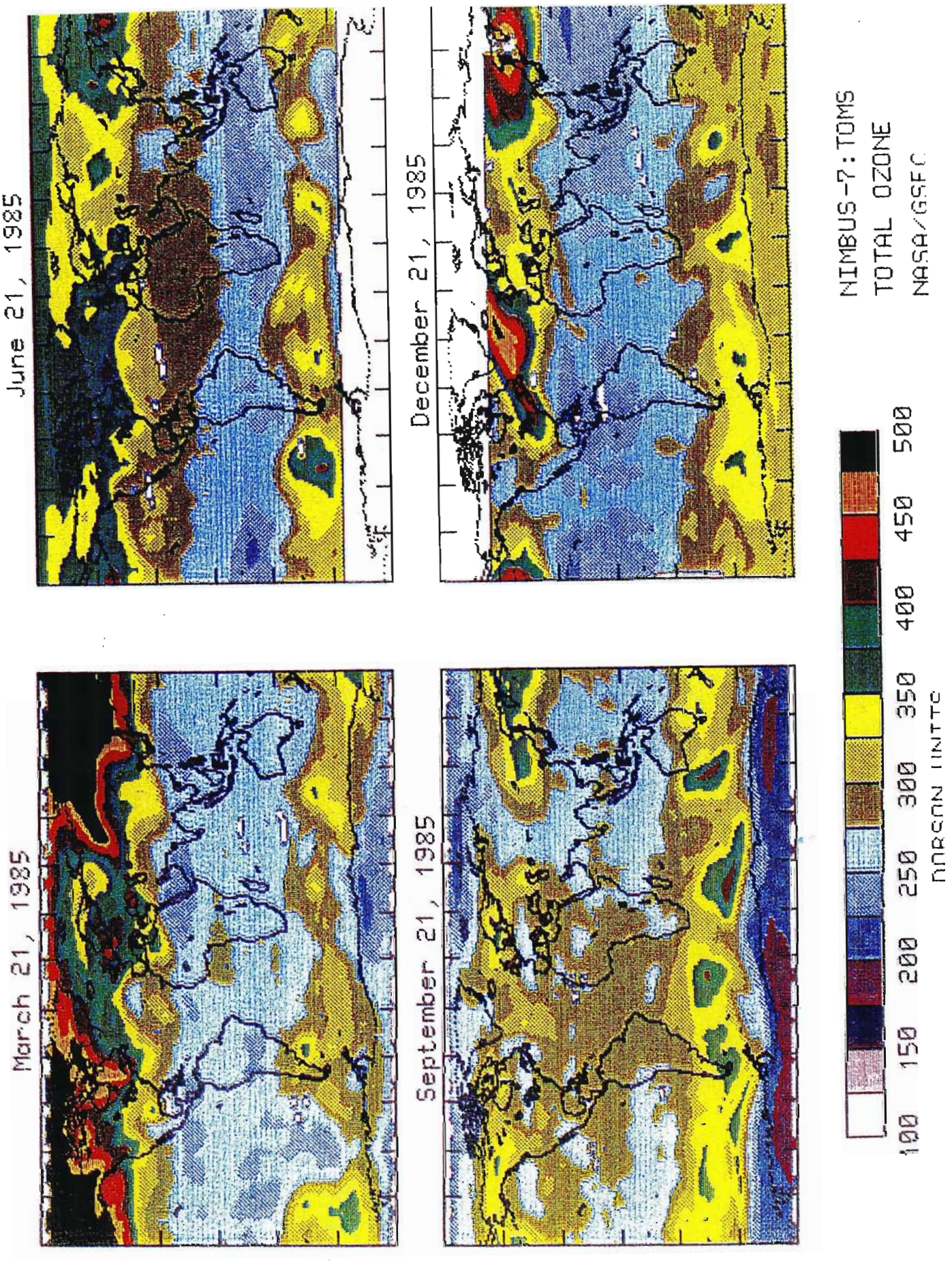

Figure 2.1: Total column ozone (in DU) measured by TOMS on 21 March 1985 (top left), 21 June 1985 (top right), 21 September 1985 (bottom left) and 21 December 1985 (bottom right). Data gaps over southern polar regions in June, and northern polar regions in December are due to lack of solar irradiance. From *Krueger (1989)*.

tropical regions. By the fall, only a thin band of higher ozone exists in the northern hemisphere near 60°N latitude.

2.2.2 Southern polar cap

It is well known that the seasonal cycle of total ozone in the southern polar region is radically different from that in the northern polar region i.e. it is not merely phase shifted by six months. *Dobson (1966)* found, using data from Halley and the South Pole stations, that total ozone decreased throughout the winter, then abruptly increased to its annual maximum in late spring. The total ozone at Resolute, on the other hand, increased through the winter, reaching a maximum in late winter. These differences are clear when inspecting Figure 2.1. By far the greatest difference is obvious during the equinoxes. The northern hemisphere on March 21 contains a near polar maximum, while the southern hemisphere on the corresponding date, September 21, has a polar minimum. This minimum is significantly lower than tropical ozone levels, and is now recognized as the Antarctic ozone hole. This feature is produced by dynamics and has, in recent years, been strongly enhanced by the action of chlorine radicals. The south polar ozone spring minimum appears to be a permanent feature of the atmosphere, as does the northern spring polar maximum. Mid-latitudes have relatively high ozone values on September 21 due to downward Hadley circulation from the tropics that is isolated from polar cap by the polar vortex. Within this mid-latitude ridge are wave-like structures which are, generally, simpler than their northern counterparts in December and March.

2.3 Interpretation of observations using modelling techniques

Models that simulate the distribution of ozone in the middle atmosphere are needed to give an interpretation of satellite observations depicted in Figure 2.1. *Perliski et al. (1989)* have used the two-dimensional photochemical model of *Garcia and Solomon (1983)* to gain insight into the chemical and dynamical factors that are important in determining the seasonal variations of ozone. This model computes the zonally averaged chemical structure of the middle atmosphere by solving the continuity equation for chemical constituents as a function of latitude and time. Chemical interactions of the O_x (O, O_3), HO_x (H, OH, HO_2), NO_x (N, NO, NO_2) and ClO_x (Cl, ClO, $HOCl$) families are included. The zonal temperature structure for each grid point is calculated from the zonal mean thermodynamic equation. Transport is represented by the

streamfunction equation for the residual Eulerian meridional circulation with Rayleigh friction introduced to balance the momentum budget. Figure 2.2(a) shows the amplitudes of the modelled annual ozone variation, and (b) the observed. The observed annual variation is computed from nine years of Nimbus 7 Solar Backscattered Ultraviolet (SBUV) data. Comparison between these figures indicates a substantial qualitative agreement between the model calculations and the observations. Attention is drawn to the absence of SBUV measurements at latitudes poleward of 65° in Figure 2.2(b). Figure 2.3(a) shows the amplitude of the semi-annual variation computed from model results while (b) shows the SBUV data. A large region of maximum semi-annual ozone amplitude present at the tropics near 3 mb, is absent in model calculations. This may be ascribed to the fact that the semi-annual temperature and wind oscillation (SAO) was not included in the model. However, regions of significant semi-annual ozone amplitude between 1 and 10 mb at mid- to high latitudes are qualitatively in good agreement with those seen in the data. The maxima of the equatorial semi-annual ozone oscillations occur during spring and fall, while the maxima of the high latitude ozone variation occur during the winter and summer. The hemispheric symmetry seen in the modelled annual and semi-annual amplitudes of ozone is a reflection of the fact that the momentum forcing in this two-dimensional model is crudely parameterized as a Rayleigh friction coefficient which is independent of latitude. Modelled transport parameters, temperatures, and therefore chemical profiles, are the same for both hemispheres during respective seasons, contrasting with asymmetry observed by TOMS in Figure 2.1. The general agreement depicted in Figures 2.2 and 2.3 suggest that the processes that are responsible for the model variations cause much of the observed variations. The model calculations at polar latitudes will be addressed below, as this region is the focus of the work in this thesis.

Modelled ozone abundances may be determined by the magnitudes of ozone loss and production rate terms calculated by the model due to photochemical production and destruction, temperature and transport. Figure 2.4(a) gives both the modelled and measured (by SBUV) ozone mixing ratio for 75°N and 42 mb. No observations are possible at high latitudes in the winter months due to lack of solar irradiance. The model clearly underestimates ozone abundances at this pressure level. Figure 2.4(b) shows chemical production ($2J[O_2]$) and loss terms each given by the equation that is the rate limiting step in each chemical family e.g.

$$O + NO_2 \rightarrow NO + O_2$$

is the rate limiting step when considering ozone destruction by the NO_x family and so has the loss term $2b[O][NO_2]$. Figure 2.4(b) clearly illustrates that the nitrogen family has the dominant effect on ozone at this latitude and altitude in summer. Figure 2.4(c) shows the sum of the chemical destruction terms and the modelled transport rate. Transport is dominant in the winter months while the calculated spring decrease and

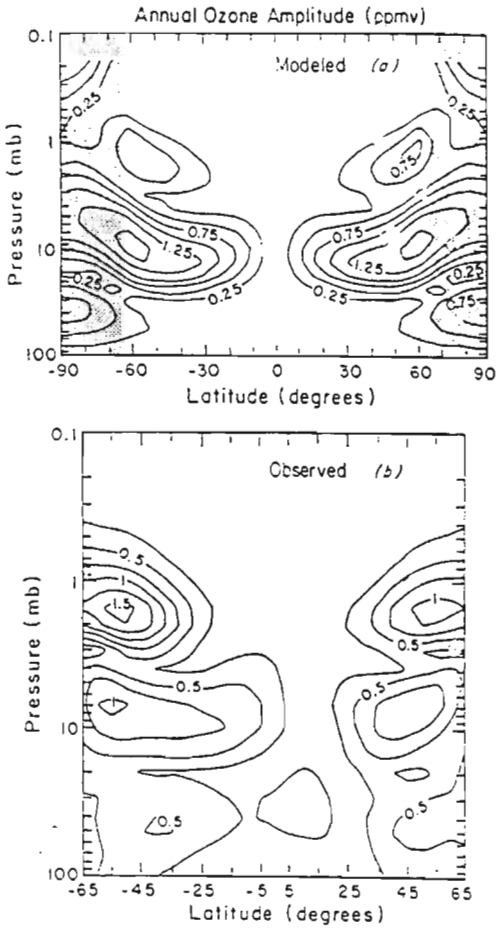

Figure 2.2: (a) Average annual ozone amplitude (in ppmv) calculated from the model of *Garcia and Solomon (1983)*. Shaded areas indicate regions where SBUV ozone profile data is unavailable. (b) Average annual ozone amplitude determined from 9 years of SBUV data. The plot extends from 65°N-65°S. Adapted from *Perliski et al. (1989)*.

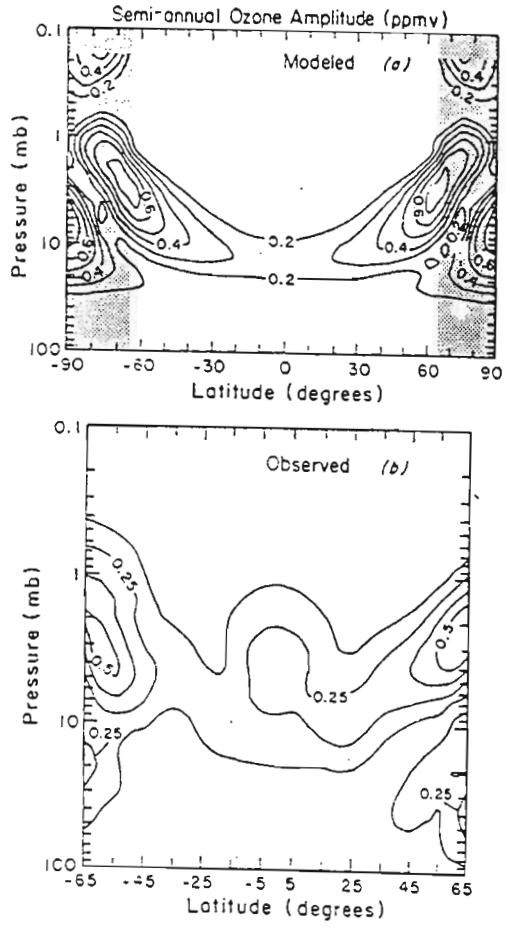

Figure 2.3: (a) Average amplitude of the semi-annual component of ozone. (b) Average semi-annual ozone amplitude determined from SBUV data. Adapted from *Perliski et al. (1989)*.

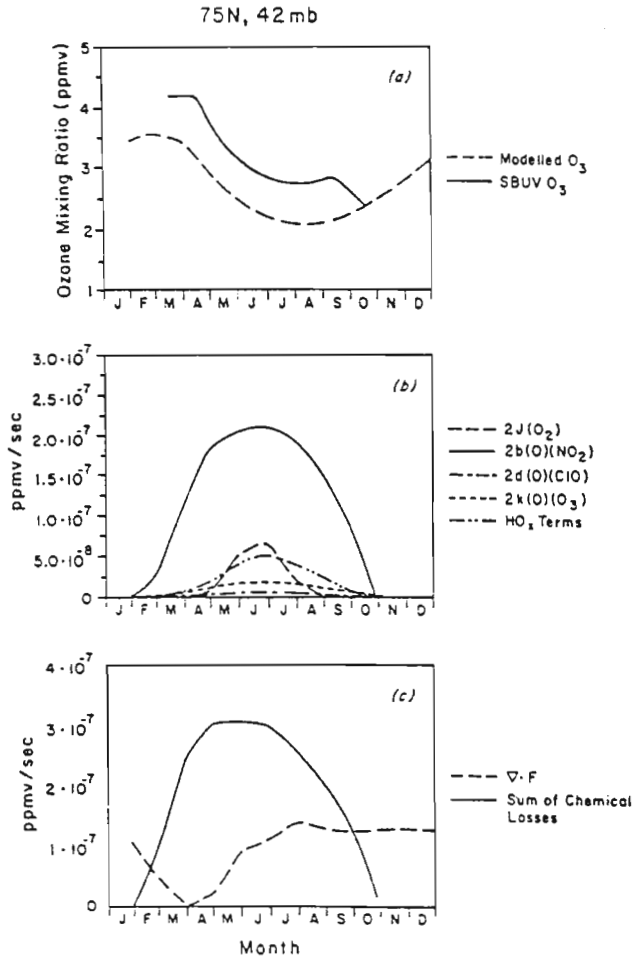

Figure 2.4: (a) The long term average SBUV (solid curve) and modelled (dashed curve) ozone mixing ratios at 75°N latitude and 42 mb for each month of the year. (b) Modelled ozone production and loss rates. (c) The modelled ozone transport rate (dashed curve) together with the sum of the calculated ozone loss rates (solid curve). From *Perliski et al. (1989).*

subsequent summer ozone minimum are caused by increased effectiveness of chemical destruction processes, particularly odd-nitrogen destruction. The winter maximum can be attributed to both the increased downward and poleward transport and decreased effectiveness of chemical destruction processes.

In the polar middle stratosphere (pressure levels around 10 mb) photochemical production becomes important. The reason that ozone production is not very rapid below 10 mb is that most of the ultraviolet radiation responsible for molecular oxygen photodissociation is attenuated at higher levels. The variation in odd oxygen production is largely responsible for the annual ozone variation in the middle stratosphere. Semi-annual increases in ozone at the equator may be attributed to two solar zenith crossings during the equinox periods. High latitude semi-annual maxima are caused by modulation of the radiatively-driven middle stratospheric ozone variation by temperature dependent chemical destruction processes, and occur in winter and summer.

Figure 2.5 illustrates summarized results from this model (*Perliski et al., 1989*) showing regions in which chemical destruction, production and transport either dominate or balance each other. The large 'dashed' area indicates that most of the model atmosphere is in photochemical equilibrium such that the ozone density adjusts to its equilibrium value due to the balance between chemical production and destruction rates. In the high latitude summer stratosphere and lower mesosphere (> 10 mb) destruction processes dominate (finely dotted area in Figure 2.5) which are initiated by radicals. The rates of these destruction reactions are temperature dependent and thus proceed quickly in summer. The tropical stratosphere is dominated by odd-oxygen production and upward transport of ozone poor air from the troposphere. Just above 30 mb production occurs more rapidly than destruction and transport. In the polar winter, ozone is dominated by transport at all altitudes since polar night increases its chemical lifetime and homogeneous chemistry can only proceed very slowly. Transport is poleward and downward in the polar winter, thus the increasing ozone abundances there. The ozone time constant due to photochemistry becomes considerably longer in winter at high latitudes than it is at lower latitudes at any given pressure level, as a direct result of the decreased insolation there. Advective transport time scale for the lower stratosphere is about 100 days for a constituent with a 5 km vertical scale height, therefore transport can dominate ozone photochemistry if the photochemical time scale is longer than 100 days. Most of the total column ozone at high latitudes is located in this dynamically controlled region. This is a well known fact which is clearly manifested by the large observed local variability in the ozone column at these latitudes and its relationship to planetary and cyclone waves (e.g. *Dobson et al., 1929, Schoeberl and Krueger, 1983*). Further, the observed monthly and zonally averaged ozone column in high latitude winter and spring is much greater than it is in summer and fall. This is generally attributed to enhanced downward, poleward motion from the

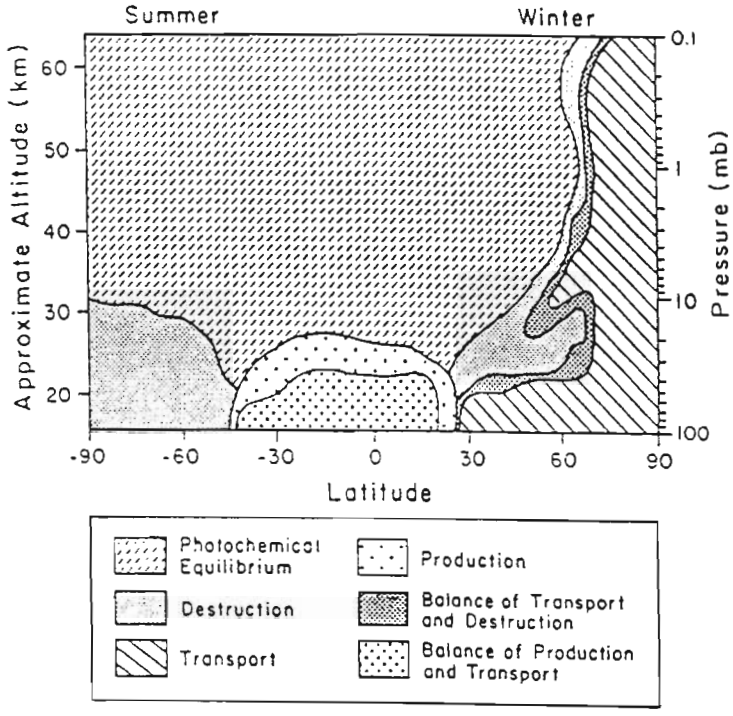

Figure 2.5: Shaded areas demarcate regions of the middle atmosphere for which seasonal ozone amounts are governed by odd oxygen production, odd oxygen destruction, transport and balances between these processes. Most of the middle atmosphere is in photochemical equilibrium. From *Perliski et al. (1989).*

chemically controlled region (where ozone mixing ratios are large) during the winter season.

Other diabatic circulation and residual Eulerian models include those of *Stordal et al. (1985)*, *Guthrie et al. (1984)* and *Ko et al. (1985)*. Some models, such as that of *Chatfield and Crutzen (1984)*, use the classical Eulerian stream function equation for the transport. The Oxford University two-dimensional model (*Harwood and Pyle, 1975, 1977*) uses the classical Eulerian mean meridional circulation computed from eddy fluxes deduced from satellite observations.

Austin and Butchart (1992) have modelled ozone levels over the polar regions by inputing different wave amplitudes into a three-dimensional model. They found that polar ozone amounts were sensitive to planetary wave dynamics and that the wave induced mean descent over the whole polar region is an important ingredient in determining the springtime behaviour of polar ozone. Two-dimensional (zonally averaged) models are often inadequate in describing the transport of ozone, as discussed in sub-section 1.3.4. In particular, cross-latitude flow, during winter and spring at high latitudes, leads to planetary wave structure, clearly visible in Figure 2.1. In addition, photochemistry depends on the local solar irradiance, which is a function of latitude, longitude, altitude and time. Only a three-dimensional model can provide a completely realistic representation of both transport processes and photochemical effects (e.g. *Tuck, 1979*). However, the atmosphere represented by such a model may not, of course, exactly correspond to the real atmosphere at a given place at any particular time.

The following sections of this chapter will describe, in some detail, the important dynamics and chemistry of polar ozone. Section 2.8 examines the possible effects of solar variations on high latitude ozone amounts.

2.4 Dynamical Influences

2.4.1 The polar vortex

The relative warmth of the stratosphere results from the absorption of solar ultraviolet radiation by ozone, which has its highest concentration there. Heating by solar absorption is balanced by cooling through emission of thermal infrared radiation, primarily from the 15 μm band of carbon dioxide. After the autumnal equinox, the polar regions fall into darkness and the solar ultraviolet heating ceases. Continuing emission of thermal radiation quickly cools the polar stratosphere to temperatures much lower than those of the mid-latitude stratosphere. A strong latitudinal pressure gradient

then develops between the pole and mid-latitudes which, combined with the Earth's radiation, (as explained in sub-section 1.3.2) produces a circumpolar belt of westerly winds referred to as the polar night jet or polar vortex. Wind speeds in this jet may exceed 100 m s^{-1}. Although temperatures within the polar vortex in wintertime are largely determined by radiative processes, they are also dependent on dynamics and transport of heat by atmospheric motions. Zonally symmetric circulations are inefficient at transporting heat in a rotating fluid such as the Earth's atmosphere. Thus, in order for temperatures within the polar vortex to depart substantially from radiatively determined values, thermal transport by waves is required. The most important waves for mixing in the stratosphere are planetary-scale Rossby waves, or planetary waves, which propagate upward from the troposphere. Stratospheric planetary waves, first described by *Charney and Drazin (1961)* are associated with large quasi-stationary features as represented by the geopotential height and temperature fields of Figure 1.8. Planetary wave transience and dissipation is responsible for the intermittent breakdowns of the polar vortex structure, called sudden stratospheric warmings (SSW's). In the northern hemisphere, SSW's are a common occurrence and can often, as early as mid-winter, lead to a temporary reversal of the north-south zonal mean temperature gradient.

Clouds in the polar stratosphere are referred to generically as polar stratospheric clouds (PSC's). Due to the dryness of the stratosphere (2 to 4 ppm mixing ratio of water) the temperature has to fall to -83°C (at ~19 km) for pure ice clouds to form. Another type of cloud composed of nitric acid trihydrate (NAT) are known to be widespread over the polar regions, requiring only -78°C (at ~ 19 km) to form (*Toon and Turco, 1991).* The formation of PSC's leads to dehydration of the polar winter stratosphere.

PSC's are the sites for a group of heterogeneous reactions that perturb the normal gas-phase chemistry in the polar region. The most important heterogeneous reaction converts the relatively unreactive species chlorine nitrate and hydrochloric acid (the dominant chlorine reservoirs) to molecular chlorine and nitric acid:

$$ClONO_{2(g)} + HCl_{(s)} \rightarrow HNO_{3(s)} + Cl_{2(g)}. \qquad (2.3)$$

Molecular chlorine is photolyzed in the spring sunlight and atomic chlorine quickly reacts with ozone to form the chlorine monoxide radical ClO. Ozone destruction begins (0.5 - 1% per day) with the formation of the chlorine dimer (*Molina et al., 1987*). Widespread ozone destruction during Antarctic spring requires that the air be chilled below -78°C for sufficient time so that the PSC's can effect the conversion of inactive reservoir species to the radical species that attack ozone. Such conditions occur over widespread regions of the Antarctic lower stratosphere but not as extensively in the Arctic. Nonetheless, high levels of ClO have been detected in the Arctic stratosphere (*Burne et al., 1990*). To prevent the conversion of ClO back to unreactive $ClONO_2$ through the reaction of ClO with NO_2, active nitrogen compounds must be suppressed.

The formation of NAT PSC's denitrifies the vortex air. The observed ozone destruction in both hemispheres is therefore contingent on the stratosphere remaining denitrified during the period of ozone loss (*Schoeberl and Hartmann, 1991*) which means that mid-latitude air, containing reactive nitrogen compounds, be not mixed into the polar vortex. Dynamical characteristics of the polar vortex provide this chemical isolation, and the ozone hole develops poleward of the latitude of strongest westerly winds.

The structure of the polar vortex is different in the two hemispheres. In the southern hemisphere the winter jet is far stronger than its northern counterpart, evident in Figure 2.6(b), with westerly wind speeds of over 80 m s^{-1} above 40 km. The associated greater poleward temperature decrease (Figure 2.6(a)) in the lower southern stratosphere can result in temperatures below -80°C there. PSC's are usually observed in the polar cold pools in the lower stratosphere (*McCormick and Trepte, 1986*). Larger cold temperature regions in the Antarctic stratosphere imply that PSC's form more extensively there than in the northern hemisphere and so there is a larger area on which heterogeneous reactions can occur. Temperatures as cold as -70°C are evident soon after the onset of polar night. The southern hemisphere thermal response is mainly radiative, and it follows seasonal changes of insolation closely. From mid-winter, the northern hemisphere can often show a departure from this path. These interhemispheric differences in stratospheric climatology may be traced to the more frequent disruption of the boreal vortex by planetary waves. Planetary waves are forced in the troposphere at lower latitudes by large scale orographic and thermal contrasts and can propagate into the stratosphere only when westerly winds are present (*Charney and Drazin, 1961*). In the northern hemisphere the great mountains of Asia and North America send large amplitude Rossby waves upward. The continent of Antarctica is surrounded by ocean and only the rather narrow band of the Andes may disrupt the zonal winds. This makes the southern hemisphere vortex symmetrically shaped and stable.

Following the autumnal equinox radiative processes cool the polar stratosphere and a symmetric overturning begins with strong downward motion, of up to 100 kg m^{-1} s^{-1} (*Garcia et al., 1984*), in the polar regions and weak rising motion elsewhere. Downward air motion in the polar region warms the air through adiabatic compression, partially offsetting the radiative cooling which would otherwise take the temperature to -90°C or below. The descending air within the vortex carries with it long lived trace species from the upper stratosphere. Below 30 km the vortex temperature approaches radiative equilibrium and the rate of descent within the vortex decreases. Outside the vortex, rapid lateral mixing by planetary waves counters descent and causes an apparent differential vertical displacement of air (by 2 to 3 km) between the vortex exterior and interior. The amplitude of upward propagating planetary waves is observed to increase with altitude as does their magnitude of heat transport. The increase in heat

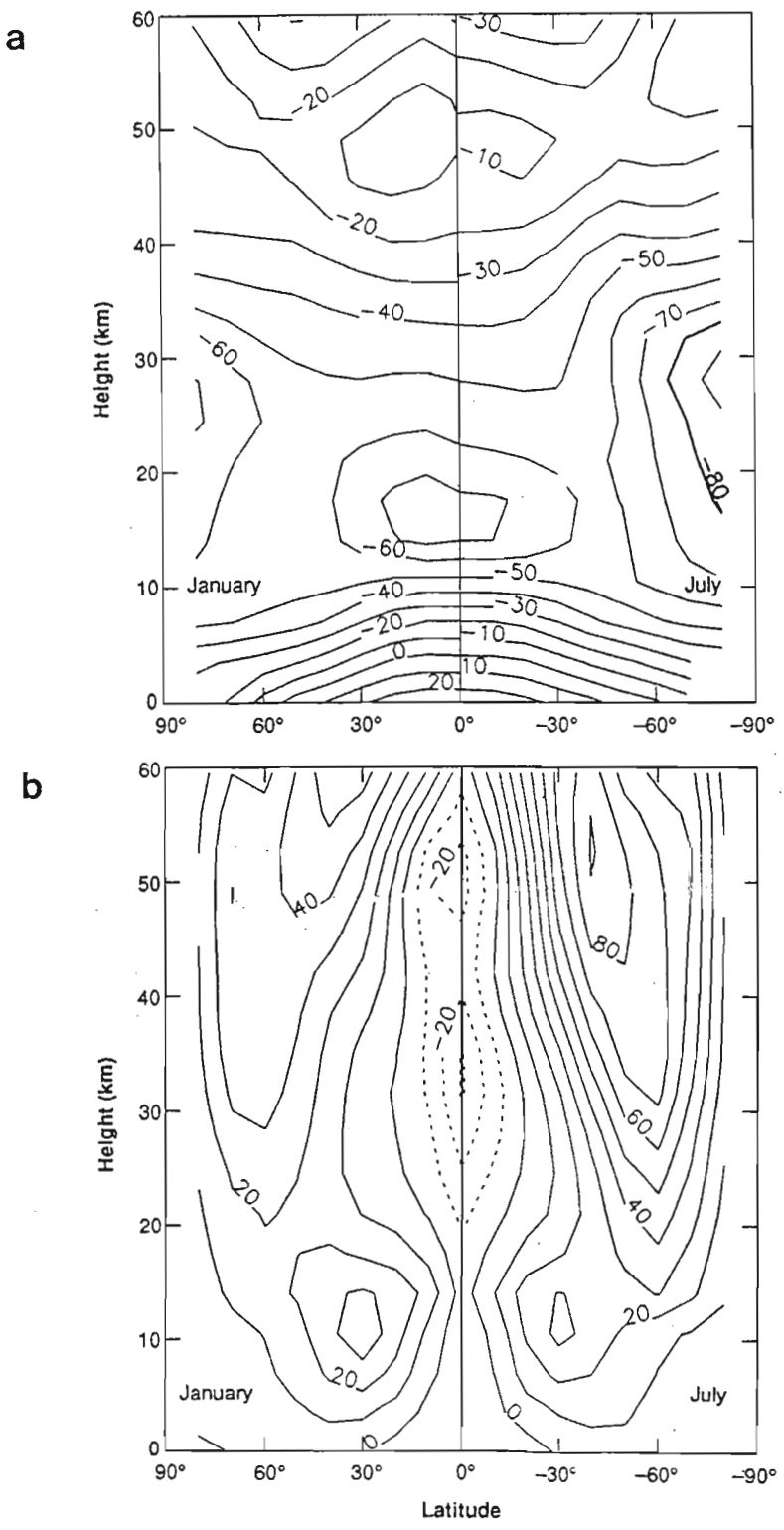

Figure 2.6: Zonally averaged temperatures (a) and westerly wind speeds (b) for winter conditions over both poles. Comparison of the two polar atmospheres, reveal that the Antarctic stratosphere is much colder than its northern counterpart. In addition, higher wind speeds characterize the southern polar vortex. From *Schoeberl and Hartmann (1991).*

flux is seen in the weakening of the pole to mid-latitude temperature gradient above 25 km. Above 30 km temperatures are too warm to support PSC's and this latitude roughly marks the top of the isolated region of the polar vortex. Meridional gradients in concentration of various chemical tracers observed in mid-winter and spring at the vortex boundary are very sharp and provide a consistent marker of the edge of the polar vortex for periods of a month or more (*Proffitt et al., 1989 (a)*). The sharpening of the meridional constituent gradient across the vortex edge appears to be directly related to the planetary wave erosion of the vortex edge.

In spring solar heating of ozone returns to the polar regions and the meridional temperature gradient begins to weaken. The vortex appears to transit abruptly to the summer circulation by way of a final stratospheric warming. During the final warming, amplitudes of planetary wave displacement increase explosively, and the vortex shatters into smaller fragments, which drift to mid-latitudes. The final warming usually develops in March in the northern hemisphere but may be as late as early December in the southern hemisphere. Hence, the southern hemisphere vortex is much longer lived. The reduced stratospheric solar heating, as a result of ozone depletion, has been proposed for the delay in the austral final warming. The stratospheric summer circulation is weak easterly winds flowing symmetrically around the pole.

Potential vorticity (PV) is used as a flow diagnostic under adiabatic and frictionless conditions, as described in sub-section 1.3.2. As the vortex spins up in early winter, PV builds up rapidly in the polar region, and a PV gradient is established between the subtropics and the pole. A positive PV gradient is required for the existence of planetary waves, which are essentially oscillations between relative and planetary components of vorticity, as described in sub-section 1.3.2. Planetary waves may be produced by air flow over large scale mountain ranges. While they are not dissipated the wave amplitudes will grow as the inverse square of the density, resulting in wave amplitude growth that is exponential with height. At some altitudes the amplitude becomes large enough to overcome the mean PV gradient, the wave becomes unstable and is said to 'break'. This causes deceleration of zonal winds and rotating of the PV gradient so that irreversible mixing of PV and chemical species occurs. The poleward transport of heat during a wave breaking event causes a stratospheric warming. Due to the equatorward propagation of planetary waves, planetary wave breaking and associated constituent mixing occur preferentially on the equatorward edge of the polar vortex, where a mid-latitude 'surf zone' is created. The surf zone flattens the PV gradient at mid-latitudes while steepening the gradient at the poleward edge of the surf zone. A major stratospheric warming takes place if the mean zonal flow is temporally reversed as low as the 30 mb level (*Holton, 1992*). If a major warming occurs sufficiently late in the winter the westerly vortex may not recover, this is the final stratospheric warming. During warmings (other than the final) erosion of PV and trace species from the edge

of the vortex sharpen the gradient, and the region of high PV decreases in area. Species eroded from the vortex are rapidly mixed into mid-latitudes.

The increased number of stratospheric warmings in the northern hemisphere compared to those in the southern hemisphere is responsible for the interhemispheric differences in the extent of the vortex and the effective isolation of the vortex air. Stratospheric warming events occur intermittently, occurring once or twice a season in the northern hemisphere but rarely in the southern hemisphere (*Schoeberl, 1978*).

In situ aircraft observations during the Airborne Antarctic Ozone Experiment (AAOE) in October 1987 are given in Figure 2.7(a) and the Airborne Arctic Stratospheric Expedition (AASE) in January and February 1989 in Figure 2.7(b). Data show clear evidence of coincident, sharp gradients in N_2O (a long lived trace gas) and PV. The sharp decline of N_2O (indicated by diamond shapes with low values towards the top of the graphs) and the increase of PV (solid, thin line) toward the pole are both indicative of the combined effects of the downward displacement of air inside the vortex and subsequent enhancement of the gradient by erosion of the outside edge of the vortex. In the polar vortex, the region where ozone can be removed most rapidly is coincident with the region of high ClO. Chlorine monoxide measurements are represented in Figure 2.7 by the connected data points. The region, called the chemically perturbed region (CPR) lies inside the vortex but does not coincide exactly with the edge of the vortex as defined by the steepening of the N_2O gradient (Figure 2.7(a)). For both Antarctic and Arctic data, ClO slightly increases just poleward of the relatively sharp decrease in N_2O (60°S in Figure 2.7(a) and 68°N in Figure 2.7(b)). The southern hemisphere vortex covers an area 8-10° larger in latitude. A second but larger increase in ClO, defining the CPR edge, occurs in the Antarctic data at 66°S, coincident with a second but smaller decrease in N_2O. This large ClO change coincides with the region where temperatures are consistently below -70°C and PSC formation is more or less continuous. Chlorine monoxide formation in the Arctic vortex (Figure 2.7(b)) appears to be associated with PSC events rather than continuous processing that occurs deep inside the Antarctic vortex. These PSC events are associated with temporary cooling of air parcels to $<$ -78°C by adiabatic expansion. Analysis of AAOE and AASE data show that N_2O and PV provide consistent information about the relative position in the vortex and can be used as surrogates (*Hartmann et al., 1989*). Furthermore, the presence of sharp gradients implies that mixing across the vortex 'wall' is limited as such gradients can only develop outside of vigorously mixed regions. This result implies that ozone depletion is not achieved by a lateral 'flow through reactor'. Recent measurements (*Randel, 1993*) from the Upper Atmosphere Research Satellite (UARS) have rekindled arguments over just how fast material does flow through the vortex. Water vapour is frozen out in the intense cold within the vortex, but UARS has observed this dehydration signature far into the middle and low latitudes. This widespread dehydration suggests that

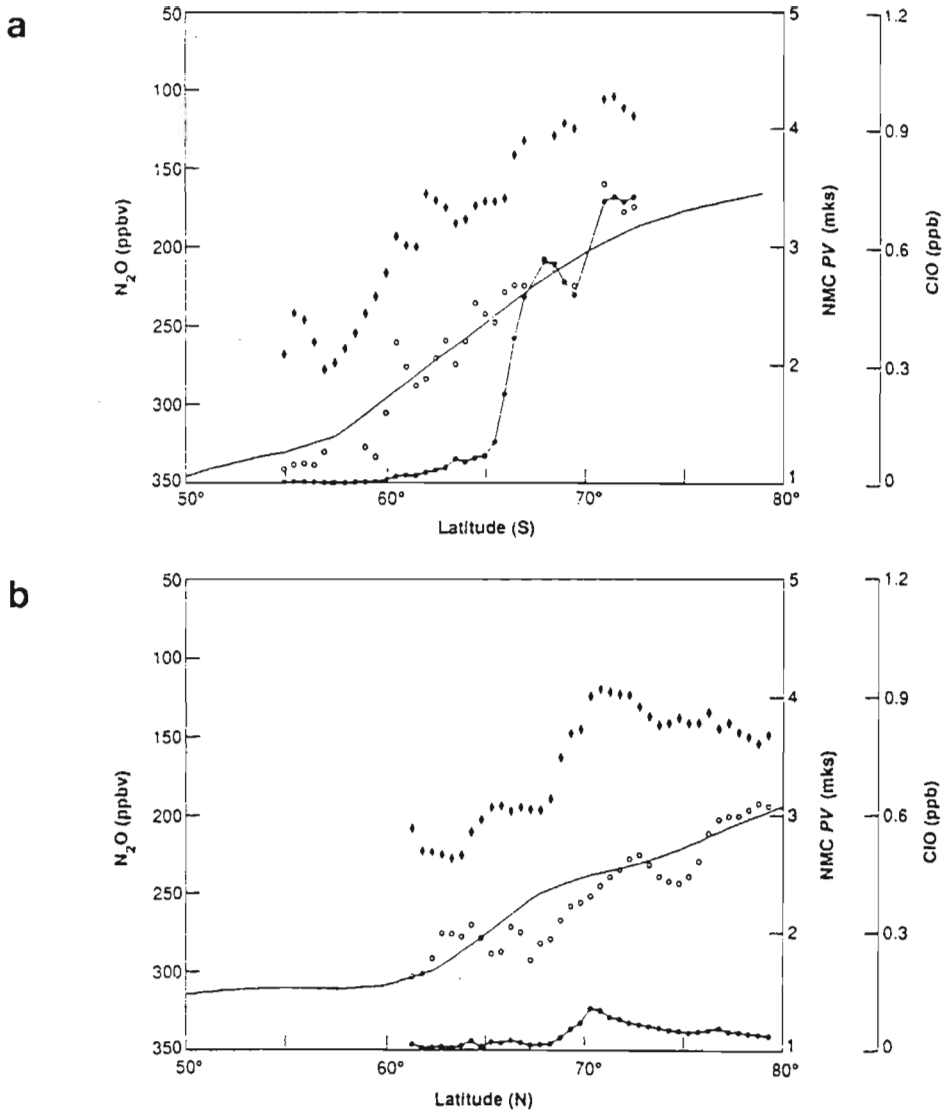

Figure 2.7: The structure of the Antarctic (a) and Arctic (b) vortex at 20km, as observed by aircraft. Diamonds represent measured N_2O amounts, low values at the top of both graphs. The sharp decrease in N_2O defines the vortex edge. Connected dots represent ClO observations. High ClO values are found within the CPR. Calculated potential vorticity values are plotted as a thin line and observed potential vorticities are plotted as open circles. From *Schoeberl and Hartmann (1991)*.

much of the air has recently traversed the polar vortex. The mechanism for flushing is thought to be transport out of the bottom of the vortex. This may explain the notable mid-latitude erosion of ozone in the northern hemisphere. Debate on the polar vortex continues.

The lifetime of the southern polar vortex may have lengthened in recent years by the negative feedback of polar temperatures due to ozone loss. In years to come, enhanced CO_2 levels, resulting in increased atmospheric thermal emission, may lead to even colder polar temperatures (*Austin et al., 1992*).

2.4.2 The quasi-biennial oscillation (QBO)

The quasi-biennial oscillation (QBO) is the large variation in the zonal wind, due to stratospheric wave transience and dissipation, in the tropical lower stratosphere (*Reed et al., 1961*). The QBO is monitored by measurements of the 50 hPa zonal wind at Singapore (1°N;104°E). This phenomenon has a mean period of 27 months, which can vary from 20 to 40 months. Theoretical studies indicate that its origin lies in the vertical transport of momentum associated with certain types of tropical waves (*Lindzen and Holton, 1968*; *Holton and Lindzen, 1972*). Total ozone exhibits a variation in the tropics which is clearly related to the QBO (e.g. *Hilsenrath and Schlesinger, 1981* and *Angell and Korshover (1973)*). In most studies the latitudinal variation of the phase of the QBO in total ozone is considered to be relatively simple, nearly in phase with 50 hPa zonal winds in the equatorial regions, and out of phase at latitudes >35° in both hemispheres (*Zerefos et al., 1992*).

Dunkerton et al. (1988) have noted that mid-winter major warmings appear to be modulated by the QBO as there is an increase in eddy activity in the QBO easterly phase. The year to year variability of planetary wave activity in both hemispheres produces a corresponding response in the strength and temperature of the polar vortex. Studies indicate that year to year variations in planetary wave activity generally follow the phase of the QBO in the tropical winds of the lower stratosphere, with more activity during easterly phase years. The ozone hole tends to exhibit the same variability (*Garcia and Solomon, 1987*). For example, in the QBO westerly year of 1987, the ozone destruction in the Antarctic was almost complete (*Hofmann et al., 1989*). In the QBO easterly year of 1988, the planetary wave activity was high, polar temperatures were warm, and the zonal mean vortex winds were weak. The ozone depletion was much less in October 1988 than at the corresponding time in 1987 (*Kanzawa and Kawaguchi, 1990*). Sensitivity of polar ozone to dynamical activity is not unexpected because slight modulation of the vortex temperatures by planetary waves could greatly modulate the regional coverage of PSC's. Indeed, any planetary wave activity at all

will tend to weaken ozone depletion by raising vortex temperatures.

The increase in eddy heat flux could also significantly reduce the degree of isolation of Antarctic stratospheric air. Chemical theories of Antarctic depletion require polar air to be highly denitrified (*Solomon, 1990*). Injection of high NO_x mid-latitude air will result in the levels of active chlorine being significantly reduced in the formation of $ClONO_2$ and there will be a subsequent reduction in ozone depletion.

2.4.3 Planetary waves and SSW's

The phenomenon of sudden stratospheric warmings (SSW's) is an important example of stratospheric wave transience and dissipation. This is shown by using the Langrangian or transformed Eulerian equations in models by *Matsuno and Nakamura (1979)* and *Dunkerton et al. (1981)*. This induces a mean meridional circulation with strong downward motion at high latitudes and upward motion in the tropics, producing observed high latitude warming by adiabatic compression. Under these circumstances $u \sim v$ and warmings affect the ozone distribution.

Austin and Butchart (1992) have used a three-dimensional model to study the influence of planetary wave dynamics on polar ozone photochemistry over the winter and springtime period. Modelled ozone amounts over the polar region showed a considerable sensitivity to planetary wave forcing. Wave-induced mean descent over the whole polar region was identified as an important ingredient for determining behaviour of polar ozone over this time period. A simple two-dimensional parameterization of planetary wave drag that leads to SSW's is not feasible because warmings are sporadic phenomena that occur during periods of strong wave transience. The latter can be simulated only by explicitly computing the evolution of the wave field and its interaction with the zonally averaged wind field. So, a model that includes equations for both the zonal mean state and perturbations (i.e. waves) in it is a three-dimensional model.

Wave number 1 pattern, exhibiting only one major ridge (high) and one trough (low) in the zonal direction, is very stable in the south polar region in winter and spring. This pattern is apparent in the geopotential height and temperature fields in Figures 1.8 (a) and (b). The trough of this planetary wave is located in the longitude region 30°E to 60°W (*Chandra and McPeters, 1986*). Wave transients are minimal in the trough region. Wave pattern 2 and 3 are common in the northern polar regions.

2.4.4 El Niño-Southern Oscillation (ENSO)

When examining a zonal strip of equatorial total ozone, there appears to exist an east-west seesaw pattern with a nodal longitude around the international date line (*Shiotani, 1992*). This east-west variation has a characteristic time scale of about four years and is clearly related to the ENSO cycle. The southern oscillation is one of the most prominent climate anomalies in the equatorial atmosphere, showing a standing variation of pressure anomalies between the Indian Ocean and the Pacific Ocean in an opposite sense. The variation is mutually coupled with the sea surface temperature (SST) variation in the eastern Pacific Ocean, particularly with warm SST anomaly events, so called El Niño events. As an index representing this fluctuation, the Southern Oscillation Index (SOI) is often used e.g. *Ropelewski and Jones (1987); Schneider and Schonwiese (1989).* This dimensionless quantity is a measure of the pressure gradient between Tahiti (18°S;150°W) and Darwin (12°S;131°E). The solid curve in Figure 2.8 represents the SOI while the east-west gradient of the equatorial total ozone field is represented by the dashed curve. The east-west gradient is defined by the differences between averages of the western (75-180°W) and eastern (60-165°E) Pacific region. El Niño events occur with a period of about four years. Figure 2.8 clearly demonstrates that these events occurred in 1982-1983 and 1986-1987, when the SOI has large negative values, there are positive anomalies in the western Pacific and negative anomalies in the eastern Pacific, resulting in a positive east-west gradient in equatorial total ozone. The anomaly pattern, and therefore the gradient, is reversed during anti-El Niño events. Because the SST's in the eastern Pacific are higher during El Niño events, an active region of convective clouds moves relatively eastward; this must bring about a change in the zonal structure of the tropopause height.

The ENSO phenomenon is a pronounced multiannual fluctuation in the tropical troposphere although its extremes can be detected both in the extratropics and the stratosphere. The primary manifestation of the southern oscillation is a seesaw in atmospheric pressure at sea level between the sub-tropical high over the south-east Pacific and a low pressure situated over the Indian Ocean between Africa and Australia. This phenomenon was described in the early 1920's by Sir Gilbert Walker and is closely linked to the interannual temperature oscillations in the eastern and central Pacific (El Niño).

The in phase relation between SOI and total ozone is confined within about 25° of latitude from the equator.

The ENSO effect, with a time scale of about four years, is not limited to the development of zonal structure. *Shiotani (1992)* used the eleven year TOMS data period to evaluate the zonal mean ozone over equatorial regions, this is plotted in Figure 2.9

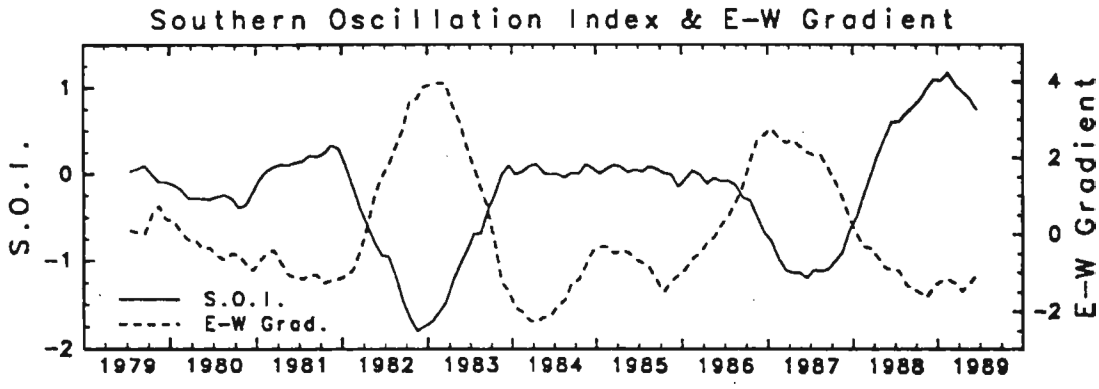

Figure 2.8: Plots of SOI (solid curve) and the east-west gradient in equatorial ozone (broken curve). From *Shiotani (1992)*.

as a solid curve. Along with the zonal mean is a dashed curve which represents, once again, the east-west gradient. The zonal mean values and the east-west gradient are strongly anti-correlated. This variation in the ozone zonal mean may be attributed to modulation of the tropical Hadley cell. The total ozone data plotted in Figure 2.9 have been filtered to eradicate high frequency components.

Therefore, the zonal mean values are lower during El Niño events than during periods of anti-El Niño levels. *Hasebe (1983)* also found a four year oscillation in equatorial total ozone using the Nimbus 4 BUV and a network of ground-based data. In addition, *Zerefos et al. (1992)* use total ozone data at 28 Dobson stations, most of which have uninterrupted records during the period 1957 to 1991. They show that large ENSO events, such as 1982-1983, are followed within a few months time lag by low total ozone values in the middle and even high latitudes. The large ENSO events are followed by negative ozone anomalies e.g. *Bojkov (1987)*.

Studies by *Trenberth and Smith (1984)* show that ENSO and stratospheric QBO are two distinct and independent phenomena. However, their effects on ozone may be interdependent.

2.5 The heterogeneous chemistry of the chlorine catalysed ozone hole

Farman et al. (1985) detected, using a Dobson spectrophotometer, an extraordinary 50% ozone column loss above Halley (75.5°S, 26.8°W). *Stolarski et al. (1986)* confirmed such decreases in total ozone from the Total Ozone Mapping Spectrometer (TOMS). There has been a decline in the minimum spring total ozone value, from 250 DU in 1980 to 125 DU in 1987 and to below 120 DU in 1991 (*Aikin, 1992)*. Figure 2.10 is a TOMS map, prepared by the author, over the southern hemisphere during spring of 1989. Details regarding the production of TOMS false colour maps can be found in section 4.2. A substantial area of the Antarctic continent is covered by blue shades indicating total ozone less than 200 DU. TOMS showed that this significant depletion, now called the 'ozone hole', in total ozone over Antarctica was apparent in the first TOMS observations in 1979 (*Chandra and McPeters, 1986*). The ozone hole occupies 7% of the area of the entire southern hemisphere and extends from roughly 100 to 30 mb in altitude. The loss of ozone mass in austral spring over Antarctica represents 3% of the total mass of global ozone (*Toon and Turco, 1991*).

By the early 1980's the chlorine content of the stratosphere had risen to about 2.5 ppbv compared to a natural level of 0.6 ppbv which is due largely to the oceanic

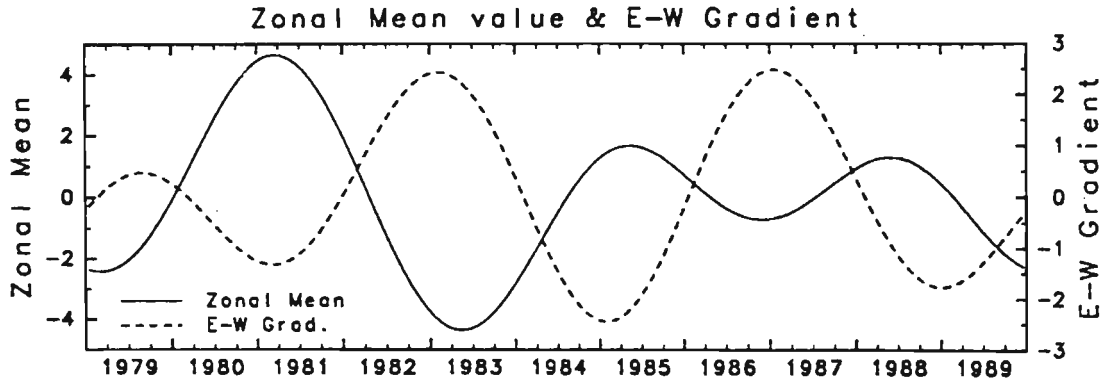

Figure 2.9: The 'out of phase' relationship is evident between the zonal mean (solid curve) and east-west gradient (broken curve) of a filtered total ozone field. From *Shiotani (1992)*.

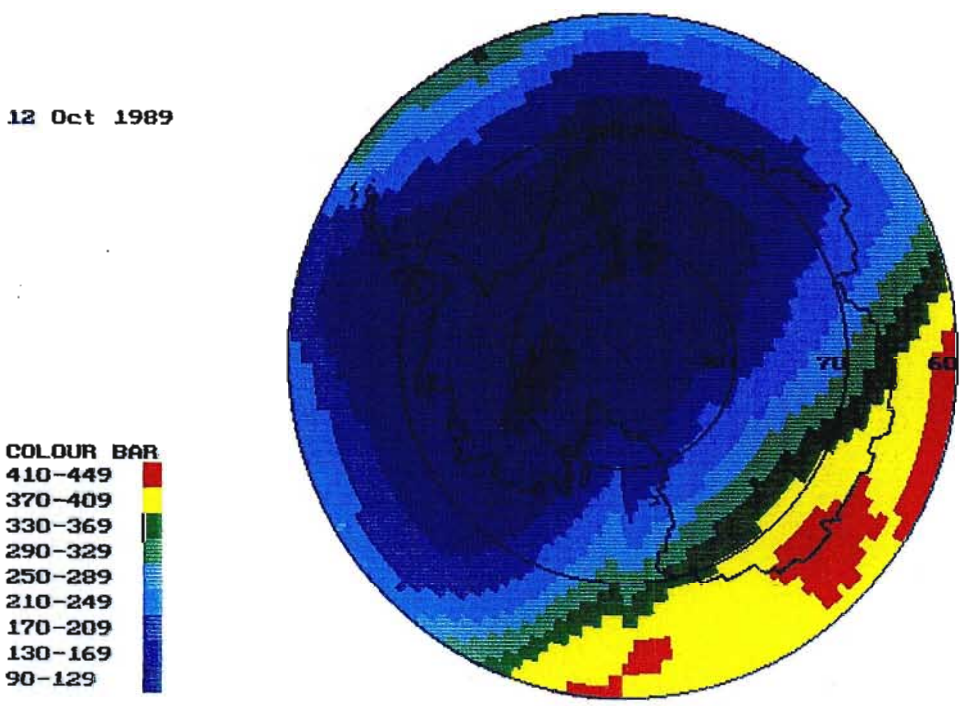

Figure 2.10: A false colour plot of total ozone, from the TOMS instrument, over Antarctica for 12 October 1989. The chlorine-catalysed 'ozone hole', represented by shades of blue colours, covers much of the Antarctic continent.

release of CH_3Cl (*Solomon, 1990*). The stratospheric photolysis of chlorofluorocarbons, most commonly CF_2Cl_2 (CFC-12) and $CFCl_3$ (CFC-11), is responsible for this large increment. These compounds have long lifetimes in the troposphere (up to 100 years) and can eventually reach the stratosphere where they are photolyzed by ultraviolet radiation, releasing reactive chlorine.

Ozone levels in Antarctica reach their lowest values in late September to early October. The stratosphere above Antarctica at this time of year is colder than anywhere else in the Earth's stratosphere. some 10-15°C colder than the Arctic stratosphere in the same season. The interhemispheric temperature difference is believed to be due to differing topographic features poleward of 50° latitude. The relatively landless strip between 50-70°S leads to fewer large scale wave disturbances at sub-polar latitudes, and so little wave induced mixing. These conditions combined with the onset of polar night, result in a strong temperature gradient between the Antarctic and southern mid-latitudes resulting in a rapid circumpolar flow called the polar vortex. The degree of isolation is critical in determining the depth and duration of the ozone hole. Polar ozone depletion is a chemical process, this chemistry appears to be highly sensitive to the dynamics of the austral winter-spring stratosphere (*Lait et al., 1989*). Ozone depletion by chlorine continues until the seasonal 'ozone hole' is filled by a rapid influx of ozone-rich air from lower latitudes that accompanies the dynamical breakdown of the polar vortex (*Schoeberl and Hartmann, 1991*) typically in late October, early November.

Prior to extensive observations in the polar regions three theories of ozone depletion were advanced, namely, the solar cycle theory (*Callis and Natarajan, 1986*), the dynamical theory (*Mahlman and Fels, 1986*), and the chlorofluorocarbon theory (*Solomon et al., 1986*).

The basis of the solar cycle theory is that reactive nitrogen compounds (NO and NO_2) are produced in abundance in the sunlit mesosphere, especially during times of high solar activity. These compounds could be transported down to the lower stratosphere during the polar night, when their chemical lifetimes are long. An unusually active phase of the 11 year solar cycle could produce an additional enhancement of reactive nitrogen compounds, normally the most important agents in destroying ozone, within the polar vortex. These species can destroy ozone catalytically (*Crutzen et al., 1975*) following the return of sunlight over the polar cap in September by the following mechanism

$$NO + O_3 \rightarrow NO_2 + O_2 \tag{2.4}$$

$$NO_2 + O \rightarrow NO + O_2. \tag{2.5}$$

This theory was eliminated as observations revealed that abundances of nitrogen oxides were exceedingly low (*Fahey et al., 1989*), not high, within the Antarctic vortex. A

fact that supported the chlorofluorocarbon theory.

The dynamical theory rested on the possibility of solar heating in an extremely cold environment such as the Antarctic lower stratosphere, leading to net radiative heating. Since heated air rises, ozone poor air would be transported upwards from the troposphere, decreasing the total column ozone. However this theory was inconsistent with observations of long lived tracers e.g. N_2O during the AAOE (*Podolske et al., 1989*). In fact, tracer observations suggest that the meridional flow pattern in spring over the Antarctic is characterized by downward motion.

Following the discovery of the ozone hole, *Solomon et al. (1986)* suggested that heterogeneous reactions (e.g. equation (2.3)) could convert inactive, reservoir species such as HCl and $ClONO_2$ into reactive chlorine species, hereafter referred to as ClO_x. In the presence of sunlight, which returns over the austral pole in September, these species are photolyzed to atomic chlorine. The chlorine can then deplete ozone by a gas phase mechanism. Chlorine reservoirs cannot react with ozone. Laboratory studies have shown that reaction (2.3) proceeds rapidly on ice surfaces (polar stratospheric clouds) but not in the gas phase (*Molina et al., 1987*). The net effect of this reaction is to release chlorine, at an altitude of maximum ozone abundance, from normally very stable hydrogen chlorine. Laboratory studies (*Molina, 1988*) also indicate that when chlorine nitrate is deposited on an ice crystal, previously doped with hydrogen chloride, nitric acid is produced that remains in the solid phase within the ice. Therefore another consequence of reaction (2.3) is to scavenge nitrogen from the air, producing inactive HNO_3 from reactive NO and NO_2 (NO_x). Other important nitrogen reservoir species are N_2O_5 and $ClONO_2$. These heterogeneous processes have the important effect of altering the altitude range over which chlorine can deplete ozone from altitudes where ozone is less concentrated (35-45 km) to lower, ozone rich altitudes yielding far larger column depletions.

Observers in the polar regions have recorded the appearance of clouds that, when viewed from the ground, glow with a seashell iridescence. These nacreous clouds extend 10 to 100 km in length and several kilometers thick at an altitude of 20 km. Another type of cloud, which consists of nitric acid trihydrate (NAT) instead of pure water, was detected by satellite. A third type is identical to nacreous clouds in chemical composition but forms in a slow cooling process that results in a larger cloud with no iridescence. When these three types of cloud form over the poles, they are broadly referred to as polar stratospheric clouds (PSC's).

At first glance, the correlation between ozone depletion and PSC's seemed unlikely since clouds in the stratosphere were thought to be uncommon. Water vapour constitutes only a few parts per million in the stratosphere (*Toon and Turco, 1991*). This extreme dryness causes temperatures to fall below 190 K (-83°C) and water ice to form.

Nacreous clouds are formed by sudden cooling and condensation, on aerosols, of water vapour downwind from mountains. Here, the ascending portion of a lee wave rapidly expands and cools. Under stable conditions this standing wave pattern can propagate into the stratosphere. Since the mountains create standing waves, the clouds remain stationary, and air rushes through them. Ice crystals collect water from these air currents, and grow to about 2 μm in size.

The Stratospheric Aerosol Measurement (SAM) II instrument aboard the Nimbus 7 satellite detects particles in the air by examining sunlight as it grazes the limb of the earth. SAM II showed that PSC's are a regular feature of the austral winter polar stratosphere (*McCormick and Trepte, 1986*) and that they existed at relatively warm temperatures of 195 K (-78°C) (*Crutzen and Arnold, 1986*). *Toon et al. (1986)* deduced that the clouds may serve as a nitrogen sink in the form of NAT. Such a compound not only accounts for nitrogen removal but condenses at temperatures higher than pure water. The slow cooling process of the polar stratosphere, radiating heat during the polar night, causes the temperature to drop below 195 K and is thought to be responsible for the formation of NAT clouds. Sulphuric acid particles (0.1 μm in size) serve as seeds onto which water and NAT can condense. These sulphur gases are produced by natural processes (e.g. volcanos) or anthropogenic sources. The slow cooling process produces geographically extensive NAT PSC's. *Browell et al. (1990)* used aircraft borne lidars during the AASE to map individual clouds. They were found to be many thousand kilometers in length and relatively thin ($\sim$ 1 km). As they are so tenuous they are very difficult to observe with the naked eye.

If the Antarctic stratosphere slowly drops to 190 K another type of PSC can form. These clouds have pure water ice coating NAT which in turn has condensed on seeds. Since they contain water ice these PSC's are classified with nacreous clouds but are distinguished by their rate of formation. Slowly cooling clouds have fewer, but larger crystals that can exceed 10 μm in size (*Toon and Turco, 1991*). These clouds do not reflect sunlight as well as their nacreous cousins, the latter have many tiny particles. All types of PSC's are key components of Antarctic ozone depletion. In fact, cloud particles containing HNO_3 may attain sizes sufficient for sedimentation (*Toon et al., 1986*) implying that reactive nitrogen can be removed from the stratosphere altogether.

In the sunlight of Antarctic spring molecular chlorine quickly dissociates into highly reactive atomic chlorine, precipitating the ClO-ClO catalytic cycle

$$Cl + O_3 \rightarrow ClO + O_2 \tag{2.6}$$

$$ClO + O \rightarrow Cl + O_2. \tag{2.7}$$

Often chlorine monoxide reacts with itself to form the dimer Cl_2O_2 (*Molina and Molina, 1987*). This is readily dissociated by sunlight freeing chlorine atoms, leading to further

ozone depletion. If odd nitrogen were present, it would quickly combine with chlorine monoxide to trap the chlorine with the inert reservoir molecule chlorine nitrate, thus halting the ClO-ClO catalytic cycle (*Toon and Turco, 1991*). However, the PSC's prevent the reaction because they convert any nitrogen present into nitric acid.

Anderson (1988) and *Anderson et al.(1989,a)* have found extremely high levels of ClO at both northern and southern polar latitudes. Chlorine radical abundances of up to 1 ppbv have been measured in Antarctic spring (*Solomon, 1990*).

Reactive nitrogen (NO_x) plays an essential role in chlorine chemistry due to the formation of $ClONO_2$. The abundance of NO_x controls the length of time during which chlorine remains available to destroy ozone and thereby plays a critical part in determining the extent and duration of the ozone hole (*Solomon, 1990*). Once chlorine has been freed by heterogeneous processes and sunlight returns to the polar cap two mechanisms are thought to be responsible for ozone loss. They involve the chlorine dimer

$$2 \times (Cl + O_3 \rightarrow ClO + O_2) \tag{2.8}$$

$$ClO + ClO + M \rightarrow Cl_2O_2 + M \tag{2.9}$$

$$Cl_2O_2 + h\nu \rightarrow Cl + ClO_2 \tag{2.10}$$

$$ClO_2 + M \rightarrow Cl + O_2 + M \tag{2.11}$$

$$\text{Net}: 2O_3 + h\nu \rightarrow 3O_2 \tag{2.12}$$

and the ClO-BrO cycle (*McElroy et al., 1986*)

$$Cl + O_3 \rightarrow ClO + O_2 \tag{2.13}$$

$$Br + O_3 \rightarrow BrO + O_2 \tag{2.14}$$

$$ClO + BrO \rightarrow Br + ClO_2 \tag{2.15}$$

$$ClO_2 + M \rightarrow Cl + O_2 + M \tag{2.16}$$

$$\text{Net}: 2O_3 \rightarrow 3O_2. \tag{2.17}$$

The ClO-BrO cycle accounts for perhaps 20% of ozone depletion (*Anderson et al., 1989, b*). The rate limiting step in both cycles is combination of two radicals (ClO;ClO and ClO;BrO) and so we can write

$$\frac{d[O_3]}{dt} = -2k_1[ClO][ClO][M] + k_2[ClO][BrO]. \tag{2.18}$$

The rate of ozone loss is therefore governed, in principle, by the abundances of ClO and BrO. Chlorine monoxide is measured by either ground-based microwave emission or a resonance fluorescence technique conducted from aircraft.

Measurements of column abundances of HCl and $ClONO_2$ reveal greatly diminished values in the polar vortex (*Toon et al., 1989*). This supports the notion that elevated abundances of ClO are due to conversion from less reactive compounds. Measurements of reactive nitrogen in both the Antarctic (e.g. *Keys and Johnston, 1986*) and the Arctic (*Fahey et al., 1990*) spring, reveal suppressed values.

In winter, polar winds are generally from west to east, due to the polar vortices, at average wind speed of tens of metres per second. This translates into approximately one week travel time for an air parcel to complete a circumpolar path. As previously discussed, during late winter and spring flow is not restricted along lines of latitude, and atmospheric waves disturb the polar vortex causing substantial north-south excursions.

Large scale vertical motion within the polar vortex must be considered. *Tuck (1989)* provided evidence of strong descent in the Antarctic vortex by monitoring the ratio of tracers $CFCl_3/N_2O$. On the other hand *Proffitt et al. (1989,b)* measured the abundance of reactive species, ClO, to deduce vertical transport, and again concluded downward flow that was particularly rapid at the vortex boundary. Downward motion could import other species (e.g. NO_x, ClO_x) from high altitudes which may also affect the rate of ozone destruction.

Vertical profiles of calculated photochemical ozone-loss rates are similar to those derived from limited observations (*Solomon, 1990*). This supports the view that heterogeneous chlorine chemistry is responsible for much, if not all, of the ozone decline over Antarctica in spring, i.e. the ozone hole. Uncertainties do remain, in particular, a more extensive knowledge of ClO abundances, kinetic rate constants and physical processes is needed.

In the northern hemisphere, the largest ozone reductions are observed around 65°N, instead of polar latitudes. The cause of this depletion is a subject of ongoing debate. Some researchers implicate a 'leaky' north polar vortex, allowing air that has been exposed to heterogeneous chlorine chemistry to spread to mid-latitudes. Others point to new chlorine chemistry in the mid-latitude region. *Toumi et al. (1993)* have suggested a mechanism whereby $ClONO_2$ can destroy ozone in the following catalytic cycle

$$ClONO_2 + h\nu \rightarrow Cl + NO_3 \quad (2.19)$$

$$NO_3 + h\nu \rightarrow NO + O_2 \quad (2.20)$$

$$Cl + O_3 \rightarrow ClO + O_2 \quad (2.21)$$

$$NO + O_3 \rightarrow NO_2 + O_2 \quad (2.22)$$

$$ClO + NO_2 + M \rightarrow ClONO_2 + M \quad (2.23)$$

$$\text{Net} : 2O_3 \rightarrow 3O_2 \quad (2.24)$$

2.6 Interhemispheric differences

Previous sections have pointed to several disparities between ozone concentrations over the northern polar regions and those of the south. Some of these interhemispheric differences are now briefly outlined.

The requisite cold stratospheric temperatures are generally neither as widespread nor as long lasting in the Arctic spring season as that of the Antarctic. On average, the austral polar stratosphere is up to 20 K colder than its northern counterpart. However, Arctic stratospheric temperatures can be comparable over limited areas and altitudes (*Nagatani et al., 1990*). Lower temperatures in the Antarctic may be attributed, in part, to the longevity and strength of the Antarctic vortex which remains intact throughout winter, sometimes into November. The Arctic vortex is less stable, often being disintegrated by SSW's, allowing influxes of ozone rich air from mid-latitudes. It can be completely broken down by these events as early as the end of February. A second factor is the feedback mechanism whereby areas of depleted ozone result in a colder stratosphere (*Shine, 1986*) which enhances the stability of the polar vortex. This leads to increased occurrences of PSC's late in the season when 'normal' ozone levels would imply temperatures that are too warm for PSC's.

2.7 The chemistry and dynamics of odd nitrogen

The study of odd nitrogen in the middle atmosphere is naturally divided between those processes occurring in the stratosphere and those appropriate in the lower thermosphere and mesosphere.

2.7.1 The stratosphere

In the stratosphere nitric oxide (NO) is produced mostly from dissociation of N_2O by reaction with an excited O (^1D) atom which is generated by photolysis at wavelengths

shorter than 310 nm (*Nicolet, 1971*) viz.

$$N_2O + O(^1D) \rightarrow 2NO \tag{2.25}$$

Nitric oxide production is therefore dependent on the vertical distribution of N_2O, which has been observed by the SAM satellite. Nitrous oxide is produced at ground level principally by bacterial processes associated with nitrification mechanisms in soils, fossil fuel and biomass burning (*Gribbin, 1988*). More recently, industrial fixation has become increasingly important, especially from sources such as fertilizers (*Gribbin, 1988*). The global budget of atmospheric N_2O is still not accurately known, but estimates lead to a chemical lifetime of 100 years in the troposphere and lower stratosphere, when only photodissociation,

$$N_2O + h\nu(\lambda < 252 \text{ nm}) \rightarrow N_2 + O(^1D), \tag{2.26}$$

is considered as the destruction mechanism. Due to the long chemical lifetime of nitrous oxide, which is comparable to that of meridional circulation, it is an excellent tracer for transport in the middle atmosphere.

Figure 2.11 shows the profile of nitric oxide production by oxidation of N_2O (*Nicolet and Peetermans, 1972*). Curves are shown for two solar zenith angles, two $O(^1D)$ distributions, and two values of the vertical eddy diffusion coefficient (k_{min} and k_{max}). It should be noted that the NO production is directly related to atmospheric dynamics; it reaches a maximum in the mid-stratosphere where it has a value of about 100 molecules cm^{-3} s^{-1}. The annual production of odd nitrogen is a result of N_2O oxidation and has been computed by *Jackman et al. (1987)* to be 2.6 $\times 10^{34}$ NO_y molecules per year.

Stratospheric nitrogen atoms can also be formed by dissociative ionization and dissociation of molecular nitrogen by galactic and solar cosmic rays. These processes generally occur through secondary electrons ejected by heavier cosmic particles. One ion pair formed by cosmic radiation leads to the production of ~1.5 atoms of nitrogen (see sub-section 2.8.2). In the stratosphere, atomic nitrogen reacts rapidly with molecular oxygen to form nitric oxide. Stratospheric production of nitrogen oxides by cosmic radiation should not be neglected, especially in polar regions. *Nicolet (1975)* indicated that background cosmic radiation leads to an integrated NO column production of 5 $\pm 1 \times 10^7$ cm^{-2} s^{-1} in polar regions. *Crutzen et al. (1975)* noted that the penetration of large amounts of protons into the stratosphere during solar proton events would also lead to intense production of atomic nitrogen. The chemistry of particle precipitation and its subsequent effect on ozone will be discussed more fully in sub-section 2.8.2.

Much of the available information on latitudinal distribution of odd nitrogen comes from total column measurements e.g. *Coffey et al. (1981)*. These observations have shown that odd nitrogen column abundance increases with increasing latitude in summer, while the winter column abundance decreases poleward of about 50°N. Gradients

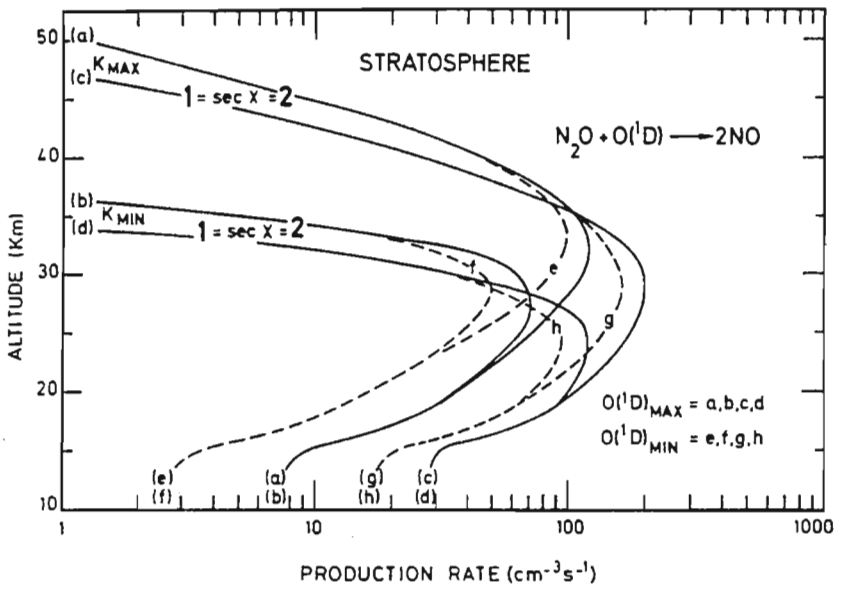

Figure 2.11: Calculated production rates of NO in the stratosphere due to the reaction of N_2O with $O^1(D)$. From *Brasseur and Solomon (1986).*

at these latitudes are often enhanced by wave breaking events which erode the vortex, and are sometimes referred to as a 'cliff'.

The tendency for NO and NO_2 to increase at high latitudes in summer can be quantitatively understood in terms of transport. Like ozone, NO_x (whose photochemical lifetime is represented by the curve marked τ_{NO_x} in Figure 2.12) is a long lived species in the lower stratosphere and is sourced in the tropical middle stratosphere. Downward, poleward transport by the meridional circulation should lead to increased total NO_x abundance at high latitudes.

The observed decrease in the winter season at high latitude is a result of the combined effects of chemistry and dynamics. A study by *Solomon and Garcia, (1983)* have shown that N_2O_5 probably provides an important reservoir for NO_2 in the winter hemisphere. Since N_2O_5 absorption cross-section is highly temperature dependent, its chemical lifetime during the polar night is very long whereas the time scale for the conversion of NO_2 and NO_3 to N_2O_5 is short. Therefore, nearly all odd nitrogen in an air parcel remaining in the polar night will be in the form of N_2O_5. The lifetime of N_2O_5 in the sunlit high latitude wintertime is a few days so that during planetary wave activity N_2O_5 can move from latitudes where its chemical lifetime is long to those where its lifetime is short. So, chemical factors can also be important in the development of the odd nitrogen cliff. Nitric acid constitutes an important reservoir for odd nitrogen, mainly in the polar, lower stratosphere. It is known to play a central role in PSC formation, as discussed in section 2.5.

The impact of high speed civil transport on nitrogen concentrations in the stratosphere has been cause for concern. Odd nitrogen in the exhaust gases could lead to ozone depletion via the homogeneous catalytic mechanism, an important consideration if there were enough aircraft. *Peter et al. (1991)* and *Weisenstein et al. (1991)* have modelled the effect of increased odd nitrogen abundance over polar regions, arguing that it could lead to a higher incidence of NAT PSC formation in winter, leading to potential greater ozone loss by chlorine radicals. It is important to be able to quantify the effect of all increases in atmospheric species that could affect the ozone hole although, because of the low density of high speed aircraft, they cannot be considered as a significant source.

2.7.2 Lower thermosphere and mesosphere

In the thermosphere, atomic nitrogen can be formed by the dissociation of N_2, either through the effects of solar radiation, or energetic particles. Atomic nitrogen is produced either in the ground state (^4S) or the excited state (^2D) where it often forms

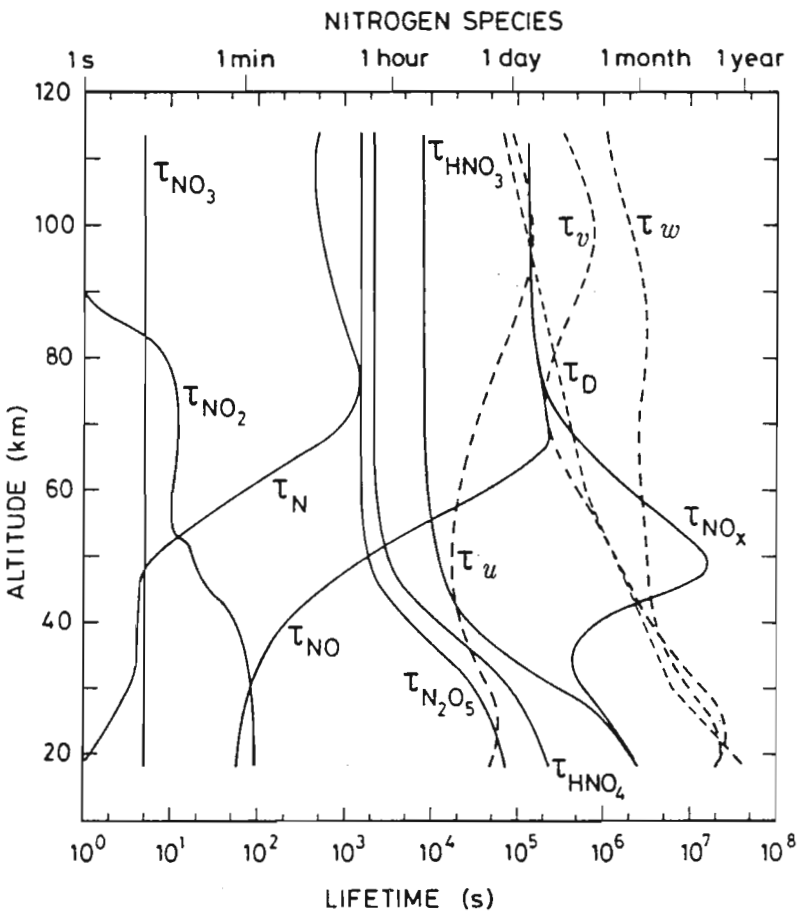

Figure 2.12: Photochemical time constants for NO_x, N, NO, NO_2, NO_3, N_2O_5, HO_2NO_2 and HNO_3, together with transport lifetimes in the zonal (u), meridional (v) and vertical (w) directions and the one-dimensional eddy diffusion coefficient. Adapted from *Brasseur and Solomon (1986).*

nitric oxide

$$N(^2D) + O_2 \rightarrow NO + O \tag{2.27}$$

$$N(^4S) + O_2 \rightarrow NO + O. \tag{2.28}$$

The reaction involving nitrogen in the excited state is extremely rapid. Ground state atomic nitrogen reacts more slowly with oxygen and the reaction is highly temperature dependent.

Figure 2.13 shows a theoretical calculation of the distribution of NO in the mesosphere and lower thermosphere. Variations with respect to latitude reflect changes in mean solar zenith angle as well as spatial variability in transport processes. Attention is drawn to the large NO densities in the high latitude winter stratosphere, particularly near polar night, where there is virtually no photochemical loss of odd nitrogen by

$$NO + h\nu \rightarrow N(^4S) + O. \tag{2.29}$$

Observations of NO in the mesosphere and lower thermosphere show high variability, which indicates strong sensitivity of NO to transport. This is also predicted in Figure 2.12 where the photochemical lifetime of nitric oxide (τ_{NO}) is shown to be approximately equal to the dynamical time constants in this altitude range. As a result of this, models predict that NO produced at thermospheric altitudes can be transported down to the mesosphere and stratosphere a these latitudes (e.g. *Frederick and Orsini, 1982; Solomon and Garcia, 1984; Brasseur, 1982*).

2.8 Possible mechanisms for a solar influence on ozone

2.8.1 Introduction

The following types of variations in solar activity can influence the middle atmosphere:

- short term solar variation: solar flares
- intermediate solar variation: active regions on the sun and solar rotation (27 days)
- long term solar variation: 11 year sunspot cycle.

The first suggestions of a connection between the middle atmosphere and solar activity were based on observations of ozone (e.g. *Willett, 1962; Paetzold, 1973; Angell and Ko-*

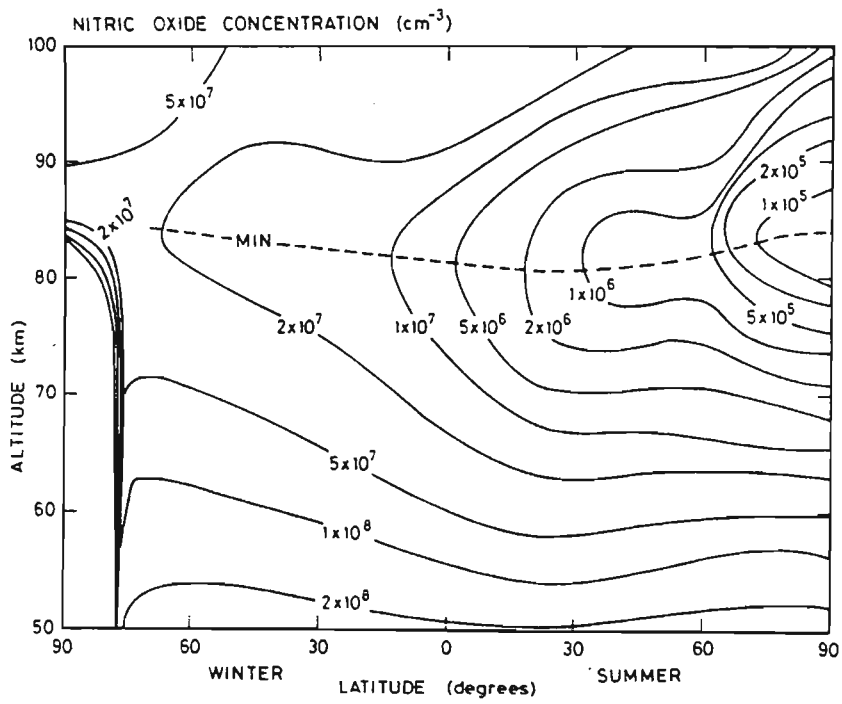

Figure 2.13: Calculated (*Brasseur, 1982*) two-dimensional NO distribution. From *Brasseur and Solomon (1986).*

rshover, 1978), temperature (*Zerefos and Crutzen, 1975*) and atmospheric circulation (*Hines, 1974*).

The polar middle atmosphere's response to short term solar variation will be more fully addressed in sub-section 2.8.2. Large solar flares release populations of energetic particles, mainly protons and electrons. Protons precipitate into the polar atmospheres (during a solar proton event) and result in large scale ionization of the atmosphere, sometimes down to 20 km in altitude. This leads to increased abundances of NO_x and HO_x radicals. Both HO_x and NO_x destroy ozone by the same catalytic mechanism i.e. NO can be replaced by HO in the following set of reactions from *(Crutzen, 1970)*

$$NO + O_3 \rightarrow NO_2 + O_2 \tag{2.30}$$

$$NO_2 + O \rightarrow NO + O_2. \tag{2.31}$$

Therefore, ozone depletions can be a signature of solar proton events (SPE's).

Variations of the solar ultraviolet flux, with the 27 day rotation period of the sun, have been observed by *Rottman (1983)*. The 27 day solar variation has a much weaker effect on ozone, it has not yet been detected by total ozone measurements. However, this 27 day periodicity is detectable in vertical profile observations (*Brasseur et al., 1987*). The temperature response has been measured by *Keckhut and Chanin (1992)*.

Long term solar variation effects on the polar middle atmosphere are examined in detail in sub-section 2.8.3. Three mechanisms by which odd nitrogen levels are substantially increased, leading to decreases in ozone, during the years near the solar maximum must be considered. Firstly, the gradual accumulation of the nitrogen radical densities as a result of repeated solar proton events over the solar maximum. The chemical lifetime of odd hydrogen below 85 km is never more than a few hours and so its effects on ozone can be discounted here. Secondly, during periods following high levels of solar activity, energetic electrons precipitate from the plasma sheet resulting in spectacular auroral displays over polar regions. These electrons are also responsible for the production of odd nitrogen, although, because of their relatively low energy, they are only effective in the thermosphere. However, during the polar night these species are transported downwards by mean meridional circulation and, with the return of solar irradiation over the polar caps in spring, are able to catalytically destroy ozone. Thirdly, solar activity enhances the levels of solar ultraviolet radiation. Radiation at these short wavelengths is very effective in ionizing and dissociating atmospheric species. The penetration of solar radiation into the earth's atmosphere depends on the absorption of each constituent in the atmosphere. Since the absorption coefficients of the constituents are functions of wavelength, the penetration depth is dependent on the shape of the spectrum which is sketched in Figure 2.14. However, radiation with wavelengths < 100 nm is almost completely absorbed above 100 km by N_2, O_2 and O (*Brasseur*

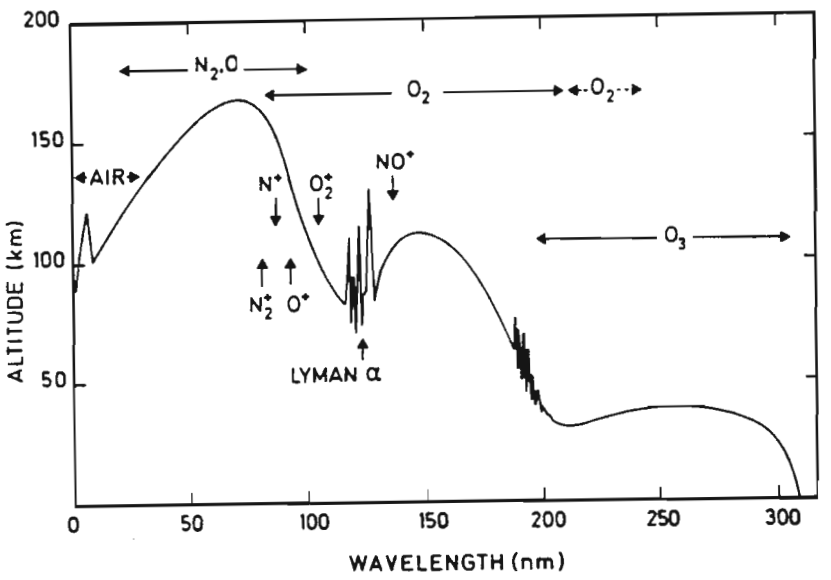

Figure 2.14: The optical depth of solar radiation as a function of wavelength. The principal absorbers and ionization limits are indicated. From *Brasseur and Solomon (1986).*

and Solomon, 1986). Solar ultraviolet, at wavelengths > 100 nm, can photodissociate atmospheric molecules, particularly O_2 and O_3. Additional ultraviolet flux therefore results in the production of additional odd nitrogen and oxygen species.

Despite the complexity of solar-ozone relations, following the NASA/WMO International Ozone Trends Panel, it is now accepted that the solar activity cycle is manifested in total ozone with an amplitude of 1.5 matm cm. or DU, per 150 sunspot number (*NASA/WMO, 1988*). It should be noted that the Ozone Trends Panel correction for the 11 year solar cycle is only for ultraviolet forcing.

2.8.2 Particle precipitation

Theory

Figure 2.15 shows a schematic noon-midnight meridional cross-section of the magnetosphere, confined by the magnetosheath. This cavity exhibits a compressed sunward side and an extensive anti-sunward side, the latter demarcating the magnetotail. This representation is believed to be sustained during 'steady state' magnetic conditions. However, an enhancement of solar activity results in the reduction of the distance to the magnetopause on the dayside from 10 $R_\ominus$ to 6 $R_\oplus$. Furthermore, the plasmasphere is compressed and open field lines must then extend to lower latitudes, distorting the dipole shape of the geomagnetic field associated with quiet magnetic conditions. Open magnetic field lines (with one 'foot' on the ground) occur over the polar regions. Charged particles move easily along magnetic field lines so that solar and galactic particles can gain easy access to the atmosphere at polar latitudes, as marked by the open arrows in Figure 2.15. To reach the equator a proton has to cross field lines right down to the atmosphere. A particle of sufficient energy could do this but, in practice equatorial regions are forbidden to protons of solar origin. In the auroral oval ($75° > \phi > 70°$) low energy electrons ($\sim$ keV) and to a lesser extent protons, precipitate from the magnetospheric plasma sheet to about 100 km altitude, indicated by black arrows in Figure 2.15. The resulting photoelectron impact, mostly on N_2, results in excited N atoms. Their decay to ground state produces green light characteristic of the aurora borealis and australis. Radiation belts are located between the plasma sheet and plasmasphere. The origin of these trapped particles, whether it be the ionosphere or magnetosphere, is uncertain.

Charged particles possessing more energy (many MeV) can penetrate deeper into the earth's atmosphere, losing energy while ionizing atmospheric species. The depth of penetration of charged particles depends on their mass and energy. Figure 2.16 depicts the approximate depth of penetration of protons and electrons and their cut off latitude.

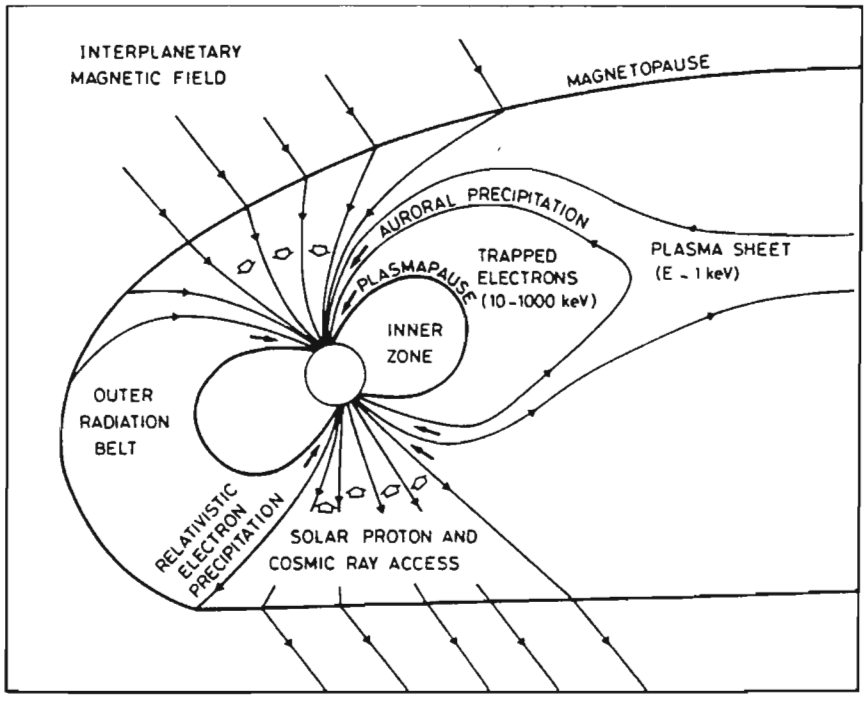

Figure 2.15: Schematic representation of the earth's magnetosphere. The principal particle populations are marked. From *Brasseur and Solomon (1986)*.

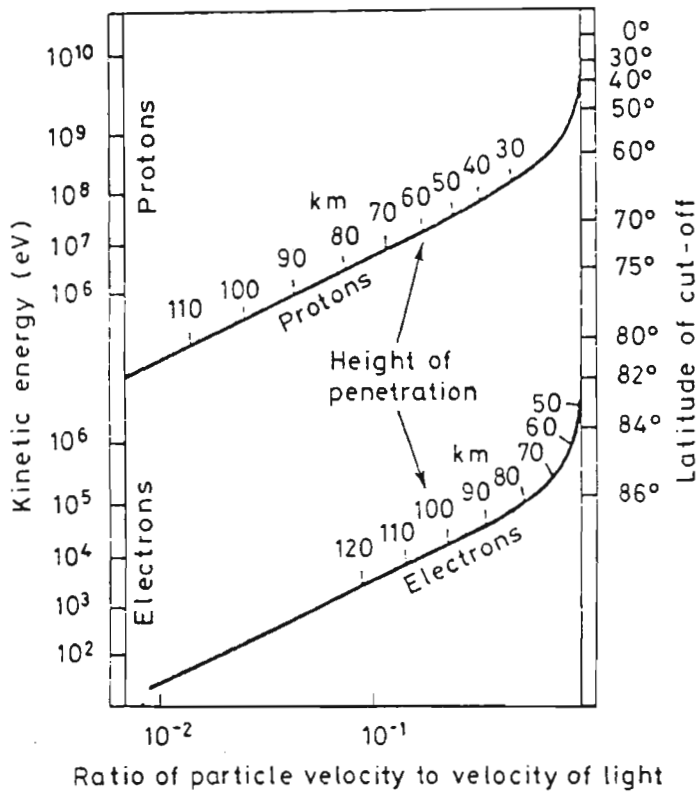

Figure 2.16: Penetration of protons and electrons into the earth's atmosphere as a function of particle energy (left axis) and cut-off latitude (right axis). From *Hargreaves (1979)*.

Figure 2.16 indicates that protons must have to have at least 10 MeV to penetrate the upper stratosphere, whereas only several keV is required for electrons to precipitate in the thermosphere resulting in auroral optical emissions.

On 23 February 1956 a large solar flare (3+) occurred, and from observations the solar proton event (SPE) was defined (*Hargreaves, 1979*). The flare was followed by radio blackouts that lasted for days, and large increases in the readings of cosmic ray monitors. By studying the effects on VHF forward-scatter radio links, D.K. Bailey showed that the blackout conditions could be attributed to additional ionization in the ionospheric D-region (*Hargreaves, 1979*). The additional ionization was produced by the arrival of energetic protons, which had been released from the sun at the time of the flare. These events can produce intense ionization and alter the composition of the neutral atmosphere. The ionization rate from these events can change its normal value of 10 cm^{-3} s^{-1} to 10^4 or 10^5 cm^{-3} s^{-1}. Figure 2.17 presents the vertical distribution of ion pair production rates associated with the solar proton events of July 1959, August 1972 and July 1982, showing that their ionization potential is much greater than that of galactic cosmic rays. The ionization of atmospheric constituents by galactic cosmic rays provides the dominant source of ions in the lower mesosphere, stratosphere and troposphere. During periods of high solar activity, galactic cosmic rays tend to be excluded by strong solar winds. Thus, although precipitation due to solar cosmic rays may be high during solar activity, galactic cosmic rays tend to be depressed, termed the Forbush decrease.

Most early studies of proton events were made using riometers, a device that measures ionospheric absorption of cosmic radio noise. Riometer studies showed that the radio absorption is confined to high latitudes. For this reason, solar proton events are often called polar cap absorption events or PCA's. In spite of the name 'solar proton event', it should be appreciated that α particles and heavier nuclei are also emitted by flares, the proportions represent the composition of the solar atmosphere. The frequency of occurrence of SPE's depend on the phase of the sunspot cycle. There might as many as 10 in an active year or none at all in a quiet one.

High levels of solar activity have serious consequences, often resulting in disruption of radio links and aircraft communication systems. In addition, solar protons degrade many elements onboard satellites and can be a potential health hazard to astronauts.

The source of radio bursts, which accompany solar flares, is synchroton emission from energetic electrons. Assuming that the burst and the proton cloud are released at the same time, the time of onset of the radio burst may be taken as the time the protons left the sun. The delay between the radio burst and the start of a SPE can be between 1 and 6 hours. The propagation of a cloud of protons will be considered in two parts, the path from the sun to the terrestrial magnetosphere, then the passage through the

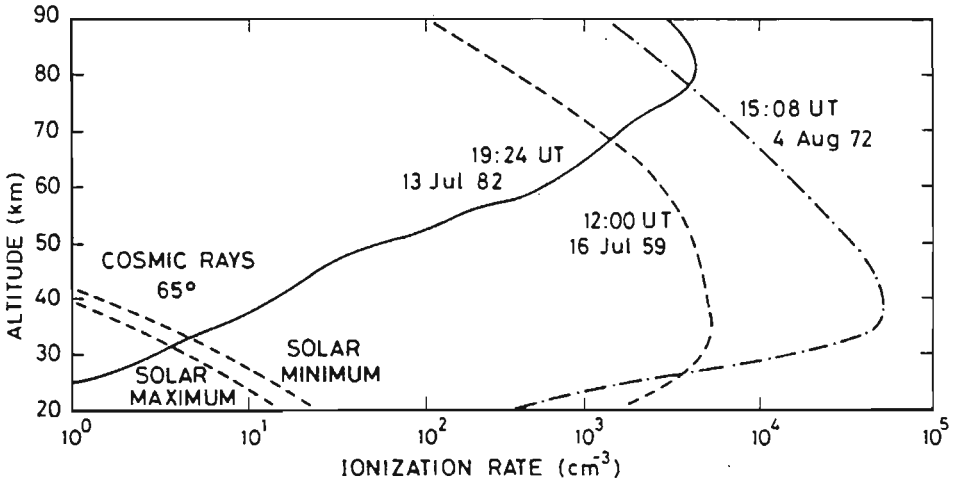

Figure 2.17: Calculated ionization rates during SPE's of July 1959, August 1972 and July 1982. A plot of ionization rates due to cosmic rays is included. From *Brasseur and Solomon (1986).*

geomagnetic field into the atmosphere.

A typical solar flare lasts for tens of minutes and the subsequent SPE can last for several days. A proton of 10 MeV travels at $\sim 4 \times 10^4$ km s^{-1} and would reach the earth in about 1 hour if it travelled in a straight line. More energetic particles should arrive sooner. In fact, most SPE events are delayed several hours after the relevant flare, implying that the propagation from the sun to the earth is not direct. The obvious explanation is that the proton cloud is 'frozen into' the magnetic field, whose archimedes spiral form would produce just the observed effect. The gyro-radius (r) of the protons around magnetic field lines must be considered

$$r = \frac{m v}{B e}, \tag{2.32}$$

where m is the mass, v the velocity and e the charge of the particle. B is the interplanetary magnetic field. Assuming a typical value of 5 γ for the interplanetary magnetic field, even a 1 GeV proton would complete over 100 gyro-radii from the sun to the earth. Protons are of course sensitive, not only to the general shape of the magnetic field, but to any irregularities or kinks in the field. The propagation of protons through interplanetary space is more like diffusion, for which the bulk velocity of the cloud is much less than that of the individual particles within the cloud. It is therefore possible to account for the delay and duration of the proton event and for the observed isotropy of the particles near the earth.

During high levels of solar activity, the interplanetary magnetic field (IMF) is often observed to be southward pointing, it is northward pointing under quiet solar conditions. Figure 2.18 depicts the geomagnetic field in polar section added to an IMF directed northward (a) and southward (b). The solar wind is incident from the sun to the left of the diagram. In the case of (b), neutral points are formed in the equatorial plane allowing for some IMF lines to be connected to lines of the geomagnetic field and thus allowing easy access of charged particle trajectories. The magnetosphere is termed 'open' in this case. In case (a) there is no connection between the two fields and the magnetosphere is 'closed'.

Direct observations of particles show that the cut-off rigidity (defined in sub-section 1.4.2) at the edge of the polar cap is significantly less than the Stormer value. Two phenomena that can reduce the cut-off are an enhanced ring current and the distortion of the magnetosphere from a simple dipole form. These two factors are known to be important during periods of high solar activity.

High energy protons produce substantial fluxes of secondary electrons (energies 10-100 eV) which are responsible for a considerable fraction of the energy transfer from the primary particle to the atmosphere. Knowledge of both primary and secondary electron fluxes as a function of energy (the particle spectrum) together with their cross-sections

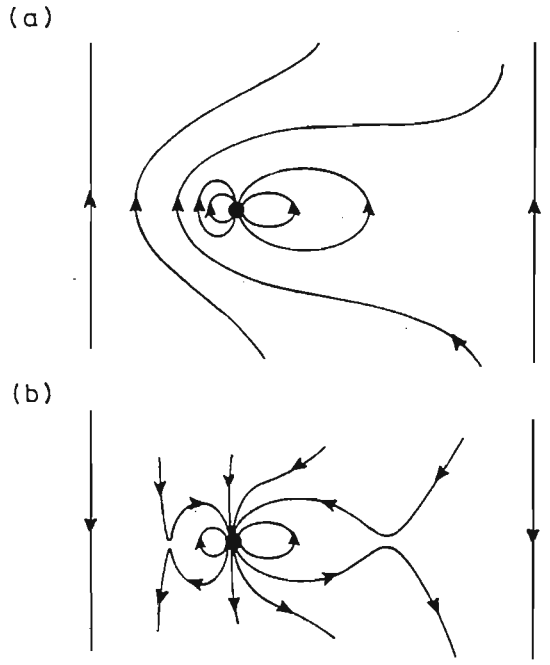

Figure 2.18: Terrestrial and solar wind magnetic fields in noon-midnight cross-section. In (a) the IMF is northward pointing, resulting in a closed magnetosphere. In (b) the IMF is southward pointing, resulting in an open magnetosphere. From *Hargreaves (1979)*.

for interaction with the principle atmospheric species, is thus an important part of evaluating the impact of particle precipitation on the atmosphere.

When an energetic proton enters the atmosphere it loses energy in collisions with the atmospheric constituents and leaves an ionized trail. The rate of energy loss from a proton travelling through air is well known from laboratory measurements, a graph of loss rate against distance travelled is known as a Bragg curve. If a proton event contains particles of 1-100 MeV we expect to see atmospheric effects in the altitude range 90-35 km (Figure 2.16). For protons in this energy range, the energy loss rate increases as the proton slows down because it spends a longer time in the vicinity of a given air molecule. Over the range 10-200 MeV, the loss rate is almost inversely proportional to the energy, a typical value for 100 MeV is 0.8 MeV m^{-1} of path of air at STP. The energy may be assumed to be used entirely in the creation of ion pairs, each requiring about 35 eV. For a proton that enters the atmosphere from space the air density also increases along the path, and therefore the ionization is very concentrated towards the end of the path in SPE events. A proton that enters the atmosphere vertically with an initial energy of 50 MeV loses only 10% of its energy down to 56 km altitude, but the last 10% is lost in only 100 m between 42.1 and 42.2 km. Half the initial energy is deposited in the final 2.5 km.

The calculation of the rate of ion pair production due to particles requires knowledge of the particle energy spectrum and the energy degradation rate as they pass through the atmosphere. If j(z,E) dE represents the flux of ionizing particles at altitude z of energies from E to E + dE (the differential flux j(E) has units cm^{-2} s^{-1} eV^{-1} $ster^{-1}$) and if $\frac{dE}{ds}$ is the energy loss per particle in an inelastic collision, then the ionization rate (Q) at altitude z is

$$Q(z) = \frac{\rho(z)}{\bar{u}} \int_{\omega} \int_{E} \frac{dE}{ds} j(z, E)\, dE\, d\omega \tag{2.33}$$

(*Dubach and Barker, 1971*) where $\bar{u}$ is the mean energy required per ion pair formation ($\sim$ 35eV), $\rho(z)$ is the air density (g cm^{-3}) and ω is the solid angle over which the equation is integrated. The ionization rate of molecular nitrogen, Q_{N_2}, may be determined from the rate of total ion pair production, Q, if it is assumed that the fraction of ionization is proportional to the mass of the target particle (*Rusch et al., 1981*)

$$Q_{N_2} = Q \frac{0.88[N_2]}{0.88[N_2] + [O_2]} = 0.77Q, \tag{2.34}$$

assuming O_2 is the only other ionized species. The ion chemistry is initiated by the impact of energetic secondary electrons, e^*, on the nitrogen molecule:

$$e^* + N_2 \rightarrow N_2^+ + 2e \tag{2.35}$$

$$e^* + N_2 \rightarrow N + N^+ + 2e \tag{2.36}$$

$$e^* + N_2 \rightarrow N + N + e. \tag{2.37}$$

Using equations (2.35) and (2.36) the ratio of the peak cross-section for simple ionization versus that for dissociative ionization is 0.76:0.24 for N_2 (*Rapp et al., 1965*). Accepting that the rate of production (P) is proportional to these cross-section ratios, we obtain the following approximate expressions (*Rusch et al., 1981*)

$$P_{N_2^+} = 0.76 \times 0.77Q = 0.585Q \tag{2.38}$$

$$P_{N}^{+} = 0.24 \times 0.77Q = 0.185Q. \tag{2.39}$$

During SPE's, the incoming particles may be of sufficiently high energy to produce appreciable amounts of ionization (and so NO_x and HO_x) in the stratosphere and mesosphere. The NO_x free radicals play an important role in the destruction of stratospheric ozone by the following catalytic mechanism

$$NO + O_3 \rightarrow NO_2 + O_2 \tag{2.40}$$

$$NO_2 + O \rightarrow NO + O_2. \tag{2.41}$$

The HO_x radicals dominate ozone destruction in the mesosphere by the same mechanism. The production of free radicals in particle precipitation events may be expected to influence the atmospheric ozone density if that production is large enough to perturb the naturally occurring abundances of these species.

The possibility that odd nitrogen would be produced by energetic particle precipitation seems to have been first presented by *Dalgarno (1971)*, but this was before the importance of odd nitrogen chemistry in the ozone balance (*Crutzen, 1970, Johnston, 1971*) was recognized. *Crutzen et al. (1975)* later suggested that SPE's could have important effects on atmospheric ozone. A few years later *Heath et al. (1977)* presented the first observations of this effect. They found that the intense, energetic SPE of August 1972 was accompanied by dramatic decreases in ozone abundances at high latitudes. The event was widely studied in other papers e.g. *Reagan et al. (1981); Solomon and Crutzen (1981)* and *Rusch et al. (1981).*

Depletion of ozone related to HO_x production during particle precipitation events was first observed by *Weeks et al. (1972)* during rocket flights. *Swider and Keneshea (1973)* later suggested that the ozone depletions were due to odd hydrogen production as a result of ion chemistry. Observations by the Solar Mesosphere Explorer (SME) satellite provided details of ozone depletion by HO_x in the mesosphere. These measurements were in good agreement with the theoretical model of *Jackman and McPeters (1985).*

Solar proton events in 1978 and 1979 have been modelled by a two-dimensional photochemical model (*Jackman and Meade, 1988*). *McPeters and Jackman (1985)* have investigated events during 1979, 1981 and 1982. Satellite observations following these

events reveal ozone depletion at high altitudes, attributed to increases in HO_x concentrations. *Shumilov et al. (1992)* report total column losses of 20-80 DU above the polar cap stations of Longyearbyen and Barentsbury (both at 78°N) after a SPE in May 1990. The total ozone was monitored by ground based M-124 filter ozonometers.

Changes in ozone concentration cause changes in heating rates and hence changes in temperatures in the stratosphere and mesosphere. The temperature variations will undoubtedly be damped by the ozone-temperature anticorrelation. However, *Kodama et al. (1992)* have investigated forty-three SPE's spanning four solar cycles for their influence on lower stratospheric temperatures. All SPE's had time integrated proton fluxes $> 10^7$ cm^{-2}. It was shown that 64% of these SPE's induce a sudden cooling of -2.4°C (on average) in the 20-30 km altitude interval. Thirteen SPE's had time integrated fluxes $> 10^8$ cm^{-2}, 69% of these resulted in a mean drop of -3.5°C. All temperature measurements were made at Syowa, Antarctica.

There are many more reports on SPE's in the literature. Two periods of intense SPE's, which have provoked much study, are chosen for a detailed review viz. August 1972 (two events) and the September, October and November events in 1989 (three events). This will enable the reader to assess the results of the case studies in chapter 4 in relation to the previous work on large SPE's. Although models can qualitatively reproduce photochemical processes for a SPE, they do not necessarily give the right quantitative changes in odd nitrogen and ozone i.e. there may be discrepancies between modelled and observed responses. In particular, the limitations of two-dimensional models must always be considered. To include dynamical components that might be effective during a SPE, a three-dimensional model would be required.

Modelling and observations of the August 1972 SPE

Crutzen et al. (1975) were the first to calculate the nitric oxide production during a SPE. They found that the production of nitric oxide during large solar proton events is comparable, and usually larger, than the total average annual production of NO by galactic cosmic rays. The mechanism by which energetic protons produce odd nitrogen is outlined here.

The primary particles (protons) are too energetic to be efficient in dissociating and ionizing molecular nitrogen, which requires 35 eV. Instead, they produce substantial fluxes of secondary electrons with energies of tens to hundreds of electron volts. These energetic electrons ionize and dissociate molecular nitrogen by equations (2.35), (2.36) and (2.37). Nitric oxide can then formed via the following reactions

$$N^+ + O_2 \rightarrow N + O_2^+ \tag{2.42}$$

$$N^+ + O_2 \rightarrow NO^+ + O \quad (2.43)$$

$$N + O_2 \rightarrow NO + O \quad (2.44)$$

$$N + O_3 \rightarrow NO + O_2. \quad (2.45)$$

A fraction of the N atoms will, however, react with NO (*Crutzen et al., 1975*)

$$N + NO \rightarrow N_2 + O. \quad (2.46)$$

Reaction (2.44) is slow and highly temperature dependent for ground state $N(^4S)$ atoms, but not for electronically excited N (^2D and ^2P) atoms. *Crutzen et al. (1975)* used two extreme assumptions, either total or no N atoms in excited states to model NO production. Later, other researchers used a more realistic ratio of excited to ground state nitrogen atoms e.g. 80:20 ratio by *Reid et al. (1991)*.

The calculation of ion-pair production rates requires the flux and spectrum of solar protons, as measured by satellite-borne particle detectors, and the standard range-energy relation for protons in air. Protons of at least 10 MeV are of prime importance as they are able to reach altitude of less than 60 km at vertical incidence. Where energies lie outside the detector range, usually 100 MeV, an exponential rigidity spectrum is assumed. This assumption can lead to errors when calculating ion-pair production rates at altitudes where ozone is most abundant. Proton flux data (29-100 MeV protons) showed two pulses, the largest on August 4 exceeding 10^4 protons cm^{-2} s^{-1} $ster^{-1}$, as measured by the Lockheed experiment aboard the polar orbiting 1971-089A satellite (*Reagan et al., 1981*). This translates to an integrated proton (> 30 MeV) flux of 8×10^9 cm^{-2} (*Kodama et al., 1992*) which resulted in calculated ionization rates as high as 6×10^4 cm^{-3} s^{-1} at 60 km and 3×10^4 between 40 and 30 km (*Crutzen et al., 1975*). *Jackman et al. (1990)* conclude from work on numerous SPE's that NO_y abundances can only be affected where ion pair production exceeds 100 pairs cm^{-3} s^{-1}. With the knowledge of ionization rates and electron impact cross-sections of N_2 (*Rapp et al., 1965*), it is possible to calculate the amount of NO produced at high latitudes (> 60°) during SPE events. The NO production profiles, as calculated by *Crutzen et al. (1975)*, of the August 1972, as well as the November 1960 and September 1966, events are shown in Figure 2.19. Curves marked $P_N = 1$, indicate that all nitrogen atoms are presumed to be in excited states, while $P_N = 0$ assume ground state. Included in Figure 2.19 are estimates of the minimum and maximum annual production by galactic cosmic rays. The assumed background concentration of NO_x ($NO+NO_2$), upper scale in Figure 2.19, has its main source in the oxidation of N_2O (described in sub-section 2.7.1). It can be seen from Figure 2.19 that the NO_x produced during a SPE, above 30 km, can be up to 10 times that due to the annual production by galactic cosmic rays. Even compared to the background source, SPE's are important.

The first observations of an ozone decrease in the stratosphere associated with a solar

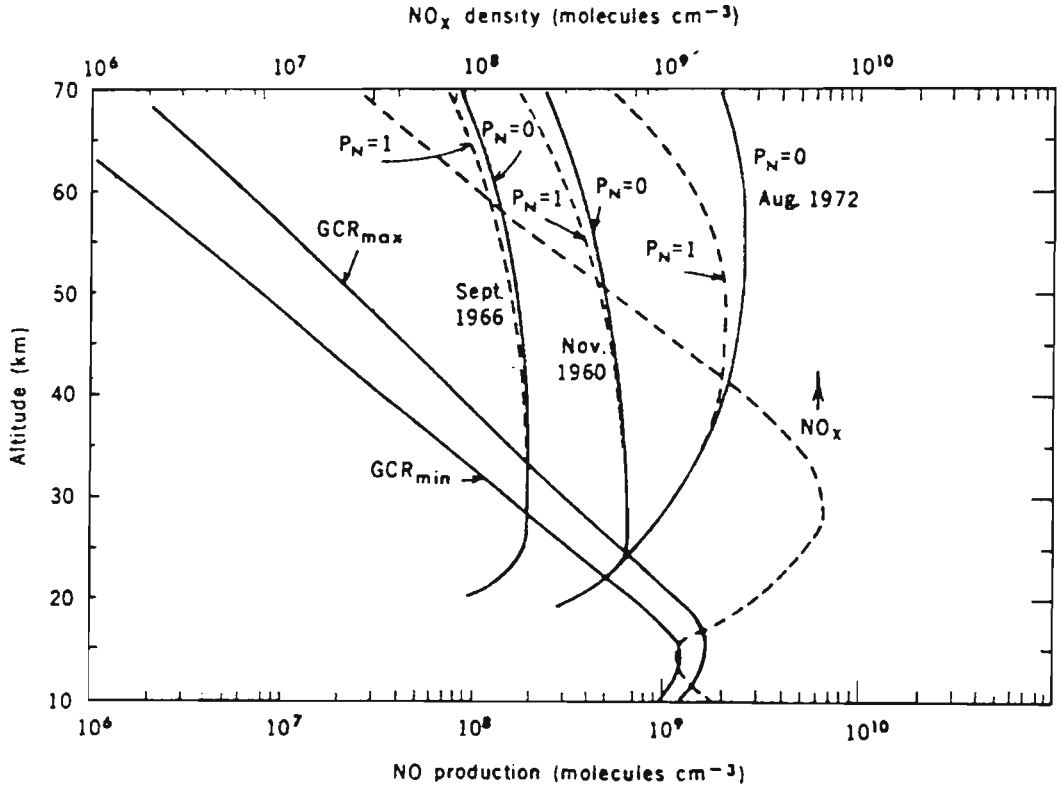

Figure 2.19: Production profiles of NO following the SPE's of November 1960, September 1966 and August 1972. The curves labelled P_N=1 represent the assumption that all N atoms were in the excited (^{2}D) state and P_N=0 all N atoms in the ground (^{4}S) state. The adopted background concentration of NO_x is given by the upper scale. In addition, the upper and lower limits of the annual production of NO due to galactic cosmic rays are included. From *Crutzen et al. (1975).*

proton event were reported by *Heath et al. (1977)* when they analyzed ozone data from the BUV experiment, aboard Nimbus 4, over the August 1972 period. Only northern polar cap data were evaluated. Data for 90-80°N were unavailable, due to high background current caused by penetrating protons. The daily, zonally averaged ozone concentrations above 4 mb pressure surface are shown in Figure 2.20 for three latitude zones (5°S - 5°N; 55°N - 65°N; 75°N - 80°N). The events produced an abrupt decrease in the 70-80°N latitude zone of 2 matm cm (or DU) in the ozone column above 4 mb after August 4. This translates to a 15% column loss above 4 mb, which persisted throughout the month of August. The latitudinal zone 55-65°N includes a range of geomagnetic latitudes and so, at certain longitudes, high energy protons were able to penetrate the atmosphere. A column decrease of 8% above 4 mb was detected with a recovery by the end of August. Protons at lower latitudes are increasingly deflected by the terrestrial magnetic field so that no there is no mechanism by which an SPE could directly effect equatorial ozone. A SPE effect was not detectable in ground-based spectrophotometer observations over the large local ozone variations. These observations were considered to be an underestimate of the depletion due to spacecraft pitch errors.

McPeters et al. (1981) have analysed the BUV data of this period, also detecting a 15% ozone reduction at 42 km over the northern polar latitudes that persisted for almost thirty days. This indicates that NO_x was produced in a quantity sufficient to alter the ozone chemistry.

The August 1972 event was further investigated by *Fabian et al. (1979)* by comparing results from a two-dimensional model with BUV ozone measurements. In their first 'run' of their model, NO production rates were assumed to be 1.5 times ionization production rates as used by *Crutzen et al. (1975)*. A factor of > 2 would imply that most of the N atoms were in electronically excited states. This high level of NO production was used in a second run of the experiment. The results from the second 'run' of the model closely matched the observations of subsequent ozone loss (*Heath et al., 1977*). Whereas the first 'run' predicted an ozone decrease that was too small. Rocket measurements concurred with this result (*Arnold, 1978*) and inferred that a NO production rate of 2 - 2.5 times the ionization rate was a better estimation than that of 1.5 used by *Crutzen et al. (1975)*. However, a theoretical study by *Porter et al. (1976)* indicated a fraction of less than 1.5. Measurements of NO for a SPE in July 1982 (*McPeters, 1986*) indicate that 1.25 nitrogen atoms per ion pair is a fair estimate for that event. In addition, this two-dimensional model tended to overestimate ozone recovery at high latitudes and underestimate it at mid-latitudes which may be attributed to the inability of the model to simulate ozone transport from high to mid-latitudes during August.

Reagan et al. (1981) reported that the August 1972 events were the most intense in

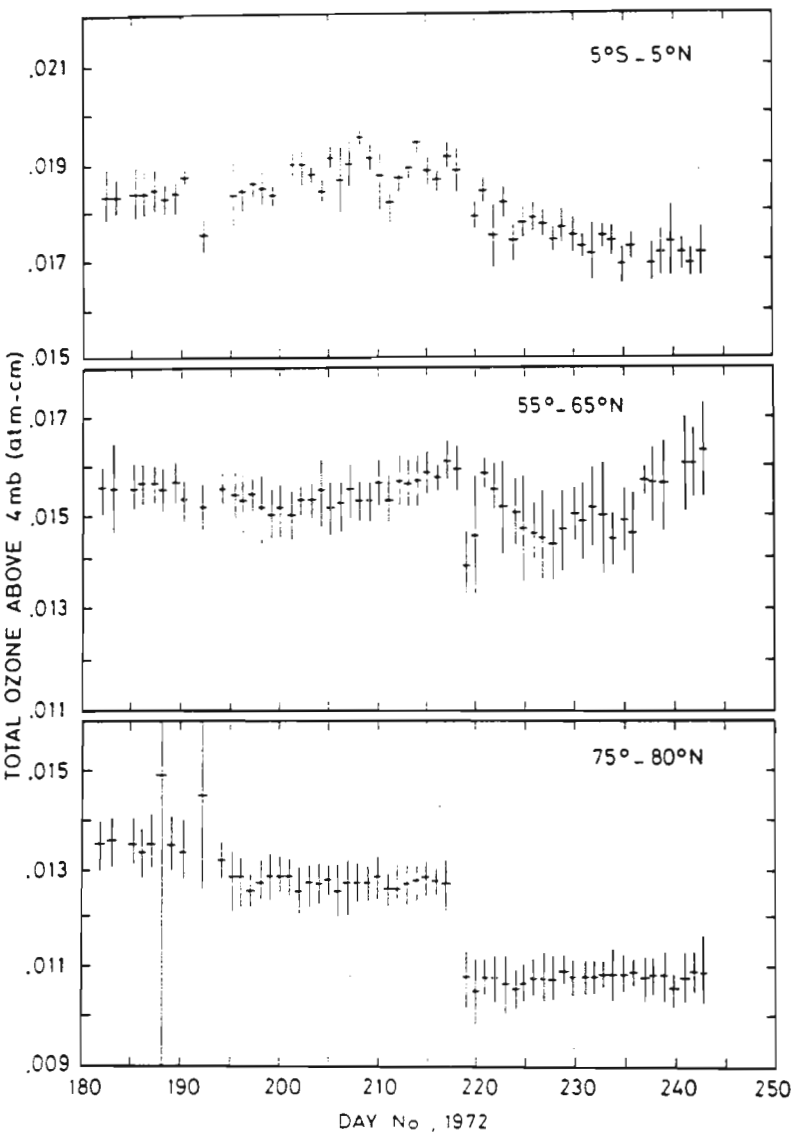

Figure 2.20: Zonally averaged column ozone above 4 mb for 5°S-5°N (top), 55°N-65°N (middle) and 75°N-80°N (bottom) latitude bands as observed by the BUV instrument during July and August 1972. A large SPE occurred on day 217. From *Heath et al. (1977).*

solar cycle 20, constituting some 85% of the integral flux of energetic solar protons in the entire solar cycle (*King, 1974*). The measured spectrum of this event was fitted by Reagan *et al.* to a single, simple analytic form, such as a power law or exponential shape. The ion production rates were calculated to peak at 5.8 $\times 10^4$ ion pairs cm^{-3} s^{-1} (38 km). At 20 km the production rate was calculated to exceed 10^3 ion pairs cm^{-3} s^{-1} at an estimated accuracy of better than 25% for altitudes >30 km. Reagan *et al.* estimated that total column ozone was reduced by 2% within a polar ozone cavity (> 60°N). However, no ozone change was detected by the BUV instrument at 2 mb, a result which could be attributed to an artifact of the mathematical inversion techniques of the earth's albedo used to calculate the preliminary ozone data.

Maeda and Heath (1980/1981) analyzed BUV data over both polar caps and reported asymmetries of both a spatial and temporal nature in ozone responses to the SPE's. They detected a 40% loss in the ozone column above 4 mb at 70°S while only a 16.5% drop was evident at 70°N. Maeda and Heath attributed the asymmetry to differing ozone dynamics during, and following, the SPE period. Large fluctuations characterize southern hemisphere polar ozone, due to planetary wave activity in the late austral winter. This would account for the fact that the depleted ozone followed irregular variations and did not show a monotonic recovery. The northern summer atmosphere is well reproduced by the general circulation mode (GCM), and no large ozone variations are expected. Depletion observed over the south pole was delayed relative to the northern polar cap, which may be ascribed to less solar irradiation over the south pole at the time of the onset of SPE's.

Maeda et al. (1984) further explored the asymmetry of this event. They summarized the possible causes to be, tilt of the IMF with respect to the earth's dipole magnetic field which influences the precipitation of energetic solar particles into the polar atmospheres, differences in ozone chemistry caused by the large change in atmospheric temperature between summer and winter hemispheres, seasonal differences of the stratosphere's dynamic states which are affected by upward propagating planetary waves in winter in contrast to the relatively undisturbed zonal flow in summer, and topographic asymmetry between the two hemispheres.

Jackman et al. (1990) have used a two-dimensional model to compute the atmosphere's response to the August 1972 SPE. Average ion pair production was calculated to be 8000 cm^{-3} s^{-1} between 40 and 50 km during the first days of the event. This ion pair production was used as the input to a two-dimensional photochemical model. The calculated output was compared with observations by the Nimbus 4 BUV experiment. Modelled ozone depletion at 75°S is >30% at 40 km altitude with a large depletion near 60 km about a month after the August 1972 SPE. The depletion around 60 km was thought to be caused by the downflux of the middle and upper mesospheric NO_y, created by the SPE, to the lower mesosphere. Modelled depletions at 75°N were

~20% at 45 km, which is less severe than it's southern counterpart. *Meada and Heath (1980/1981)* reported that southern hemisphere did indeed experience a larger ozone loss, although the observed depletion was greater than that predicted by their model. The prediction by the model that southern depletions would be deeper than their northern counterpart was vindicated by an observed 16.5% drop in total column ozone above 4 mb. Jackman *et al.* ascribe the interhemispheric ozone depletion difference to the different seasons of the two hemispheres during and after the SPE of August 1972. SPE produced NO_y in winter has a greater effect on ambient NO_y amounts. Not only are the background NO_y amounts less in winter, but the lifetime of NO_y is longer in winter, and the transport is directed downward to regions of even longer lifetimes for NO_y.

In summary, the stratospheric effects of the August 1972 SPE lasted about a year past the event (*Jackman et al., 1990*). The two-dimensional model (*Jackman et al., 1990*) predicts that the largest ozone depletion in the northern hemisphere, following the August 1972 SPE, occurs near to 85°N during, and subsequent to, the event. The largest ozone depletion in the south hemisphere, which was calculated to be greater than the losses over the northern polar latitudes, occurs around 85°S about a month after the SPE.

An analysis of a SPE at a similar time of year (July 1982), which forms case study 3 in chapter 4, reveals similar asymmetries of the ozone response to that of the August 1972 event.

Modelling and observations of the SPE's in the latter part of 1989

Three solar proton events occurred in the latter months of 1989. Together, these events form the basis of our analysis in case study 2 of chapter 4. The following integrated fluxes of > 30 MeV protons were measured for these events (from *Kodama et al., 1992*)

29 September to 12 October	$>8 \times 10^8$ cm^{-2};
19 October to 13 November	2.7×10^9 cm^{-2};
30 November to 14 December	$>7.8 \times 10^7$ cm^{-2}.

Reid et al. (1991) modelled the response of the middle atmosphere to these events with a two-dimensional dynamical/chemical model. Input to the model was derived from proton fluxes measured by the GOES-7 satellite. The satellite data consist of binned protons of different energy, the highest energy bin is > 100 MeV. To translate this information into a smooth spectrum, an exponential rigidity relation (*Freier and Webber, 1963*) is assumed in order to interpolate intermediate energies and hence the

shape of the spectrum for protons > 100 MeV. In this type of relationship the flux j of protons above rigidity R is given by

$$j(> R) = j_0 e^{-\frac{R}{R_0}}. \tag{2.47}$$

Protons > 100 MeV can penetrate below 30 km in altitude, so the computed ionization rates at lower levels are strongly influenced by this extrapolation, which must be used with caution. The angular distribution of protons was assumed to be isotropic over the upward looking hemisphere. Heavier nuclei, α particles and auroral electrons were ignored since their fluxes are at least an order of magnitude less than those of the protons.

The rate of production of N atoms has been shown to have a simple linear relationship to the rate of ionization (e.g. *Porter et al., 1976*) and 1.3 NO_y molecules were produced by an ion pair, calculated for each hour. The calculation used the standard proton energy loss relation for standard air (e.g. *Bethe and Ashkin, 1953*), and computed the ion pair production rate at 51 pressure levels in the atmosphere between the altitudes of 115 to 15 km. Figure 2.21 shows computed daily average ionization rates at 5 levels for the entire period of August to December 1989. Considerable ionization has been calculated to occur at 40 km.

Response of NO_y and ozone to the intense solar events is studied using a coupled model. The reader is reminded that in order for the increased NO_y to be effective for ozone depletion, it must occur in the stratosphere where ozone is most abundant. In addition, the mechanism by which NO_y can deplete ozone effectively requires sunlit conditions so that the response of ozone to a SPE may be delayed until the return of sunlight over the polar cap. The calculated NO_y concentrations produced by the SPE's were superimposed on the model atmospheres of both hemispheres at latitudes greater than 68°. Contour plots of the calculated percentage changes in NO_y and ozone for October 1989 are presented on the left side panels of Figure 2.22 and on the right side panels for March 1990. Large, but different, percentage increases are expected over the polar regions. Any interhemispheric differences in this model can only be ascribed to differences in photochemistry. Stratospheric ozone depletions over the southern polar cap are calculated to be 20% near 40 km, whereas only a 10% depletion is estimated for the northern polar regions. These effects will be long lived, as the lifetime of NO_y in the stratosphere is months and so predictions of persisting ozone loss up to March 1990 are shown on the right-hand side of Figure 2.22. Calculated losses are 0% for southern latitudes and up to 10% near 30 km for high northern latitudes. These ozone losses translate to approximately 2% change in column ozone. Measurements from the SAGE II satellite have measured significant NO_2 enhancements (60-70%) and ozone depletions (10-20%) near 20-30 km over the Arctic in late March 1990. These observed perturbations were larger and extend deeper into the atmosphere than those predicted by Figure 2.22. Furthermore, the NOAA-11 SBUV/2 records a greater depletion in the

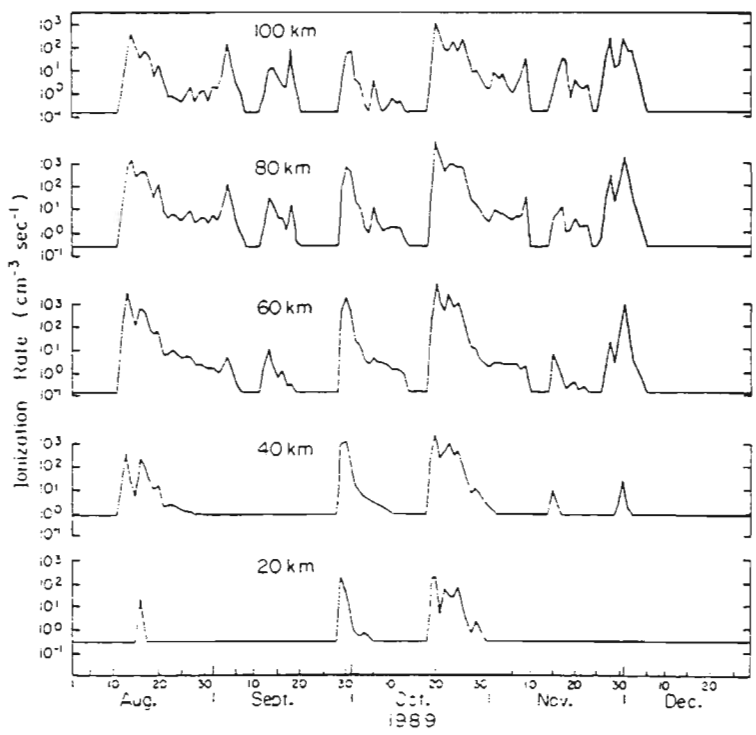

Figure 2.21: Ionization rates calculated at 20 km intervals, starting at 100 km in the top panel, and working down to 20 km in the bottom panel, for the SPE's during the months August to December 1989. From *Reid et al. (1991).*

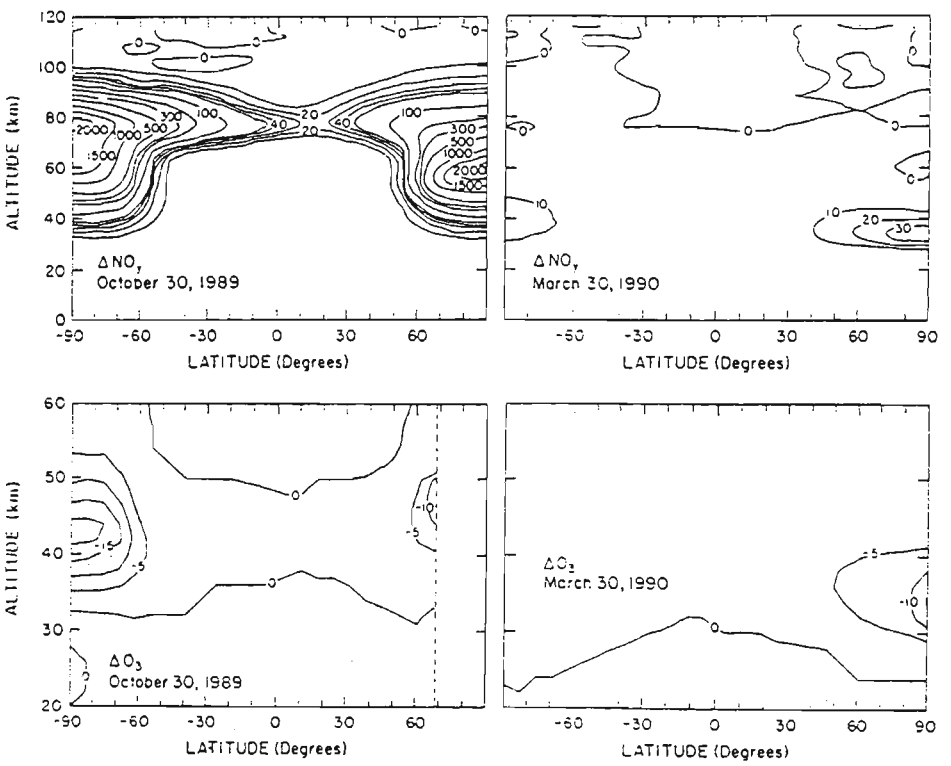

Figure 2.22: Calculated percentage changes of NO_y (top graphs) and O_3 (bottom graphs) by 30 October 1989 (left side) and 30 March 1990 (right side). From *Reid et al. (1991)*.

northern hemisphere than the south (*Jackman et al., 1993*), the opposite of what is predicted by this model. Clearly, the inability of this two-dimensional model to simulate planetary wave dynamics and chemical reactions involving heterogeneous processes must be a factor. Further, uncertainties in the fluxes of high energy particles capable of penetrating to the lower stratosphere (> 100 MeV) could incur errors.

Jackman et al. (1993) have addressed the discrepancy between the two-dimensional modelling and the observations of ozone due to the SPE's in late 1989 by using a three-dimensional chemistry and transport model to simulate the distribution of NO_x and ozone. The inclusion of meridional mixing and wave transience would, it was speculated, achieve predictions that were closer to observations, especially in regards to interhemispheric differences. This model only included effects of SPE's between 19-27 October 1989, and therefore will underestimate the NO_x production due to SPE's over this period. Proton fluxes from IMP-8 satellite were utilized which comprise proton counts in intervals between the energy thresholds of 0.29 and 440 MeV. Ozone data, to be compared with calculated ozone losses, were measured by NOAA-11 SBUV/2 instrument. These measurements confirmed the long term ozone depletion following SPE's in both hemispheres. Despite the fact that proton flux entering the two hemispheres should be roughly equal (*McPeters et al., 1981*) only a 1% decrease was observed in the southern hemisphere at 4 hPa compared with a 12% decrease at the same altitude in the northern hemisphere by December 1989. The three-dimensional model incorporates a simplified homogeneous photochemistry which includes complete O_x, HO_x and NO_x chemistry. Transport of these chemical families and other long lived nitrogen species is considered. Chlorine and bromine chemistry is omitted, thereby excluding the important links between Cl_x and NO_y chemistry. In particular, the reaction

$$ClO + NO_2 + M \rightarrow ClONO_2 + M$$

is excluded. This is a key reaction when considering the ozone loss processes during winter and early spring over polar regions. Without sunlight there is no mechanism for the NO_y to directly affect ozone. Hence, in the southern hemisphere the three-dimensional model predicts a strong correlation between features of increased NO_y and decreased ozone, illustrated graphically in the collage Figure 2.23 (a) This correlation holds throughout the simulation (until January 1990), even when ozone rich and NO_y poor air is mixed from lower latitudes during the final warming. The model predicts 130% increases of NO_y at 4 hPa with a corresponding -30% ozone decrease by November 13, as plotted in the top two false colour plots of Figure 2.23 (a). The strong winter vortex breaks down due to dynamical and radiative forcing in the southern hemisphere in November. This irreversible final warming causes large scale mixing of polar and middle latitude air and so, by January 3, enhancements in NO_y of only 3% are estimated with an ozone depletion of -10%, illustrated in the bottom colour plots of Figure 2.23 (a). In the northern hemisphere, the NO_y depletion of ozone is less effective as the

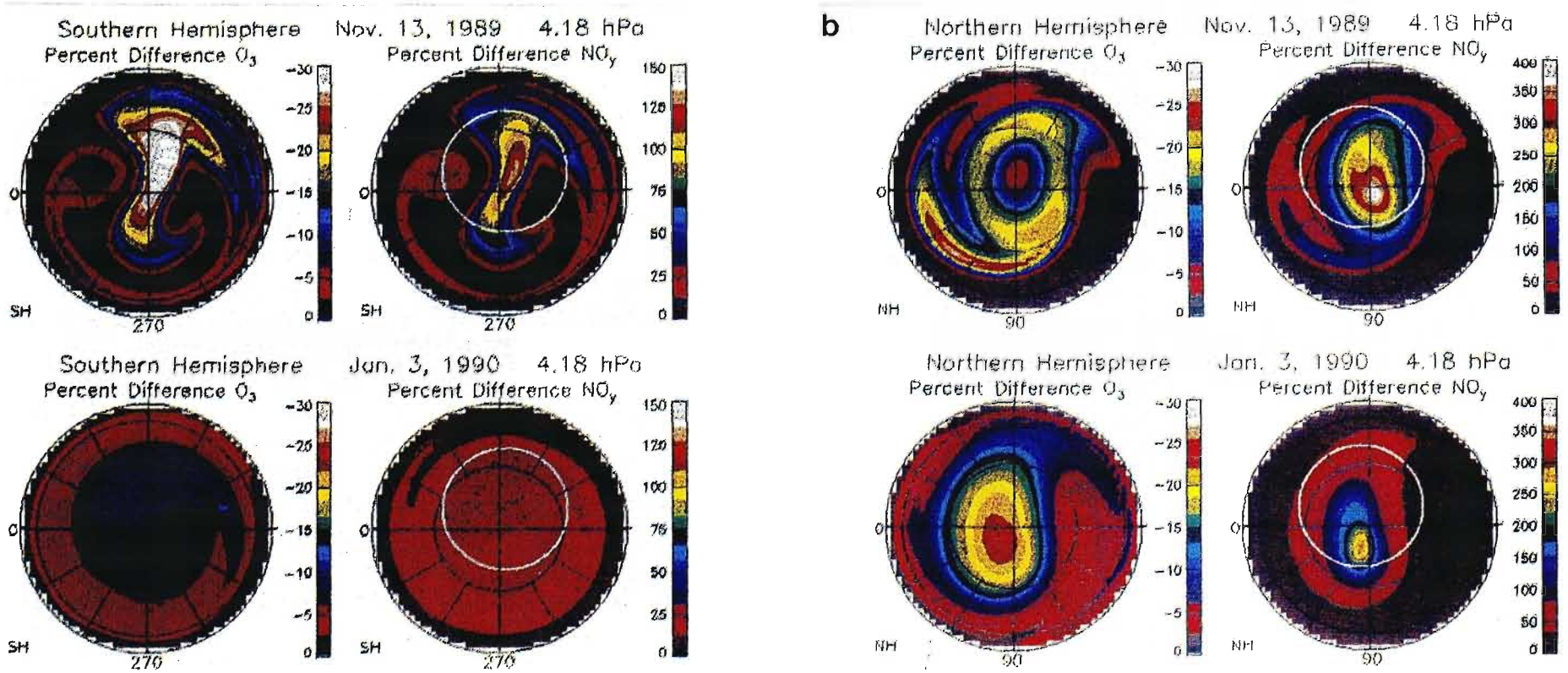

Figure 2.23: (a) Top graphs show the *southern* polar cap O_3 and NO_y percentage responses for November 13, 1989 at 4 hPa. Bottom graphs for January 3, 1990. (b) Top graphs show the *northern* polar cap O_3 and NO_y percentage responses for November 13, 1989 at 4 hPa. Bottom graphs for January 3, 1990.

polar cap is in darkness, so that transport of air into and out of sunlit regions allows the chemistry to proceed. For this reason, regions of highest NO_y increases (yellows, reds and white) of up to 370% over the darkened pole at 4 hPa do not correspond directly to areas of maximum ozone decrease (yellows, reds and white) of more than -27%, which lie well off the pole by November 13. The wintertime polar vortex is forming in the northern hemisphere in November. The tremendous radiative cooling associated with the northern hemisphere fall, leads to descent of air in the polar regions and spin up of the vortex. This is expected to isolate the polar regions from the middle latitudes for a few months. By January 3 large ozone losses (-25%) and high NO_y abundances (+270%) are still predicted (lower colour plots in Figure 2.23 (b)), due to little mixing of polar and mid-latitude air in the presence of the polar vortex.

It is clear that the three-dimensional simulations present a very different picture of ozone loss over each hemisphere. While the prediction for the northern hemisphere is similar to that of two-dimensional simulations, southern hemisphere results are quite different. Neglect, in two-dimensional models, of the influence of planetary wave dynamics in the austral spring is considered to be responsible for the incorrect predictions. This model underlines the importance of the prevailing conditions of the middle atmosphere in determining the ozone response to an SPE.

2.8.3 Solar cycle influences

Figure 2.5 (as determined by the model of *Garcia and Solomon (1983)*) summarizes the regions of the middle atmosphere in which seasonal ozone variations are controlled by chemical production or destruction, or transport, or a combination thereof. The polar regions were found to be controlled by transport in the winter and by chemical destruction (chiefly due to NO_x) in the summer. Therefore, polar ozone levels critically depend on the production of NO_x in the high latitude regions. An increase in odd nitrogen species during periods of solar activity would therefore imply a loss in polar ozone. On the other hand, at low latitudes chemical production controls the stratospheric ozone, so that the tropical regions would be sensitive to increasing levels of O_x, in particular atomic oxygen. Enhancements of O_x would lead to ozone increments in the tropics. Clearly, the solar cycle invokes a very different ozone response in polar latitudes to that in tropical regions.

Modern photochemical and dynamical theories point to three different mechanisms that can influence the ozone layer over the period of a solar cycle. Large variations in the ultraviolet and extreme ultraviolet portions of the solar spectrum during periods of intense solar activity provide the first mechanism. These wavelengths ($\lambda < 310$ nm) have been shown to control the temperature and photochemistry of the middle

atmosphere (*Brasseur and Solomon, 1986*). Secondly, cumulative effects of energetic proton precipitation on the budget of polar NO_x must be considered. As mentioned previously, *Crutzen et al. (1975)* suggested that the production of NO_x in SPE's could be an important part of the budget of odd nitrogen in the middle atmosphere, since the odd nitrogen input provided by a single SPE can be as large as that produced by oxidation of atmospheric N_2O in an entire year for latitudes poleward of 50°. This concept is explored in a model which provides a budget analysis of NO_x (*Jackman et al., 1990*). The model predicts that NO_x amounts would not be changed substantially over a solar cycle. However, the one-dimensional model study of *Orsini and Frederick (1982)* indicated that the impact of SPE's is an important parameter when determining NO_x production in the mesosphere. Thirdly, *Solomon et al. (1982)* found that the production of NO_x in aurorae near 100 km could also provide an important source of NO_x in the mesosphere and stratosphere. The effect in the stratosphere is due to large scale downward transport during the polar night. Calculations, from the streamfunction equation for Eulerian circulation, predict that downward air motions over the poles during winter could be as much as 100 kg m^{-1} s^{-1} (*Garcia et al., 1984*). Such a connection had long been suspected based on an observed correlation between aurorae and winter anomaly events (e.g. *Schwentek, 1971*). These three mechanisms suggest an ozone reduction over polar regions (due to NO_x enhancements) during, and around, solar maximum. They are addressed, in turn, below.

Variations in solar UV radiation

Solar ultraviolet radiation (λ <300 nm) is responsible for the photolysis and ionization of oxygen and nitrogen species above 50 km, as illustrated by Figure 2.14. The variability of ultraviolet flux over the eleven year solar cycle is difficult due to degradation and drift of the instruments used over such a long period of time. The ratio of the solar minimum to solar maximum flux at ultraviolet wavelengths employed by several authors is given in Figure 2.24. Observations by *Heath and Thekaekara (1977)* indicate that fluxes at $\lambda < 200$ nm are halved during solar minima. Estimates in other studies are somewhat more conservative.

The variability of the emission of electromagnetic radiation from the sun has two direct consequences on the middle atmosphere. Firstly, it produces a change in the atmospheric heating rate, primarily attributed to the absorption of short wavelength radiation by molecular oxygen and ozone. The change in temperature will also depend on the ozone-temperature feedback. The photodissociation of ozone

$$O_3 + h\nu \rightarrow O_2 + O, \tag{2.48}$$

which leads to the heating of the stratosphere, is highly temperature dependent reaction. Therefore, an ozone decrease leads to a corresponding temperature reduction

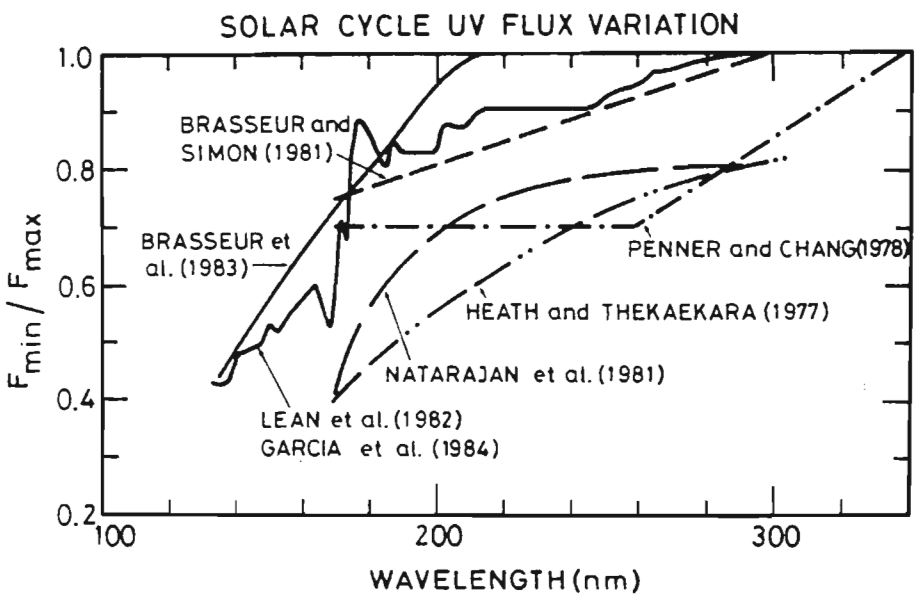

Figure 2.24: Ratios of solar ultraviolet flux at solar maximum to that at solar minimum adopted by several authors. The ratio employed by *Garcia et al. (1984)* is indicated. From *Brasseur and Solomon (1986).*

(due to less ozone). The lower temperatures cause the rate of ozone reduction to lessen and so provides the stabilizing feedback between temperature and ozone. Nevertheless, model calculations (*Garcia et al., 1984*) indicates a temperature increase of 2 to 3 K near the stratopause over the solar maximum.

The second effect is the modification of photodissociation and photoionization rates of atmospheric constituents. Radiation in the Schumann Runge bands can penetrate to the stratopause and its variability can therefore influence the entire mesosphere. The rate of photodissociation of molecular oxygen in the γ(0-0) and γ(0-1) bands can vary from 15 to 20% over a solar cycle. The radiation of the Hertzberg continuum penetrates into the stratosphere, an important part of which constitutes an optical window, may induce an enhancement of 15% in the photodissociation rate of O_2. The variation applies not only to the photodissociation of molecular oxygen but other source gases as well e.g. N_2O, $CFCl_3$, etc. Variation in ultraviolet radiation can therefore effect changes in production of odd oxygen, odd nitrogen and odd chlorine.

In the case of ozone, its concentration depends on the ratio $\frac{J_{O_2}}{J_{O_3}}$. Although both of these photolysis rates increase with increasing solar activity, J_{O_2} is more sensitive to variations in the shorter wavelengths than J_{O_3}. The net effect of enhanced solar activity is an increase of ozone in the upper stratosphere of the tropics. In addition, temperature and dynamical feedbacks must be considered. Since certain ultraviolet wavelengths are absorbed exclusively by ozone, any ozone depletion at higher altitudes will allow penetration of ultraviolet radiation to lower altitudes.

Increased photolysis of nitrous oxide can enhance levels of odd nitrogen in the middle atmosphere. An increase in NO_x can be particularly effective in destroying polar ozone since it has a long chemical lifetime, as photolysis does not occur during the polar night, and it may be transported down to the stratosphere.

In addition, *Hood and Jirikowic (1991)* have reported, from cross-spectral analysis, that there are dynamical perturbations associated with increased 27 day ozone variations near solar maximum. This has lead to speculation that the solar cycle may effect total ozone amounts dynamically. A suggested mechanism is the modulation of eddy activity. A temporally lengthy global satellite data set is needed to test this hypothesis. However, a regression analysis of 13.2 years of TOMS data (*Hood and McCormack, 1992*) suggest that both changes in dynamics as well as direct photochemical changes are primary forcing mechanisms.

Solar proton events

The mechanism by which these events can produce odd nitrogen and odd hydrogen in the middle atmosphere has been addressed in sub-section 2.8.2. When considering long term effects, HO_x production is not relevant, it has a very short chemical lifetime of only a few hours. However, an increase in the longer lived (several months, up to a year) NO_x species is important after an SPE. The frequency of occurrence of optical and X ray flares and SPE's correlates well with the 11 year sunspot cycle (*Hirman et al., 1988*). During periods when SPE's are more numerous, the excess NO_x will amass in the mesosphere and, during the polar night when their chemical lifetimes are further extended, be transported to lower altitudes where ozone is more abundant.

Auroral electrons

The mechanism of odd nitrogen production by energetic protons involves the production of many energized secondary electrons, which are effective in forming odd nitrogen. In the case of auroral electrons, they are the primary particles, but are only effective in the thermosphere at an altitude of around 100 km. Thermospheric NO_x is expected to vary greatly (at least by a factor of 2) with solar activity, due to both the ultraviolet radiation and auroral particles variation over the solar cycle. Geomagnetic activity leads to intense thermospheric warming in the polar and auroral regions. This may be attributed to kinetic heating caused by the precipitation of energetic charged particles and Joule heating caused by enhanced ionospheric currents. This warming leads to more efficient production of odd nitrogen (see sub-section 2.8.2). In particular, large increases in nitric oxide in the thermosphere and mesosphere are expected to propagate to lower altitudes, over polar regions during the winter season. Therefore, increased amounts of NO_x, produced during solar maximum periods, are transported to the stratosphere where they can be more effective at ozone destruction.

Modelling of ozone responses to the solar cycle

The potential response of the atmosphere to solar activity has been studied with one-dimensional (*Penner and Chang, 1980*) and two-dimensional models (*Brasseur et al., 1983*; *Garcia et al., 1984*). These models adopt different solar flux variations, illustrated in Figure 2.24, as inputs to their models.

Results from the numerical model study of *Garcia et al. (1984)*, which estimates the response of the middle atmosphere to the 11 year solar cycle, are given below. Briefly, this model which simulates coupling between dynamics and chemistry inputs the vari-

ations of extreme ultraviolet, ultraviolet and visible radiation and auroral particles over a solar cycle. The cumulative effect of SPE's was not considered here. *Garcia et al. (1984)* adopted conservative estimates of the ultraviolet flux ratio for their model, marked as *Garcia et al. (1984)* in Figure 2.24. The degree of the ultraviolet variance chosen as input can, obviously, lead to a different calculated response by the middle atmosphere. This is due, in part, to fluctuations of the photolysis rates that control O_x production and the $\frac{O}{O_3}$ equilibrium which lead to increases in ozone total column over low to mid-latitude regions. Furthermore, NO_x is produced in the thermosphere by ultraviolet radiation and particle precipitation in aurorae, which can then be transported to the stratosphere by mean meridional circulation during the polar night.

The *Garcia et al. (1984)* model incorporates mean meridional motions which have been calculated from the streamfunction equation for the residual Eulerian circulation. The results are compared under solar minimum and maximum conditions at the end of northern hemisphere summer/southern hemisphere winter. This period of the year illustrates most clearly the effects of solar cycle variability. The computed global mean meridional circulation is thermally direct linking the equatorial and polar stratospheres and furthermore it couples the summer winter pole at meso- and lower thermospheric heights. The NO production rate due to particle precipitation in aurorae is computed.

Figure 2.25, from *Garcia et al. (1984)*, shows the calculated thermal response to the 11 year solar cycle for September 1. Modelled temperature increases from solar minimum to maximum are evident in the lower thermosphere, due to heating by molecular oxygen and the Joule heating by precipitating energetic particles. Large temperature increases ($>$ 10 K) are predicted throughout the thermosphere. Smaller changes are found in the mesosphere and stratosphere, where ozone heating dominates. Attention is drawn to the negative thermal response in the austral spring (60-80°S) associated with the ozone hole. Changes in stratospheric ozone will affect heating rates there which, in turn, influence the temperature distribution and circulation e.g. a 3 m s^{-1} increase speed of the polar vortex is predicted at solar maximum. Finally, temperature and ozone fluctuations can perhaps influence planetary wave propagation (*Geller and Alpert, 1980*).

Calculated enhancements of NO_x, associated with the 11 year solar cycle, are shown in Figure 2.26. Clearly, the largest photochemical responses to changing solar activity are at thermospheric heights. Large increases (80%) of NO_x are calculated in the tropics above 90 km, these enhancements are not effective in altering ozone concentrations due to their high altitude and the characteristic upwelling there. Very large enhancements in the spring polar latitudes of 160% are predicted down to the stratosphere due to downward advection in the polar night. At high latitudes in late summer (northern hemisphere) the thermospheric source can not penetrate to lower altitudes as the mean meridional motion at this time is upward.

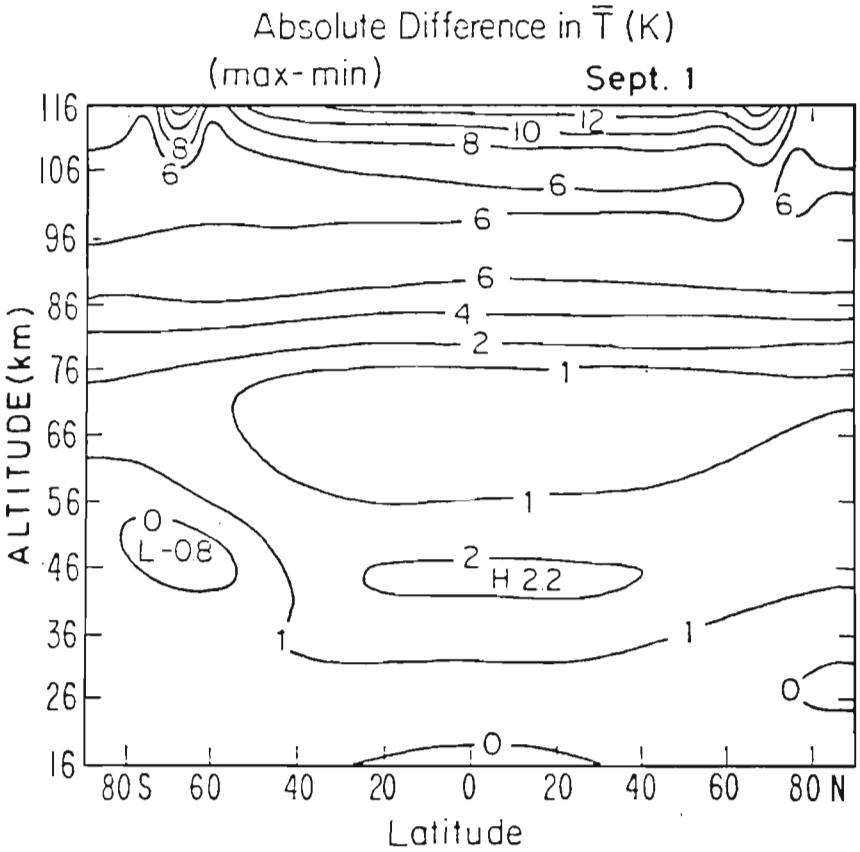

Figure 2.25: Computed change in the zonally averaged temperature between solar maximum and solar minimum. From *Garcia et al. (1984)*.

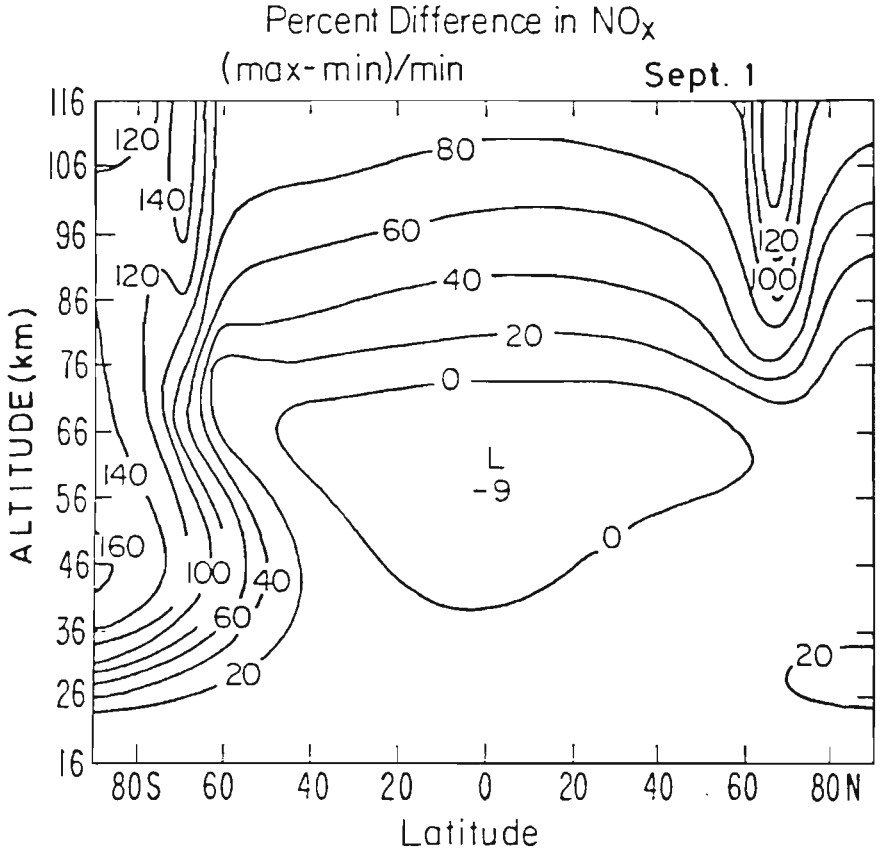

Figure 2.26: Computed percentage change of the zonally averaged NO_x distribution over a solar cycle. From *Garcia et al. (1984)*.

Modelled ozone density responses are shown in Figure 2.27. In general, ozone increases are calculated for solar maximum due to increased atomic oxygen abundance. This is particularly so at higher altitudes where shorter wavelengths, and thus larger solar ultraviolet variation, play the most important role. The largest ozone increases (up to 80%) occur in the upper mesosphere and thermosphere as a result of greater O_x production by photolysis of O_2 in the SRC and SRB. Decreases in ozone around 70 km may be attributed to the increase of OH radicals (by the photodissociation of water vapour) which scavenge ozone. The calculated ozone reduction obtained in the upper stratosphere in high latitude spring reflects the effects of increased odd nitrogen produced at higher altitudes that has been transported downwards. The model results show that it is possible to obtain a large variation in stratospheric NO_x at high latitudes which is positively correlated with the solar cycle, resulting in decreases in ozone of up to -40% over the springtime polar stratospheres. The ozone depletion at stratospheric heights leads to greater column decreases than the column increases predicted at middle and low latitudes.

The Garcia *et al.* model predictions were tested against observations of upper stratospheric ozone spanning at least half of an eleven year solar cycle (*Solomon and Garcia, 1984*). Data were used from the Nimbus 4 BUV experiment to monitor relatively solar quiet years and the Nimbus 7 BUV to monitor the solar active years; at 4, 2 and 1 mb pressures. Effects of thermospheric NO were expected to be manifested at specific latitudes (>60°), altitudes (40-52 km) and seasons (winter and spring). In looking for these specific effects in ozone reduction, one is able to gain indirect evidence of NO_x increases. Comparison of observations and model calculations suggest that the trends predicted by the model are present in the data, although the magnitude of the observed variations is generally smaller than predicted.

Jackman et al. (1990) have used a two-dimensional time dependent photochemical model to determine whether NO_y produced by SPE's can build up to maximum abundances around solar maximum. The solar cycle ultraviolet flux variation or auroral electron precipitation was not included in any of the computer simulations. Figure 2.28 illustrates calculated values of odd nitrogen production by SPE's from 1955 through 1985. Clearly, more NO_x is produced during solar maxima. Although SPE's were found to provide a substantial annual addition of odd nitrogen, accumulative effects could not be seen in this model. However, *Callis and Natarajan (1986)* report satellite data which detected significant increased levels of stratospheric odd nitrogen for the 1979-1984 (solar maximum) year period.

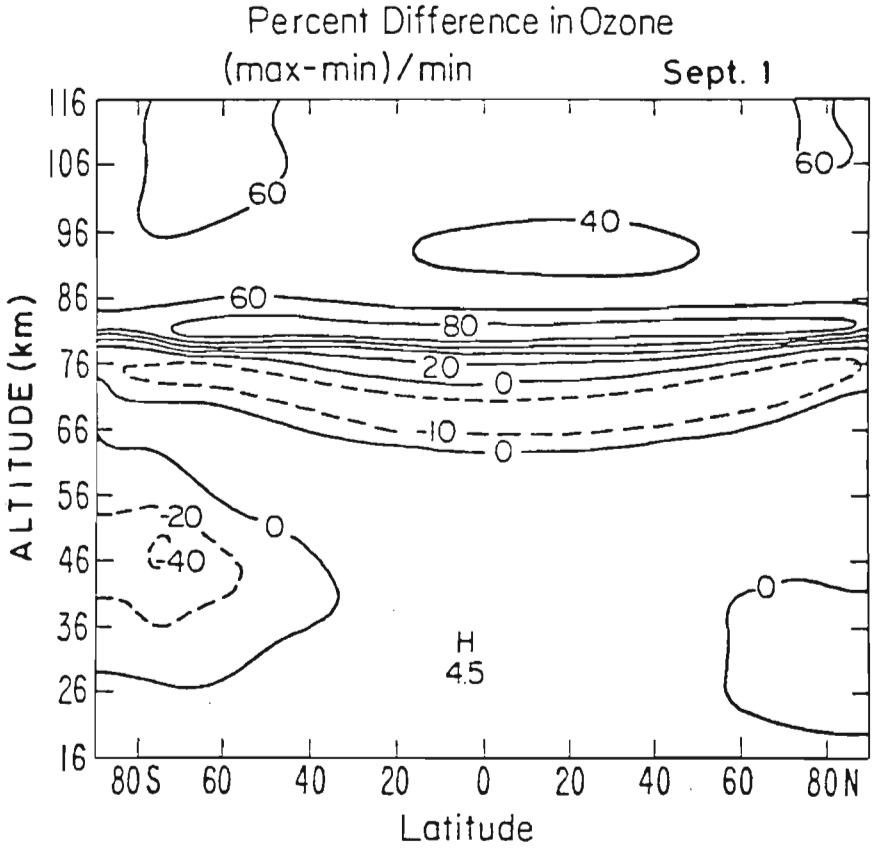

Figure 2.27: Calculated percentage ozone variation over a solar cycle. From *Garcia et al. (1984)*.

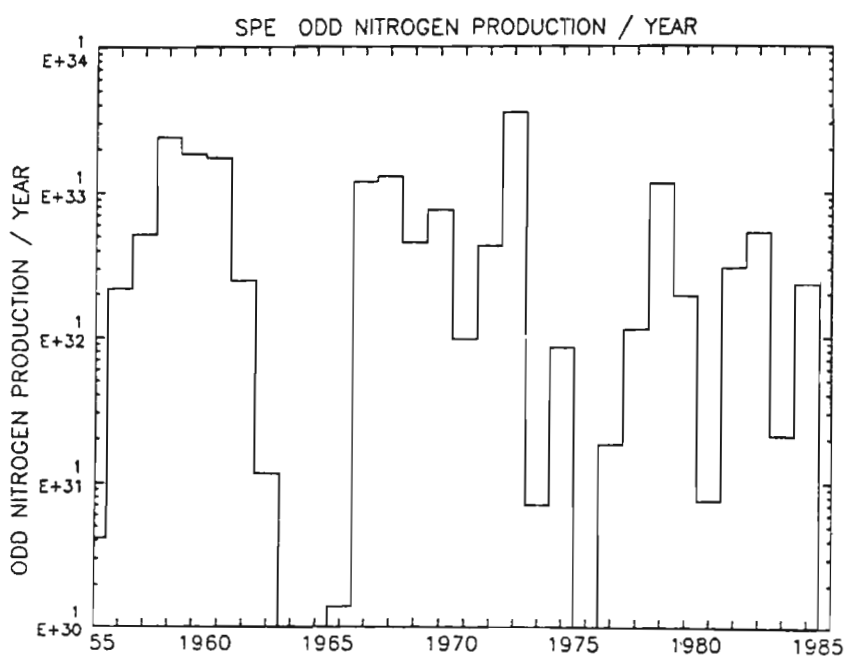

Figure 2.28: Calculated odd nitrogen production (molecules year^{-1}) for all SPE's in the period 1955 to 1984. From *Jackman et al. (1990).*

Solar cycle trends detected in ozone data

Data from several ground based and satellite instruments have been analysed and have been found to exhibit solar cycle periodicities. Some results are summarized below:

By analyzing 20 years of soundings with an optical sonde over central Europe, *Paetzold (1973)* claims a high correlation between sunspot number (and a magnetic activity index) and ozone content between 20 and 30 km. This relationship is illustrated in Figure 2.29 in which periods of high magnetic activity (dashed line) correspond to ozone (solid line) increases in the stratosphere.

Gille et al. (1984) have detected increases (0.25 to 0.60%) in zonal mean ozone values over the tropics following a period in the winter of 1979 when the solar ultraviolet irradiance varied by 1%. These zonal decreases were measured by the Limb Infrared Monitor of the Stratosphere (LIMS) experiment on the Nimbus 7 spacecraft. These increases imply changes of up to 12% at an altitude of 48 km over a solar cycle.

Hood and McCormack (1992) have used a multiple regression statistical model to estimate the components of interannual ozone change of TOMS ozone data. The existence of a solar cycle component is indicated by empirical studies of the model residuals.

Blackshear and Tolson (1978) have found high correlations between monthly averages of global total ozone and solar activity over a 9 month period. While work by *Angell and Korshover (1978)* suggests ozone column increases of a few percent over the first half of solar cycle 20.

Labitzke and van Loon (1988) have found that correlations between high latitude temperatures and geopotential heights with solar activity may vary according to the phase of the QBO. This would imply that the magnitude of the effect of solar activity on ozone could depend on the phase of the QBO.

2.9 Summary and conclusions

To conclude this chapter, a brief summary is given of the pertinent points that are important to the interpretations of the case studies in chapter 4 and the Fourier analysis of the 11 year ozone data series in chapter 5.

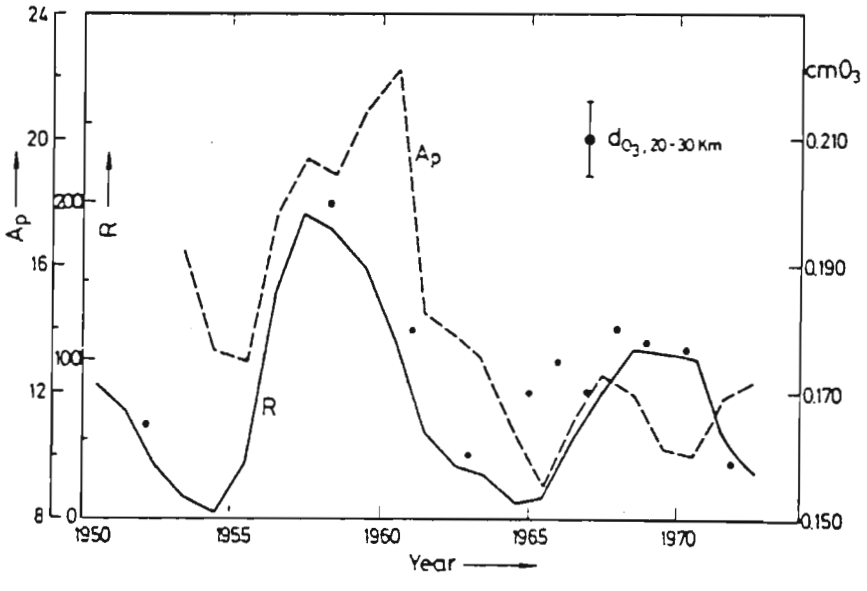

Figure 2.29: Variation of ozone between 20 and 30 km altitude during two sunspot cycles. From *Paetzold (1973)*.

- Ozone is in chemical equilibrium throughout most of the middle atmosphere. However, there are certain latitudes and altitudes in which other processes may be dominant. The tropical stratosphere, which receives the most insolation, is dominated by chemical production. Ozone is therefore sensitive to any changes of O_x in this region. In winter, high latitudes are dominated at all altitudes by transport, due to the extended chemical lifetimes of atmospheric species. The polar summer stratosphere is dominated by chemical destruction, particularly by NO_x.

- Background circulation is directed from the summer to the winter hemisphere. Downward motion is characteristic of the winter high latitude and weak upward motion in the summer. Other important dynamical features which affect polar ozone distribution include the strong mid-latitude zonal winds of the polar vortex which isolate the polar caps in winter and early spring and the upward propagating planetary waves that erode and, finally break down the polar vortex. Furthermore, the biennial variation of tropical stratospheric winds (QBO) modulate the number of wave breaking events that occur during winter and spring.

- SPE's which are initiated by solar flares result in the production of NO_x at high latitudes. Odd nitrogen is able to catalytically destroy ozone and thus can induce ozone loss.

- Prediction of the quantitative ozone reduction following a large SPE requires, not only a knowledge of the primary proton flux, but the inclusion of the prevailing dynamical and chemical conditions in the middle atmosphere. This may be achieved with a three-dimensional model. Two-dimensional models have proved to be inadequate.

- There are three possible mechanisms for solar cycle effects on the middle atmosphere.

    1. Enhancement of ultraviolet flux during high levels of solar activity leads to increased O_x, HO_x and NO_x abundances. Higher ozone abundances are predicted for middle and low latitudes, whereas ozone losses will occur over polar regions.
    2. Precipitation of auroral electrons which generate NO_x in the polar thermosphere.
    3. Cumulative NO_x amounts due to numerous SPE's occurring over solar maximum.

    The last two mechanisms are rendered more effective by downward advection of NO_x during the polar night, when the chemical lifetime of NO_x is extended, to lower altitudes where ozone is more abundant.

Chapter 3

Solar and ozone data acquisition

3.1 Introduction

The first observations of atmospheric ozone were made by ground-based spectrometers. Later, these were supplemented by *in situ* measurements from instruments carried through the atmosphere on balloons and instruments on aircraft which sample the air through which they travel. Satellite platforms provided an opportunity to monitor ozone over vast spatial extents. The earliest satellite ozone measurement was recorded by an experiment aboard the Echo 1 satellite (*Venkateswaran et al., 1961*). Instruments aboard satellites utilize a variety of remote sensing techniques to provide global measurements of many atmospheric constituents and constitute some of the most important data sets in atmospheric research. The basic scientific premise of ozone observations from space is that ozone molecules absorb radiation strongly in the middle ultraviolet Hartley-Huggins bands (220-320 nm), moderately in the infrared (9.6 μm) and weakly in the visible Chappuis bands (0.6 μm). Utilizing these basic principles, ozone measurement techniques have been developed based on, atmospheric backscatter in the ultraviolet, radiative emission in the infrared and occulation in the visible portion of the electromagnetic spectrum.

These three techniques employed by satellites to determine atmospheric ozone amounts will be briefly reviewed by summarizing the most important missions. Following this, the TOMS instrument aboard the Nimbus 7 satellite, from which the total column ozone data are extensively used in the work described in this thesis, will be described. A subsequent section will detail the energetic particle, magnetometer and X-ray data, used during SPE's studied in chapter 4, from the GOES-6 and -7 satellites. Finally, a description of the solar radio flux monitored in Ottawa, Canada is given. These data are used as a proxy for ultraviolet forcing over a solar cycle period in the Fourier

analysis performed in chapter 5.

3.2 An overview of satellite observations of ozone

The first satellite-borne ozone instruments were designed to measure the vertical distribution through the stratosphere. The techniques made use of absorption of ultraviolet sunlight in the atmosphere below the spacecraft or the emission of infrared light from the atmosphere at the limb of the earth. *Dave and Mateer (1967)* suggested that the total ozone column above the ground could also be measured from space with respectable accuracy. Their technique emulated the wavelength pair methodology used by Dobson to derive total ozone from differential absorption. However, the path of sunlight through the atmosphere to the satellite is no longer a simple function of the solar zenith angle but rather a complex multiple Rayleigh scattering atmosphere. *Dave and Mateer (1967)* chose an indirect method wherein the albedo of model atmospheres, containing climatological ozone distributions, different surface reflectivities, and surface pressures, is computed for the selected instrument wavelengths over a range of solar zenith and view angles. When this technique was first tested by the Nimbus 4 BUV instrument (*Mateer et al., 1971*) it became clear that accurate measurements were possible and that global total ozone measurements were practical. An instrument that could measure total column ozone with a higher resolution was proposed. The Total Ozone Mapping Spectrometer (TOMS) aboard the Nimbus 7 satellite was launched in 1978. This instrument will be described in detail in section 3.3.

3.2.1 Solar backscattered ultraviolet technique

The principle experiments employing this technique over the last decade are the Solar Backscattered Ultraviolet Radiometer (SBUV) experiment on-board Nimbus 7, the Total Ozone Mapping Spectrometer (TOMS) on Nimbus 7, the Solar Backscattered Ultraviolet radiometer Version 2 (SBUV/2) which flies on the National Oceanic and Atmospheric Administration (NOAA) operational afternoon satellite series, the Shuttle SBUV (SSBUV) designated to fly on the Space Shuttle system and the Solar Mesospheric Explorer (SME). General operational information on these instruments is listed in Table 3.1.

- **SBUV** The SBUV is a nadir viewing double monochromator which measures solar radiances backscattered from the atmosphere at 12 discrete wavelengths from 250 to 340 nm with a 1 nm bandpass (*Heath et al., 1975*; *Fleig et al., 1982*;

instrument	platform	operation	parameters
SBUV	Nimbus 7	November 1978-February 1987	total ozone ozone profiles 22-55 km
SBUV/2	NOAA Operational Afternoon Satellite	March 1985-failed	total ozone ozone profile 22-55 km
SSBUV	schuttle	intermittent October 1989-	total ozone ozone profile 22-55 km
TOMS	Nimbus 7	November 1978-May 1993	total ozone
SME	SME	January 1982-December 1986	ozone profiles 45-65 km
TOMS	Meteor-3	August 1991-present	total ozone

Table 3.1: The instruments that employ the solar backscatter technique to determine ozone profiles and/or total ozone. Their lifetimes are tabulated in the third column. Adapted from *Miller, (1989)*.

McPeters et al., 1984). It is an extension, with modification, of the BUV measurement system flown on Nimbus 4 (1970-1977). Atmospheric radiances between 250 and 306 nm are used to infer the ozone vertical profile distribution, while radiances between 312 and 340 nm are used to calculate total ozone. The method requires that the backscattered radiance be referenced to the solar irradiance at each wavelength which is measured by deploying a diffuser plate.

Inferral of an ozone profile is possible because backscattered radiation at a given wavelength originates mostly in a limited altitude region of the atmosphere, and this altitude region varies with wavelength. Therefore, a wavelength scan is equivalent to an altitude scan. As sunlight penetrates the atmosphere, the scattering term increases exponentially with increasing density of air molecules, but the increasing depth of ozone causes the transmission of direct and back-scattered sunlight to exponentially decrease. The balance between an exponentially increasing source term and exponentially decreasing transmission term produces a well defined scattering layer of about 14 km half-width. The wavelength of maximum ozone absorption, 255 nm, produces a scattering layer (contribution function) at the maximum possible altitude, 50 to 55 km, depending on solar zenith angle. Radiation at wavelengths longer than 310 nm penetrates the ozone layer to be scattered by the troposphere and reflected by the ground and clouds. These wavelengths are useful for inferring the total ozone content of the atmosphere.

The inferral of an ozone profile from a set of measured backscattered albedos is performed by using a partial derivative inversion algorithm described by *Schneider et al. (1981)*. The optimum statistical concepts of *Rodgers (1976)* are used in the algorithm; because of the width of the contribution functions, the altitude resolution that can be obtained in the retrieved profile is limited to about 8 km.

The SBUV makes only daytime measurements since it uses backscattered solar radiation. Measurements are made continuously over a broad latitude range at a spacing of about 200 km along the orbit track and at the orbit spacing of about 26° longitude. The latitude limits vary with season depending on the sun angle however, continuous spatial coverage is possible from 60°N to 60°S.

- **SBUV/2** The SBUV/2 is descendant of the SBUV instrument, and incorporates several engineering improvements including an on-board Hg lamp to serve as a stability monitor for the diffuser plate. Unfortunately, this mechanism on the NOAA-9 instrument failed to function properly and a modification has been included on the NOAA-11 instrument.

- **TOMS** The TOMS is similar in concept to the SBUV but has several distinct differences. It has only six discrete wavelength channels, and incorporates a side scan feature with an approximately 50 × 50 km field of view. These differences

result in the TOMS instrument being limited to total ozone measurements, but with true global daylight coverage and a significantly enhanced capability to observe small spatial scales features. As the SBUV and TOMS measurements are restricted to the sunlit portion of the Earth this means that continuous spatial coverage throughout the year is from 60°N to 60°S, with coverage to the poles by TOMS only in the summer season. A more detailed description of the TOMS instrument is given in the following section.

- **SSBUV** The purpose of the Shuttle Solar Backscatter Ultraviolet Radiometer (SSBUV) instrument is to provide regular in-orbit calibration checks of the SBUV/2 ozone monitoring instruments being flown routinely on NOAA satellites. The in-orbit calibration transfer will be accomplished by comparing the observations of the Shuttle and NOAA satellite instruments.

- **SME** Instruments on the Solar Mesosphere Explorer (*Rusch et al., 1984*) have been used to determine the ozone density in the earth's atmosphere from about 1.0 to 0.001 mb and the NO_2 density from about 10.0 to 2.0 mb starting 1 January 1982, until December 1986. A description of the mission objectives can be found in *Barth et al. (1983)*.

3.2.2 Infrared emission technique

The only system currently employing the infrared emission technique is the Tiros Operational Vertical Sounder (TOVS) aboard the NOAA operational satellite series. An experiment to measure the infrared emission at the earth's limb, the Limb Infrared Monitor of the Stratosphere (LIMS), is aboard the Nimbus 7 spacecraft. General operational information on these instruments is listed in Table 3.2.

The underlying premise of this technique (except for the Solar Mesospheric Explorer (SME)) is that if the temperature profile of the atmosphere is known, then measurement of the radiance emitted in a spectral interval of ozone absorption provides information on the ozone amount. In principle, this is relatively straightforward. In practice, one generally needs to derive the temperature profile from ancillary data which is an additional source of error. SME operated on a very different principle described below.

- **TOVS** The actual technique for predicting total ozone from infrared emission measurements, is by multivariate regression with three independent radiance measurements. The radiance at 9.71 μm, which is the ozone absorption channel, is determined. In addition, the radiance in the window channel 11.11 μm since the transmittance from the 9.71 μm channel is broad enough that a significant portion of observed changes are due to changes in surface temperature. Finally, the

instrument	platform	operation	parameters
TOVS	NOAA all Operational Satellites	November 1978-present	total ozone
LIMS	Nimbus 7	November 1978-May 1979	ozone profiles 10-65 km
SME	SME	January 1982-December 1986	ozone profiles 45-90 km

Table 3.2: The instruments that employ the infrared emission technique to determine ozone profiles or total ozone. The only instrument that is still operational is TOVS. Adapted from *Miller (1989)*.

radiance at 14.99 μm to correct for the temperature effect. In practice, the regression coefficients are calculated monthly utilizing total ozone data from about 40 Dobson stations worldwide. Total ozone retrievals and instrumentation are fully described by *Planet et al. (1984)*.

- **LIMS** Determination of the ozone profile employing measured radiances at the earth's limb is exemplified by the LIMS experiment. It's limited lifetime (~7 months) was due to the requirement that the six detectors be cooled to 64 K by a solid cryogen cooler. The six channels included two in the 15 μm carbon dioxide band for temperature measurements, one at 9.6 μm for ozone profiles, and others at 11.3 μm for nitric acid, 6.9 μm for water vapour and 6.2 μm for nitrogen dioxide. Firstly, by measuring emission in the band of gas whose mixing ratio is known or assumed (i.e. CO_2), the temperature profile was inferred. Then, by measuring emission in the other bands the unknown mixing ratios were retrieved (*Gordley and Russell, 1981*). Since the observed parameter was horizon thermal emission, data were collected both night and day.

  The LIMS radiometer scanned the atmospheric horizon vertically once every 12 seconds obtaining radiance profiles in each of the six spectral bands as a function of tangent height. Tangent height is defined as the point of closest approach of a ray path to the earth's surface. The orbit parameters of the spacecraft provides uniform coverage of the earth's surface from 64°S to 84°N. The altitude coverage was determined largely by signal to noise ratios. The range for temperature and ozone was about 10-65 km, and for NO_2, HNO_3 and H_2O about 10-50 km. The vertical resolution of the ozone measurements was 2.8 km however, horizontal resolution is much more coarse (~300 km) and is dictated by limb geometry and atmospheric absorption characteristics.

- **SME** This instrument observed the earth's limb and measured radiation from excited molecular oxygen at 1.27 μm resulting from the photolysis of ozone. A knowledge of the mechanisms of this state of molecular oxygen production and loss allowed the ozone density to be inferred in the 45-90 km altitude range (*Thomas et al., 1984*).

3.2.3 Solar occulation technique

The current instrument utilizing this method of ozone profile determination is the Stratospheric Aerosol and Gas Experiment II (SAGE II) which is a follow on to the SAGE I instrument and the Solar Maximum Mission (SMM). General characteristics are depicted in Table 3.3. As this technique is virtually self calibrating, the instruments have assumed a very strong position regarding ozone trend evaluation.

instrument	platform	operation	parameters
SAGE I	AEM-2	February 1979-December 1981	ozone profiles 10-55 km
SAGE II	ERBS	October 1984-present	ozone profiles 10-65 km
SMM	SME	November 1984-present	ozone profiles 55-85 km

Table 3.3: The instruments that employ the solar occulation technique to determine ozone profiles. Both SAGE II and SMM are presently operational. Adapted from *Miller (1989)*.

The SAGE instrument is a four channeled sun photometer. Spectral discrimination for SAGE is achieved by using a holographic diffraction grating which disperses the incoming sunlight in different directions depending on wavelength. By using four detectors at appropriate locations along the dispersed spectrum, the visible sunlight intensity at four wavelengths, 0.385, 0.45, 0.60 and 1.00 μm is measured. Absorption at 0.385, 0.45 and 1.00 μm by stratospheric gases is quite small below 25 km so that solar extinction in these channels is almost entirely due to scattering by aerosol particles and air molecules (*McCormick et al., 1979*). At higher stratospheric altitudes, attenuation at 0.60 μm is primarily due to ozone. Above an altitude of about 25 km, the extinction at 0.385 and 0.45 μm is mainly due to absorption by nitrogen dioxide.

In operation, the instrument is activated just before a sunrise or sunset is encountered by the satellite. The instrument searches for the sun and nulls the centre of intensity of the solar image. A scan mirror the begins scanning up and down across the face of the sun. The mirror reverses in direction each time a limb crossing occurs. Solar light is reflected from the scan mirror to the aperture of a small Cassegrainian telescope which defines about a $\frac{1}{2}$ km instantaneous field of view on the horizon and focuses this light onto the diffraction grating. The intensity of light dispersed by the grating at the four wavelengths of interest is measured by the four sensors. The data are inverted in ground processing to yield extinction as a function of altitude for each spectral channel at each location and time of the SAGE instrument.

3.3 The TOMS experiment on Nimbus 7

3.3.1 Introduction

In 1972 a 'Total Ozone Mapping Spectrometer' was proposed for flight on the Nimbus 7 satellite. This was envisaged as a variation of the BUV design but would only make use of the longer, higher radiance wavelengths useful for measurement of total ozone. This permitted much briefer signal integration times than required for the short, profiling wavelengths of the BUV (a factor of 60) and could be accomplished with a simple, single monochromator. The narrow $3^\circ \times 3^\circ$ field-of-view of the entrance optics of the spectrometer was swept at a right angle to the motion of the spacecraft moving in a polar orbit. By scanning out to 51° from the nadir, it was possible to view all locations between adjacent orbits. By repeating the scan at 8 second intervals, the scans were contiguous.

This approach generated a swath of ozone observations extending from pole to pole on each orbit. With a sun-synchronous orbit, all points of the sunlit earth could be viewed

near local noon during the 13 to 14 orbits in a day. The field-of-view was designed to resolve the upper air fronts associated with jet streams and corresponded to 50 km at the nadir or a swath average of 66 km from the 950 km Nimbus orbit. Because of the high albedo of the atmosphere at total ozone wavelengths it is possible to interleave the observations of the wavelengths in wavelength pairs to remove the noise introduced by motion of the field-of-view across cloud edges, and thus retain high precision in the ozone determination.

Nimbus 7 was launched on 24 October 1978 and TOMS data were obtained on November 1. The orbit of the Nimbus 7 satellite is sun-synchronous such that the equator transit time is always near local noon. In spite of the concern for the adequacy of an atmospheric model for generation of tables for retrieval of total ozone values, the data exhibited no significant scan angle dependence, thus demonstrating convincingly the efficacy of the Dave multiple Rayleigh scattering computer code. In addition, the model-generated albedos were segregated for clear and cloud sky conditions with assumed surface pressures for both. The methodology for discrimination of the two cases in actual albedo data worked so well that cloud errors in the total ozone were invisible in the broader scale structure of the meteorological forcing.

This experiment failed on 6 May 1993, it therefore provided daily global mapping of total column ozone over the entire surface of the earth outside the regions of polar night, for over 13 years. The TOMS instrument shared the front end optics and some supporting electronics with the SBUV instrument. The SBUV provided vertical ozone profiles over the altitude range 25-50 km, and was in operation continuously from October 1978 until 13 February 1987 when the dark current chopper went out of synchronization with the measuring optics. For a period after February 1987 correction algorithms could be applied to correct this non-synchronization, extending the SBUV data set to June 1990.

A second TOMS instrument was launched on the Russian Meteor-3 satellite on 15 August 1991. However, satellite orbit parameters prevent the Meteor-3 TOMS providing daily global coverage of total column ozone. With the failure of TOMS on Nimbus 7 and SBUV/2 on NOAA-13, the TOMS instrument aboard Meteor-3 is the only backscatter ultraviolet instrument providing total column ozone data. However, TOMS instruments are scheduled for launch on Earth-Probe-94 in the middle of 1994, the ADEOS satellite in early 1996 and Earth-Probe-98 in the middle 1998.

The following sub-section provides details of the TOMS instrument. Once the TOMS instrument has measured the required spectra, the data must be extensively processed before total column ozone values are determined. The data retrieval algorithm will be described briefly in sub-section 3.3.3, followed by some comments on shortfalls in the algorithm and corrections that have been applied. After the raw data have been

reduced, the resultant data set is made available to users in a gridded format. A short discussion on data quality follows in sub-section 3.3.4.

Several TOMS data versions have been released by NASA. A new version is released whenever improvements in data reduction algorithm has been applied or, corrections to the instrument drift. Versions 1 to 3 were not released as these were used in the original development of the data reduction algorithm. Version 4 was the first operational algorithm used for providing global total column ozone data. The version 5 data retrieval algorithm was updated after it was discovered that the TOMS data had drifted by 3% with respect to the Dobson network. This resulted in version 6 which is the most recent version and is discussed in some detail in sub-section 3.3.3. The author has made extensive use of the version 6 TOMS data in the work of this thesis. However, some studies, performed before the version 6 data release, use version 5 data. It is envisaged that version 7 data will be released by NASA at the end of 1994.

3.3.2 The TOMS instrument

TOMS is an Ebert-Fastie monochromator that scans across the nadir track to provide a global map of deduced ozone each day. It measures the ultraviolet albedo of the earth at six wavelengths. Four of these (312.5, 317.5, 331.2 and 339.8 nm) are used in pairs for the measurement of total ozone in combination with surface reflectivity measurements at 360 and 380 nm (*Stolarski et al., 1991*).

Figure 3.1 is a schematic illustrating the mode of operation of the TOMS instrument. Total column ozone measurements by TOMS are achieved by means of a cross-course mapping. The ultraviolet spectrometer, which $3^{o} \times 3^{o}$ instantaneous field of view, is swept at right angles to the direction of spacecraft motion (*Heath et al., 1975*), as depicted in Figure 3.1. In this field of view, which translates to a spatial extent of 50×50 km in the nadir and 250×250 km at extreme off-nadir, up to 200 000 measurements are recorded. Irradiances are sampled in 3^{o} steps in the $\pm 51^{o}$ cross-scan from the nadir. This means that the ground track is sampled 35 times every 7.95 seconds (*Heath et al., 1975*), following which there is a 1 second period for retrace. The orbital period of the Nimbus 7 spacecraft is 104 minutes, and so, successive cross-course scan lines are displaced about 50 km apart along the orbital path. As the earth rotates underneath the satellite, each sun-synchronous orbit is located 26^{o}W of the preceding orbital path. This scan geometry provides coverage over the entire earth in under one day.

The optics diagram for the TOMS instrument is given in Figure 3.2. The scanning mirror scans out to 51^{o} across the orbital path. Backscattered ultraviolet radiation enters the photomultiplier tube. The monochromator has a fixed grating and an array

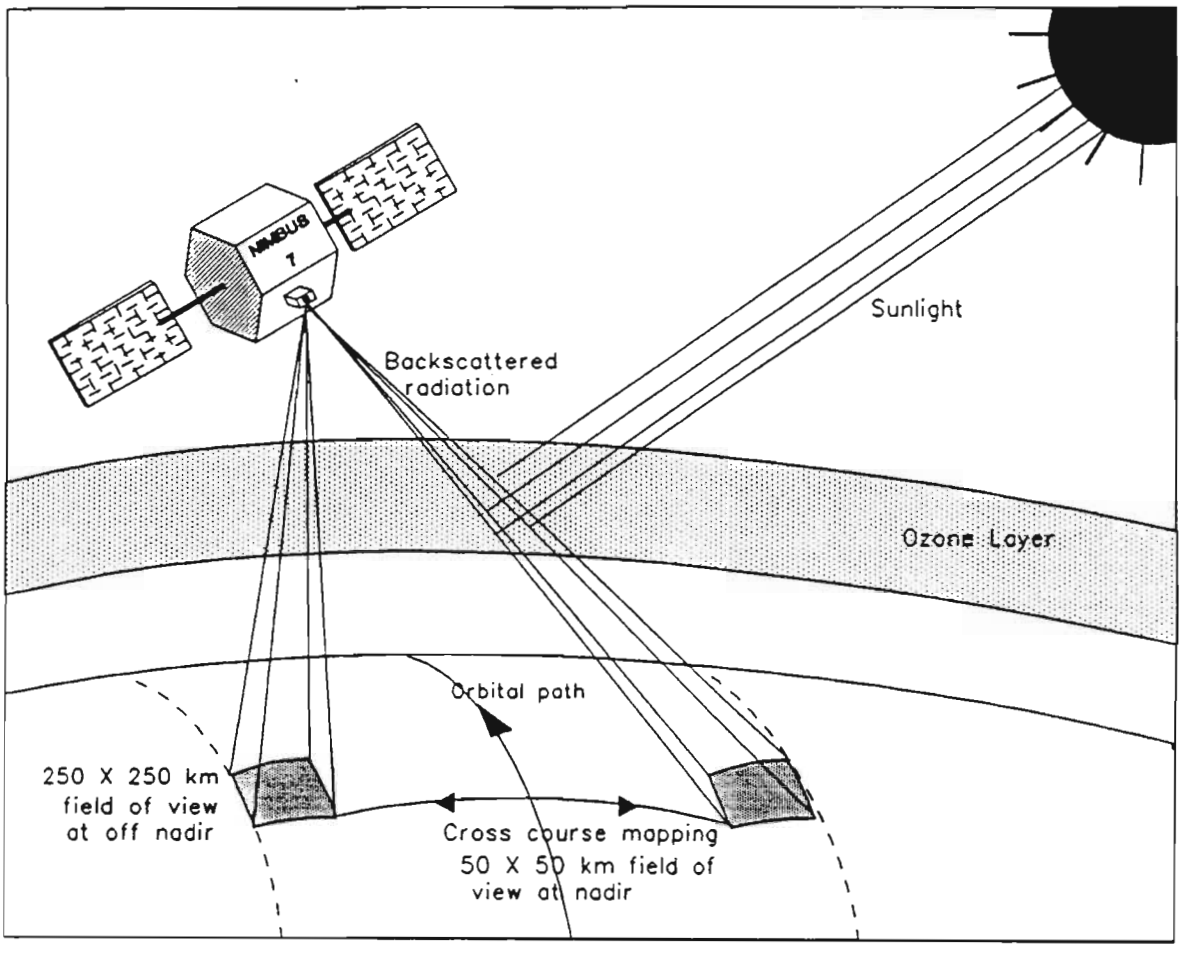

Figure 3.1: The viewing geometry of the TOMS instrument aboard the Nimbus 7 spacecraft. From *Bodeker (1992)*.

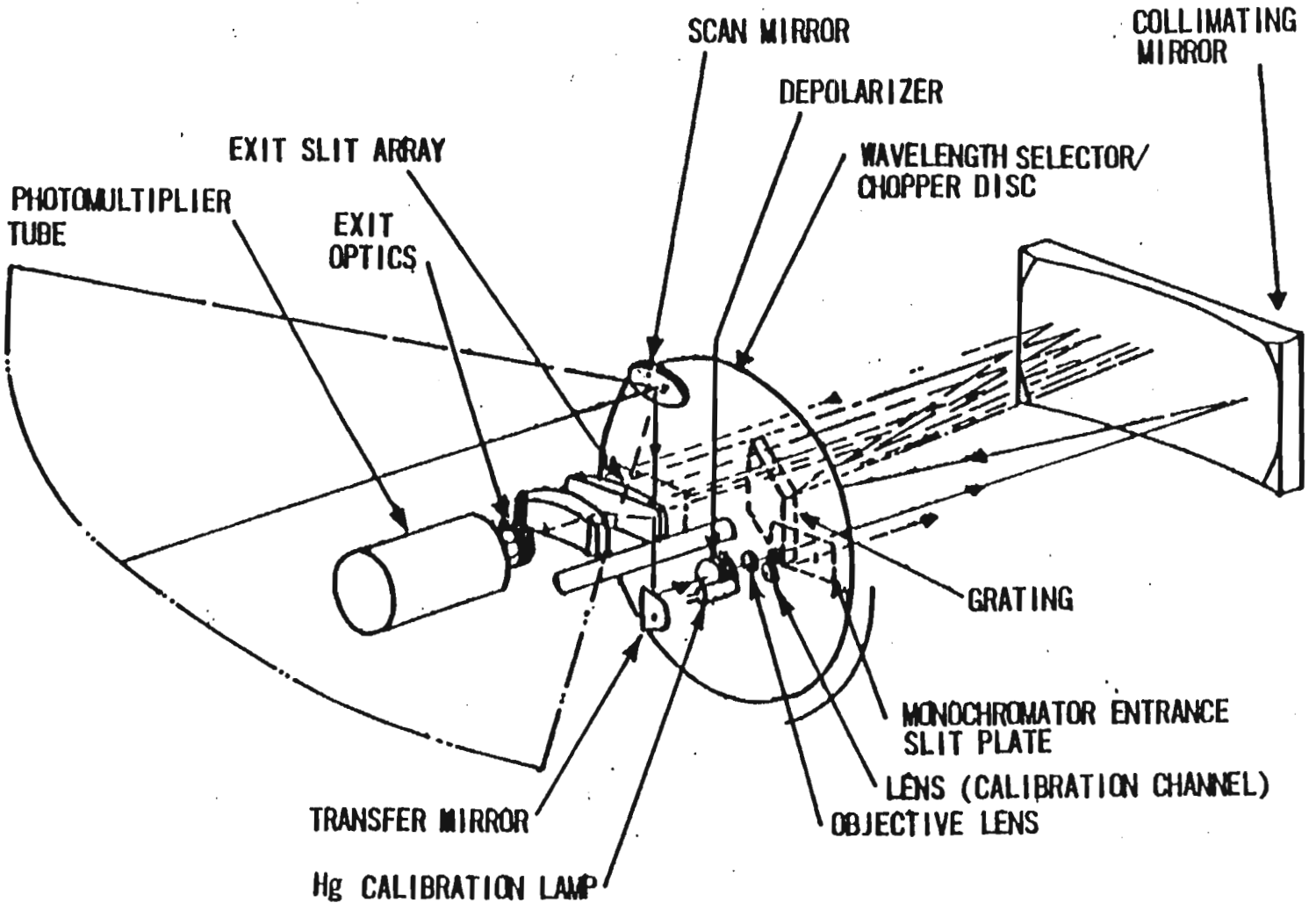

Figure 3.2: The TOMS optics. From *Nimbus 7 users' guide (1978)*.

of exit slits. A rotating wavelength selector is used to guide the dispersed light to the desired exit slit and on to the detector. The wavelength selector disc also has a set of entrance slits for wavelength calibrations using a Hg calibration lamp. A set of holes (exit slits for the 317.5 nm band) in the wavelength selector serve as a fixed exit during wavelength calibration. The TOMS scanner is stepped up to six degrees over a horizontal position during solar irradiance measurement to view the diffuser plate (not shown in Figure 3.2). The cloud cover photometer is separate from the monochromator, but uses the same chopper disc and monitors the same field of view.

3.3.3 TOMS data retrieval algorithm

The TOMS total ozone retrieval algorithm is a table look-up and interpolation process. The pre-computed table contains backscattered radiance as a function of total ozone, optical slant path length, surface pressure, surface reflectivity and latitude. Table values are computed based on a set of assumed climatological ozone and temperature profiles. The TOMS data retrieval algorithm has been updated with the release of each new version of TOMS data. In addition, the diffuser plate is exposed and as a result is subject to bombardment of by energetic particles which leads to degradation of the plate's reflectivity. This results in an underestimation of ozone amounts and corrections for this drift are incorporated in new data versions. In particular, a wavelength pair technique, described below, was the major motivation for releasing version 6 data.

The wavelength pair justification method (PJM)

The main disadvantage of TOMS, or any other satellite, is the difficulty in maintaining calibration over time. A method has been developed to determine the calibration drift of the TOMS instrument (*Herman et al., 1991*) based on a requirement of internal consistency in ozone measured with different wavelength pairs. The method produces a calibrated total ozone data set which is independent of the Dobson network. The re-processed data, called version 6, were available since 1991 from National Space Science Data Centre (NSSDC). Evaluation of the propagation of errors through the calibration and analysis indicates that the ozone measurements are precise to $\pm 1.3\%$ (2σ) at the end of the record relative to the beginning of the record (*Herman et al., 1991*). This is confirmed by comparison to the World Standard Dobson Instrument (#83) during satellite overpasses, and by comparison to a composite of stations in the Dobson network (*McPeters and Komhyr, 1991*).

Since different wavelengths used to measure ozone show different sensitivity to calibration error, the long term drift between total ozone measured at two different wavelength

pairs was used to determine the relative calibration error in each pair. By assuming that the sea surface reflectivity is stable in the long term, the less critical absolute calibration of the longer reflectivity wavelengths can be inferred. This approach is the Pair Justification Method (PJM) and provides long term calibration for TOMS that is accurate to about 1% in total ozone over a period of a decade.

The four shorter wavelengths monitored by TOMS (312.5 nm, 317.5 nm. 331.2 nm and 339.8 nm) are combined into pairs (*Herman et al., 1991*), listed in Table 3.4. Ratios of the radiances for each pair provide 'N values'. These N values are compared with the table of theoretical N values in the TOMS total ozone retrieval algorithm, calculated from *à priori* ozone and temperature profiles. Measured N values can be converted into total ozone amounts by interpolation between adjacent values in the table. Data quality may be ensured by the requirement that calculations from each wavelength pair yield the same total column ozone amount. Based on the need to minimize errors in the reduction algorithm, the A and B′ pairs were selected for total column ozone estimation over most of the earth's surface (*Herman et al., 1991*). However, under large solar zenith angle conditions the C pair is preferential since at these slightly longer wavelengths greater penetration into the atmosphere is achieved.

Errors in data retrieval

- **Satellite attitude** Errors have been identified in the computation of the Nimbus 7 spacecraft attitude. These errors affect the determination of the viewing geometry associated with the TOMS instrument, but they are so small that significant error in total ozone determination occurs only at the extreme off-nadir scan positions where any miscalculation of spacecraft roll angle affects the path length determined for the backscattered radiation. The average equatorial cross-track bias has been computed to be ~2 DU, which added to computed sensitivities in spacecraft roll angle determination, is used in version 6 to estimate a roll angle correction to the original attitude determination. Therefore, attitude associated errors in the version 6 product should be small.

- **Climatology** Due to the extreme conditions which have developed over the Antarctic continent during the latter part of TOMS lifetime, it has been necessary to expand the set of standard profiles to improve the TOMS retrieval during ozone hole conditions. The version 5 ozone profiles misrepresent the shape of the typical ozone hole profile and over-predict the 'depth' of the ozone hole, by 4-7 DU at 76° solar zenith angle and 4-10 DU at 86° solar zenith angle. The version 5 temperature profiles are too warm and have the effect of over-predicting the 'depth' of the ozone hole by 3-3.5DU at 76° solar zenith angle and 0.5-1.5 DU at 86°. In addition, the drift of the version 5 calibration also tends to over-

wavelength pair (nm)	name	seperation (nm)	toggling error (DU)
312.5/331.2	A	18.8	0.3
317.5/331.2	B	13.8	5.0
331.2/339.8	C	8.5	13.0
312.5/317.5	A'	5.0	
317.5/339.8	B'	22.3	3.0

Table 3.4: The wavelength pairs used in the determination of total ozone by TOMS. The impact of toggling on each pair is listed in the extreme right column of the table.

estimate the depth of the ozone hole by ~6 DU. The PJM in version 6 corrects the calibration error. Furthermore, the version 6 look-up tables have been extended to include typical ozone hole and temperature profiles, so that the errors outlined above are minimized.

- **Polar stratospheric clouds** A significant error source, which remains in the version 6 data set, is the effect of the presence of anomalously high clouds in the Antarctic region. PSC's above the ozone maximum may cause an underestimation of the TOMS retrieved total ozone at large solar zenith angles ($\chi > 70^{o}$). Preliminary modelling results indicate that NAT PSC's may produce an underestimation of up to 2% at solar zenith angles greater than 80°. Larger errors (up to 6%) may be introduced by water ice PSC's, especially those formed by lee waves. Errors due to PSC's have not been corrected in version 6 however, they tend to be localized in time and space.

- **Synchronization** Starting in early 1984, the TOMS instrument began to develop lack of synchronization between its wavelength selection/chopper wheel and the photon counting electronics (*Fleig et al., 1986*). This condition is identified by a threshold sensor as 'out of sync', and all of the TOMS data flagged in this way are removed from the data set. The rate of rejection is very low, and so no significant effect on TOMS data is predicted. However, after 1984 the TOMS instrument was found to operate in two distinct modes, one of which is normal and the other appears to be associated with a sub-threshold, non-sync condition, termed toggling. This condition occurs sporadically on short time scales (less than one day) and results in a calibration error of the earth's albedo, as determined by TOMS. No attempt has been made in the version 6 algorithm to correct individual measurements for this effect. However, some wavelengths are more seriously affected by toggling than others. The toggling induced errors on the different wavelength pairs A, B, C, B′ are listed on the extreme right of Table 3.4. The B′ pair, last entry in Table 3.4, is less affected than the B pair used in version 5, and so the B′ pair is used in the version 6 processing. The reader's attention is drawn to the relatively large error (13 DU) on the C pair used in total ozone estimation during high solar zenith angle conditions.

- **Large solar zenith angles** Comparisons of total ozone measured by the Système d'Analyse par Observations Zénithales (SAOZ) instrument at Dumont d'Urville, Antarctica and TOMS reveal some discrepancies at large solar zenith angles (*Pommereau et al., 1989*). A TOMS error of this type is caused by sensitivity of the TOMS algorithm to profile shape at large solar zenith angles were penetration to the ground is not good (*Klenk et al., 1982*). Zenith angle dependent problems in the TOMS data should be confined to the region where measurements are made at solar zenith angles greater than 80° for moderate to large total column

ozone amounts. These errors are estimated using sensitivities derived from radiative transfer calculations and measurements of the actual ozone profile from Solar Backscatter Ultraviolet (SBUV) and balloonsonde. At the maximum retrieval solar zenith angle of 88°, these calculations indicate that TOMS long term ozone depletions are overestimated by 5% per decade (*Wellemeyer et al., 1993*). However, the differential impact of the profile shape error at the various TOMS wavelength pairs indicates that profile shape information is present in the TOMS measurements at high solar zenith angles (*Wellemeyer et al., 1993*). This information can be used to improve the TOMS retrieval algorithm at high latitude leading to, perhaps, a version 7 data set.

3.3.4 Data quality

The primary means of quality control of the TOMS data set is by means of comparison with a network of 39 ground-based Dobson spectrophotometers. Such a comparison in 1984 revealed that the precision of TOMS data retrievals to be better than 2% (*Bhartia et al., 1984*). *Fleig et al. (1986)* reported that TOMS version 5 data indicated no detectable degradation (r.m.s. error less than 1%) when compared with the Dobson network. However, similar comparisons revealed that by 1988 the TOMS data had declined by 3% with respect to the Dobson network (*Heath, 1988; Reinsel et al., 1988*). This calibration drift is attributed to the degradation of the diffuser plate. The PJM corrected for this and the version 6 data was released. Analysis indicates that version 6 ozone measurements are precise to ± 1.3% over the lifetime of the TOMS instrument (*Herman et al., 1991*). *McPeters and Komhyr (1991)* have confirmed that the version 6 data is stable with respect to the world standard Dobson spectrophotometer.

3.4 The GOES satellite system

The Geostationary Operational Environmental satellites (GOES-1, GOES-2, etc.) all carry on board the Space Environment Monitor (SEM) instrument package. The SEM has magnetometer, energetic particle, and soft X-ray detectors, each of these will be described, in turn, in subsequent sub-sections. Two of the GOES spacecraft usually operate simultaneously in earth's equatorial plane. They travel in geosynchronous orbit (6.67 $R_{\oplus}$) and are generally located at or between 75° and 135°W. This trajectory allows the detectors an unobstructed view of the sun for all but the few dozen hours per year when the earth eclipses the sun.

The data are transmitted via direct telemetry to the Space Environment Laboratory

Data Acquisition and Display System (SELDADS) in Boulder, Colorado. and are available through the Solar-Terrestrial Physics Division of the National Geophysical Data Centre in Boulder, known internationally as World Data Centre A for Solar-Terrestrial Physics.

The time of these observations has not been corrected for the down-link and preprocessing delays. The Space Environment Laboratory has not made an accurate determination of this delay, but it may be assumed to be within the range 1-5 seconds.

Figures 4.3, 4.12, 4.13, 4.14 and 4.15 are plots, for selected periods, of the data provided by these instruments aboard the GOES-6 satellite.

3.4.1 The magnetometer

A twin-fluxgate spinning sensor allows earth's magnetic field to be described by three mutually perpendicular components: H_p, H_e and H_n. H_p is parallel to the satellite spin axis, which is in turn perpendicular to the satellite's orbital plane. H_e lies parallel to the satellite-earth centre line and points earthward. H_n is perpendicular to both H_p and H_e, and points eastward for GOES-6 and -7. Field strength changes as small as 0.2 nT can be measured.

The magnetometer samples the field every 0.75 seconds. Four of these values constitute a frame and are sent to the ground station together. The high and low values in the frame are discarded, and the average of the two remaining values is recorded. No record is kept of which of the four values are used in the archive and therefore some uncertainty exists in the exact time of the recorded observation.

3.4.2 The X-ray sensor (XRS)

Ion chamber detectors provide entire sun X-ray fluxes for the 0.5-4 Åand 1-8 Åwavelength bands. These observations provide a sensitive means of detecting the start of solar flares. Two bands are measured to allow the hardness of the solar spectrum to be estimated.

X-ray photons pass through a collimator, which defines the view aperture, followed by a thin metallic window, which defines the low energy threshold, before entering the ion chamber. The XRS viewing direction is in the meridian of the spacecraft spin axis. Dynamic positioning of the XRS elevation provides for maintaining the sun in the swept field. The X-ray emission of the sun is determined once during each spin.

The spin period is 0.6 seconds and the data for both bands are given in W cm^{-2} s^{-1}.

3.4.3 The energetic particle sensor (EPS)

Solid-state detectors with pulse-height discrimination measure proton, alpha particle, and electron fluxes. The field of view of the EPS is perpendicular to the GOES spin axis which is approximately aligned with earth's rotation axis. Since the satellite spin period (0.6 seconds) is much shorter than the accumulation times, the EPS provides a spin-averaged estimate of the local high pitch angle particle fluxes. The integral electron channel data are given for an energy threshold of >2 MeV with units of counts cm^{-2} s^{-1} $ster^{-1}$. Proton channels detect protons in the following categories, >1 MeV, >5 MeV, >10 MeV, >30 MeV, >50 MeV, >100 MeV.

Because the spacecraft travel in a geostationary orbit, the electron and the lowest energy threshold (>1 MeV) proton channels are responding primarily to trapped outer-zone particles. The >5 MeV proton channel may occasionally respond to trapped particles as well, during magnetically disturbed conditions. The geomagnetic cut-off at geostationary orbit is typically of the order of a few MeV as indicated by the lack of trapped particle >5 MeV proton channel response except under disturbed conditions. Therefore, the remaining, higher energy threshold, proton channels measure fluxes originating outside the magnetosphere i.e solar and galactic cosmic particles.

Significant secondary responses have been found to exist in the particle data, i.e. responses from other particles and energies. However, the 5-minute averaged data, employed in the work in chapter 4, have been corrected for these responses.

3.5 Daily solar flux values observed at Ottawa

Daily observations of the 2800 MHz (10.7 cm) radio emissions that originate from the solar disk and from any active regions are made at the Algonquin Radio Observatory (ARO) of the National Research Council of Canada in Ottawa (45.5°N;75.7°W), with a reflector of 1.8 m diameter. These data are published regularly in the *Solar-Geophysical Data prompt reports*. Numerical values of flux in the tables are in units of 10^{-22} W m^{-2} Hz^{-1} and refer to a single calibration made near local noon. When the flux changes rapidly, or when there is a burst in progress at that time, the reported value is flagged and the best estimate of the undisturbed level, provides the reference level for measuring the burst intensity.

The observed flux values have variations resulting from the eccentric orbit of the earth

in its annual path around the sun. Although these radio values are suitable to use with observed ionospheric and other data, an adjustment must be introduced when the observations are used in studies of the absolute or intrinsic variation of the solar radio flux. Therefore, the flux data that are adjusted to 1 AU are used for the work presented in chapter 5. The observations are made for a single north-south polarization but are reduced under the assumption of two equal orthogonal polarizations.

A graph showing the monthly mean adjusted flux for the period 1947 to 1986 is shown in Figure 3.3. The solar cycle number appears below the corresponding eleven year interval. Relative errors over long periods are believed to be $\pm 2\%$. The characteristics of the observations are surveyed by *Covington (1969)*. Values of the quiet sun for the minima are around 60-70 W m^{-2} Hz^{-1}. During solar maximum years the solar radio flux is very variable, values over 300 W m^{-2} Hz^{-1} correspond to exceptional solar activity levels.

Experiments have indicated that a multiplying factor of 0.90 should be applied to the reported flux values in order to derive the absolute flux values, the published flux values are not corrected by this factor because of the number of computerized data series listing these values. Maintaining homogeneity of the published series is considered more important than having the absolute flux values published. A review of the history of the absolute calibration of the Ottawa series, as well as a number of other series of observations made within the microwave region, was prepared by H. Tanaka of the Research Institute of Atmospherics, Nagoya University, as convener of a Working Group of then Comm. 5 of URSI.

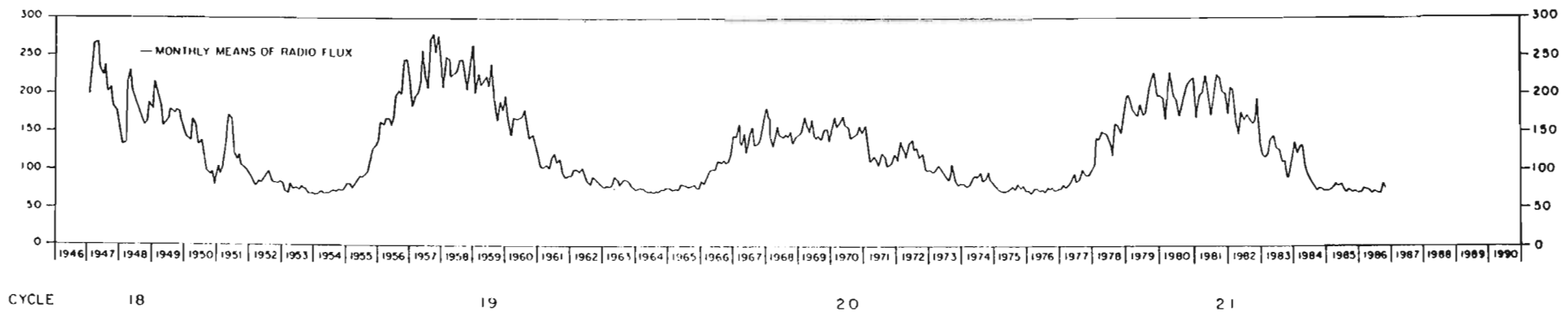

Figure 3.3: The solar radio flux (19.7 cm) adjusted to 1 AU, for the period 1967 to 1986. From *Solar geophysical data (1987).*

3.6 Conclusion

The concepts of the chapter that are important in the work described in chapters 4 and 5 of this thesis are now summarized.

- The time and spatial resolution of the TOMS data is ideal for monitoring short (days to months) or long (months to years) term variations in total ozone over large areas of the earth. In addition, simultaneous observations over both polar regions during equinox provides a unique opportunity to monitor responses over these regions to well defined events e.g. SPE's. However, vertical ozone profile measurements would be very important, if available. The correction for calibration drift made to the version 6 data allows us to use the TOMS data with confidence for the evaluation of long term ozone variations, as is the case of the Fourier analysis performed in chapter 5.

- The magnetometer, X-ray and energetic particle data from the GOES-6 and -7 spacecraft provide valuable parameters of conditions in the near-earth environment during the SPE's that form the case studies in chapter 4. Unfortunately, the particle energy spectrum above the energy threshold 100 MeV, which is crucial in estimating ozone loss during SPE's, is unavailable.

- The solar radio flux data has been shown to correlate particularly well with solar ultraviolet output and is often used as a proxy to characterize the solar ultraviolet flux for computation of atmospheric effects (*Walterscheid, 1989*). Furthermore, it's accuracy over long periods is very good and is estimated at about 2%.

Chapter 4

Satellite observations of polar ozone following solar proton events

4.1 Introduction

One of the most direct influences of solar activity on the near-earth space environment occurs when intense solar flares produce very energetic (10 MeV to over 1 GeV) solar particle emissions. The first observations of solar particles reaching the Earth's surface were made by *Forbush (1946)*.

The SPE's of 1989, and associated geomagnetic activity, represent some of the most extreme recorded effects in the near-earth environment (*Gorney, 1990*). A power outage in eastern Canada caused by the extraordinary geomagnetic event in March 1989 affected approximately 6 million people for over 9 hours, underlining the practical importance of these phenomena. The maximum in solar cycle 22 is one of the most significant in the space age with a mean value of the 10.7 cm solar radio flux of 213.5 $\times 10^{-22}$ W m^{-2} Hz^{-1} in 1989, its highest value since the International Geophysical Year (IGY) 1957.

The frequency of occurrence of optical and X-ray flares, which are a prerequisite for the production of SPE's, correlates with the 11 year solar cycle (*Hirman et al., 1988*). SPE occurrence tends to peak within a period extending from two years before to four years after the sunspot maximum, and is greatly diminished during the few years surrounding sunspot minimum. In 1989, just prior to the solar maximum of solar cycle 22, twenty three SPE's were recorded, the most ever in a single year (*Ranta et al., 1993*). Six of these events have proton (>10 MeV) fluxes of >1000 particles cm^{-2} s^{-1} ster^{-1}. These events are clearly visible as sharp peaks of proton flux in Figure 4.1.

Onset day numbers are 67, 76, 224, 272, 292 and 334. The six SPE's in 1989, combined with the August 1972 SPE represent the most severe events ever observed. Together, their total proton fluences (time integrated flux) exceed the sum of all other events which occurred over the last twenty years. For many years, until 1989, the August 1972 SPE's were considered as representative of a worst case.

The overall importance of an individual solar proton event depends on the maximum flux intensity of the event (especially that of the higher energy protons which penetrate to lower altitudes where ozone is most abundant), and the length of the event (which establishes the fluence). The middle panels of Figures 4.3, 4.12, 4.13, 4.14 and 4.15 show the time evolution of the 1989 SPE's observed by a geosynchronous satellite. Details of the time evolution profile of each event depend on characteristics of the originating solar flare(s), the time scale associated with diffusion of the energetic particles within the solar corona, and the propagation of the particles within the interplanetary medium, described in sub-section 2.8.2. Although energetic particles can be detected for over a week during a SPE, the temporal response of the middle atmosphere will be different, depending on the chemical lifetime of the species produced by the SPE. SPE's produce, via processes described in sub-section 2.8.2, long-lived enhancements of up to one year in NO_x species which may result in ozone destruction. The chemical lifetime of NO_x (and all other species that can be photolyzed) depends on the amount of solar irradiation it receives and will therefore be shorter in the summer season and longer in winter. This effect is particularly important over the poles where solar irradiation levels can vary from complete darkness to complete daylight conditions. In addition, downward meridional circulation over the poles in winter will further extend the chemical lifetime of the NO_x enhancements (Figure 2.12 shows NO_x chemical lifetimes as a function of altitude) and transport to lower altitudes where ozone is more abundant. Conversely, meridional circulation is directed upwards over the summer pole, so reducing the ozone destruction potential of NO_x species.

The IMF can only merge or connect with the geomagnetic field when it has a large component in a direction opposite (i.e. south pointing) to that of the geomagnetic field (see Figure 2.18 (a) and (b)). Thus, a southward IMF is a prerequisite for effective deposition of charged particles over the polar caps. The sustained intervals of an IMF with a significant southward component is well correlated with the onset of geomagnetic storms (*Gorney, 1990*). The more intense the geomagnetic activity, the greater the region over which the open field lines extend over the polar caps. Equatorward extension of open field lines during periods of high geomagnetic activity can be visually traced by monitoring latitudes at which discrete auroral forms appear. The open field lines, which are connected to the IMF (as long as it is has a southward pointing component) are frozen into the solar wind and carried along with it. Consequently the open field lines are 'dragged' over the polar regions from the day side to the night side. This

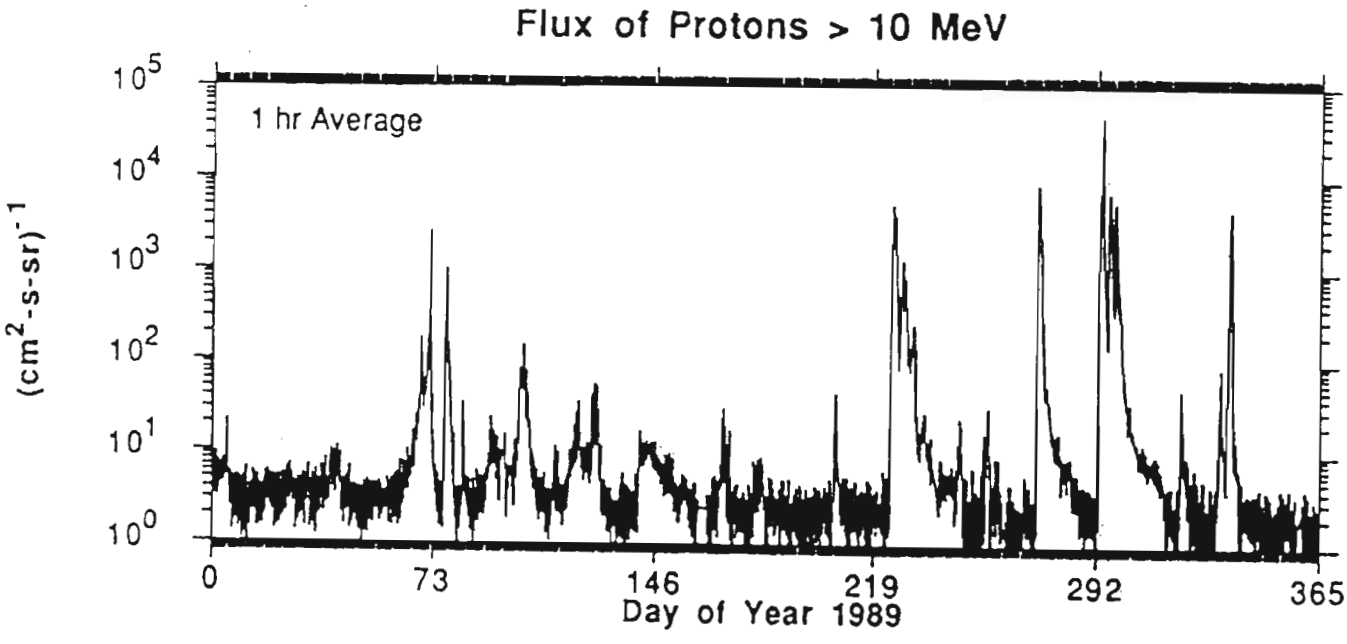

Figure 4.1: Hourly proton (> 10 MeV) flux measurements by the 1987-097 spacecraft during 1989. There are six periods over which the flux exceeded 10^3 cm^{-2} s^{-1} ster^{-1}. From *Reeves et al. (1992).*

is schematically represented in Figure 2.18(b) by assuming that each open field line represents successive positions of that field line. Therefore, the greater the geomagnetic activity, the greater the distortion of the geomagnetic field, such that open field lines cover a more extensive area over the poles. This allows for particle deposition to lower latitudes. Furthermore, this allows additional time for particle precipitation since the open field lines have to be 'dragged' over an effectively larger polar cap. Geomagnetic activity is therefore crucial in determining the effective precipitation of particles into the polar atmospheres.

This chapter reports on total ozone observations following all major SPE's in 1989. The two SPE's of March 1989 are analysed as case study 1 and the four SPE's occurring during the last months of 1989 are examined as case study 2. In addition, the total column ozone in the SPE of July 1982, following which measurements of NO enhancements (*McPeters, 1986*) suggest that ozone loss occurred only over the southern polar cap, the basis for case study 3. All ozone data are derived from the TOMS instrument, aboard the Nimbus 7 satellite, the format of which is described in Section 3.3. Case studies 1 and 2 use version 5 of the TOMS data, as this was the current version at the time the analysis was performed. Case study 3 employs the updated version 6 total column ozone data.

In all the case studies, the method of detecting ozone response to SPE's is original and new. A detailed description of the procedure is described in the following section. It is the first time, to the best of our knowledge, that a quantitative measure of ozone mass loss over a large area, following a SPE, has been determined. A region extending from the pole down to 70° is selected, being the area over which most particle precipitation occurs. This approach was utilized to average out any small scale, or localized, transport and chemistry which may otherwise mask the ozone response to a SPE, as would be the case for ground-based detection of ozone depletion during these events.

Wind and temperature data above Sanae, Antarctica are used to ascertain background atmospheric conditions over the southern polar cap at the time of SPE onset. Where pertinent, temperature responses to SPE's are sought. Sanae, located at 70.3°S; 2.4°W (geomagnetic latitude 60.1°S) is at the equatorward edge of the analysis region. In addition it may be assumed to be, for the most part, located within the polar vortex.

The modelled ozone response to the SPE's in the latter months of 1989 has been calculated by others e.g. *Reid et al. (1991)* and *Jackman et al. (1993)* in sub-section 2.8.2. These models are fundamental tools that can qualitatively predict ozone behaviour. However, because of their incompleteness e.g. a lack of a complete set of chemical reactions (especially relating to chlorine) and an incomplete description of polar dynamics in two-dimensional models, they cannot provide exact quantitative estimates. Three-dimensional simulations provide a closer reproduction of the photochemical and

dynamical processes which influence ozone behaviour.

4.2 Data analysis of TOMS total column ozone

4.2.1 Introduction

The operation of the TOMS experiment aboard the sun-synchronous Nimbus 7 satellite has been discussed in chapter 3. The observations of the TOMS instrument were collected and processed by the NASA Goddard Space Flight Centre. The daily global ozone data provided by NASA consist of an array of cells 1° long in latitude (180 latitude cells) and 1.25° wide in longitude (288 longitude cells), each providing a total column ozone measurement in DU. The version 5 data exhibit a 3% decline with respect to the Dobson network during the period 1979 to 1988. These data are utilized in some of the case studies that follow in this chapter as this was the current version at the time the analysis was performed. At this time the data were available from NSSDC via a telephonic link. Following the development of a correction technique, version 6 data were provided with an estimated accuracy of 1.3% over 12 years, on a compact disc prepared by NSSDC. These data are utilized in some of the case studies here and in the long term analysis described in chapter 5.

4.2.2 Ozone mass calculation

This sub-section outlines the method employed to calculate the daily ozone mass over the pole (90-70°) from TOMS total column ozone data. Calculations were performed by a Pascal program, written by the author, that runs on an IBM 486 PC.

The Ideal Gas Law, expressed by equation (1.18), has been addressed in chapter 1 of this thesis. This law may be applied to determine the density of ozone

$$\rho = \frac{\mathrm{p\,m_m}}{\mathrm{R\,T}}, \tag{4.1}$$

where ρ is the density of ozone and R is the gas constant. The pressure (p) under conditions of STP is 1 atm=1.01325 x 10^5 Pa and the temperature (T) is 273.15 K. The molar mass (m_m) of ozone is 47.997 x 10^{-3} kg mol^{-1}. By substituting these values into the Ideal Gas Law equation, the density of ozone is

$$\rho = \frac{1.01325 \times 10^5\,\mathrm{Pa} \times 47.997 \times 10^{-3}\,\mathrm{kg\,mol^{-1}}}{8.31441\,\mathrm{J\,K^{-1}\,mol^{-1}} \times 273.15\,\mathrm{K}} \tag{4.2}$$

$$\rho = 2.1414\,\mathrm{kg\,m^{-3}}. \tag{4.3}$$

The Dobson unit corresponds to the height (matm cm) of the ozone column if all the ozone within the column was at standard pressure and temperature. Therefore, we can define a new quantity, the spatial density (σ), such that

$$\sigma = \text{total column ozone value (units matm cm)} \times \rho. \tag{4.4}$$

The units of σ are kg m^{-2}. Therefore, once we have determined the spatial density of ozone above a TOMS data cell we can calculate its ozone mass by multiplying it by the area of that TOMS cell. For convenient computing, the total ozone mass of a zonal strip 1° wide is determined by adding the ozone mass contributions from each of the 288 TOM cells in that strip. The ozone mass of each zonal strip, of which there are 20 for the region 90-70° latitude, are then added.

The method for determining the spatial extent of each TOMS data cell is outlined below. Figure 4.2 is a schematic diagram of the earth. The radius of the earth, $R_\oplus$ is marked together with the radial distance, r, at latitude ϕ. The arc length, Δs, of a 1° zonal band can be determined from Euclidean geometry viz.

$$\Delta s = R_\oplus \, \Delta \phi, \tag{4.5}$$

where $\Delta\phi = 1^o = \frac{\pi}{180}$ radians. The area of a zonal band (A) at latitude ϕ is

$$A = 2\pi r \, \Delta s. \tag{4.6}$$

Substituting equation (4.5) into equation (4.6) we can write

$$A = \frac{\pi^2 R_\oplus^2}{90} \cos\phi. \tag{4.7}$$

The area of one TOMS cell (A_{TOMS}) in this zonal strip is $\frac{A}{288}$ and so

$$A_{TOMS} = \frac{\pi^2 R_\oplus^2}{90 \times 288} \cos\phi. \tag{4.8}$$

Finally, the ozone mass contained in one TOMS cell (M_{TOMS}) can be determined

$$M_{TOMS} = \text{total column ozone value (units matm cm)} \times \rho \times A_{TOMS}. \tag{4.9}$$

In order to give the reader an appreciation of the typical ozone mass values, the total ozone mass in the earth's atmosphere is estimated below. Assuming an average total column ozone value of 300 DU, the spatial density is

$$\sigma = \rho \times 300 \times 10^{-3} \text{atm} \, 10^{-2} \text{m}. \tag{4.10}$$

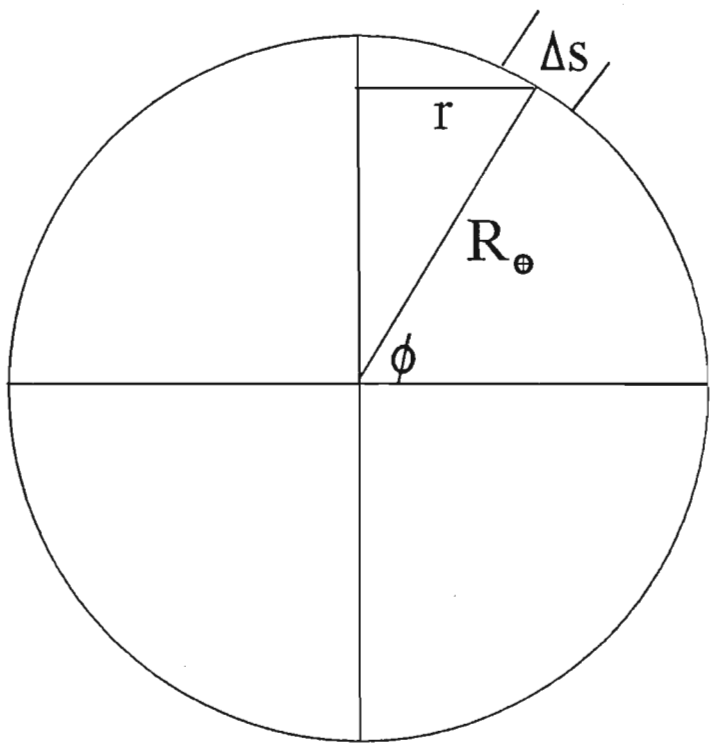

Figure 4.2: A schematic diagram illustrating the geometry employed in the ozone mass calculation. The radius of the earth is labelled $R_\oplus$ at a latitude ϕ. A curve segment ($\triangle$s) extends 1° in latitude.

Therefore, the global ozone mass (M) is

$$M = 4\pi R_{\oplus}^2 \times \sigma$$

$$M = 4\pi \times (6.37 \times 10^6\,\mathrm{m})^2 \times 2.1414\,\mathrm{kg\,m^{-3}} \times 300 \times 10^{-3}\mathrm{atm} \times 10^{-2}\mathrm{m}$$

$$M = 3.27 \times 10^{12}\,\mathrm{kg}.$$

The inability of the TOMS instrument to monitor ozone poleward of the day-night terminator, due to lack of irradiance, results in ozone data gaps over the polar regions during the winter season. These data were estimated by assuming that the zonally averaged ozone values poleward of the day-night terminator were equal to the average value of those near to the terminator. This technique has been applied by *Schoeberl et al. (1989)* to estimate polar ozone data during the dark winter period. However, additional caution is required when applying this technique to northern polar data as compared to southern polar data. There is a more frequent occurrence of large ozone maxima and minima, due to planetary wave activity, in boreal polar ozone.

4.2.3 Percentage area determination

Another method of monitoring total column ozone over the entire polar cap is the calculation of the area covered by selected total ozone contours, and expressing this as a percentage of the total area down to 70° latitude. A Pascal program was written by the author which allows the user to specify as many as 6 threshold total ozone values. The program then assesses each TOMS cell, determining whether its value is less than or equal to the threshold. If it is, it calculates the spatial extent of that TOMS cell. The areas of all TOMS cells with total column ozone values less than the threshold, are then added. Finally, it is expressed as a percentage of the area 90-70° latitude.

In instances of the unavailability of ozone during winter periods, the program will not assess these TOMS cells, but will subtract this TOMS cell area from the total area 90-70° latitude.

4.2.4 False colour maps

Computer software for the presentation of TOMS false colour maps was written by *G. E. Bodeker* of the Space Physics Research Institute (SPRI) of the University of Natal. Documentation of these programs may be found in *Bodeker (1992)*. The software, which utilizes Pascal Graphics routines, may be run on any PC that has a Video Graphics Adapter (VGA) card. The false colour maps produced by this software cover

the entire southern hemisphere (90-0°S) or, alternatively, the entire northern hemisphere. Colours associated with a range of total ozone values are determined by the software. Outlines of major landmasses are overlaid onto the colour-coded total ozone levels. Important modifications by the author, in order to customize the display of TOMS data for work described in this thesis are listed below.

- An option which allows the user to interactively select a latitude range e.g. 50-10° was included. In addition, the user can then input a 'zoom factor' which determines the physical size of the map. Map outlines and coordinates are automatically scaled. Examples of this include Figures 4.4 and 4.9 where a latitude range 90-60° has been selected.
- All aspects of the colour bar can be specified by the user. This not only includes colour choice but the size of the interval to be represented by each colour.
- The resolution of the map outlines was improved.

The TOMS false colour images were printed on a Hewlett-Packard 500C paintjet printer. Software that produces print-ready files was written by *J. Doull* of SPRI.

4.3 Case study 1: March 1989

4.3.1 Solar activity

Solar activity in March 1989 produced an historically 'great' magnetic storm (the K_p geomagnetic index attained a maximum value of 9) accompanied by some of the largest solar flares ever recorded. Solar particle flux and geomagnetic parameters were disturbed for the whole of March, allowing access of high energy particles over a very large area of the polar atmosphere for a very extended time period. Therefore data for this period in March 1989 were selected in order to assess the relationship between solar protons and ozone depletion.

Important solar parameters during the month of March period are summarized in Table 4.1. The March 1989 geomagnetic storm was preceded by a series of very intense solar flares from NOAA/USAF region 5395 viz. a 3B/X15 flare on March 6, a 4B/X4.0 flare on March 9 and a 3B/X4.5 flare on March 10 (*Solar Geophysical Data, 1989*). In addition, one or more flares were detected daily near or surpassing the X1.0 level from March 5 until March 18 (*Allen et al., 1989*). Large fluxes of high energy (> 10 MeV) protons resulted in two SPE's where up to 3500 (onset 8 March) and 2000 (onset 17

onset date of SPE	max k_p	major flare date class	proton flux > 10 MeV	fluence > 30 MeV
08 March (>9)	9	06 March 3B/X15 09 March 4B/X4.0 10 March 3B/X4.5	3500	> 9.1 X 10^6
17 March (>13)	6	17 March 2B/X6.5	2000	> 9.6 X 10^6

Table 4.1: The SPE's, and associated solar flares, during March 1989. Data from *Solar Geophysical Data (1989) and (1990, a).*

March) particles cm^{-2} s^{-1} $ster^{-1}$ were reported by the GOES-7 satellite. The duration, in days, of these events is given in brackets in the leftmost column of Table 4.1, and the estimated fluences (cm^{-2}) for these events appear in the rightmost column.

Figure 4.3 displays three data sets from the geosynchronous satellite GOES-6, located at 6.6 $R_\oplus$, for the month of March. The following description of the structure of Figure 4.3 also applies to Figures 4.12, 4.13, 4.14 and 4.15. The top panel displays X-ray emissions, originating from solar flares. On the right-hand side of this panel are the letters 'X', 'M' and 'C', which correspond to the flare descriptions in sub-section 1.4.2. Dates are marked within the top panel in which the extraordinary flares of 6, 9 and 10 of March are clearly visible. Furthermore, other 'X' flares which occurred can be identified in the upper panel of Figure 4.3, indicating the extreme activity of the sun during this period. The logarithm of the particle flux is plotted in the middle panel of Figure 4.3. Each curve represents binned counts of protons according to their detected energy threshold. The lowest flux curve represents >100 MeV, successive curves represent >60 MeV, >50MeV, >30MeV, >10MeV, >5 MeV and >1 MeV while the uppermost curve represents the flux of electrons >2 MeV. Clearly, the middle panel reveals that from about March 7 the flux of protons increases. On March 8 a SPE was officially in progress. The proton flux continued at high levels through March 14 (sustained by fresh injections from subsequent flares) with a maximum on March 13. Another SPE was announced on March 17, whose particle source was the March 10 flare and further maxima in the proton flux were measured by GOES-6 on March 18 and 19. Examination of the spacing between proton flux profiles, particularly during the first SPE, suggests that the spectra of these events cannot be accurately represented by an exponential function (e.g. equation (2.47)) as is customary when calculating ion pair production for these events. The proton flux profile for protons of energies > 100 MeV is not determined by the monitor aboard the satellite. This profile is deduced by simple extrapolation of the exponential-rigidity relation. The ozone column depletion is strongly influenced by the flux of protons with energies above the highest measured threshold of 100 MeV so that if this were significantly harder than the spectrum between 60 and 100 MeV, the ozone column depletion would be considerably larger (*Reid et al., 1991*). The bottom panel in Figure 4.3 shows the GOES-6 observation of the Earth's magnetic field (H). The H_p component (approximately parallel to the earth's rotation axis) is seen to be reversed many times during March, especially around the 13 and 14. Such negative values of H_p indicate that the magnetopause, typically located at 10 $R_\oplus$, moved inside the geostationary orbit (6.6 $R_\oplus$). Such rare events are called Geostationary Magnetopause Crossings (GMC) and are caused by extreme conditions of solar wind pressure, coupled with a strongly southward IMF. This implies that the area over which solar particles could be deposited was extensive ensuring effective precipitation of the proton population on open field lines into the polar caps.

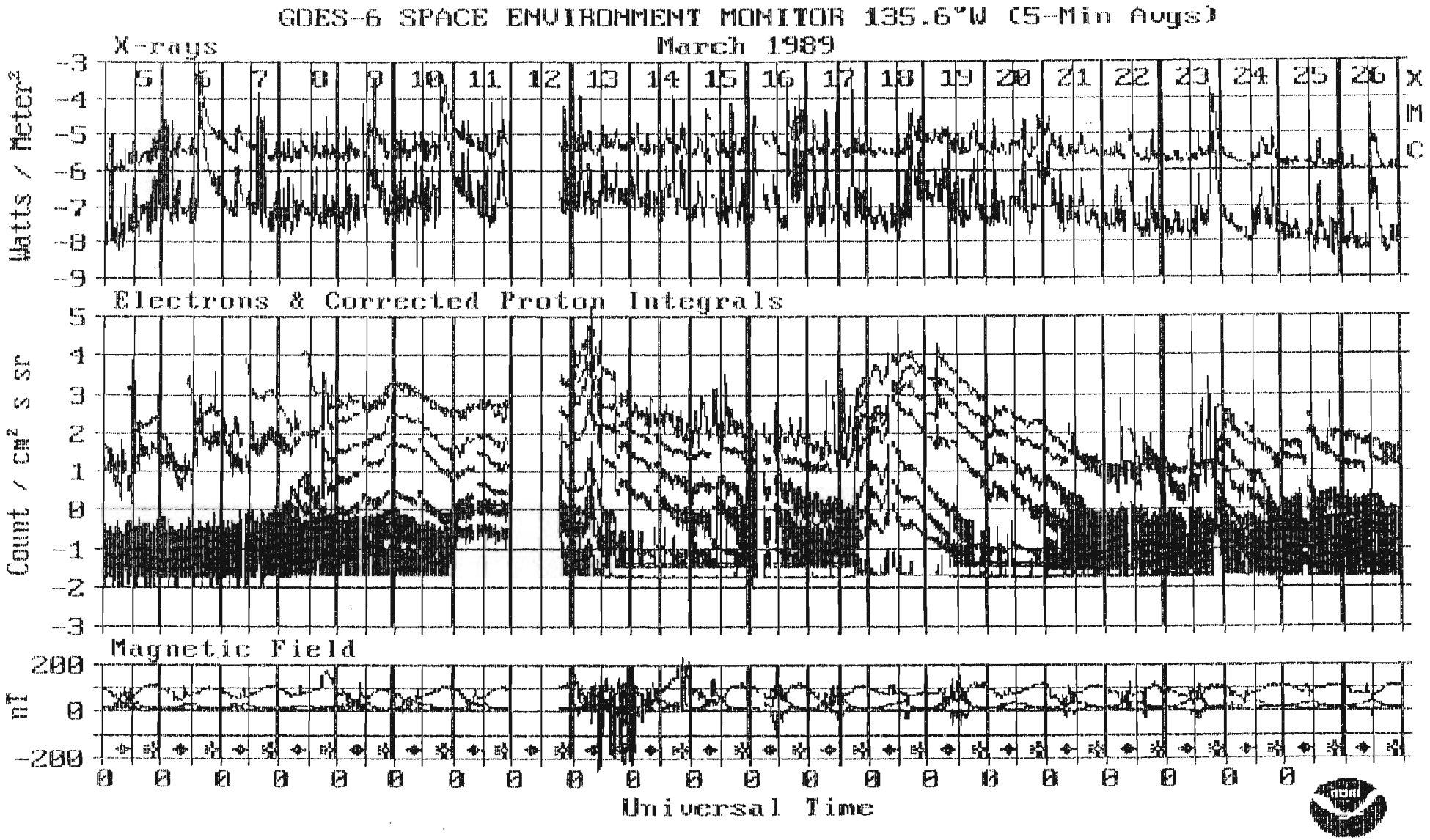

Figure 4.3: X-ray (top panel), integrated particle flux (middle panel) and magnetic components (bottom panel) observed by the SEM instrument package aboard the GOES-6 spacecraft for the period 5 to 26 March 1989.

4.3.2 Estimation of NO production

The method of determining odd nitrogen production by solar protons and the modelled atmospheric response to several SPE's are addressed in sub-section 2.8.2. Ideally, ionization rates should be determined in the same fashion for this case study, but we do not have the relevant computer programs at this stage. Ion pair production calculations have been preformed for this event (*C. H. Jackman*, private communication) but these profiles do not correspond well with observations. Therefore some doubt may be cast on the accuracy of the particle spectrum for this event. Nevertheless, the evidence presented in chapter 2 demonstrates the importance of large SPE's when considering NO_x abundances. *Crutzen et al. (1975)* showed that 6×10^{15} molecules cm^{-2} of nitric oxide were produced during the SPE's of August 1972. This figure is comparable to the total annual production due to oxidation of nitrous oxide (N_2O). Comparisons with previous events should be viewed with caution as high solar activity levels can produce, quite apart from solar protons, features of importance for ozone depletion. Natural phenomena, such as SPE's, are not reproducible. Therefore, the results of this case study are important, despite the unavailability of some parameters. Furthermore, the SPE's appeared during equinox (March) which allowed the rare opportunity to monitor the responses of both hemispheres in sunlight. For any other time of the year, except September, there are large gaps in ozone data over one of the polar regions due to lack of solar irradiance which is the basic requirement for TOMS measurements. It has been calculated that proton flux into the two hemispheres should be roughly equal (*McPeters et al., 1981*), so that symmetric ozone responses over the two hemispheres might be expected. The favourable time of occurrence of the large SPE's in March allows us to examine this claim.

4.3.3 Southern polar cap ozone response

To investigate any possible association between total column ozone and solar protons, the satellite ozone data were processed from March 5, the day before the two SPEs, through to March 31. In the results that follow, the total column ozone for March 5 is used as a baseline from which to gauge any depletion effects from the SPEs beginning on March 8. Figure 4.4 shows a selection of colour plots of total column ozone, over an area extending from the south pole to 60°S, for March 5 (top), 14 (middle) and 21 (bottom). Each hue of the colour bar represents a range of 20 Dobson units, the lowest range being 180 to 199 DU. Attention is drawn to the region down to latitude 70°S, in which there is a marked increase after 5 March of the area covered by the deep blue colour corresponding to decreasing total column ozone. The other days between 5 and 21 March were also analysed and are consistent with a gradual increase in the area of

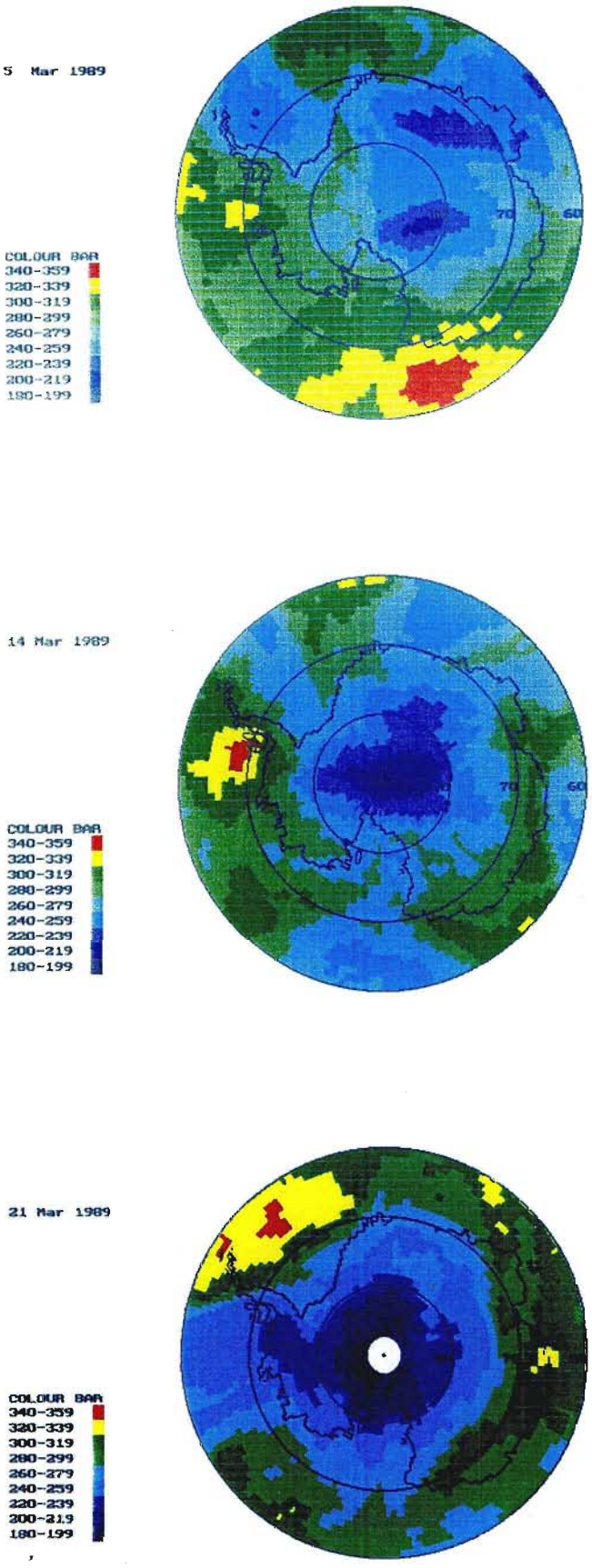

Figure 4.4: Colour plots of total column ozone from the south pole to 60°S. Upper panel, before solar activity (5 March); middle and lower panels, during SPE's (14 and 21 March respectively). The colour bar shows low total column ozone as blue, increasing through green and yellow to red. Each hue represents 20 DU. Latitudes 80°S, 70°S and 60°S are shown, together with an outline of Antarctica.

ozone depletion.

This general decrease in total column ozone can be shown quantitatively by calculating the area covered by low total column ozone and expressing this as a percentage of the total area down to 70°S. These percentages are shown in Figure 4.5 for low total column ozone in the ranges < 200 DU, 200-219 DU and 220-239 DU. The increase in percentage area covered by low total ozone is sustained until the 26 March for the two lowest total column ozone ranges. We have also calculated the total mass of ozone (at STP), from the pole down to 70°S, taking into account the cosine dependence of the cell size of TOMS with latitude. The total mass decreased from 8.85 $\times 10^{10}$ kg on 5 March to 8.11 $\times 10^{10}$ kg on 21 March, representing a decrease of 7.4 $\times 10^{9}$ kg, or about 9%. An average total mass of ozone down to 70°S has also been calculated for the period 1 to 21 March in 1987 and 1988. These averages with standard deviations are, respectively, 9.26 $\pm$ 0.24 $\times 10^{10}$ kg and 9.12 $\pm$ 0.29 $\times 10^{10}$ kg. The decrease of 7.4 $\times 10^{9}$ kg is therefore 2 to 3 times the standard deviation during these quiet times, and the observed decrease cannot be ascribed to natural variability.

Observations made from Sanae, Antarctica show that prior to the March 1989 SPE's, the temperatures at the 30 mb level were above -50°C, as shown in Figure 4.6. On no occasion did the temperature fall anywhere near to -78°C, which is the condition required for the formation of polar stratospheric clouds (*Toon and Turco, 1991*). Sanae was within the region of low total column ozone on March 21, so the observed depletion cannot be due to the presence of such clouds. Heterogeneous reactions on the surface of ice particles in polar stratospheric clouds, with subsequent release of chlorine, can therefore be ruled out as a possible source of ozone depletion. This supports the idea that the depletion observed in March is due to solar protons. Temperatures at 30 mb altitude dropped throughout March, as shown in Figure 4.6, indicating the loss of ozone at this level.

For comparison, we draw the reader's attention to the upper colour plot in Figure 4.19, again from the south pole to 60°S, for the springtime ozone hole on 12 October 1989. The colour bar is a little different from those of Figure 4.4 because we have used an increased dynamic range of total column ozone values. By defining the region of ozone depletion as that area where the total column ozone is < 240 DU (the blue shades of Figures 4.4) and the upper plot of 4.19 we have estimated the mass of ozone per unit area for 21 March 1989 and 12 October 1989. This was achieved by dividing the ozone mass confined within the < 240 DU contour by the spatial extent covered by it. These estimates are 4.6 $\times 10^{-3}$ kg m^{-2} for SPE-associated depletion and 3.6 $\times 10^{-3}$ kg m^{-2} for the chlorine-catalysed depletion. Chlorine-catalysed depletion, however, occurs over an area four times as great as that due to SPEs. Nevertheless, the influence of SPE's should not be ignored in an assessment of the Earth's ozone budget.

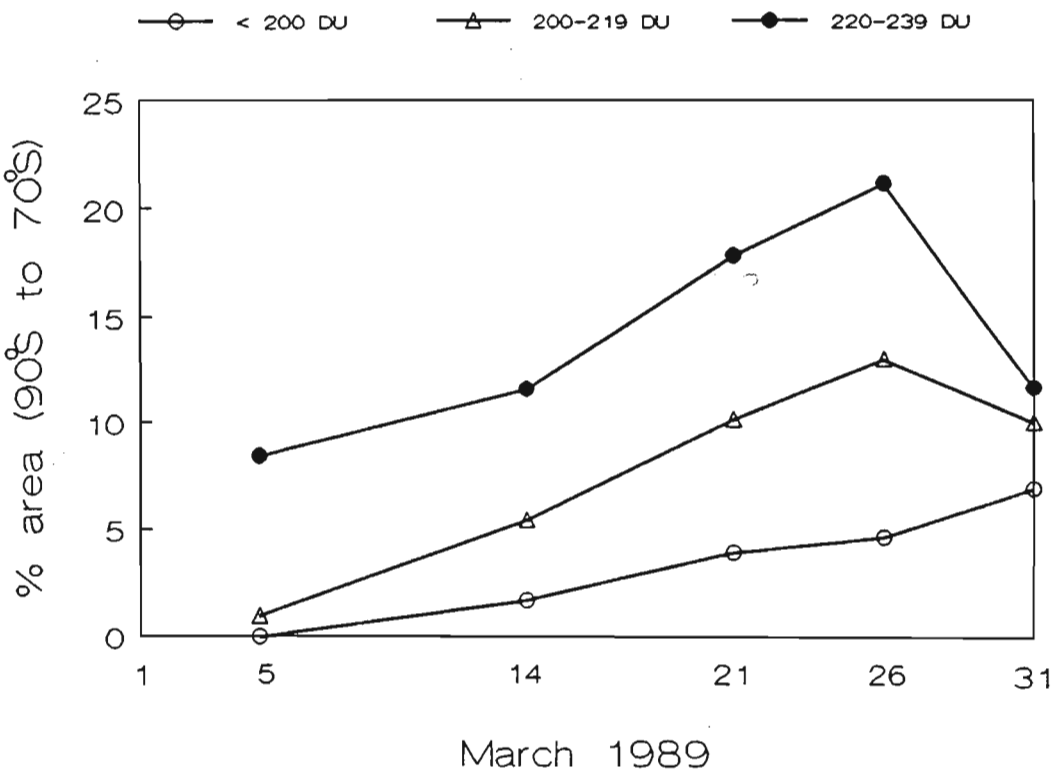

Figure 4.5: The percentage area from the south pole to 70°S occupied by each of the three lower ranges of total column ozone (dark blues in Figure 4.4) , during selected days between 5 March and 31 March. Open circles, <200 DU; open triangles, 200-219 DU; full circles, 220-239 DU.

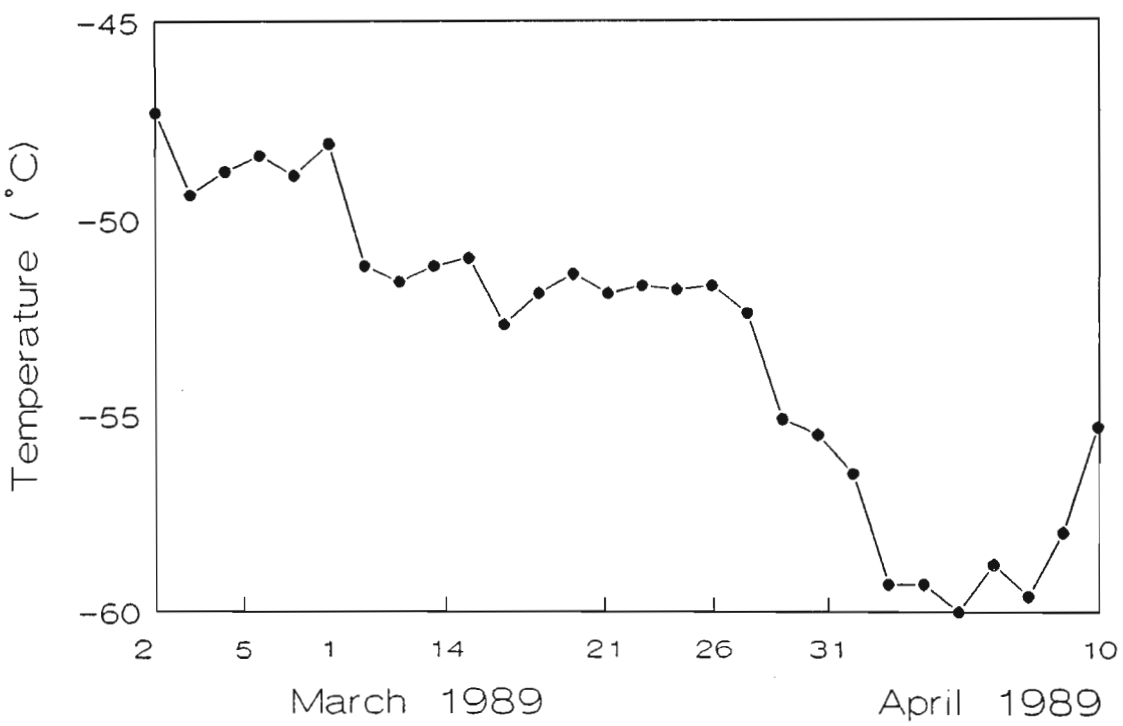

Figure 4.6: The temperatures at the 30 mb pressure level at Sanae, Antarctica from 2 March to 10 April 1989.

4.3.4 Comparison of northern and southern polar cap ozone response

The northern hemisphere satellite ozone data were also processed for the period March 5 to 31. Again we stress that March 5 is prior to the onset of a series of large solar flares that occurred throughout March 1989. The initial flares on March 6 were responsible for the first SPE on March 8. March 5 can therefore be used as a baseline from which to gauge any effects of SPE's.

We have calculated the total mass of ozone (at STP), from the south and north pole down to a latitude of 70°, taking into account the cosine dependence of the cell size of TOMS with latitude. A profile of total ozone mass throughout March over the southern polar cap is shown in Figure 4.7 for 1989 (thin broken line), 1988 (broad broken line) and 1987 (thickest broken line). The months of March 1987 and 1988 may be considered as 'solar quiet' as no large solar flares were recorded, and as such the ozone levels during this time may be considered unaffected by solar protons. Following the SPE onset (8 March 1989), total ozone masses during March 1989 exhibit small variations, after which, there is consistent decline to a minimum of 8.11 $\times 10^{10}$ kg on 21 March 1989. This represents a large decrease of 7.4 $\times 10^{9}$ kg or ~9%. In contrast, during the same period in 1988 and 1987 total ozone mass maxima are visible between 11 and 16 March. Although, by March 21 (1987, 1988) total ozone masses were down 4% and 3% on pre-event values, this represents a much smaller percentage ozone mass loss than in 1989. Clearly, compared with 1987 and 1988, the depletion during 1989 is unusual.

The mass of ozone for the same month in 1989, between north pole and 70°N (thin solid line), 1988 (broad solid line) and 1987 (thickest solid line) are shown in Figure 4.8. Variations, evident in each year, which show an increase of ozone in the earlier part of March may be ascribed to sudden stratospheric warmings (SSW's). Observations of SSW's are described below. Such warmings must cause the total mass over the northern polar cap in 1989 to increase from 1.52 $\times 10^{11}$ kg on March 5 to 1.56 $\times 10^{11}$ kg on March 14 (see discussion below). Smoother curves in the latter part of March reveal less dynamic activity. In Figure 4.8, the ozone mass (90-70°N) profile of 1989 exhibits a consistent loss in ozone mass, dropping to 1.44 $\times 10^{11}$ kg by March 26 1989. This represents a loss of 8.0 $\times 10^{9}$ kg from March 5 to 26. In 1988 and 1987 Figure 4.8 reveals no such loss, on the contrary, consistently increasing values are evident. Only two other years (1987 and 1988) of version 5 TOMS data were accessible for comparison with 1989.

Having established an approximate depletion symmetry (7.4 $\times 10^{9}$ kg ozone mass loss over the southern polar cap and 8.0 $\times 10^{9}$ kg for the north) we now draw attention to

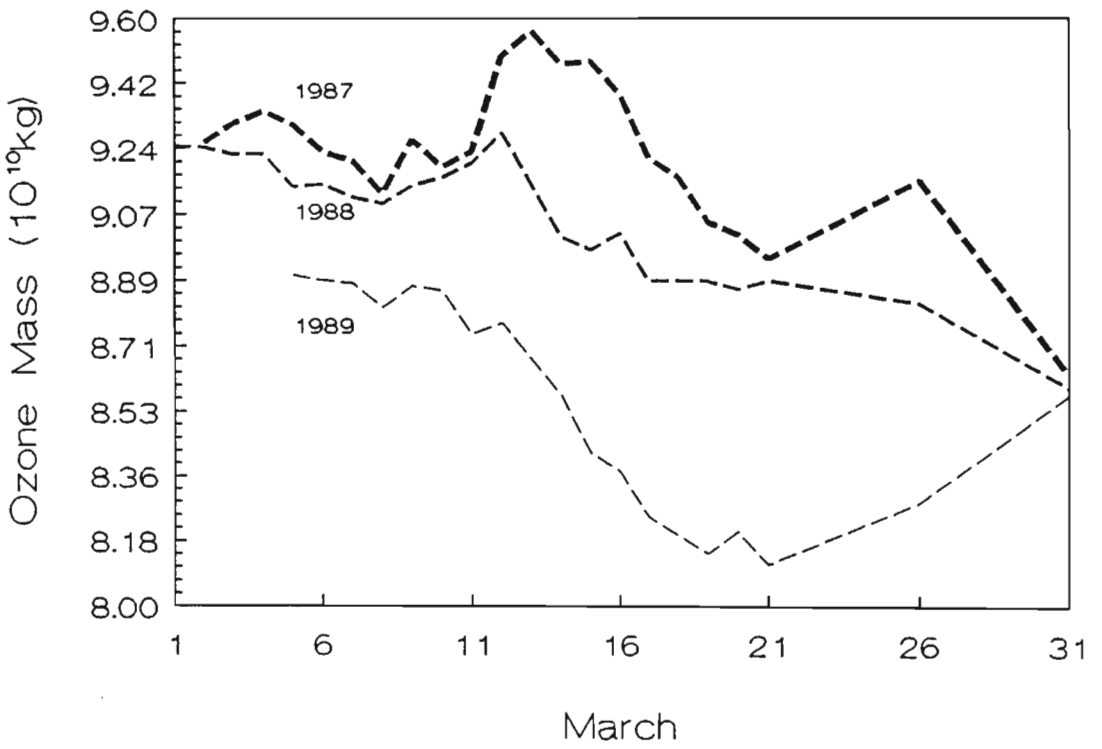

Figure 4.7: Daily total ozone mass values over the southern polar cap during March 1989 (thin broken line) and, for comparison, profiles of 1988 (broad broken line) and 1987 (thickest broken line).

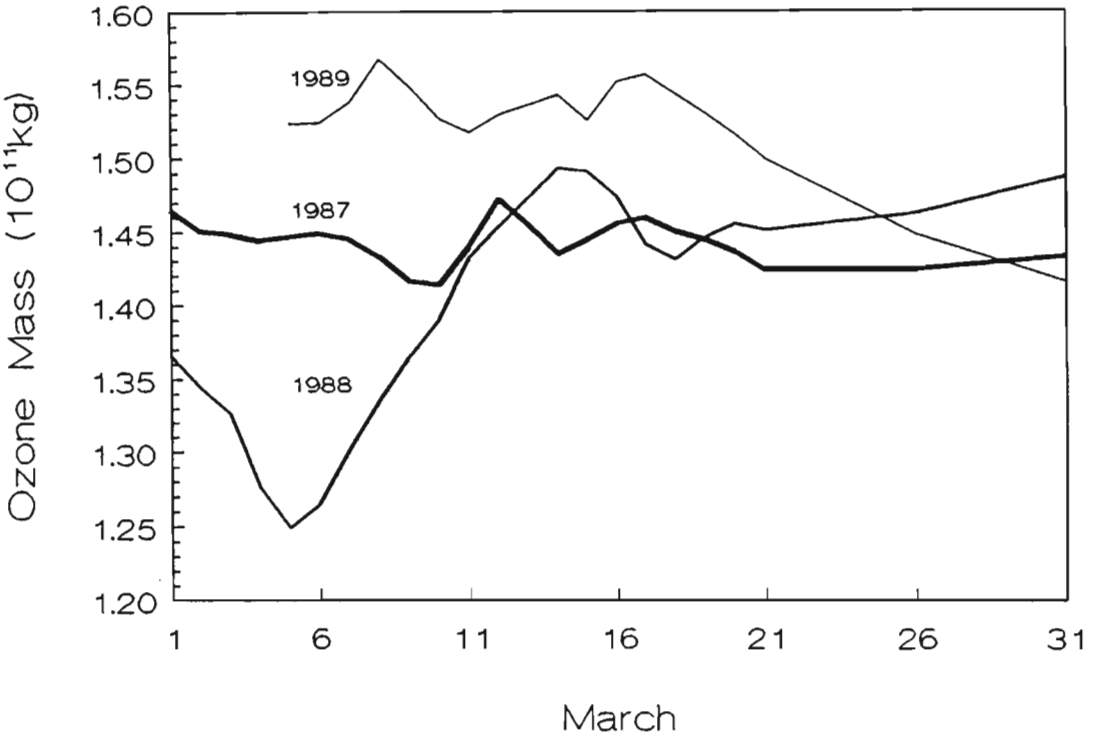

Figure 4.8: A plot comparing northern polar cap ozone masses during March 1989 (thin solid line) with two previous years, 1988 (broad solid line) and 1987 (thickest solid line).

some temporal and spatial differences in the ozone distributions.

For purposes of interhemispheric comparison, the reader's attention is drawn to the three colour plots in Figure 4.4 for 5, 14 and 21 March for the region extending from the south pole to 60°S. The colour bar shows low total ozone as deep blue with total ozone increasing through green, yellow the red. Each hue of the colour bar represents a range of 20 DU, the lowest range being 180 to 199 DU. Within the region down to 70°S on March 21 (bottom) there is a significant area covered by deep blue colour corresponding to decreased total column ozone. The same region on March 5 (top) is characterized by green and pale blue tones i.e. total column ozone range of 240-320 DU. The days between March 5 and 21 were also analysed and are consistent with a gradual increase in the area of ozone depletion.

In Figure 4.9 we show colour plots of the total column ozone distribution for March 14, 26 and 31 for an area covering the polar cap down to a latitude of 60°N. In contrast to Figure 4.4 each hue in the colour bar now represents 40 DU and reflects the greater dynamic range of ozone values in the northern hemisphere compared with those of the southern hemisphere. Hence the colour bar in Figure 4.9 covers values from 300 to 659 DU and that in Figure 4.4 from 180 to 359 DU. The clear visual evidence of systematic ozone depletion in Figure 4.4 is not evident in the colour plots of Figure 4.9. In fact there is an increase in total column ozone by March 14 as indicated by the extended areas covered by green and the appearance of some yellow and red hues on the uppermost colour plot of Figure 4.9. However these are replaced by light blue hues by March 26 (middle plot of Figure 4.9) and several dark blue patches on a light blue background appear by March 31 (lowermost plot of Figure 4.9). These losses were sustained well into April and this is why we have shown the distribution on March 14, 26 and 31 over the north polar cap as opposed to March 5, 14 and 21 in Figure 4.4 for the south polar cap.

As mentioned earlier, the temporal decreases in total column ozone over the polar caps can be shown quantitatively by calculating, for selected days, the percentage area covered by low total column ozone and expressing this as a percentage of the total area down to a latitude of 70°. These percentages are shown in Figure 4.10 for low total column ozone in the ranges < 460 DU (north polar cap) and < 260 DU (south polar cap). These ranges are different since the background ozone levels in the north polar cap greatly exceed those for the south polar cap. These upper limits were chosen to include all blue hues in Figures 4.4 (< 260 DU) and 4.9 (< 460 DU). There is an increase in the area covered by low total ozone over the southern polar cap (dashed line in Figure 4.10), with some recovery to higher ozone values following March 21. For the north polar cap (solid line in Figure 4.10) the area covered by the low range of < 460 DU decreased until March 14, indicating an increase in north polar cap ozone. However, following March 14 this trend is reversed and ozone values decrease. By

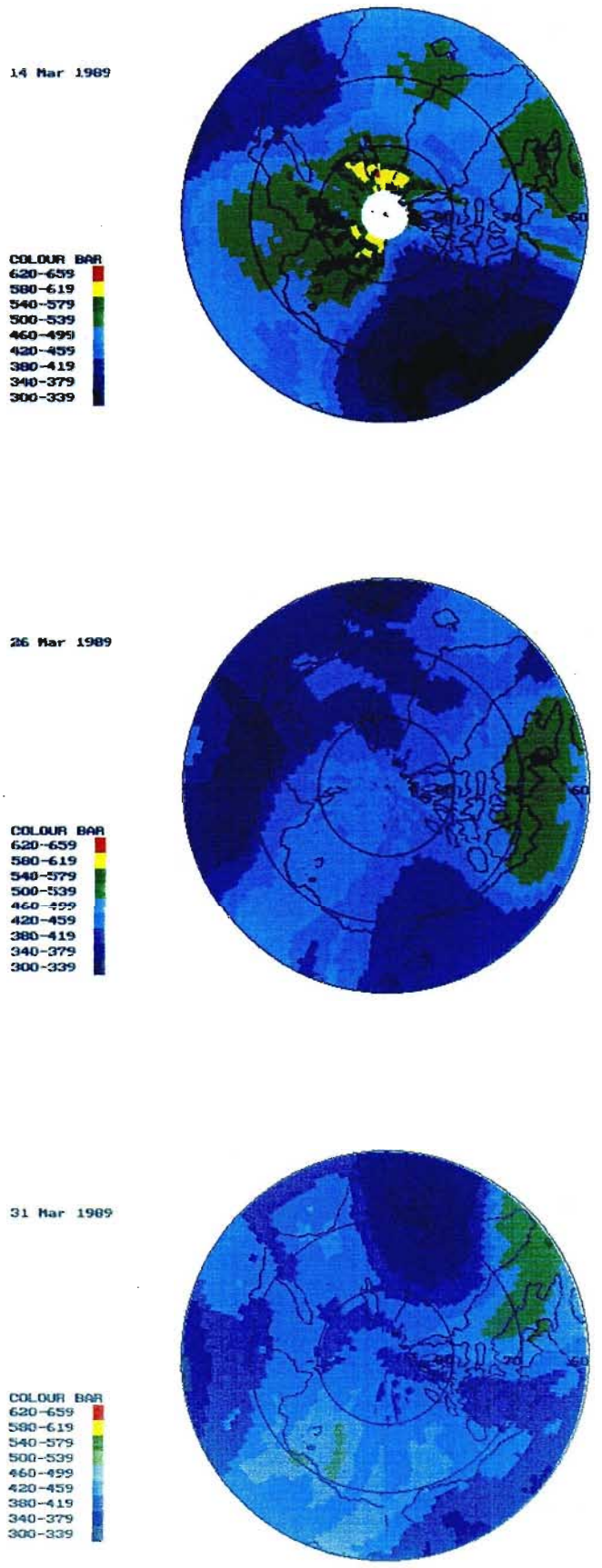

Figure 4.9: Colour plots of total column ozone from the north pole to 60°N. Upper panel for 14 March, middle panel for 26 March and bottom panel for 31 March. All plots are subsequent to SPE onset. Due to increased dynamic range in total ozone over high northern latitudes at this time each hue of the colour bar represents a 40 DU range, unlike the 20 DU interval in Figure 4.4.

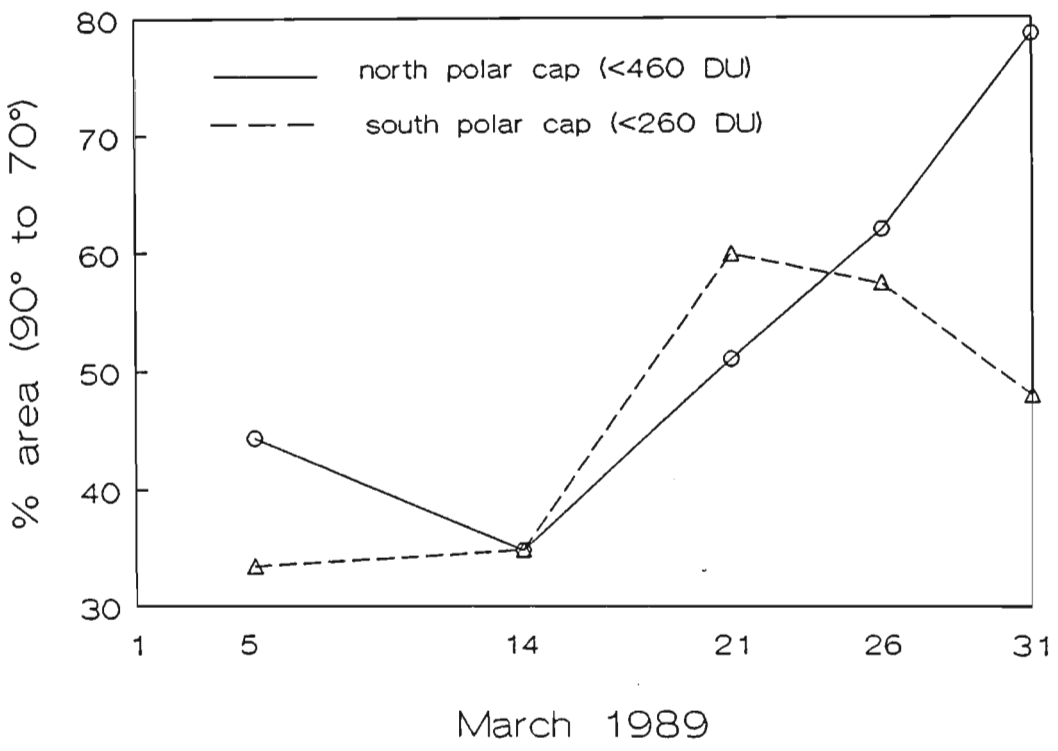

Figure 4.10: The percentage area from each pole to 70°, occupied by the lowest ranges of total column ozone (blues in Figures 4.4 and 4.9), for preferred days between March 5 and March 31. Broken line with open triangles, <260 DU and solid line with open circles, <460 DU.

March 31 about 80% of the area from north pole to 70°N is covered by ozone of < 460 DU compared to about 40% before the SPE. This contrasts a northern hemisphere ten year average (1979-1988) where ozone values monotonically increased throughout March (*Newman et al., 1990*).

In Figure 4.11 we show the ozone mass differences for both south (dashed line) and north (solid line) polar caps, over an area extending to 70°S and 70°N respectively. These differences are measured with respect to the ozone values on March 5. Two features can be clearly identified in Figure 4.11. Between March 5 and 14 the ozone mass decreases over the south polar cap but, contrary to expectations following an SPE, increases over the north polar cap. In addition, the depletion for the north polar cap, after March 14, is more prolonged than that for the south polar cap.

The most important aspect of this work is the quantitative determination of the depletion of ozone, over both polar caps, due to solar protons. Similarity of the magnitudes of the depletion (not conspicuous from visual examination of the colour plots of satellite ozone measurements) corroborates, at least for the present conditions, the often assumed symmetry of precipitating solar protons (*McPeters et al., 1981*) between the two hemispheres. We now attempt to identify possible sources of the deviation, albeit in some fine detail, from the observed symmetry in ozone depletion.

Contrasting background ozone values (~250 DU for southern autumn polar regions and ~450 DU for northern spring) may be ascribed to seasonal variations which are governed by transport and chemical production and destruction (*Perliski et al., 1989*).

During March 1989, the autumn stratosphere above the southern polar regions was relatively unperturbed. Temperatures at the 30 mb level (~24 km altitude) recorded at Sanae, Antarctica showed small fluctuations around -50°C, thus ruling out the possibility of PSC's. Wind speeds at all altitudes were low (several m s^{-1}). Hence, the stratosphere was conducive to homogeneous gas phase chemical destruction, particularly by nitric oxide produced by solar protons. The net effect was that the precipitations of solar protons resulted in a well defined ozone depletion signature (dashed line in Figure 4.11) over the southern polar cap.

Ozone distribution over the northern hemisphere at this time is quite different. Springtime ozone values exhibit large spatial and temporal variability, occasionally exceeding 600 DU, due to transport of ozone rich air from lower latitudes. At this time chlorine chemistry, which destroys ozone, is proceeding. Hence, the ozone distribution is far from quiescent and additional changes due to solar protons are not clearly apparent. This could account for the depletion in Figure 4.9 being not nearly as well defined as that for the south polar cap in Figure 4.4.

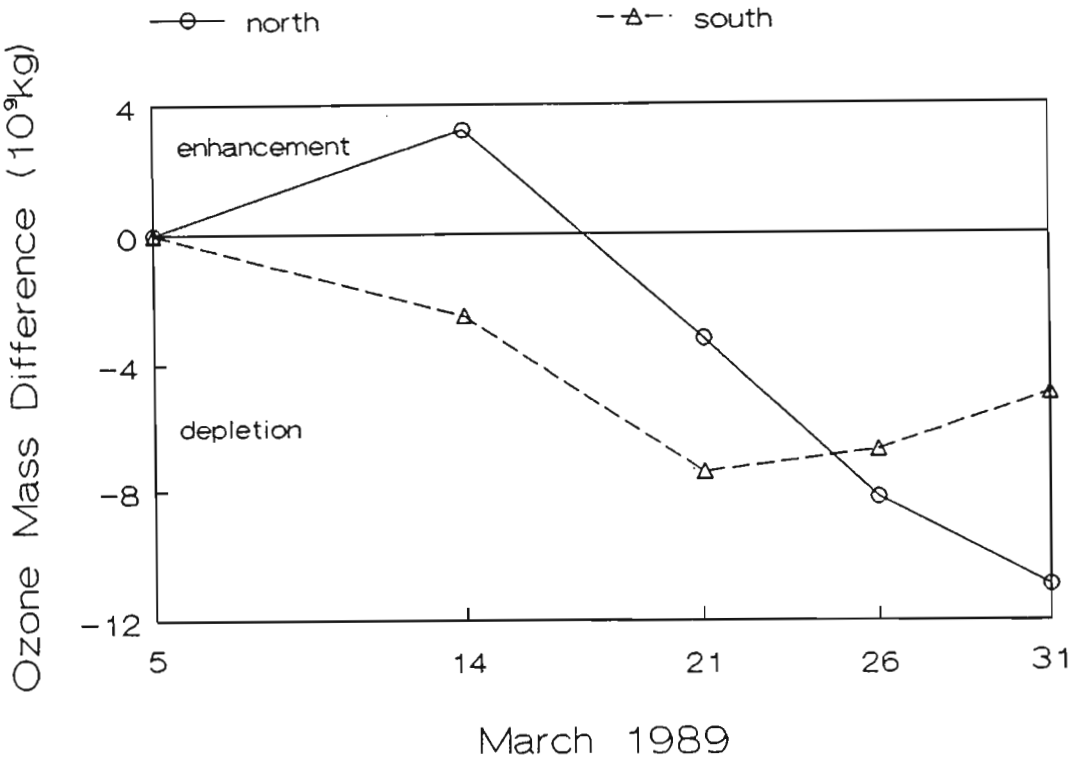

Figure 4.11: Ozone mass difference between selected days (14, 21, 26 and 31 March) and the pre-event day March 5. Ozone masses were calculated from each pole to 70°S (broken line with open triangles) and 70°N (solid line with open circles).

Each year, in late winter, sudden stratospheric warmings (SSW's) are reported in the northern hemisphere. Vertical propagation by planetary waves incites poleward transport of both heat and ozone (*Matsuno and Nakamura, 1979*). In mid-February 1989 there was a major stratospheric warming event and around March 6 a second, and final, SSW event caused further increases in ozone and temperature (*Neuber and Krüger, 1990*). This increase in ozone might well be responsible for the enhanced ozone mass between March 5 and 14 in Figure 4.11. On the assumption that we expect the same mass change due to the SPE in both polar caps, the SSW near March 6 would be required to increase the total ozone mass by about 5.5 $\times 10^9$ kg between March 6 and 14. This translates into the SSW generating an ozone increase of about 5% over the north polar cap.

Cold conditions over northern polar regions in January and February of 1989 were favourable for the formation of PSC's over large areas (*Krüger, 1990*). PSC's provide surfaces on which inert chlorine reservoirs i.e. $ClONO_2$ and HCl are converted into reactive chlorine species. They also remove reactive nitrogen by absorption into, and sedimentation of, ice particles containing nitrogen oxides (most commonly nitric acid). This leaves the stratospheric air denitrified, and in doing so prevents reactive chlorine from being reconverted into inert reservoirs of $ClONO_2$. A model (*Isaksen et al., 1990*) that allows for the effect of heterogeneous reactions on PSC's predicts ozone losses in late March and might account for the prolongation of the north polar cap ozone depletion which is borne out by the ever decreasing ozone mass values in Figure 4.11.

Conclusions regarding ozone depletion, obtained from visual inspection of satellite measured ozone distributions, such as those in Figures 4.4 and 4.9, can be misleading. Determination of the total masses of ozone over the two polar caps, as we have done here, is of prime importance and reveals similar large scale ozone mass loss over both polar caps during March 1989.

4.4 Case study 2: Final months of 1989

4.4.1 Solar activity

Intense solar activity during 1989 prompted six major particle events, four of these occurred between August and December. The ground level events (GLE's) in September and October caused several operational problems with spacecraft in deep space, in geosynchronous orbit, and in low earth orbit. These problems included enhanced background noise in star trackers and rapid degradation of solar arrays (up to 5 years' equivalent aging in less than one week). Further, significant radiation levels were mea-

sured by dosimeters aboard transpolar airline flights at high altitudes.

Characteristics of the four events are summarized in Table 4.2. Each SPE produced substantial fluxes of high energy protons, which were recorded by the GOES-7 satellite. The onset dates and fluxes (particles $cm^{-2}s^{-1}ster^{-1}$), tabulated in Table 4.2 are 12 August 9200, 29 September 4500, 19 October 73000 and 30 November 7300. Maximum values of K_p (second column of Table 4.2) during the SPE's indicate that only the August and October events were accompanied by intense geomagnetic activity, neither being as severe as the March event. The lower panels of Figures 4.12, 4.13, 4.14 and 4.15 which display geomagnetic field components for the four SPE's reveal an enhanced, but steady geomagnetic field. Only during the October SPE are negative values detected by GOES-6, implying a very distorted geomagnetic field which would produce an extended region for particle precipitation. The time evolution of the particle fluxes of these events is displayed in the middle panels of Figures 4.12 (August event); 4.13 (September event); 4.14 (October event) and 4.15 (November event).

Inspection of the profiles of all four events reveals that maximum fluxes were registered the day after onset and are therefore characterized by short rise times (especially the September event). The event on September 29 represents the largest ground level effect observed in over 30 years (*Gorney, 1990*). The solar region responsible for the 29 September event was NOAA region 5698 which rotated past the western limb of the sun on 28 September. The GOES-6 spacecraft reported the onset of an intense X-ray event at 1047 UT on 29 September, evident in the top panel of Figure 4.13. *Humble et al., 1991* detected 0.5-15 GeV protons, inferred from responses of neutron monitors in Australia. The most significant event is the SPE in October which has a flux of 73 000 (> 10 MeV) protons cm^{-2} s^{-1} $ster^{-1}$. Riometer observations at Sanae (*Stoker and van Wyk, 1993*) revealed linear increases in cosmic radio noise following these SPE's.

4.4.2 Modelled estimations and *in situ* observations of NO enhancements

The first *in situ* measurements of NO concentration enhancements subsequent to a SPE were achieved, following October 1989 SPE's, by *Zadorozhny et al. (1992)*. The nitric oxide profiles were measured by rockets launched from a Soviet research vessel from 50-60°S. The NO concentration was measured by means of the photoionization method based on the selective ionization of NO molecules by vacuum ultraviolet radiation with the subsequent collection of the ions created. Results of these observations are given as the solid profiles in Figure 4.16. The curves marked '2' and '3' are pre-event profiles and the curve marked '7' was determined on 23 October 1989 at 59°S. The observed NO increases (one to two orders of magnitude) are larger (~500%) than those esti-

onset date of SPE	max k_p	major flare date class	proton flux > 10 MeV	proton fluence > 30 MeV
12 August (3)	7^-	12 August 2B/X2	9200	$> 2.1 \times 10^9$
29 September (4)	4	29 September X9.8	4500	$> 8.2 \times 10^8$
19 October (11)	8^+	19 October 4B/X13	73000	2.7×10^9
30 November (3)	4^+	30 November 3B/X2	7300	$> 7.8 \times 10^7$

Table 4.2: The SPE's, and associated solar flares, during the final months of 1989. Data from *Solar Geophysical Data (1990, a, b, c, d).*

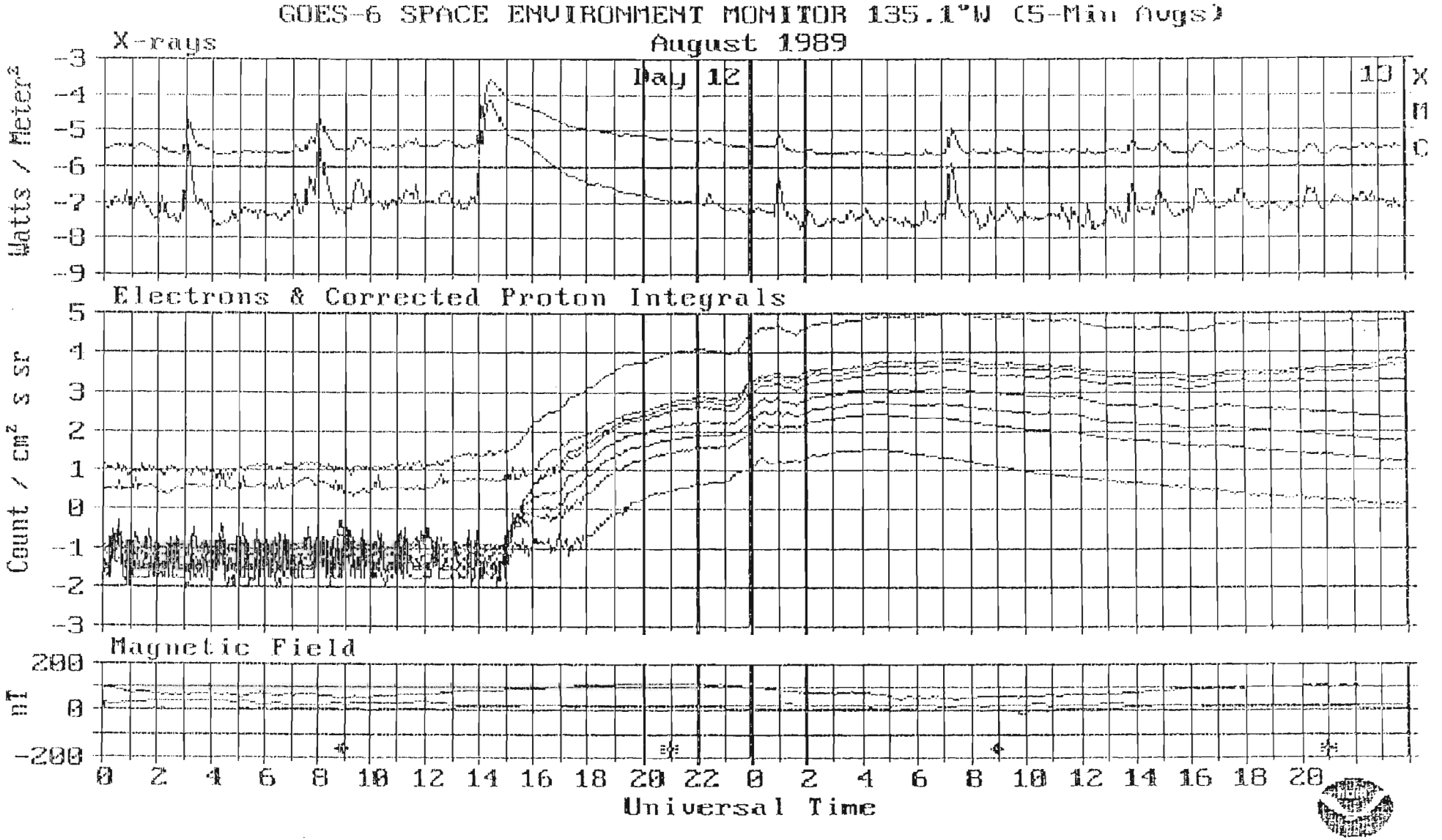

Figure 4.12: X-ray (top panel), integrated particle flux (middle panel) and magnetic components (bottom panel) observed by the SEM instrument package aboard the GOES-6 spacecraft for 12 and 13 August 1989.

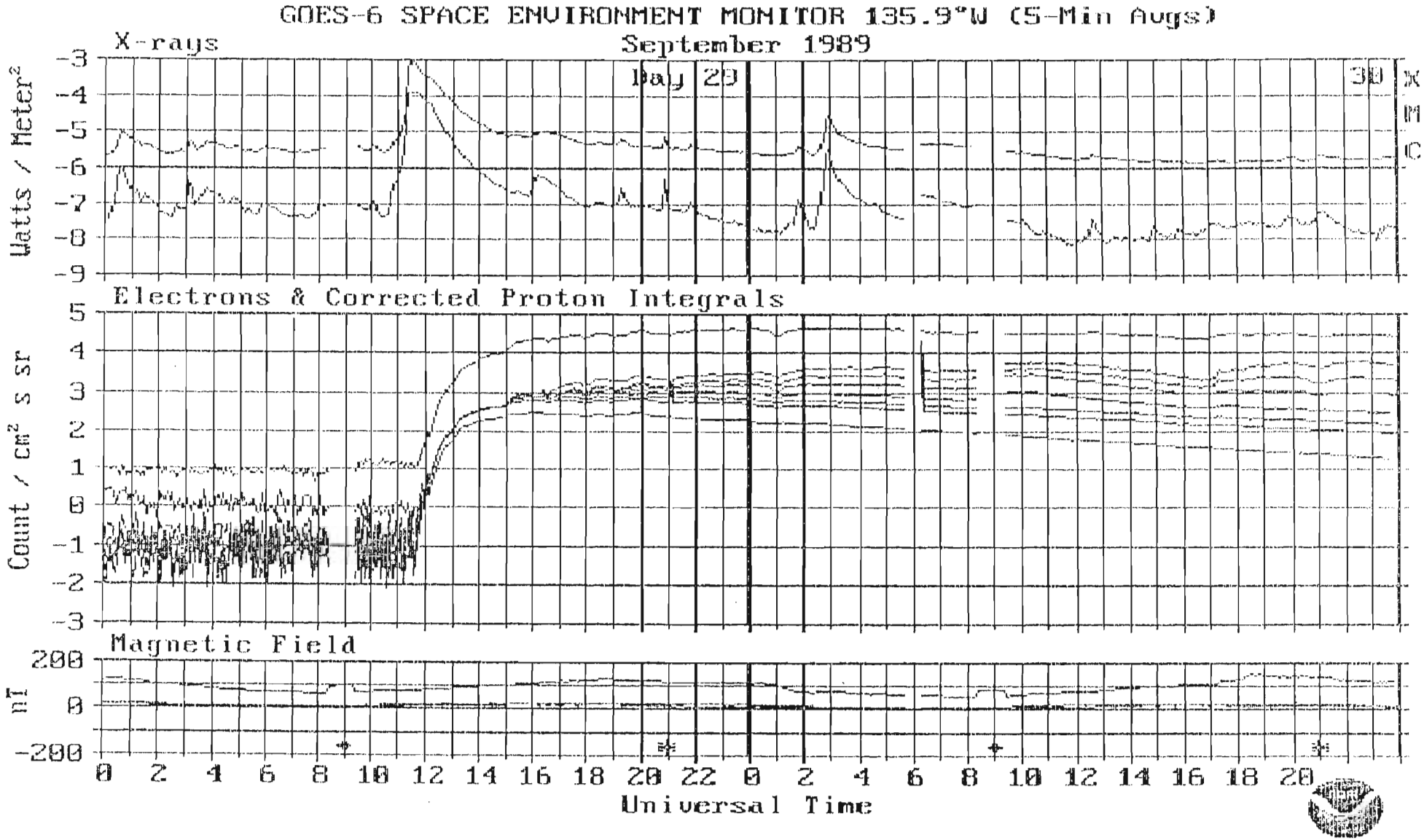

Figure 4.13: X-ray (top panel), integrated particle flux (middle panel) and magnetic components (bottom panel) observed by the SEM instrument package aboard the GOES-6 spacecraft for the 29 and 30 September 1989.

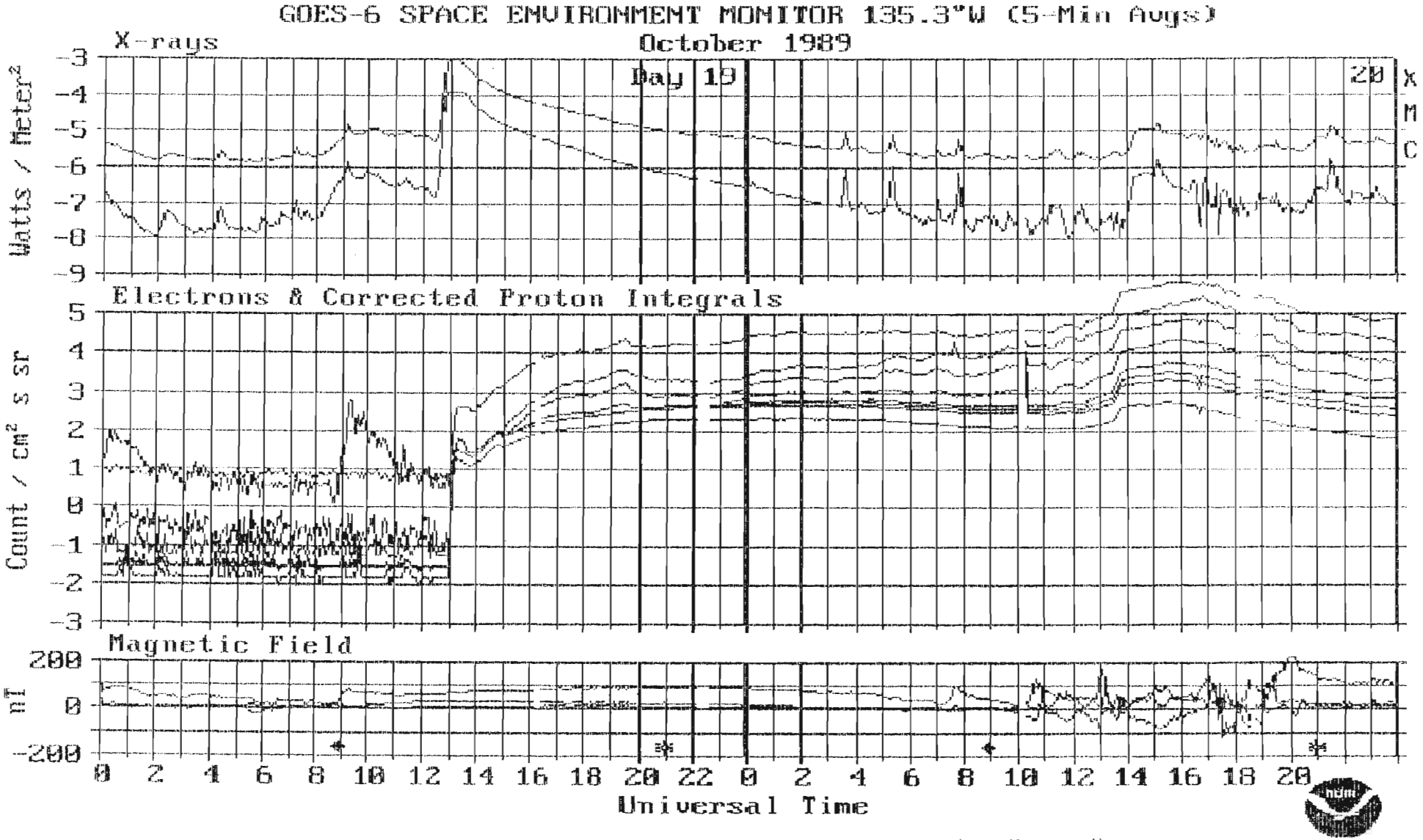

Figure 4.14: X-ray (top panel), integrated particle flux (middle panel) and magnetic components (bottom panel) observed by the SEM instrument package aboard the GOES-6 spacecraft for the period 19 and 20 October 1989.

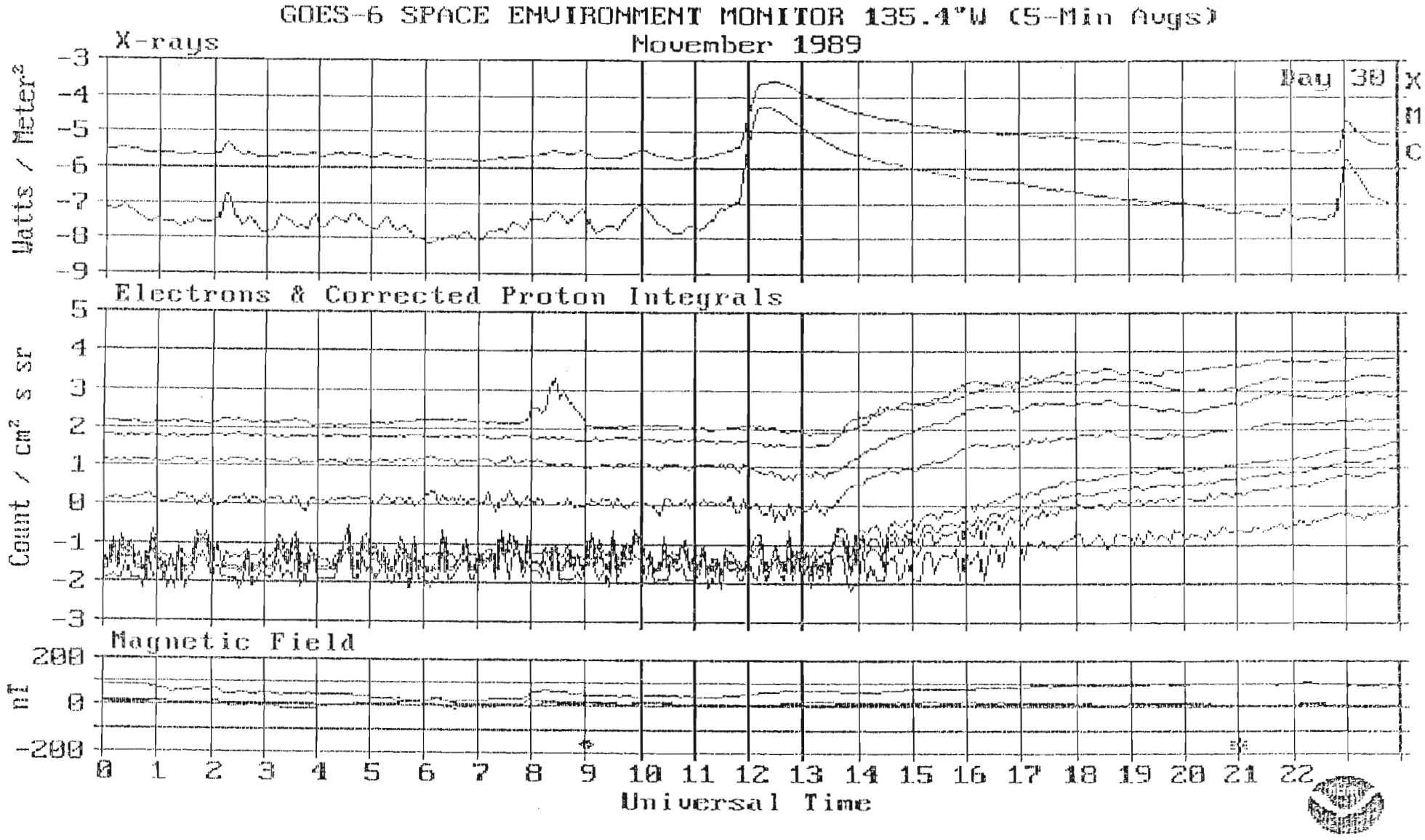

Figure 4.15: X-ray (top panel), integrated particle flux (middle panel) and magnetic components (bottom panel) observed by the SEM instrument package aboard the GOES-6 spacecraft on 30 November 1989.

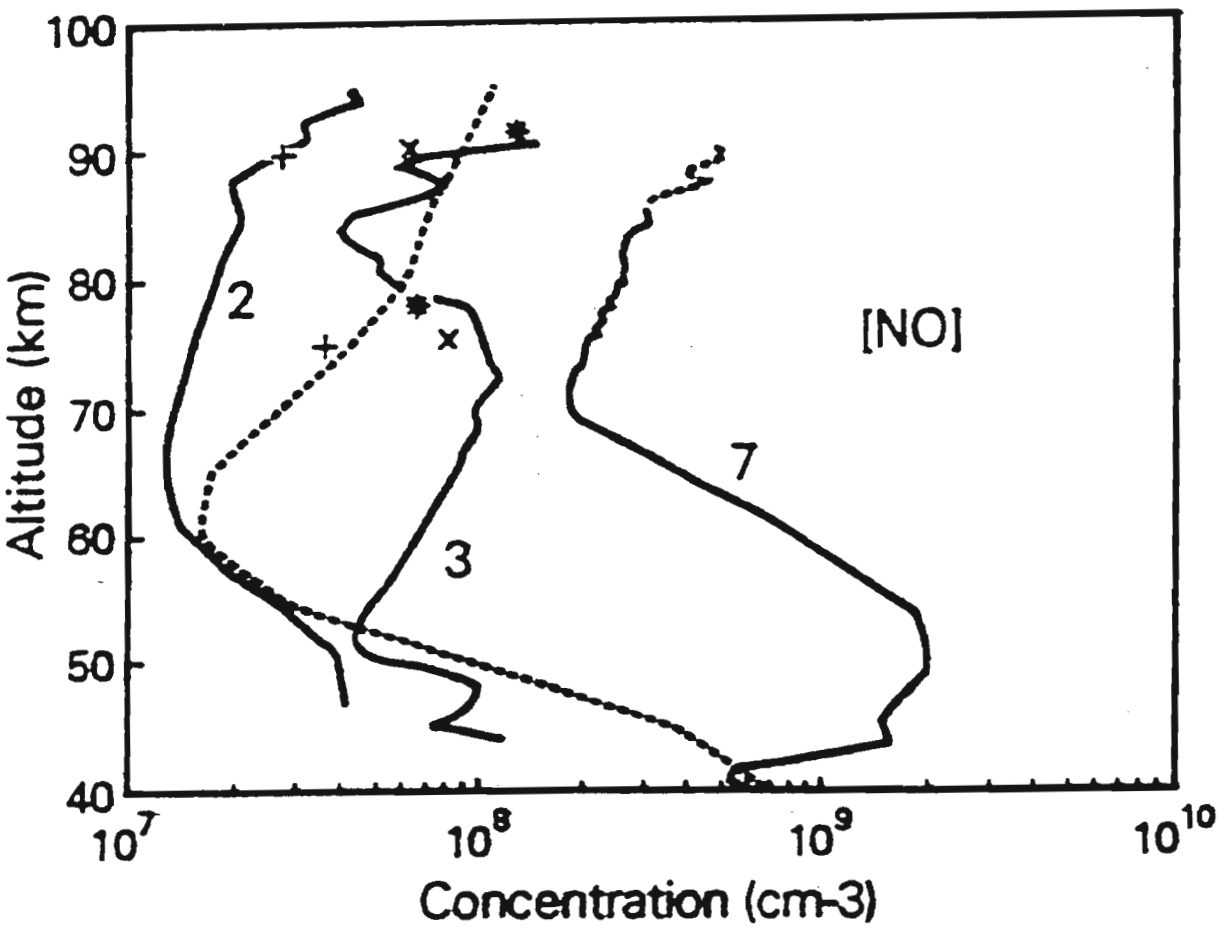

Figure 4.16: Nitric oxide profiles measured during quiet ('2' and '3') and SPE ('7') conditions measured by a rocket experiment. Dashed lines are radio absorption observations from Spacelab 1. From *Zadorozhny et al. (1992).*

mated from ion-pair production rates by *Reid et al. (1991)* at 50-60°S. In particular, measurements at lower altitudes (around 40 km) reveal significantly more NO production (~1000%) than that predicted in Figure 2.21 (from *Reid et al., 1991*) of ~100%, possibly due to downward meridional circulation in winter over the southern polar regions. *Reid et al. (1991)* forecast 55% column density enhancements in nitric oxide for these four events for which their two-dimensional model predicts > 20% depletion in upper stratospheric ozone at southern polar latitudes. Conjugate northern hemisphere ozone losses were calculated to be only ~10%. We remind the reader that this model only applies to quiescent atmospheric conditions and cannot simulate planetary wave dynamics. *Jackman et al. (1993)* analyzed ozone data from the NOAA-11 SBUV/2 instrument and found evidence for long term ozone depletion following the SPE's in both hemispheres. Furthermore, they detected a significantly larger ozone depletion in the northern hemisphere than that in the south, contrasting to the prediction by *Reid et al., 1991)* that the largest depletions would be found in the southern polar regions. In an attempt to achieve predictions that would match the observations more closely *Jackman et al. (1993)* used a three-dimensional model which included planetary wave activity during October/November. Results of this model, despite chemical limitations with regard to chlorine chemistry, indicate that the southern hemisphere response should indeed be significantly less than that of the north. These two models are discussed, in more detail, in sub-section 2.8.2.

4.4.3 Southern polar cap ozone response

The impact of odd nitrogen enhancements on the spatial extent of low total column ozone and of the total ozone mass, over a region extending from 90°S to 70°S, is determined for the period August to December 1989. The northern polar cap's response was not measurable due to the lack of solar irradiance at this time of year. Comparisons are made with a moderate solar activity baseline of previous years (1984 to 1988). The effect, if any, of the SPE's on ozone during times of heterogeneous chlorine chemistry and dynamic processes (important aspects of which are described in chapter 2) is discussed.

This case study investigates a means by which odd nitrogen species, produced by solar protons, could enhance or retard chlorine-catalysed ozone depletion in the austral spring.

TOMS satellite ozone data were processed for the period from late August (day number 240) 1989 to the end of that year. We have again calculated the total mass of ozone (at STP), taking into account the cosine dependence of the cell size of TOMS with latitude, from the south pole to latitude 70°S. High zenith angles, as is the case for

polar latitudes in winter, can result in measurement errors by TOMS. However, in austral spring when ozone levels are very low (facilitating ultraviolet penetration), these errors are minimized.

A plot of daily ozone mass, from 90°S-70°S, for day number 240 to 365 (i.e from the end of August to the end of the year) of 1989 is shown as a thin solid line in Figure 4.17. This time period, although excluding the immediate effects of the SPE on 12 August, was selected because all latitudes down to at least 80°S are illuminated. Arrows indicate the onset of SPE's, the effects of which, since the lifetime of odd nitrogen in the stratosphere is days to months (*Reid et al., 1991*), may be cumulative.

In order to assess any effect the SPE's may have had on ozone, a typical mass profile for previous periods over which the chlorine-catalysed ozone hole appears, must be computed. A baseline comprising of five years of daily ozone masses averaged over 1984, 1985, 1986, 1987 and 1988 (all solar quiet years) was evaluated. This is represented by the thick line in Figure 4.17. The error bars represent the standard deviation, for consecutive 20 day periods, of all the daily ozone masses over five years. These standard deviations represent interannual differences in dynamics (e.g QBO and SSW's) and chemistry. Until day number 300 the 1989 ozone mass profile is always well below the lower limit of these standard deviations. The interrupted line in Figure 4.17 represents the ozone mass profile of 1987. This profile is included since 1987 is certainly the 'worst case' ozone hole prior to 1989. Radiative transfer computations indicated that the 1987 Antarctic vortex approached radiative equilibrium, suggesting that dynamical heat transport was very weak. Analysis of trace constituent data from aircraft also indicates that mixing was very weak in 1987 (*Hartmann et al., 1989*). It is therefore doubtful that the spatial extent of the pool of cold air in the vortex could increase beyond that observed in 1987 through purely dynamical mechanisms (*Toon and Turco, 1991*).

The ozone mass reached all time low values of 4.7 $\times 10^{10}$ kg in both 1987 and 1989 around mid-October (day numbers 270 - 285). These low ozone levels can be seen visually in the upper colour plots in Figures 4.18 and 4.19 which show colour coded total column ozone levels over Antarctica in October 1987 and 1989 respectively. The spatial extent and depth of these two ozone holes, indicated by shades of blue, are clearly very similar. In both 1987 and 1989, the prevailing quiescent dynamics resulted in an isolated vortex within which the ozone depletion occurred at a rate which was chemically determined.

The year-to-year variability of planetary wave activity in both hemispheres produces a corresponding response in the strength and temperature of the polar vortex. These year-to-year fluctuations in the vortex strength and temperature appear to dominate over the slow increase in available stratospheric chlorine in determining the year-to-

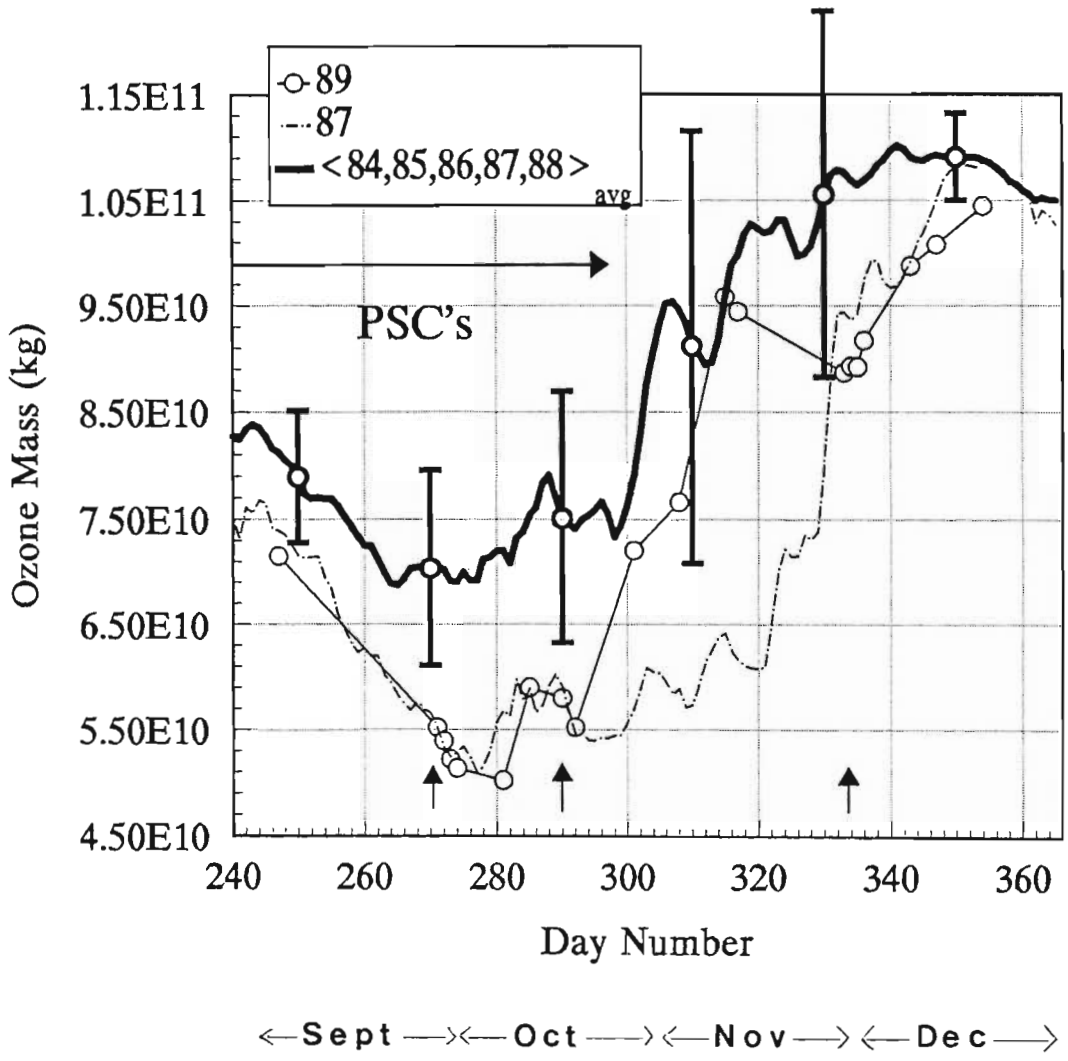

Figure 4.17: The ozone mass, calculated from TOMS total column ozone data, from the south pole to 70°S for the final months of 1989 (thin solid line), 1987 (thin broken line), and a 5 year average (thick solid line).

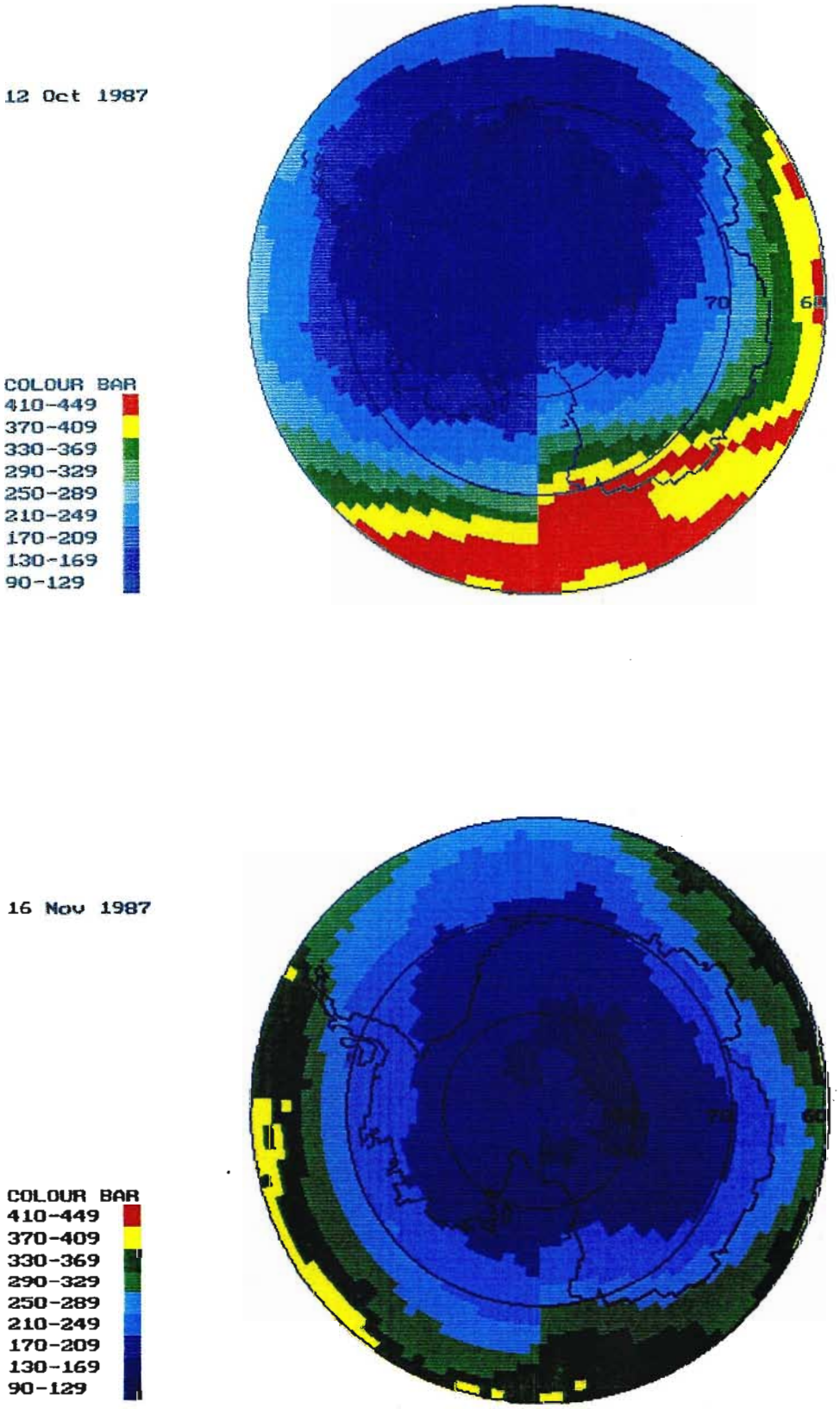

Figure 4.18: Colour plots of total column ozone over Antarctica for 12 October 1987 (top) and 16 November 1987 (bottom). Blue colours represent low total column amounts. Each colour shade represents 40 DU.

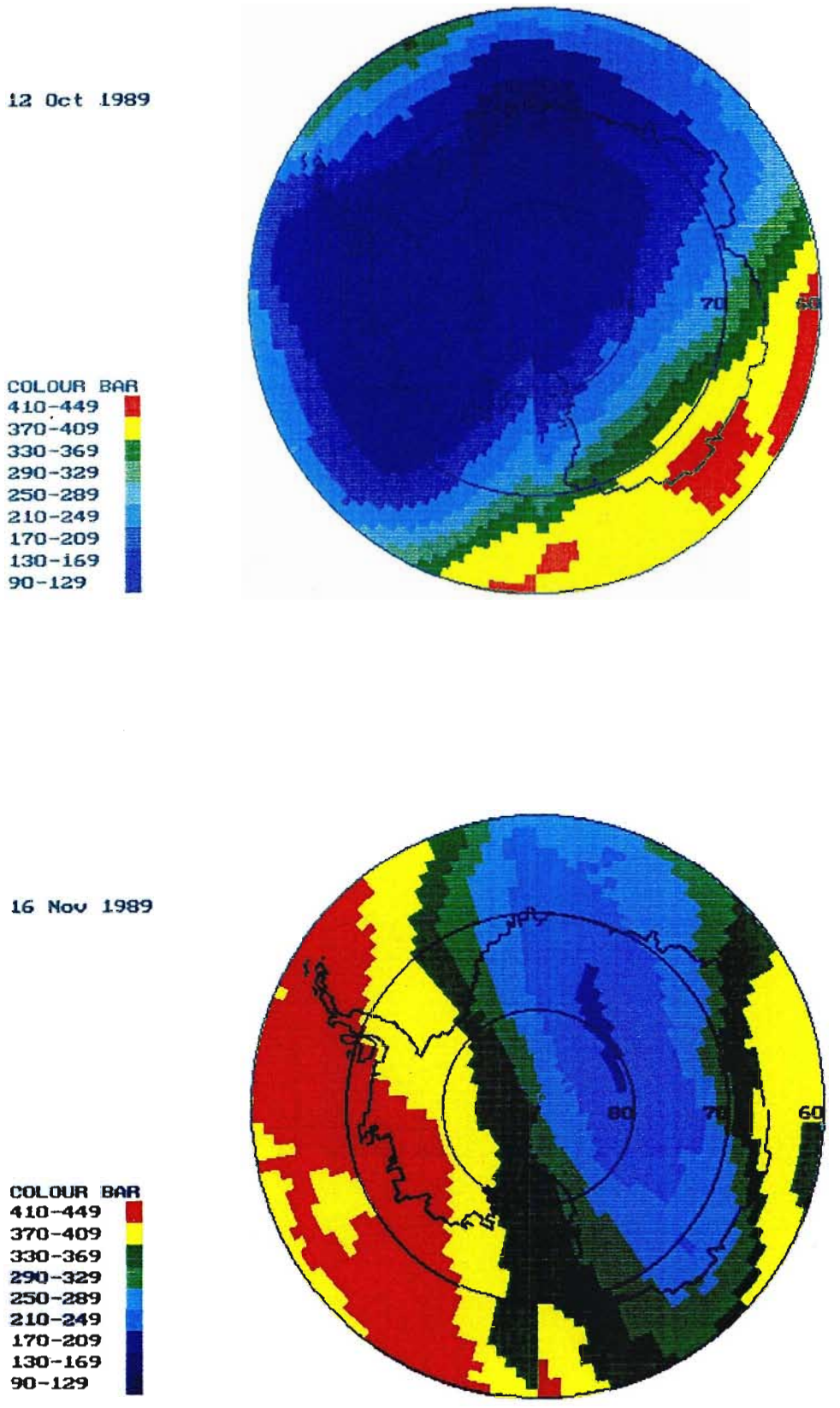

Figure 4.19: Colour plots of total column ozone over Antarctica for 12 October 1989 (top) and 16 November 1989 (bottom). Blue colours represent low total column amounts. Each colour shade represents 40 DU.

year variability of the severity of the ozone hole. Studies indicate that year-to-year variations in planetary wave activity generally follow the phase of the QBO (described more fully in subsections 2.4.2 and 2.4.3) in the tropical winds of the lower stratosphere, with more activity during easterly phase years. The ozone hole tends to exhibit the same variability (*Garcia and Solomon, 1987*). For example, in the QBO westerly year of 1987, the ozone destruction in the Antarctic lower stratosphere was almost complete (*Hofmann et al., 1989*). In the QBO easterly year of 1988, the planetary wave activity was high, polar temperatures were warm, and the zonal mean vortex winds were weak. In October 1988 the ozone depletion was much less than in 1987 (*Schoeberl et al., 1989*). The sensitivity of the ozone hole to dynamical activity is not unexpected because slight modulation of the vortex temperatures by planetary waves could greatly modulate the regional coverage of PSC's. Indeed, any planetary wave activity at all will tend to weaken the ozone depletion by raising the vortex temperatures (*Poole et al., 1989*). Cooling associated with ozone depletion may serve to further prolong the longevity of the vortex. The 1989 austral ozone depletion occurred under the easterly phase QBO conditions, yet the mid-latitude stratosphere was relatively quiet in the early spring and polar temperatures were colder than average (*Stolarski et al., 1990*), and so corresponded on the surface more to a westerly condition. Thus the high latitude response to the tropical QBO is less than straightforward.

Ozone masses are most variable at the time of vortex breakup (the commencement of which may vary from year to year), reflected in the high standard deviations, from day number 300 to 340, throughout October and November in Figure 4.17. Two distinct differences between the 1989 ozone mass profile (thin line) and the 5 year average profile (thick line) are evident. Between day numbers 260 and 305 (end of October) the 1989 ozone hole is considerably deeper than the average profile. A mass difference of 2.2 $\times 10^{10}$ kg around day 280 is over twice the standard deviation. In addition there is a further deviation between the lines corresponding to 1989 and the five year average in Figure 4.17 from day number 305 to the end of the year. Ozone masses during December 1989 do not recover to the average five year value. A difference of up to $1.9 \pm 1.7 \times 10^{10}$ kg, around day number 333, is evident in this period. Although, as already noted, the polar ozone mass of 1989 (thin solid line) corresponds well with that of 1987 (thin dashed line) during late September and early October, large differences are noticeable for late October and early November. This is the time of the vortex breakup which leads to warmer temperatures (leading to evaporation of PSC's) and mixing of lower latitude, ozone rich air.

Colour coded temperature profiles above Sanae for 1987 are given in Figure 4.20 (some data were unavailable) and those for 1989 in Figure 4.21. The coldest temperatures are represented by blue shades. At altitudes of ~50 mb, the three darkest blues represent a high probability of PSC formation. The pattern of warming associated with the

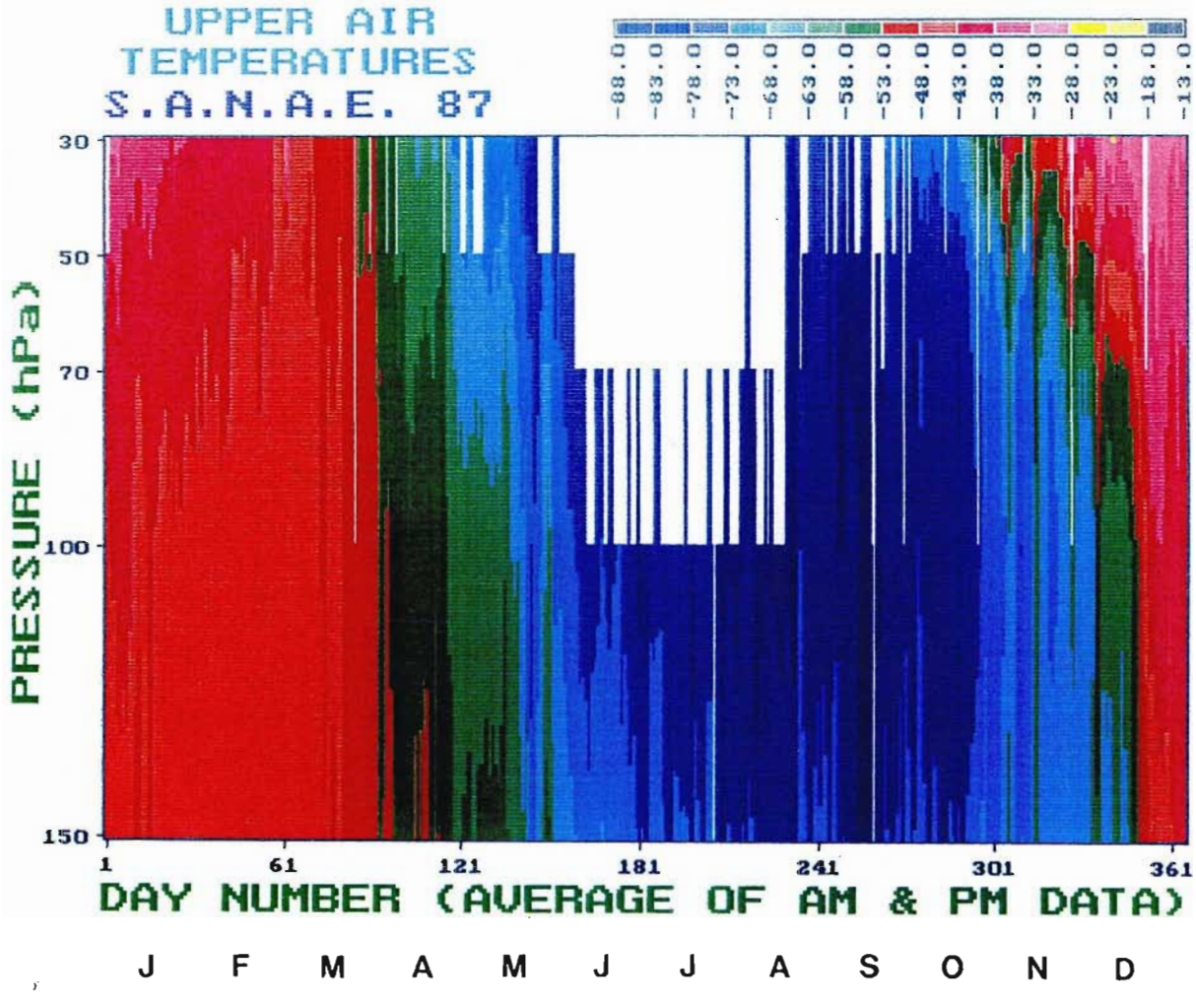

Figure 4.20: Temperatures between the 150 to 30 mb pressure surfaces above Sanae, Antarctica for 1987. Blue colours represent the lowest temperature ranges, increasing through green, red and yellow. Data gaps are due to inclement weather.

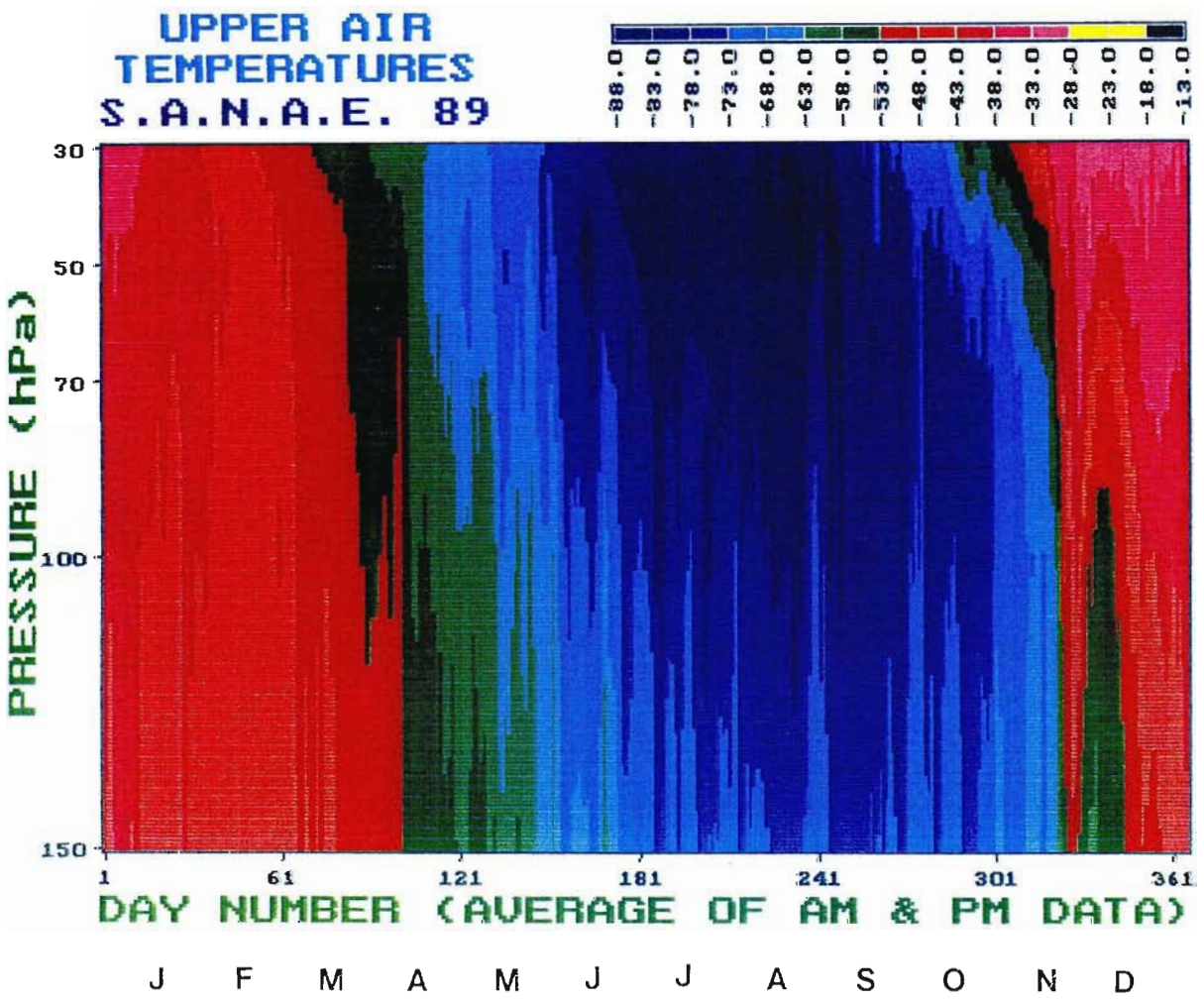

Figure 4.21: Temperatures between the 150 to 30 mb pressure surfaces above Sanae, Antarctica for 1989. Blue colours represent the lowest temperature ranges, increasing through green, red and yellow.

vortex breakdown illustrated by the appearance of green, and then red, shades around day number 280 is fairly similar in 1987 and 1989. This suggests that the 1987-1989 difference in the recovery of the ozone hole may not be solely dynamical. Additional NO_x at this time may, instead of leading to homogeneous gas phase destruction of ozone, lead to the increased rate of the confinement of liberated chlorine radicals into inert reservoirs such as $ClONO_2$. Although the models of *Reid et al.* and *Jackman et al.* do not include chlorine and bromine chemistry, *Jackman et al. (1993)* postulate that its inclusion would lead to reducing ozone's sensitivity to NO_y at a time when $ClONO_2$ can build up in low sun conditions.

Clearly, since background chemical and dynamical characteristics are crucial in determining the ozone response to NO_x enhancement, the period from August to December should be divided into intervals of similar atmospheric conditions. Thus, the possible effect of the first two of the SPE's (day numbers 224 and 272) which were coincident with dark conditions and PSC formation will be discussed separately to the final two SPE's (day numbers 292 and 334) which occurred during the vortex breakup and sunlit conditions. The effect of the first two SPE's is considered below and is followed by discussion of the possible ozone response to the last two SPE's.

During the winter and spring PSC's (see section 2.5), which form in the extreme cold of the Antarctic middle atmosphere, are surfaces for the heterogeneous conversion of passive chlorine reservoirs into reactive chlorine species. The vast majority of PSC's observed are not pure water ice or nacreous clouds since they have extinctions less than 10^{-2}/km (*Hamill et al., 1986*). They probably consist of a frozen form of nitric acid with three water molecules ($HNO_3.3H_2O$) called nitric acid trihydrate (NAT). These clouds form at a higher temperature ($-78°C$) than their nacreous counterparts ($-83°C$). The formation of NAT PSC's results in a highly denitrified stratosphere as the reactive nitrogen, which can trap chlorine into a reservoir molecule $ClONO_2$, is converted into nitric acid. During this time period there is still significant radiative cooling, leading to a descent of air in the polar vortex, thus increasing NO_x abundances at 16 to 32 km (*Callis and Natarajan, 1986*) as they are transported to altitudes where PSC's are present.

Peter et al., 1991 have modelled the effect that odd nitrogen from exhaust gases of 600 high-flying aircraft may have on NAT PSC formation probabilities. Since increased concentrations of odd nitrogen lead to a higher saturation temperature (in their case an increment of 5°C), PSC's are able to form at higher temperatures. They predict a doubling in occurrence of NAT PSC's and an even stronger increase of ice condensing on NAT particles for northern polar latitudes. Coincidentally, the mass of odd nitrogen produced by the aircraft is similar to that deposited by a large SPE. The purpose of this work is to ascertain what effect odd nitrogen species may have when added directly to low altitudes in the polar stratosphere (thousands of protons with energies $>$ 100

MeV reaching 30 km and below were detected) during the occurrence of the ozone hole.

A study by *Steele et al. (1983)* showed that PSC observations are highly correlated with low temperatures. Cloud is continuous in regions where temperatures are around or below $-83°C$ (*Toon and Turco, 1991*), inferring that the primary prerequisite for PSC formation is temperature. Assuming that NAT PSC's may form at temperatures of $-78°C$, the Antarctic ozone depletion is principally confined to the region inside the polar vortex where the lower stratospheric temperatures remain below about -78°C for several months and PSC's are frequently observed (*Toon and Turco, 1991*). The physical size of the ozone hole appears to be largely determined by the spatial extent of the cold temperature region, which in turn controls the area of potential PSC formation. We infer from temperature data for Sanae (70°S, 2°W), Antarctica in Figure 4.21 that days of possible PSC formation are from day number 155 to 290 in 1989 as indicated by the darker shades of blue at approximately 50 hPa. The corresponding time interval of possible PSC formation in 1989 is marked by a horizontal arrow in Figure 4.17. Of course, the underlying assumption here is that Sanae, being located on the periphery of the area under consideration viz 90°S-70°S, represents an upper limit of prevailing temperatures over the south polar cap. The effect on the northern polar ozone might be quite different as moonlight observations suggest that denitrification is a much smaller effect there (*Gernandt*, private communication).

In order to quantify any effects on ozone concentrations due to additional NAT PSC's we calculated the extent in area of low ozone, < 250 DU (used to define the ozone hole by *Proffitt et al. (1989), a*) within the latitude regions 90°S-70°S and 80°S-70°S. The difference in the area covered by the < 250 DU contour in 1989 and the area covered by this contour during five year average (1988, 1987, 1986, 1985 and 1984) are shown as thick (solid and broken) lines in Figure 4.22. Solid lines represent area difference calculated within the latitude zone 90°S-70°S and the broken line 80°S-70°S.

A noticeable feature of the thick lines in Figure 4.22 is that they are very similar in variance and magnitude, from which we may infer that any changes in area are largely in the zonal band 80°S-70°S. This is what we might expect since lower stratospheric ozone poleward of 80°S is almost totally absent throughout the ozone hole period (*Stolarski et al., 1990*) and the processing by PSC's is therefore saturated. These very low total column values indicate only residual tropospheric and upper stratospheric ozone (*Stolarski et al., 1990*). The increase in area of the < 250 DU contour between 80°S and 70°S in 1989, compared with the mean of previous years, varies between 1.3 and 6.0 $\times 10^6$ km^2, well outside any error that maybe incurred by the difference in data versions.

To confirm this increase, a similar analysis between 1989 and 1987 (thin broken and solid lines in Figure 4.22) was performed. Again, the solid line is for the region 90°S-

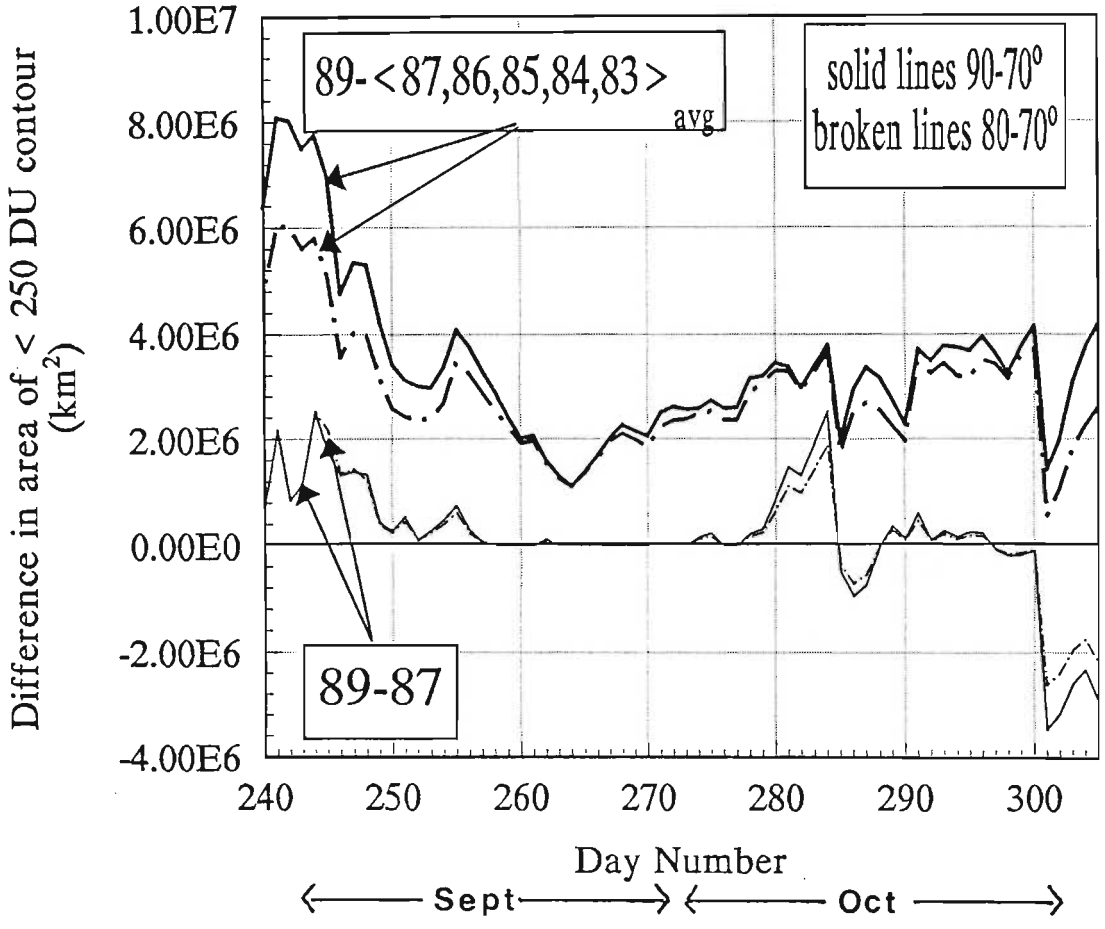

Figure 4.22: The difference in spatial extent of the <250 DU contour between 1989 and 1987 (thin curves) and 1989 and a five year average (thick curves). Solid lines indicate that calculations were performed within 90°-70°S latitude. Broken lines indicate that calculations were performed within 80°-70°S latitude.

70°S and the broken for 80°S-70°S, which once again match each other in variance and magnitude. Areas of increased ozone depletion can be up to 2.632 $\times 10^6$ km^2. The accuracy of the results ($\sim 10^3$ km^2) is limited only by the TOMS polar latitude cell size since version 5 TOMS data were used for both 1987 and 1989 data.

The effect of the final two SPE's in 1989 (day numbers 292 and 334) is hard to quantify as vortex erosion in the last two weeks of October 1989 (*Stolarski et al., 1990*) implies high ozone temporal and spatial variations. However, the thin lines in Figure 4.22 (which show differences between the previous worst case hole of 1987 and 1989) show a negative change for this time of year, indicating the more rapid recovery of the 1989 ozone hole. The 1989 ozone hole (defined as areas over which total ozone < 250 DU) was 3.47 $\times 10^6$ km^2 smaller at the end of October than that of 1987. The rapid ozone hole dissolution in 1989, as compared with 1987, is clearly obvious when inspecting the 1989 and 1987 mass profiles following day number 300 in Figure 4.17. This can also be detected by on inspection of the colour plots in the lower panels of Figures 4.18 and 4.19. Since dynamics were similar for these two years, it seems likely that differences in ozone recovery may be attributed to chemical factors. The increased abundances of NO_x produced by the SPE's in October and November, may have contributed to the 1989 ozone hole's relative quick healing by effectively tying up the chlorine radicals into $ClONO_2$ reservoirs. This has been suggested by *Jackman et al. (1993)* in speculating the effect of the inclusion chlorine chemistry in their three-dimensional modelling of the October 1989 SPE.

In summary, an increase in the spatial extent of the 1989 ozone hole compared to previous years, even 1987, was detected. The enhancement of NO_x concentrations due to SPE's, occurring simultaneously with the ozone hole, may be seen as a likely source to increase the probability of NAT PSC formation, and so enhancing chlorine catalysed ozone destruction. However, the addition of NO_x may, in some instances, result in larger ice particles rather than increasing the area of coverage of PSC's and additional odd nitrogen will have no effect in regions where processing of ozone is complete. Analysis of total ozone data later in the year suggests that the further NO_x enhancements produced by SPE's during the breakup of the polar vortex and subsequent warming of the polar region, could lead to rapid repair of the 1989 ozone hole as they are able to nullify active chlorine radicals.

4.5 Case study 3: July 1982

4.5.1 Solar data and subsequent satellite NO observations

Solar parameters for the July 1982 SPE are set out, in the same format as the SPE's in the previous case studies, in Table 4.3. The GOES-7 measured proton (> 10 MeV) flux of 2900 particles cm^{-2} s^{-1} $ster^{-1}$, makes it the largest SPE of solar cycle 21. Geomagnetic conditions were extreme, with the K_p index attaining a maximum value of 9.

McPeters (1986) used the SBUV instrument aboard Nimbus 7 to measure NO concentrations following this SPE. The SBUV scans, once every 24 days, all wavelengths in the interval 160 to 400 nm, with a resolution of 0.2 nm. This facilitates accurate measurement of the vibrational emission spectrum of nitric oxide. The (10), (01) and (02) Γ bands appear at 215, 236 and 246 nm respectively. Results from this analysis are plotted as the cumulative amount of NO above 1 mb (10^{14} molecules cm^{-2}) for the 20 July 1982, a week following the SPE, as the dot-dash-dot curve in Figure 4.23. For comparison, data from the 25 July 1981 (solid curve) and the 15 July 1983 (dashed curve) which represent solar quiet conditions are also shown in Figure 4.23. Measurements at high southern latitudes indicate that there was far more NO over this region in 1982 than in either the previous or following years. An increase amounting to 5 $\times 10^{14}$ molecules cm^{-2} above 50 km was detected by analysis by *McPeters (1986)* of the spectrum recorded by the SBUV instrument over southern polar regions. This translates, by theoretical calculation, to $\sim$7 $\times 10^{14}$ molecules cm^{-2} in the area above 60°S. The NO enhancements following the SPE were still detectable on September 6. In the northern hemisphere, on the other hand, there is no evidence, as shown in Figure 4.23, of prolonged NO increments. The interhemispheric difference in NO production subsequent to the SPE may be attributed to seasonal differences. The chemical lifetime of NO near the austral winter pole is extended due to lack of solar irradiation, whereas its lifetime is considerably shortened over the summer pole.

4.5.2 Southern polar cap ozone response

Again, we calculate the total ozone mass within 90-70°S which is plotted in Figure 4.24. The 1982 ozone mass profile is represented by a solid curve, the 1981 by the dot-dash-dot curve and the 1980 by the dotted curve. The vertical arrow marks the SPE onset. No ozone masses can be estimated before day number 214 (beginning of August) since there is no solar illumination above 70°S. The ozone mass profiles extend well past the last detection of NO enhancements on September 6 (day number 250)

onset date of SPE	max k_p	major flare date	major flare class	proton flux > 10 MeV	proton fluence > 30 MeV
11 July (>4)	9	09 July	3B/X9	2900	> 2.2 X 10^7

Table 4.3: The SPE, and associated solar flare, during July 1982. Data from *Solar Geophysical Data (1985)*.

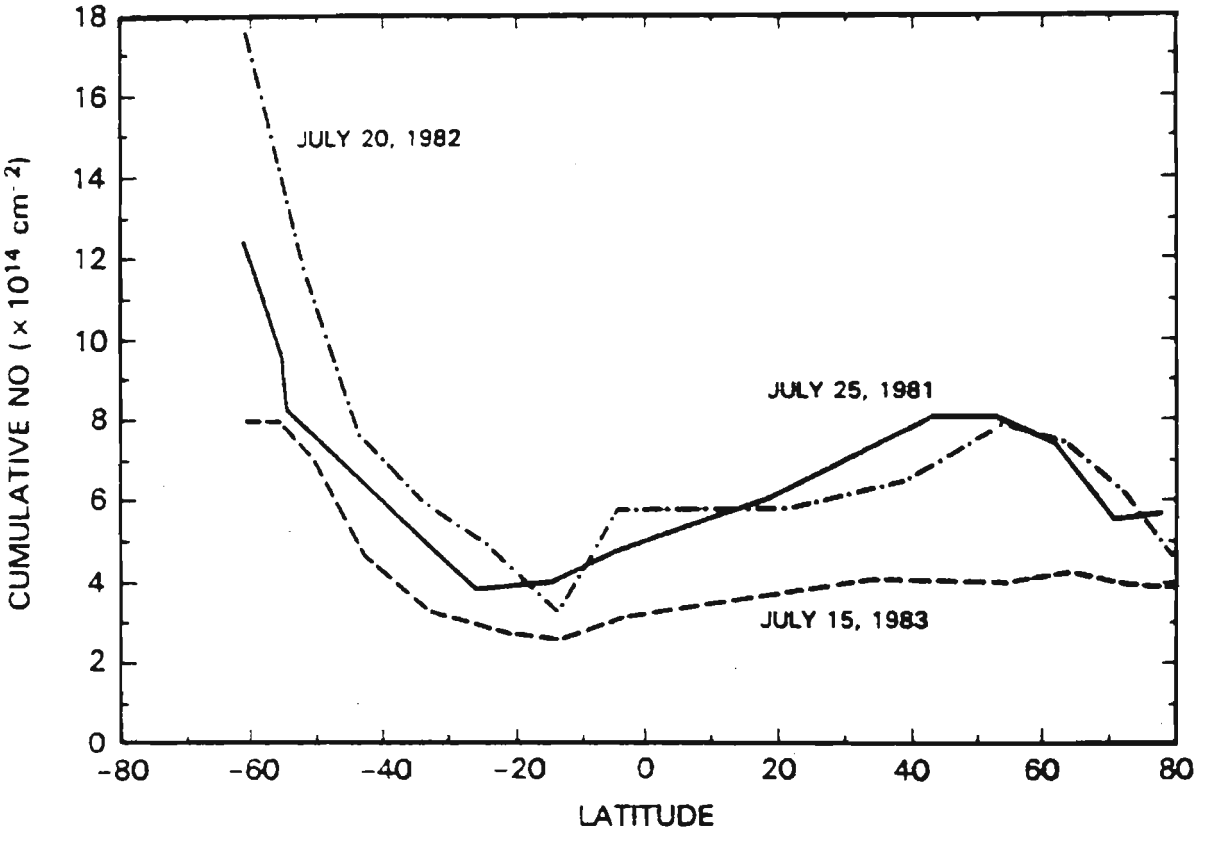

Figure 4.23: The meridional distribution of the NO column above 1mb following the July 1982 SPE (dot-dash-dot curve) and solar quiet years 1981 (solid curve) and 1983 (dashed curve). From *McPeters (1986)*.

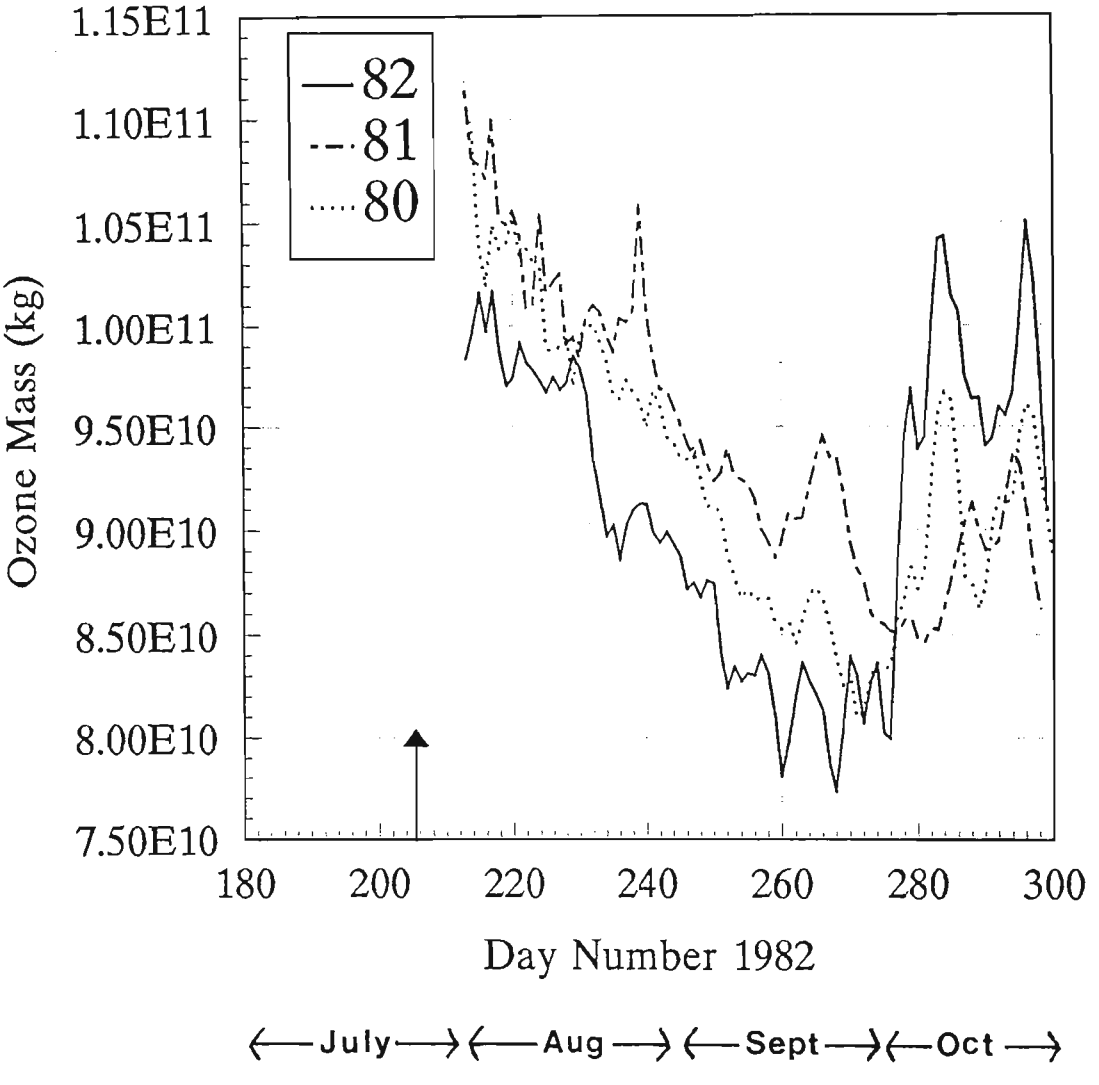

Figure 4.24: The ozone mass, calculated from TOMS total column ozone, over 90°-70°S latitude following the SPE in 1982 (solid curve). Quiet year austral polar ozone masses are represented by the dot-dash-dot curve (1981) and the dotted curve (1980). A vertical arrow indicates SPE onset.

to allow assessment of the SPE within the context of the development of the ozone hole. The ozone mass loss that is common to all years, during the period up to day 260 can be ascribed to the natural austral winter decline *Dobson (1966)* over austral polar latitudes, as discussed in section 2.2. Continuing ozone decline after day 260 in each year is due to chlorine catalysed depletion.

We now address any differences between the ozone mass profile of 1982 and those of previous years. Since we are unable to estimate the ozone mass before the SPE we calculate the amount of maximum ozone loss in 1982 compared to a 1980/81 average. By day 234 (12 August) 1982 there was 9.2 $\times 10^9$ kg less ozone than the 1980/81 average. An average total mass of ozone down to 70°S has also been calculated over the interval from day number 214 to day number 245 in 1980 and 1981. These averages with standard deviations are, respectively, 9.98 $\pm$ 0.38 $\times 10^{10}$ kg and 10.24 $\pm$ 0.39 $\times 10^{10}$ kg. The additional ozone loss in 1982 is two to three times standard deviation during the solar quiet years, implying that the observed decrease is not due to natural variability of ozone at this time of year. Colour plots of total column ozone for day number 234 of 1980 are shown in the upper image of Figure 4.25 and day number 235 of 1982 in the lower image. Comparison, by visual inspection (which is deemed to be representative of the ozone amounts for the dynamically quiet month of July) of these plots, suggests that there are indeed lower ozone levels over Antarctic regions following the July 1982 SPE. Prevailing winter conditions imply a longer chemical lifetime of NO and downward transport to lower altitudes. However, without sunlight there is no mechanism for NO gas phase destruction of ozone so that any ozone loss associated with a SPE over the southern polar cap would not be expected until some illumination returns there. Therefore it may be likely that this ozone depletion cannot be attributed to a gas phase mechanism but, as in case study 2, the additional NO_x may rather be effective by increasing the probability of NAT PSC formation. Chemistry at this time of year has proved to be very sensitive to dynamics. During August-September the QBO was in the west phase for 1980 and 1982 and the east phase for 1981 (*Lait et al., 1989*) inferring similar dynamic conditions for 1980 and 1982. Comparison of only 1982 and 1980 mass profiles reveals that 1982 is still significantly lower. However, small differences in the QBO between these two years could account for decline rate deviations in September.

4.5.3 Northern polar cap ozone response

Ozone masses calculated over 90-70°N are shown in Figure 4.26. The horizontal scale is different to that of Figure 4.24, showing only the calculated ozone masses until mid-September. The decline in ozone mass, common to each year, over this period may be attributed to summer ozone decline over northern polar cap, as discussed in section 2.2.

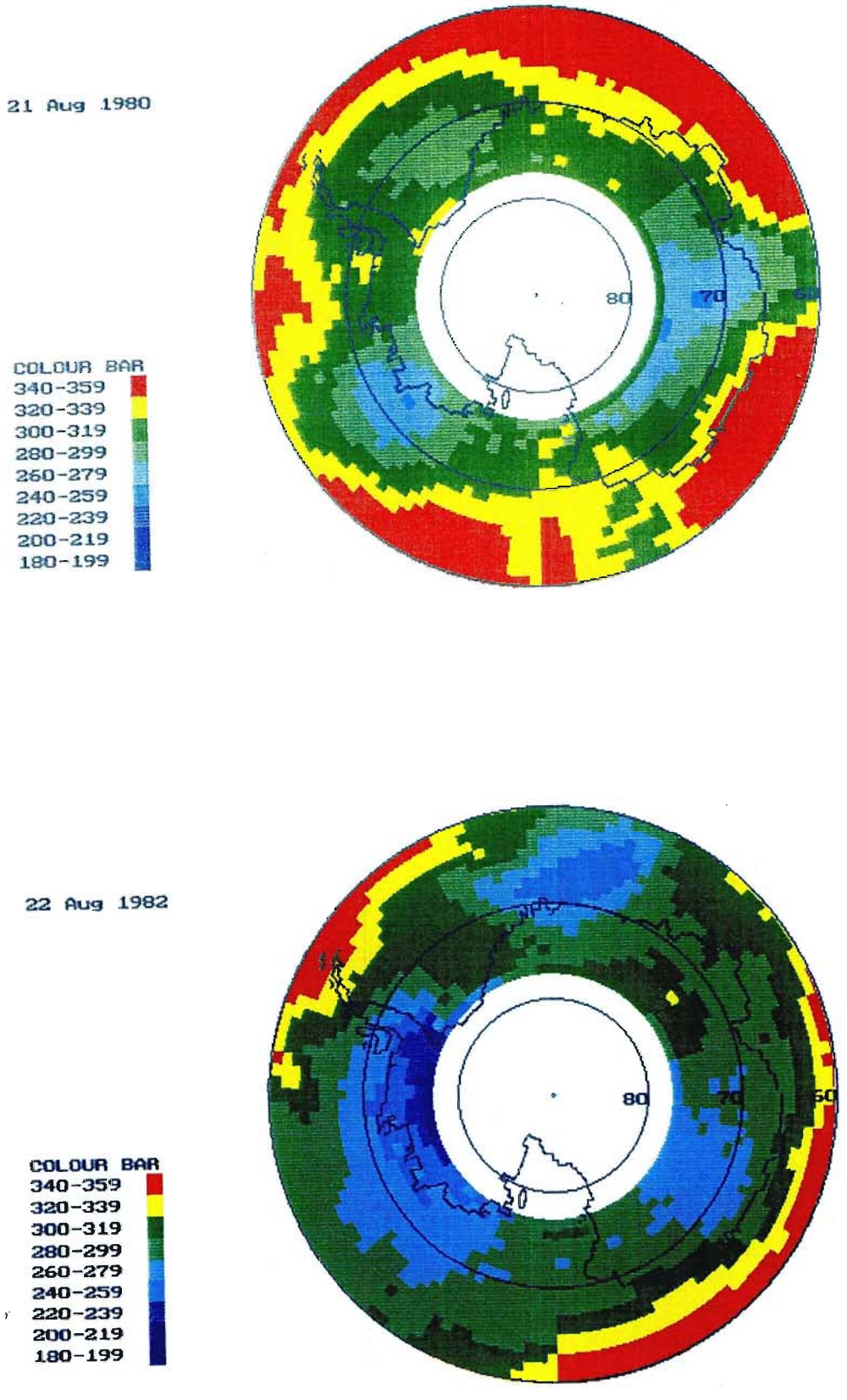

Figure 4.25: Colour plots of total column ozone over Antarctica for 21 August 1980(top) and 22 August 1982 (bottom). Each colour represents 20 DU.

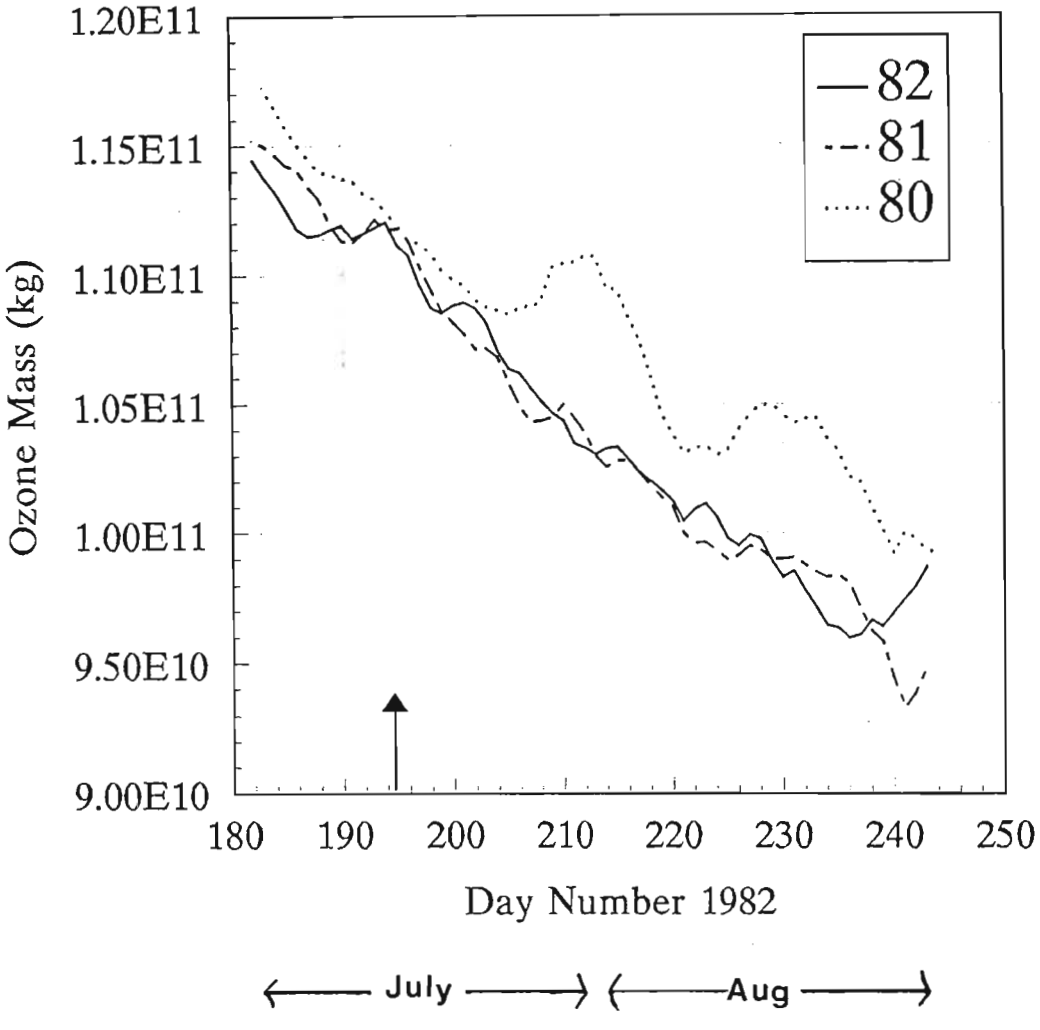

Figure 4.26: The ozone mass, calculated from TOMS total column ozone, over 90°-70°N latitude following the SPE in 1982 (solid curve). Quiet year boreal polar ozone masses are represented by the dot-dash-dot curve (1981) and the dotted curve (1980). A vertical arrow indicates SPE onset.

Clearly, there is no marked difference between the decline rate of ozone masses between 1982 (solid curve) and previous years. The absence of additional ozone depletion in 1982, that might be associated with a SPE, is corroborated by the NO observations of *McPeters (1986)* which indicate that there were no measurable NO enhancements over high northern latitudes. The interhemispheric asymmetry of the ozone loss is similar to that following the SPE of August 1972, indicating that atmospheric conditions are of prime importance when evaluating ozone loss resulting from SPE's.

4.6 Summary

For all three case studies, techniques were employed to estimate, quantitatively, the ozone loss over the entire polar cap region (90-70°) following large SPE's.

The results of the analysis of the TOMS data during March 1989, which allowed an assessment of the simultaneous ozone response over both polar caps show:

- over the region 90-70° similar ozone mass depletions of 7.4 $\times 10^9$ kg for the south polar cap and 8.0 $\times 10^9$ kg for the north polar cap, indicating some degree of symmetry for this event.

- there was deviation, albeit in detail, from the observed ozone depletion. The possible source of this is the contrasting autumn conditions over the south polar cap with the spring conditions prevalent in the north. Springtime ozone values exhibit large spatial and temporal variability due largely to planetary wave dynamics at this time. Meanwhile, autumn conditions are relatively quiescent over the austral pole.

Due to lack of solar irradiance only the southern polar cap ozone response was observable following the SPE's in the latter months of 1989. Great caution was exercised when interpreting the analysis of total ozone as the SPE's were concurrent with the complex dynamics and chemistry that accompany the formation and dissolution of the ozone hole. However, results suggest that:

- the SPE's in August and September enhanced the NAT PSC formation within the zonal strip 80 to 70°S. Total ozone < 250 DU covered a greater area (up to 2.63 million km^2) than the previous 'worst case' ozone hole of 1987

- the SPE's in October and November retarded the chlorine radical destruction of ozone (as proposed by *Jackman et al. (1993)*) by providing large quantities of NO_x which are able to react with chlorine monoxide and result in the formation of the inert reservoir $ClONO_2$. The 1989 ozone hole was 3.47 $\times 10^6$ km^2 smaller than that of 1987 by the end of October.

Total ozone mass observations following the July 1982 SPE agree with satellite observations of NO by *McPeters (1986)*. Ozone mass profiles during the period following the SPE reveal:

- an ozone mass depletion of 9.2 $\times 10^9$ kg by August 12 over the austral winter polar cap compared with previous years. The mechanism of this depletion is uncertain. Although gas phase NO_x destruction cannot be completely ruled out as solar irradiation returns to the polar regions in August, it is thought more likely that this depletion could be ascribed to increased NAT PSC formation.

- no ozone mass loss was detected which substantiates the absence of NO_x enhancements in satellite observations of NO. Odd nitrogen chemical lifetimes are thought to be too short in summer polar cap to be effective in ozone depletion.

It is clear from these case studies that the ozone response can not be determined simply by a knowledge of the particle profile of an SPE and subsequent ion pair production estimates. Although important, these studies show that the prevailing dynamics (especially planetary wave and QBO effects), meridional circulation patterns, photochemistry and competing heterogeneous chlorine chemistry can play the most crucial role in ultimately determining the quantitative ozone loss following a SPE. This is reflected by the inability of some models (in particular two-dimensional versions) to accurately predict these depletions.

4.7 Suggestions for future work

Computer simulations, performed by other authors, of the ozone response to SPE's have been discussed in chapter 2 of this thesis. We do not have such a model. Clearly,

it would be very beneficial if the author could gain some experience in performing such calculations.

The analysis performed in the case studies in this chapter calculates the ozone response to SPE's within a spatial extent defined by geographic latitude. However, spatial constraints such as geomagnetic latitude or potential vorticity contours, could have been applied. Clearly, it would be advantageous to determine whether these alternative spatial coordinates could provide further, and maybe different, information on the relationship between enhanced NO_x levels, due to SPE's, and ozone.

Satellite vertical profile ozone data would be very valuable in assessing the effects of SPE's. The SBUV experiment failed in February 1987 (*Miller, 1989*). The Stratospheric Aerosol and Gas Experiment (SAGE II) aboard the Earth Radiation Budget Satellite (ERBS) employs the solar occulation technique to determine ozone vertical profiles between altitudes of 10 and 65 km. This satellite has a highly precessing 55° inclined orbit (*McCormick et al., 1989*) providing observations of limited spatial extent each day. Although it does, on occasion, make observations near 80°, during periods of SPE's in 1989, the satellite was over middle or low latitudes.

Atmospheric parameters (temperature and wind speed) were observed at Sanae during the SPE's, and were used to establish prevailing conditions over the southern polar cap at these times. However, since Sanae is located on the periphery of the analysis region, it would be advantageous to have further data at higher latitudes. These data were not immediately available.

SPRI has installed a Système d'Analyse par Observations Zénithales (SAOZ) spectrometer at Sanae, Antarctica (70.3°S;2.4°W). The ground-based ozone response, if any, during future SPE's will be monitored by this instrument.

Chapter 5

Long term variations in total ozone over Antarctica

5.1 Introduction

The purpose of this study is to investigate the long term (months to years) variations in the austral polar ozone field. In order to achieve this, a data set is required that offers long periods of uninterrupted measurements and has extensive coverage over the southern polar cap. Continuous total ozone measurements, extending over many years, are relatively rare, especially over Antarctica. The first Antarctic ozone measuring stations were established during the International Geophysical Year (IGY), 1957 (*MacDowall, 1960*). Single station measurements are strongly influenced by regional weather systems and so, exhibit large fluctuations. Therefore, ground-based measurements are not ideally suited for the detection of the more subtle long term influences (*Dütsch, 1979*). The TOMS instrument, aboard the Nimbus 7 satellite, has monitored global ozone on a daily basis throughout the period from November 1978 to May 1993. This constitutes the longest, uninterrupted satellite ozone data set. The acquisition and format of TOMS total ozone data is described in chapter 3 of this thesis. These data are therefore unique in that they permit detection of long term oscillations of the large scale structure of total ozone over Antarctica, not just above one station. The area of interest chosen for this thesis is the southern polar cap (90-70°S). In order to provide one value per day to represent the amount of ozone over the south pole, the ozone mass was calculated by the method described in section 4.2.

An eleven year data set, 1981 through to 1991, was selected from the TOMS version 6 data records. It was fortuitous that this interval incorporates two strong solar maxima, providing the strongest possible solar cycle signal in the polar ozone data. Furthermore,

some of the largest SPE's ever recorded occurred over this period. The analysis of some of these SPE events was described in chapter 4. Dynamical conditions were also pronounced in this interval which includes two strong ENSO events.

A Fast Fourier Transform (FFT) was applied, using a Hewlett-Packard 9000 mainframe computer, to the diurnal polar ozone mass values, derived from the TOMS instrument, in order to determine the principal periodicities of Antarctic total column ozone.

Solar radio (10.7 cm) flux data were selected as a yardstick for solar ultraviolet forcing during the eleven year interval. These data correlate particularly well with solar extreme ultraviolet output, and are often used as a proxy to characterize the solar extreme ultraviolet flux for computation of atmospheric effects (*Walterscheid, 1989*). An in-depth review of the data acquisition and format of the solar radio flux data can be found in chapter 3. A FFT was applied to this data in exactly the same manner as it was applied to the ozone data. This procedure is described in section 5.3.

Interpretation of the resulting spectra requires some caution, as the data sets extend over only one solar cycle. In addition, a plausible mechanism must be demonstrated when assessing the significance of the frequency spikes. The current understanding of these mechanisms has been outlined in chapters 1 and 2 of this thesis. They will be highlighted again in sub-section 5.4.1 entitled 'Interpretation and discussion of significant peaks' in this chapter.

Although the FFT was only applied to the southern polar cap ozone data in the work of this thesis it can, of course, be applied in the same way to northern high latitude ozone data. However, the burden in terms of computer time did not allow for the inclusion of the northern hemisphere long term variation study in this thesis. This is intended to be a future project by the author. The southern hemisphere was chosen in preference to the northern hemisphere as it was thought that solar effects would be more dominant there due to the reduced amount of mixing associated with planetary wave activity. *Labitzke and van Loon (1988)* found that the effect of solar activity on ozone may be influenced by the phase of the QBO. This result implies a complex picture of solar-ozone relations which is confirmed by the Fourier analysis performed here.

5.2 Fourier theory

5.2.1 Groundwork

The application of a Fourier transform to any waveform identifies the unique set of sine functions which, when combined, can recreate the original waveform. Mathematically, this relationship is stated as

$$H(f) = \int_{-\infty}^{\infty} h(t)\, e^{-i2\pi ft}\, dt, \tag{5.1}$$

where h(t) is the waveform to be decomposed into a sum of sine functions, H(f) is the Fourier transform of h(t), and $i = \sqrt{-1}$. The Fourier transform, or -i transform, is a frequency domain representation of a function h(t). Throughout this chapter, a lower case symbol will represent a function of time while its Fourier transform will be represented by the corresponding upper case symbol, which is a function of frequency. In general, the Fourier transform of a function is complex

$$H(f) = R(f) + i\, I(f) = |\, H(f)\, |\, e^{i\theta(f)}, \tag{5.2}$$

where R(f) is the real part of the Fourier transform, I(f) is the imaginary part of the Fourier transform, | H(f) | is the amplitude such that

$$|\, H(f)\, | = \sqrt{R^2(f) + I^2(f)} \tag{5.3}$$

and $\theta(f)$ is the phase angle given by

$$\theta(f) = \tan^{-1}\left[\frac{I(f)}{R(f)}\right]. \tag{5.4}$$

The inverse Fourier, or $+i$, transform is defined as

$$h(t) = \int_{-\infty}^{\infty} H(f)\, e^{i2\pi ft}\, df. \tag{5.5}$$

Inverse transformation determines a function of time from its Fourier transform. If two functions h(t) and H(f) are related by the equations (5.1) and (5.5), they are termed a Fourier transform pair. This relationship is denoted by a double arrow,

$$h(t) \Leftrightarrow H(f). \tag{5.6}$$

5.2.2 Some useful transform pairs

The most frequently encountered transform pairs are illustrated in Figure 5.1. Functions

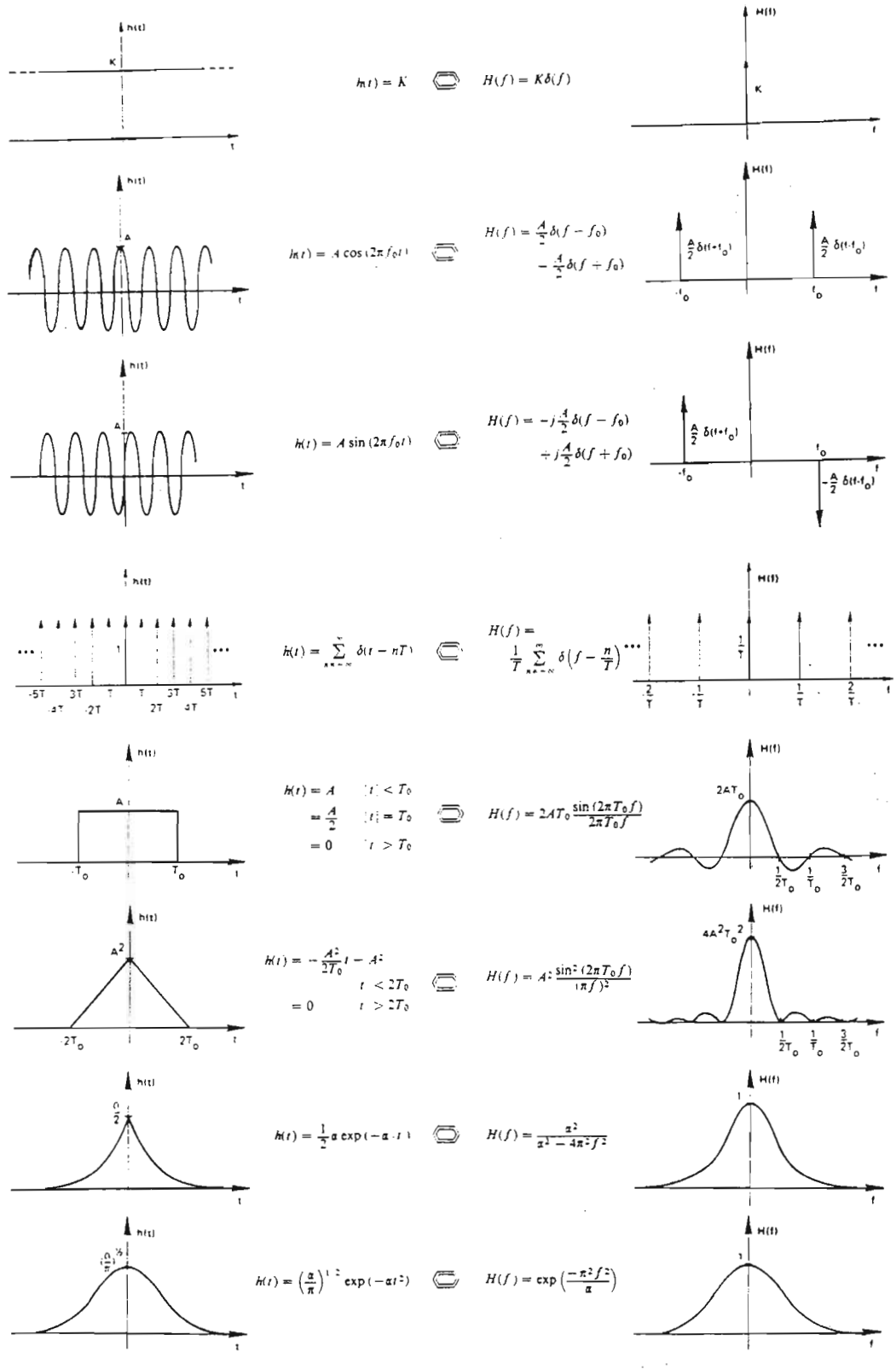

Figure 5.1: Frequently encountered Fourier transform pairs. From *Brigham (1974)*.

that are time dependent are sketched on the left side of Figure 5.1, and the -i Fourier transform of each function is plotted on the right side. Each transform pair will be addressed, starting with the uppermost pair and proceeding down to the last.

- **dc function**
  The dc function h(t) = K, where K is a constant, transforms to a frequency spike of amplitude K at f=0. A frequency spike is defined so that its integral is equal to its amplitude i.e. it is infinitely narrow. When the spike's amplitude, and so area, is of unitary value the spike is termed a δ, or impulse, function.

- **cosine function**
  The cosine function $A\cos(2\pi f_0 t)$ is plotted on the left and it's associated Fourier transform consisting of two δ functions, at f_0 and $-f_0$, corresponding to the frequency of the cosine function. Each δ function has an amplitude $\frac{A}{2}$. This transform is called an even impulse pair.

- **sine function**
  The Fourier transform of $A\sin(2\pi f_0 t)$ is an odd impulse pair. The impulse at $-f_0$ has an amplitude $\frac{A}{2}$, while that at f_0 has an amplitude $-\frac{A}{2}$.

- **shah function III(t)**
  The shah function, plotted on the left side of the fourth panel, is composed of a row of unit amplitude impulses occurring at an interval T. This function transforms to a row of impulses of amplitude $\frac{1}{T}$ spaced at $\frac{n}{T}$ frequency intervals.

- **top hat function Π(t)**
  The top hat function has a value of A for $[|\ t\ | \leq T_0]$ and 0 elsewhere so that the 'top hat' has a width of $2T_0$ and a height A. It transforms to a sinc function where

$$\mathrm{sinc}(2\pi f_0 t) = 2\,A\,T_0\,\frac{\sin(2\pi T_0 f)}{2\pi T_0 f}. \tag{5.7}$$

  By inspection, the sinc function has an amplitude of $2\,A\,T_0$ at f=0 and has zero value at the frequency node points $f = \frac{n}{2T_0}$. This transform implies that a square wave can be constructed by simply adding many cosine functions, each having their own characteristic amplitude, frequency and phase. The sharp edges of the top hat function are represented by the high frequency components of the transform.

- **Chinese hat function Λ(t)**
  A Chinese hat function, which has a triangular shape, is plotted on the right side of the sixth panel, in Figure 5.1. It's transform is the sinc^2 function with frequency nodes at $f = \frac{n}{2T_0}$ and a maximum amplitude of $4\,A^2\,T_0^2$ at f=0.

- **exponential function**
  This function consists of an exponential function, $h(t) = \frac{\alpha}{2} e^{-\frac{\alpha}{t}}$ and its mirror image in the t=0 axis. The transform of this function is $H(f) = \frac{\alpha^2}{\alpha^2 + 4\pi^2 f^2}$ is plotted on the left side of the penultimate panel.

- **Gaussian function**
  A typical Gaussian function of amplitude $\left[\frac{\alpha}{\pi}\right]^{\frac{1}{2}}$ at t=0 is transformed in the last panel of Figure 5.1.

These transform pairs have been derived by application of the Fourier integral equation (5.1) to a function of time h(t). However, it is not necessary to resort to mathematical integration each time a schematic transformation is required. By knowing these fundamental transforms and, by applying certain properties of the Fourier transform, it is usually easy to estimate even the most complicated of transforms, as illustrated in the following sub-section.

5.2.3 Fourier transform properties

For completeness, the fundamental properties of the Fourier transform are mathematically derived here. However, it is not easy to conceptualize the properties in this format, and so, each property will be illustrated by a graphical example.

Linearity

If x(t) and y(t) have the Fourier transforms X(f) and Y(f) respectively, then the Fourier transform of x(t) + y(t) is

$$\begin{aligned} \int_{-\infty}^{\infty} [x(t) + y(t)]\, e^{-i2\pi ft}\, dt &= \int_{-\infty}^{\infty} x(t)\, e^{-i2\pi ft}\, dt + \int_{-\infty}^{\infty} y(t)\, e^{-i2\pi ft}\, dt \\ &= X(f) + Y(f) \end{aligned}$$

so that

$$x(t) + y(t) \Leftrightarrow X(f) + Y(f). \tag{5.8}$$

To illustrate this concept, consider the determination of the Fourier transform of an infinite cosine function, frequency f_0, with an offset of K. This function is sketched in the lower left corner of Figure 5.2. Obviously, this function is a composite waveform, made up of the functions $x(t) = K$ and $y(t) = \cos(2\pi f_0 t)$. The transforms of both x(t) and y(t), which have been graphically determined in the previous sub-section, are

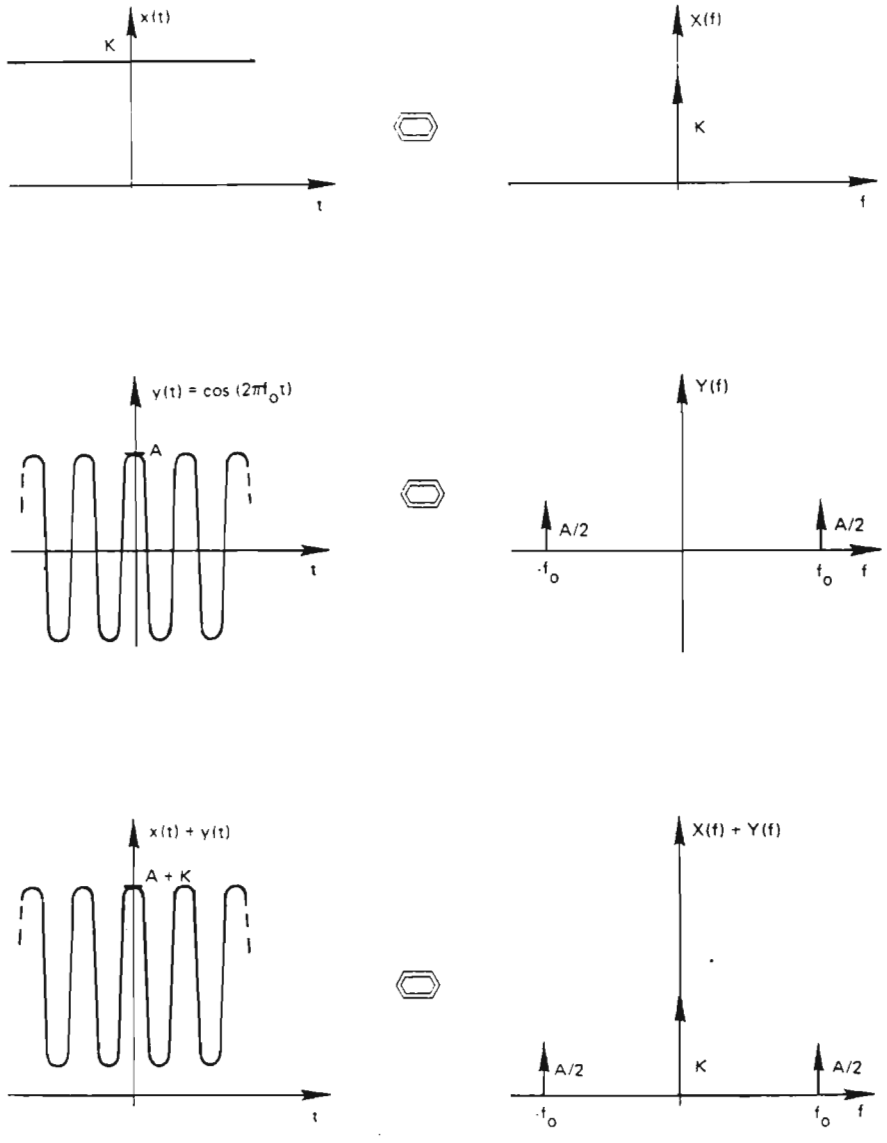

Figure 5.2: The linearity property of the Fourier transform. From *Brigham (1974)*.

shown in the top and middle panels respectively of Figure 5.2. Applying the linearity theorem, x(t) + y(t) = X(f) + Y(f), we can determine the Fourier transform of the offset cosine function by simply adding the two known transforms. This result, which is shown on the right side of the final panel, comprises an impulse function of amplitude K at f=0 and an even impulse pair of amplitude $\frac{A}{2}$ at frequencies f_0 and $-f_0$.

Symmetry

If h(t) and H(f) are a transform pair then

$$H(t) \Leftrightarrow h(-f). \tag{5.9}$$

This symmetry property is established by replacing t with -t in equation (5.5) so that

$$h(-t) = \int_{-\infty}^{\infty} H(f)\, e^{-i2\pi ft}\, df \tag{5.10}$$

and then, by interchanging t with f,

$$h(-f) = \int_{-\infty}^{\infty} H(t)\, e^{-i2\pi ft}\, dt. \tag{5.11}$$

This theorem ensures that once we have computed 'one way' of the Fourier transform, that the determination of the inverse transform is trivial. A function such as $\mathbf{\Pi}(t)$, where h(f) = h(-f), the transform pair then becomes H(t) $\Leftrightarrow$ h(f) and so, both the direct and inverse transform pairs are identical, as illustrated in Figure 5.3.

Time scaling

If the Fourier transform of h(t) is H(f) then the transform of h(kt), where k is a real constant > 0, can be determined by substitution of $t' = kt$ into the Fourier integral equation

$$\int_{-\infty}^{\infty} h(kt)\, e^{-i2\pi ft}\, dt = \int_{-\infty}^{\infty} h(t')\, e^{-i2\pi t' \frac{f}{k}}\, d\frac{t'}{k} = \frac{1}{k} H(\frac{f}{k}). \tag{5.12}$$

A more general form of the time scaling transform pair, which includes values of k < 0, is

$$h(kt) \Leftrightarrow \frac{1}{|k|} H(\frac{f}{k}). \tag{5.13}$$

An illustration of the application of time scaling theorem to $\mathbf{\Pi}(t)$ is given in Figure 5.4. The top panel depicts the standard transform pair. The middle panel shows a time scale expansion by a factor of 2 applied so that $h(\frac{t}{2}) = A$ for $[|\ t\ | \leq 2T_0]$.

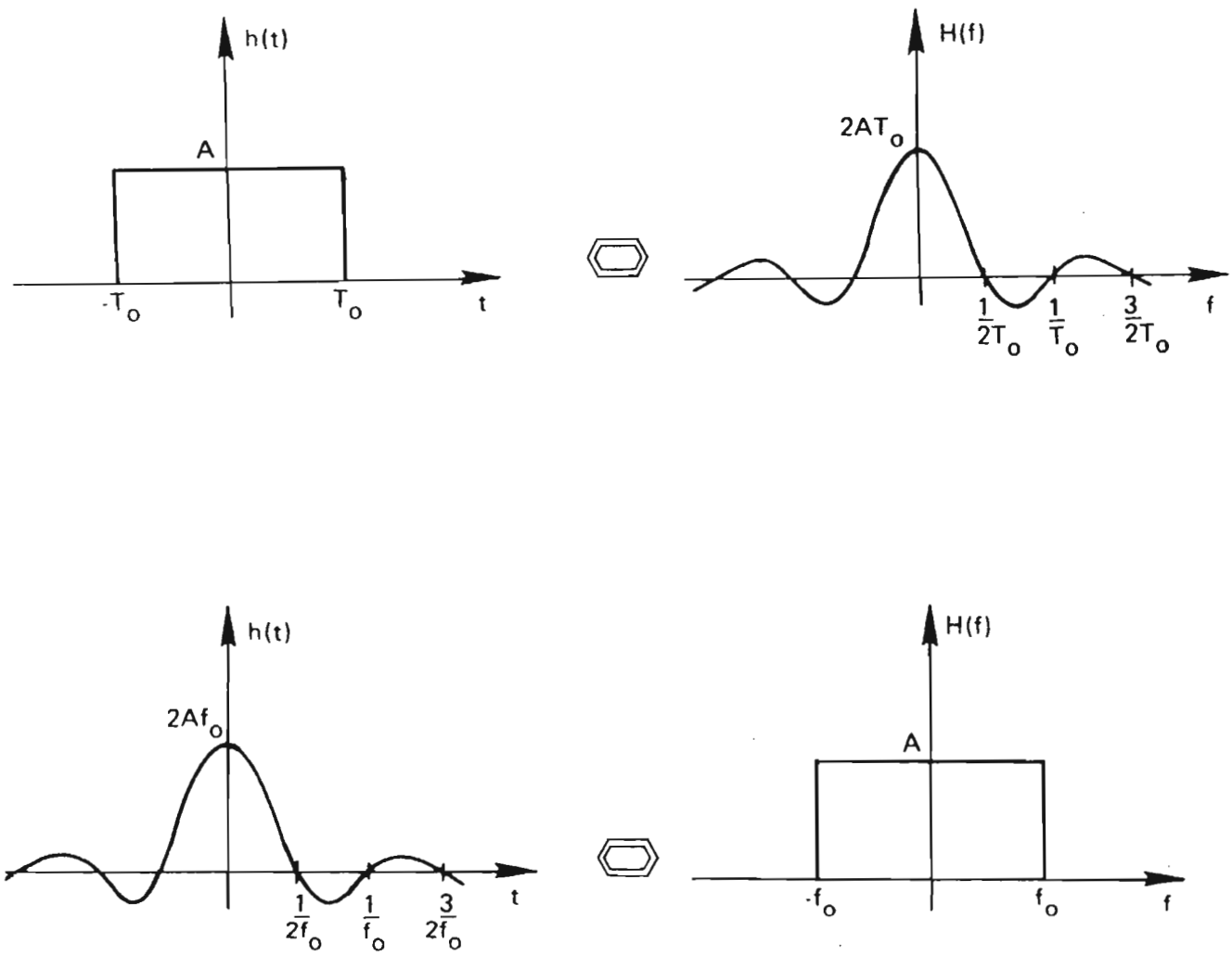

Figure 5.3: The Fourier transform of $\Pi(t)$ and the associated inverse transform of $\Pi(f)$. Adapted from *Brigham (1974)*.

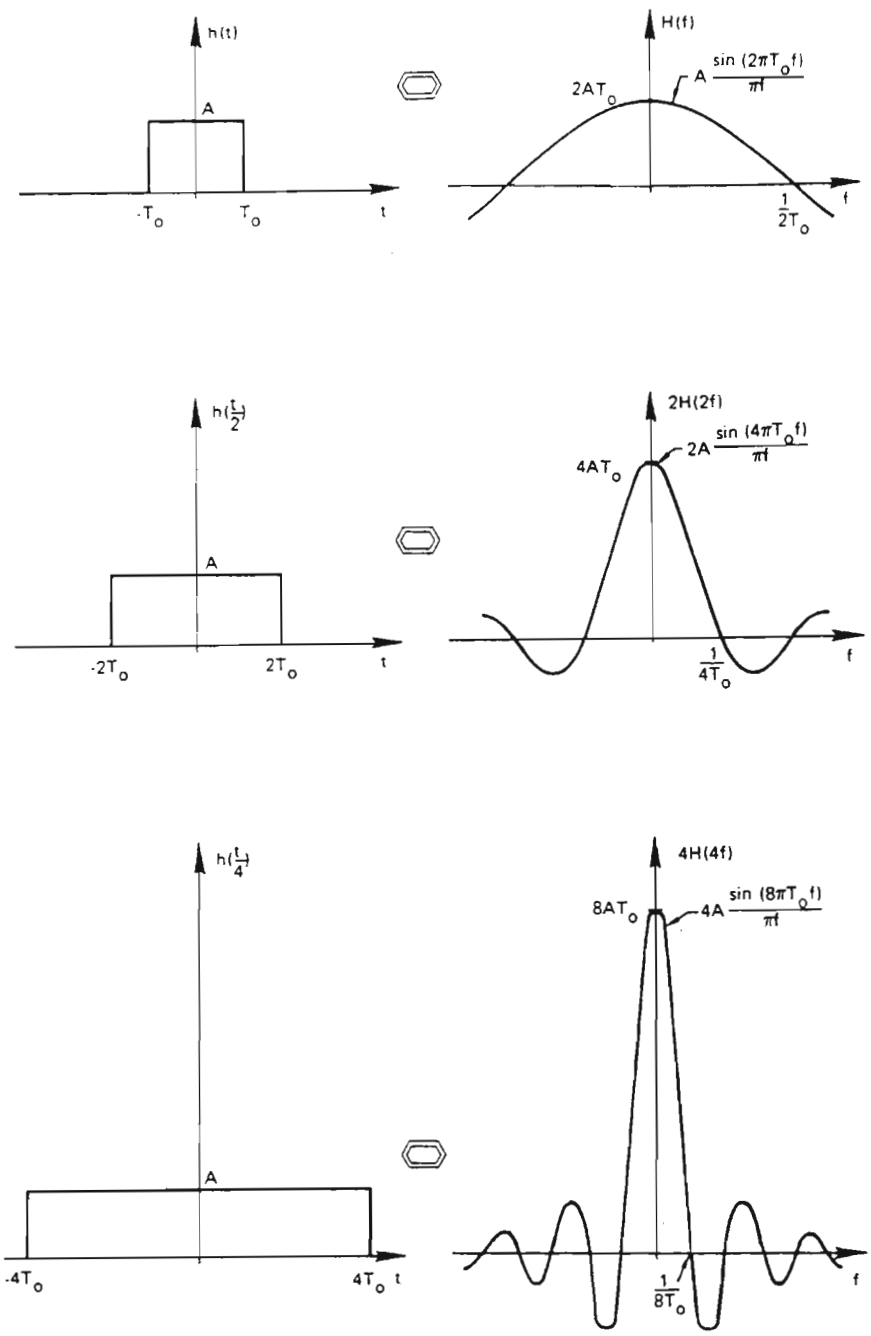

Figure 5.4: The time scaling property of the Fourier transform. From *Brigham (1974)*.

The Fourier transform of $h(\frac{t}{2})$, shown on its right, is compressed compared with the standard transform shown above it, with frequency nodes at $\frac{n}{4T_0}$. Associated with the frequency compression is an amplitude enhancement so that the amplitude, at f=0, is doubled to $4AT_0$. Despite an alteration in the time scale, the integral of the time function, h(t), must remain unaltered. The lowest panel displays the transform pair where the top hat function has a time scale that is expanded four fold.

Caution must be exercised when applying the time scaling theorem to impulses since they are defined to be infinitely narrow and so,

$$\delta(kt) \Leftrightarrow \frac{1}{|k|}\delta(f). \tag{5.14}$$

Frequency scaling

If the inverse Fourier transform of H(f) is h(t), the inverse Fourier transform of H(kf) where k is a real constant, is given by

$$\frac{1}{|k|}h(\frac{t}{k}) \Leftrightarrow H(kf). \tag{5.15}$$

The frequency scaling theorem can be mathematically established in a similar way to that of its time counterpart derived above, by substitution of $f' = kf$.

The effect of frequency scaling on the Fourier transform of the top hat function is illustrated in Figure 5.5. Once again, the top panel depicts the standard transform pair. In the middle panel the frequency scale has been expanded by a factor of 2. This implies a top hat function whose width is halved and whose height is therefore doubled. Of course, by expanding the frequency by a factor of 4, depicted in the last panel, the inverse transform is a top hat function that is a quarter of the width and four times larger than the original function.

Time shifting

If a function h(t) is shifted by a constant amount, t' its Fourier integral can be evaluated by letting $s = t - t'$. The time shifted Fourier transform pair becomes

$$\begin{aligned}\int_{-\infty}^{\infty} h(t-t')\,e^{-i2\pi ft}\,dt &= \int_{-\infty}^{\infty} h(s)\,e^{-i2\pi f(s+t')}\,ds \\ &= e^{-i2\pi ft'}\int_{-\infty}^{\infty} h(s)\,e^{-i2\pi fs}\,ds \\ &= e^{-i2\pi ft'}\,H(f).\end{aligned} \tag{5.16}$$

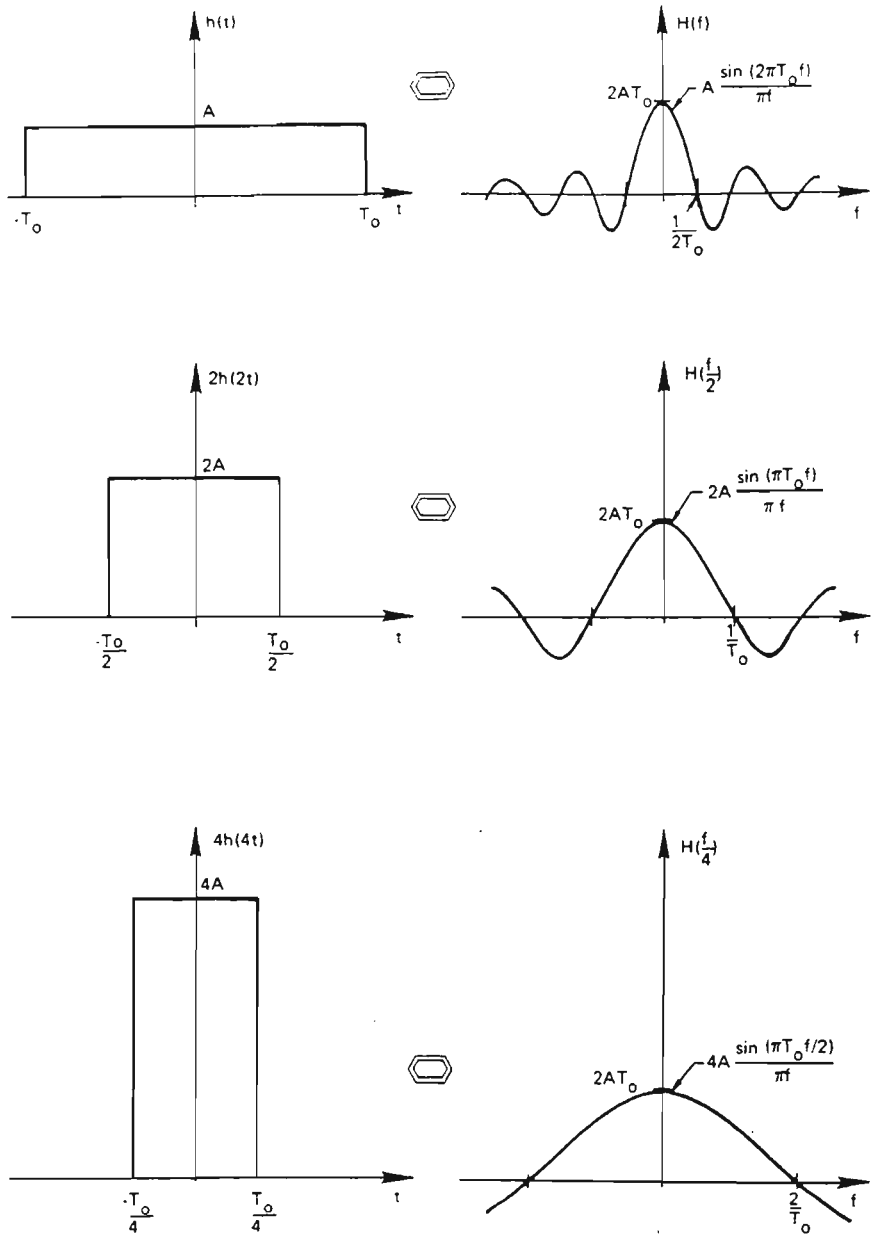

Figure 3-3. Frequency scaling property.

Figure 5.5: The frequency scaling property of the Fourier transform. From *Brigham (1974)*.

Or, in a more concise notation

$$h(t - t') \Leftrightarrow H(f)\, e^{-i2\pi f t'}. \tag{5.17}$$

Time shifting results in a change of phase angle, $\theta(f) = \tan^{-1}[\frac{I(f)}{R(f)}]$, so that the real, R(f), and imaginary, I(f), parts of the Fourier transform must be plotted to illustrate its effects. The magnitude of the Fourier transform, |H(f)|, remains unchanged. Figure 5.6 shows four transform pairs where the time functions have the form h(t - nt'). In the top panel n=0, n=1 in the second, n=2 in the third and, n=4 in the last panel. The cosine function on the left side of each panel has an amplitude of 2A and a period of 8t'. Time shifting h(t) alters the amplitudes of the impulse pairs in the real and imaginary parts of the frequency domain. However, the sum of their amplitudes is always |2A|. Attention is drawn to the third panel where h(t) is a sine function with a transform that is purely imaginary. On the other hand, the first and last panel which depict cosine curves, have transforms that are purely real.

Frequency shifting

If H(f) is shifted by a constant frequency f_0, by the symmetry theorem, its inverse transform must be multiplied by $e^{i2\pi t f_0}$ and so,

$$h(t)\, e^{i2\pi t f_0} \Leftrightarrow h(f - f_0). \tag{5.18}$$

This property can be derived in a similar fashion to its time counterpart, by substituting $s = f - f_0$.

A transform pair of an arbitrary function h(t) is plotted in the top panel of Figure 5.7. In order to shift the frequency spectrum, H(f), by f_0 as shown in the middle panel, it is necessary to multiply h(t) by a cosine function whose frequency is f_0. This process, called modulation, will be addressed in more detail in the next sub-section.

Oddness and evenness

The importance of symmetry in Fourier theory has already been suggested. However, more alertness is required to ensure full exploitation of symmetry properties in Fourier transforms.

A function e(t) such that e(-t) = e(t) is a symmetrical, or *even*, function. A function o(t) such that o(-t) = -o(t) is an anti-symmetrical, or *odd*, function. The sum of even

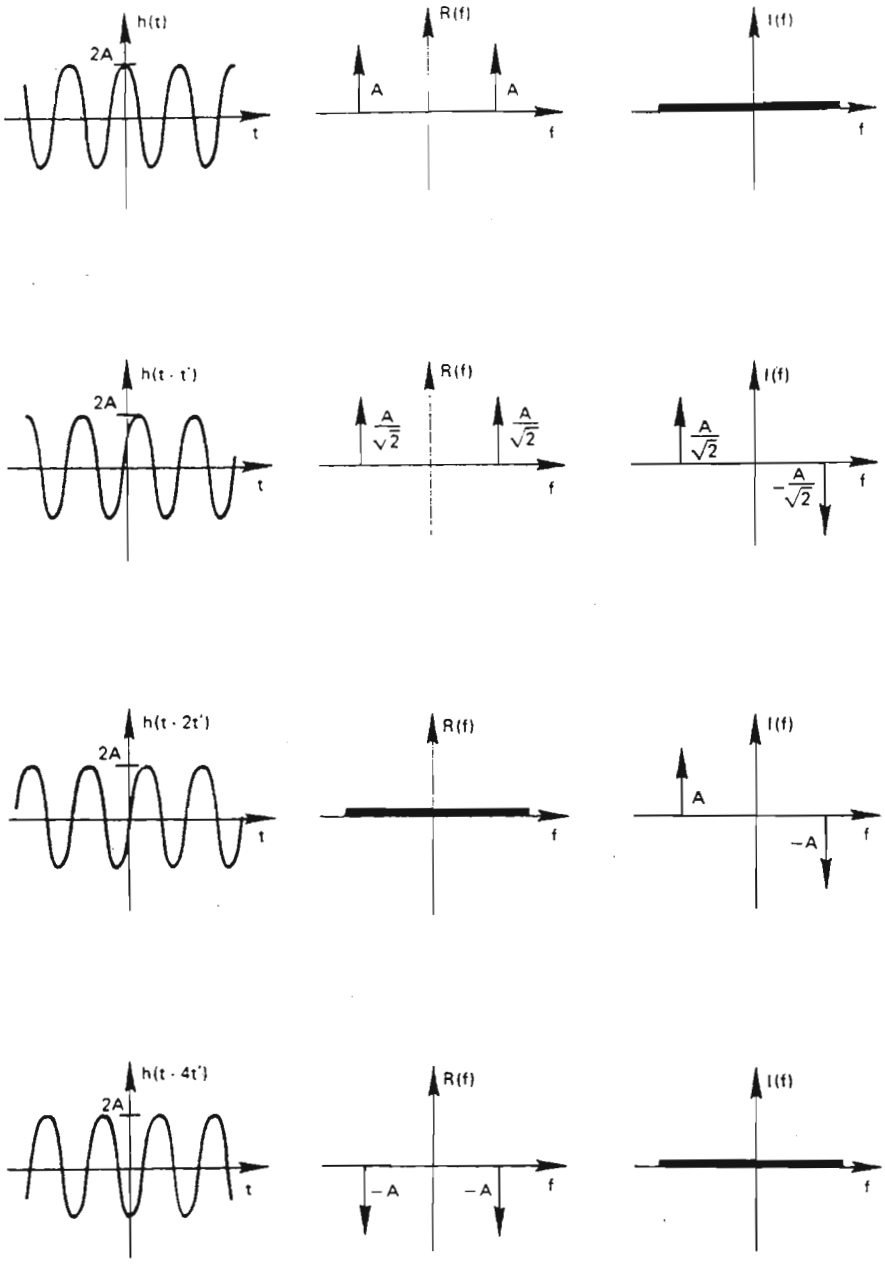

Figure 5.6: The time shifting property of the Fourier transform. From *Brigham (1974)*.

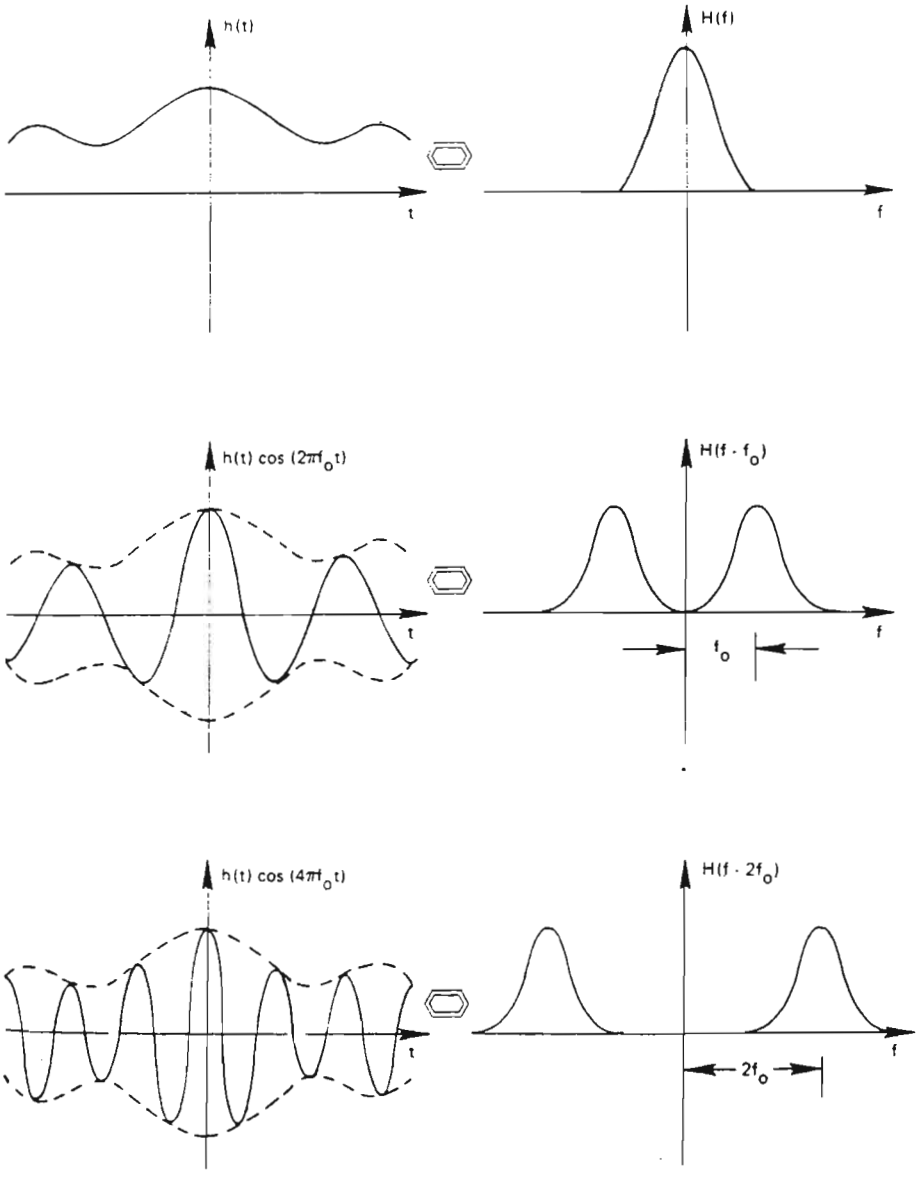

Figure 5.7: The frequency shifting property (modulation) of the Fourier transform. From *Brigham (1974)*.

and odd functions is, in general, neither even nor odd. Any function h(t) can be split unambiguously into odd and even parts

$$h(t) = e(t) + o(t) \tag{5.19}$$

where e(t) and o(t) are, in general, complex. The even part of any function is the mean of the function and its reflection in the vertical axis while the odd part is the mean of the function and its negative reflection i.e.

$$e(t) = \frac{1}{2}\left[h(t) + h(-t)\right] \tag{5.20}$$

$$o(t) = \frac{1}{2}\left[h(t) - h(-t)\right]. \tag{5.21}$$

Time shifting can alter the magnitude of odd and even components. Figure 5.6 illustrates how a purely even function, such as $2A\cos(2\pi ft)$, can be converted to a purely odd function, $2A\sin(2\pi ft)$, by a time shift of $-\frac{\pi}{2}$.

The Fourier integral, defined by equation (5.1), may be expressed

$$H(f) = \int_{-\infty}^{\infty} \left[e(t) + o(t)\right]\left[\cos(2\pi ft) - i\sin(2\pi ft)\right] dt \tag{5.22}$$

since

$$e^{-i2\pi ft} = \cos(2\pi ft) - i\sin(2\pi ft), \tag{5.23}$$

H(f) becomes

$$H(f) = 2\int_0^{\infty} e(t)\cos(2\pi ft)\, dt - 2i\int_0^{\infty} o(t)\sin(2\pi ft)\, dt. \tag{5.24}$$

Time functions of differing symmetry and reality are shown on the left side of Figure 5.8 and, on the right, the symmetry and reality of their corresponding Fourier transforms. Arguments from symmetry show that the integral of odd functions vanish and so, inspection of equation (5.24) reveals that if a time function is even, its transform is even and its reality does not change. Whereas, if it is odd, its transform is odd and its reality changes. Of particular importance to the work presented in this chapter, is that a real and asymmetrical function (penultimate pair in Figure 5.8) has a transform that is Hermitian i.e. the spectrum is composed of an even real part and an odd imaginary part.

5.2.4 Convolution

The convolution of two functions x(t) and h(t) is

$$x(t) * h(t) = \int_{-\infty}^{\infty} x(\tau)\, h(t-\tau)\, d\tau. \tag{5.25}$$

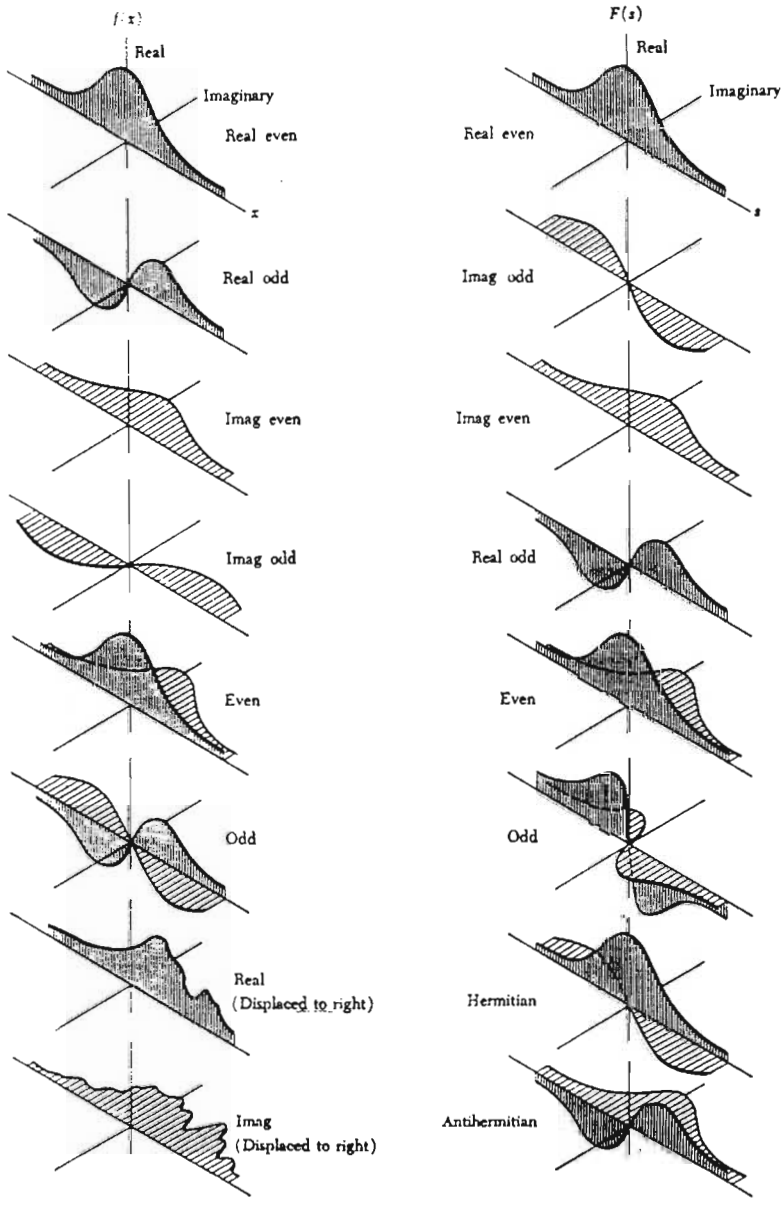

Figure 5.8: The symmetry properties of a function and its Fourier transform. From *Bracewell, 1978.*

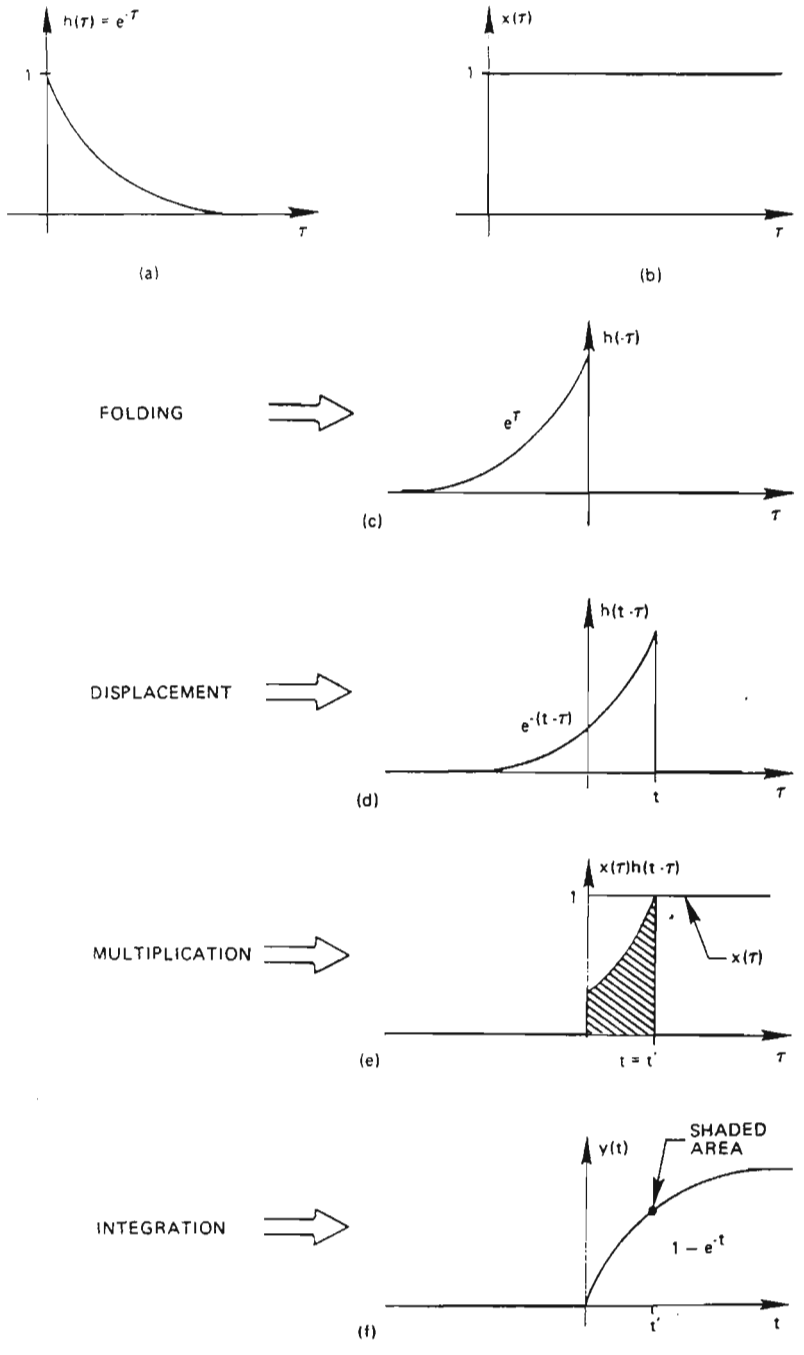

Figure 5.9: Graphical illustration of the convolution procedure (c) folding, (d) displacement, (e) multiplication and (f) integration. From *Brigham (1974)*.

This mathematical operation is extremely difficult to visualize, but is much clearer when illustrated graphically. Figure 5.9 summarizes the steps of this graphical technique. To evaluate equation (5.25), functions $x(\tau)$ and $h(t - \tau)$ are required. The functions $h(\tau) = e^{-\tau}$ in Figure 5.9(a) and $x(\tau) = 1$ in (b) are simply x(t) and h(t) respectively, where the variable t has been replaced by the variable τ. The first step is to fold, or take the mirror image, of $h(\tau)$ to obtain $h(-\tau)$ as illustrated in (c). Next, shift $h(-\tau)$ by the amount t as shown in (d). Then multiply the functions $h(t-\tau)$ and $x(\tau)$ shown in (e). Finally, integrate under the multiplied function. The value of this integration is represented by the dot in (f). To obtain the curve in (f) the integral must be evaluated for all values of t.

The simplest type of convolution integral to evaluate is one in which either x(t) or h(t) is an impulse function. To illustrate this point, let h(t) be an even impulse pair of unit magnitude at T and -T, depicted in Figure 5.10(a). Let $x(t) = A$ for $[a \geq t \geq 0]$ shown in (b) of this figure. By performing the technique outlined above, it is clear that x(t) is replicated each time it encounters a δ function and so, the δ function is often referred to as the replicating symbol.

Possibly one of the most important tools in modern scientific analysis, is the relationship between convolution and its Fourier transform. The relationship is stated in the convolution theorem which provides complete freedom to determine the convolution mathematically, or visually, in the time domain by simple multiplication of the corresponding spectra in the frequency domain. That is, if h(t) has the Fourier transform H(f) and x(t) the transform X(f) then

$$h(t) * x(t) \Leftrightarrow H(f)\, X(f). \tag{5.26}$$

Equivalently, the frequency convolution theorem states

$$h(t)\, x(t) \Leftrightarrow H(f) * X(f). \tag{5.27}$$

Some examples follow which illustrate the power and simplicity of these theorems.

Figure 5.11 illustrates how convolution of two top hat functions in the time domain, shown at the top of this figure, is equivalent to the multiplication the two sinc functions in the frequency domain. Therefore, the Fourier transform of the Chinese hat function can be easily determined without resorting to mathematical integration.

The frequency convolution theorem is illustrated in Figure 5.12. In order to determine the Fourier transform of a cosine function that is time limited between $-T_0$ and T_0, depicted in the middle graph of this figure, we note that this function is achieved by multiplying an infinite cosine function with an appropriate top hat function. Therefore, the transform must be equivalent to the convolution of the spectra of an infinite cosine function (an even impulse pair) with that of the top hat function (a sinc function).

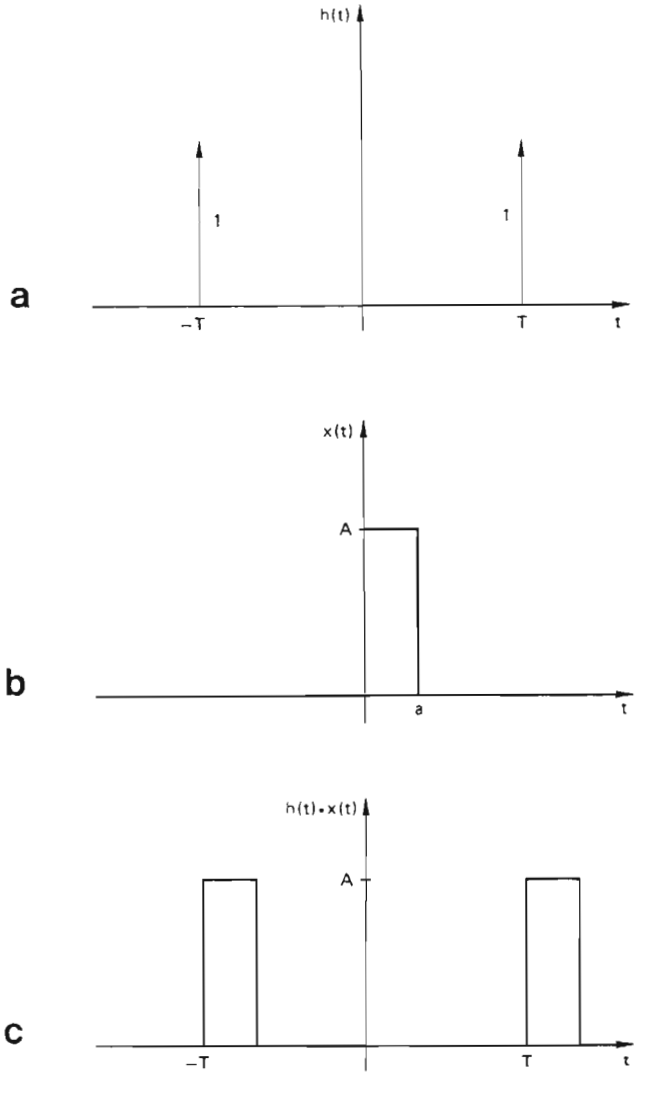

Figure 5.10: The special case of convolution involving impulse functions, in this case referred to as replicating symbols. From *Brigham (1974)*.

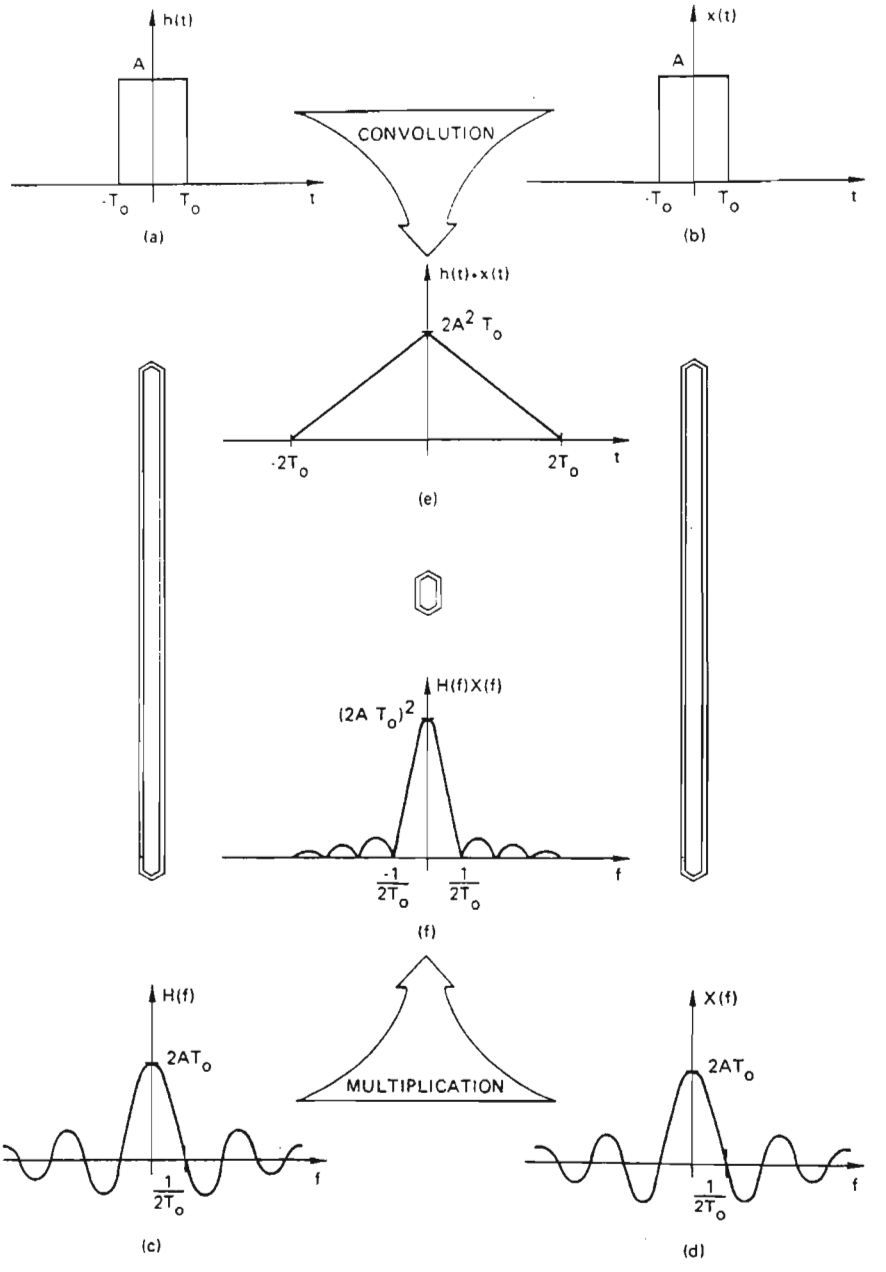

Figure 5.11: Graphical example of the convolution theorem. From *Brigham (1974)*.

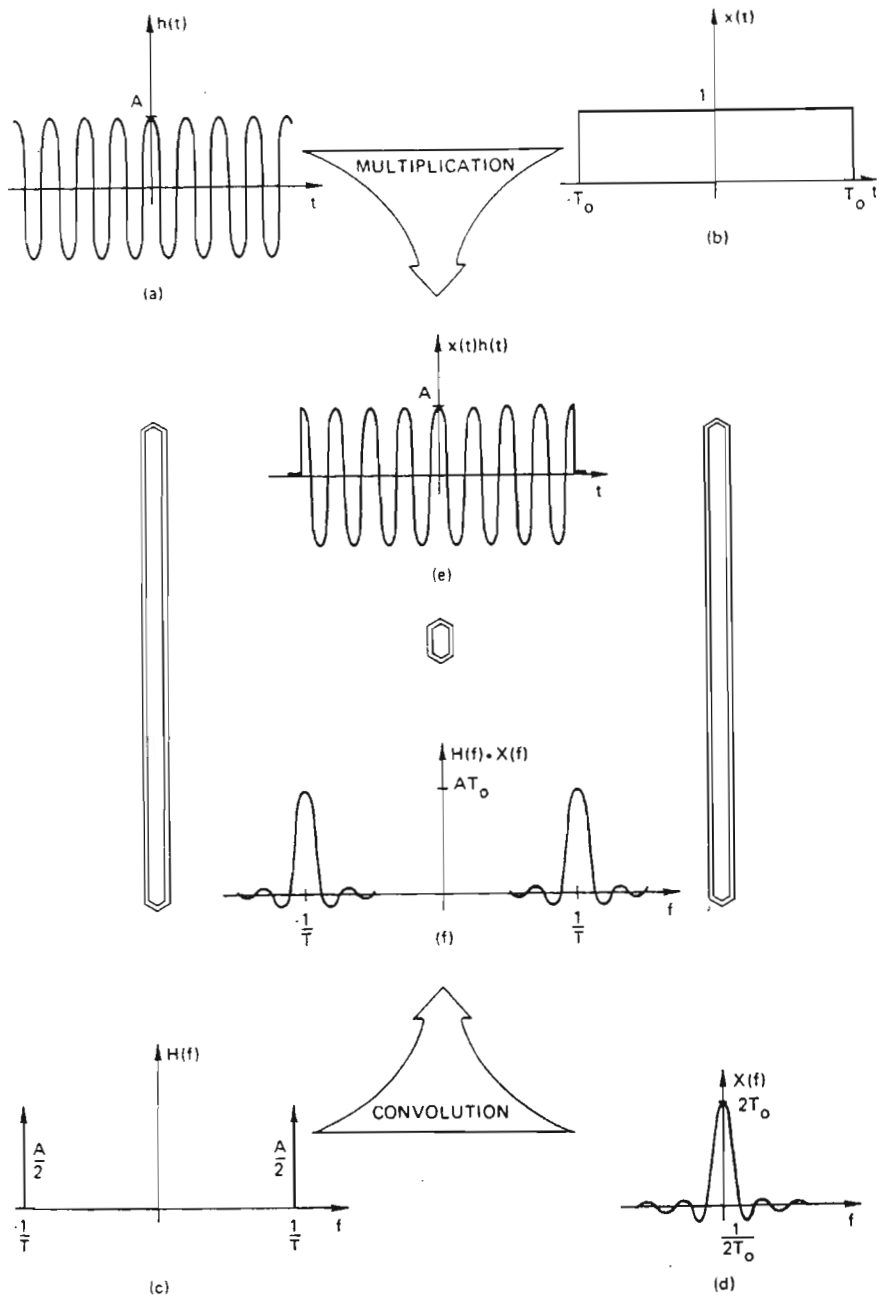

Figure 5.12: Graphical example of the frequency convolution theorem. From *Brigham (1974)*.

This result is plotted in the lower graph in the middle of Figure 5.12. This example illustrates the consequences of transforming a function that is finite in time, the result of which is the presence of high frequency ripples on the spectrum.

A special case of the frequency convolution theorem is modulation, a concept which has already been alluded to in the sub-section on frequency shifting. If h(t) has the Fourier transform H(f), then the multiplication of h(t) by a cosine function, is equivalent to the convolution of the spectrum H(f) with the appropriate impulse pair i.e.

$$h(t)\cos(2\pi f_0 t) \Leftrightarrow \frac{1}{2}H(f - f_0) + \frac{1}{2}H(f + f_0). \tag{5.28}$$

The modulation of h(t) on a higher carrier frequency f_0 is illustrated in the middle panel of Figure 5.7. The modulated spectrum is clearly the result of a convolution of H(f) with an impulse pair which is the Fourier transform of $\cos(2\pi f_0 t)$.

5.2.5 Correlation

The correlation of two functions x(t) and h(t) is defined by

$$x(t) \star h(t) = \int_{-\infty}^{\infty} x(\tau)\, h(t + \tau)\, d\tau. \tag{5.29}$$

A graphical comparison between correlation and convolution is given in Figure 5.13. The steps to be preformed to convolve two functions are shown on the left and to correlate on the right. The difference is that folding is not necessary when performing a correlation. In addition, the function $h(\tau)$ is displaced by -t in order to form the function $h(t+\tau)$ for correlation whereas $h(-\tau)$ is displaced by t to form the function $h(t-\tau)$ for convolution.

Once again, correlation in the time domain is related to multiplication in the frequency domain. The correlation theorem states that

$$h(t) \star x(t) \Leftrightarrow H(f)\, X^*(f), \tag{5.30}$$

where $X^*(f)$ is the complex conjugate of X(f).

5.2.6 The Discrete Fourier Transform (DFT)

In developing the background Fourier theory, functions that are infinite and continuous in both the time and frequency domains have been used. In practice, Fourier transforms are applied to finite data sets. In addition, the data are usually sampled or discrete.

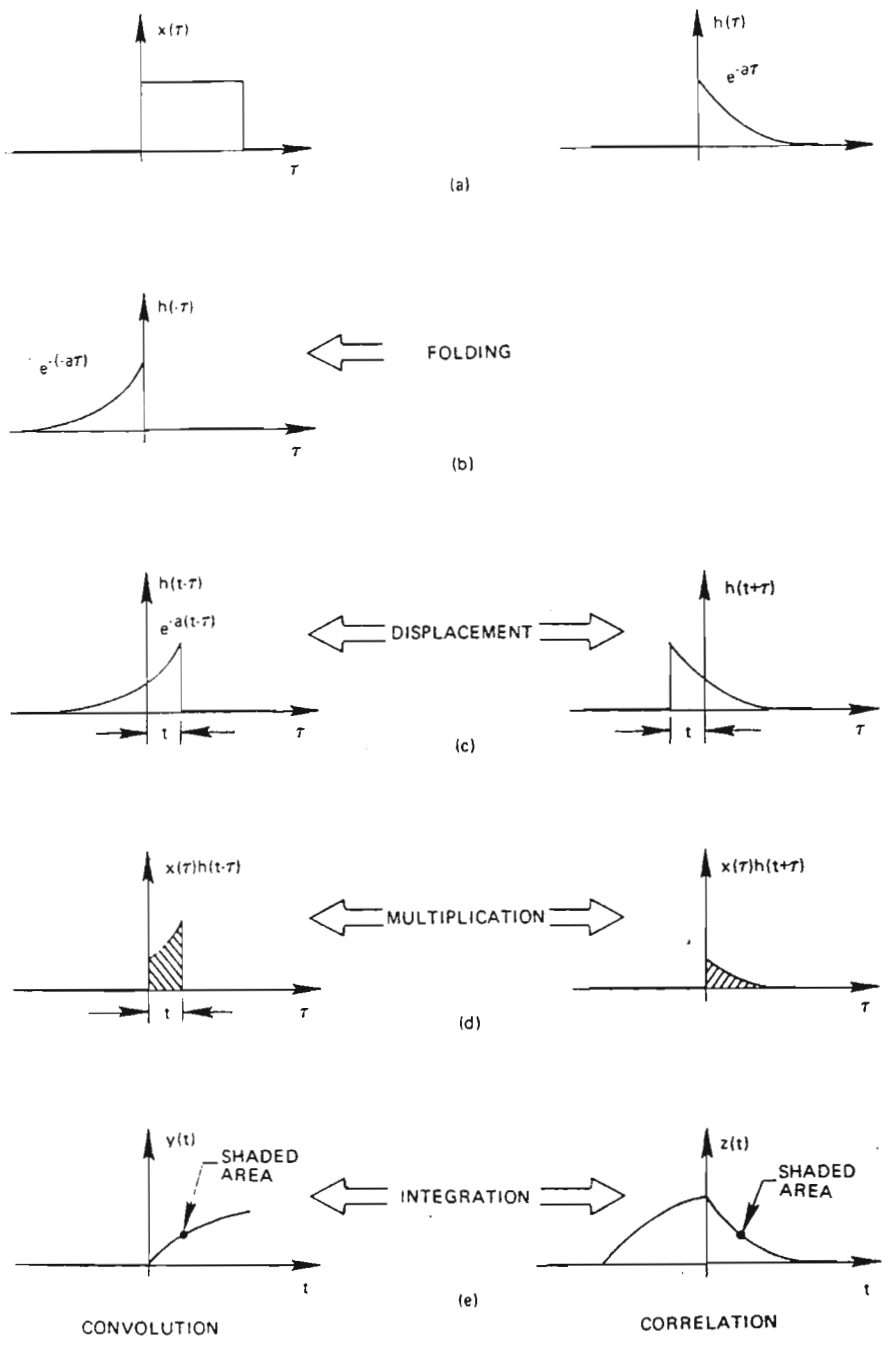

Figure 5.13: A graphical comparison of the convolution and correlation procedures. From *Brigham (1974)*.

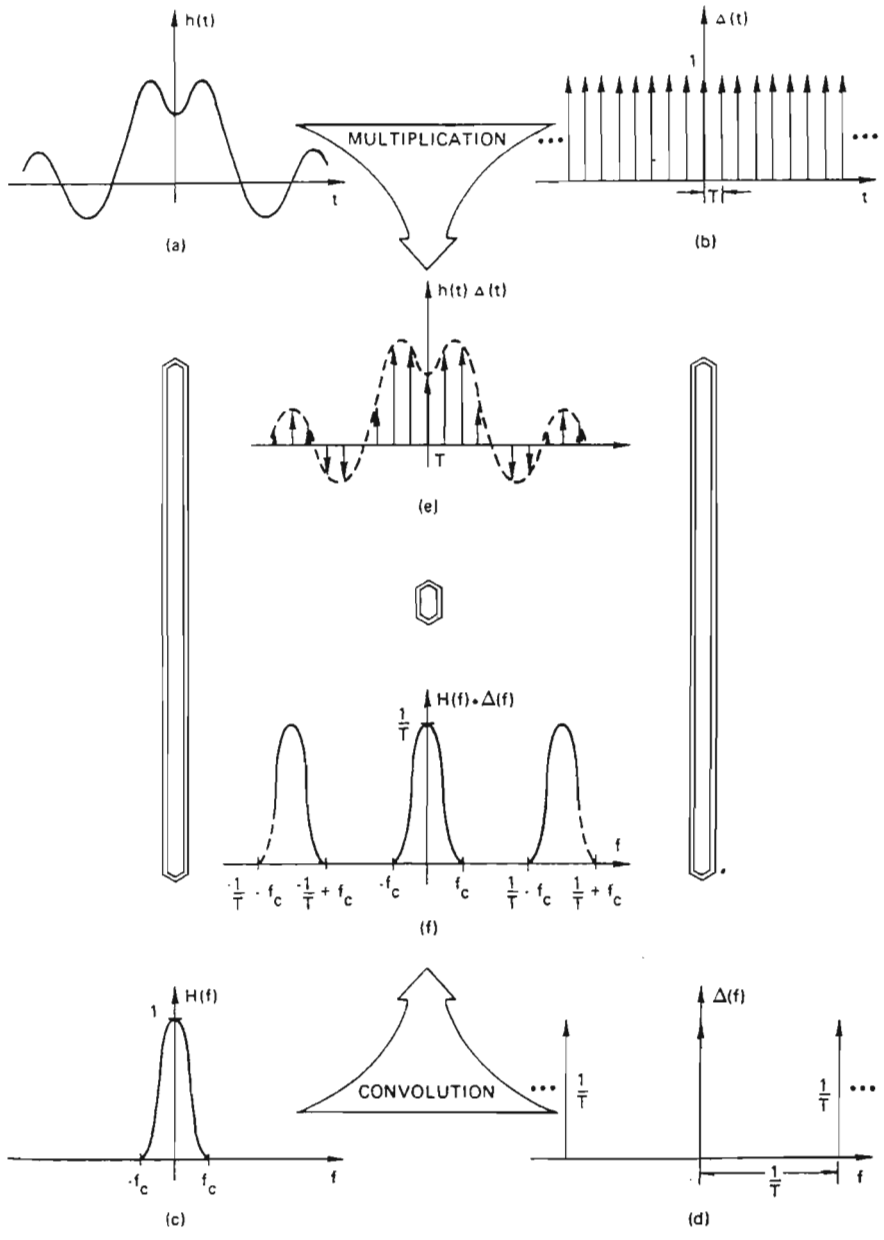

Figure 5.14: A graphical illustration of the application of the frequency convolution theorem when determining the Fourier transform of a sampled waveform. From *Brigham (1974)*.

How does a discrete transform relate to the continuous Fourier transform? To answer this question, the development of a DFT is graphically illustrated in Figure 5.14. A discrete, or sampled. function is the product of a continuous function h(t), in Figure 5.14(a). and the sequence of impulses, called the sampling function $\Delta(t)$, in Figure 5.14(b). The notation $\Delta(t)$ implies an infinite sequence of unit impulses separated by a sampling interval T. The Fourier transforms of h(t) and $\Delta(t)$ are H(f) and $\Delta(f)$ shown in Figures 5.14(c) and (d) respectively. In order to derive the Fourier transform of the sampled waveform we apply the frequency convolution theorem. The product $h(t)\Delta(t)$ is equivalent to the convolution of H(f) with $\Delta(f)$, illustrated in Figure 5.14(c). The reader's attention is drawn to the periodic, or repetitive, nature of $H(f)*\Delta(f)$, which is directly attributed to the sampling of h(t). A discrete waveform implies a periodic transform in which one period is equal, within a constant, to the Fourier transform of the continuous function h(t). However, this statement can only be valid if the sampling interval T is sufficiently small. The result of T being too large is graphically illustrated in Figure 5.15. A larger sampling interval T implies that the impulses in the frequency sampling spectrum, which are spaced at intervals of $\frac{1}{T}$, move closer together. This results in successive spectra, H(f), overlapping and allows low frequencies to masquerade as high frequencies. This distortion of the desired Fourier transform is called aliasing. Careful inspection of Figures 5.15(c) and (d) reveals that convolution overlap will not occur if the separation of the impulses in $\Delta(f)$ is increased so that

$$\frac{1}{T} = 2f_c, \tag{5.31}$$

where f_c is the highest frequency component of H(f). The sampling theorem states that 'a function whose Fourier transform is zero for $|H(f)| > f_c$ is fully specified by values spaced at equal intervals not exceeding $\frac{1}{2f_c}$ save for any harmonic term with zeros at the sampling points' (*Bracewell, 1978*). The minimum sampling frequency, that prevents aliasing, is also known as the Nyquist sampling frequency.

The graphical development of a Fourier transform of a function that is both discrete and finite, is illustrated in Figure 5.16. Once again, the continuous function h(t), Figure 5.16(a), is sampled by multiplication with a row of impulses denoted by $\Delta_0(t)$, in Figure 5.16(b). This process is equivalent to the convolution of H(f) with $\Delta_0(f)$ shown on the right side of Figure 5.16(c). The reader's attention is drawn to the aliasing that is apparent here, as is the case in most applications. In order to make h(t) finite, it is multiplied with a top hat function, of width T_0. This process is imitated in the frequency domain by the convolution of a sinc function, having central lobe of width $\frac{2}{T_0}$, with the spectrum $H(f)*\Delta_0(f)$. This process induces a ripple on the spectrum, depicted in Figure 5.16(e), due to the effect of the high frequency side lobes in the sinc function. This effect, called leakage, is inherent in all Fourier transforms of truncated functions. To reduce this effect, the width of $\Pi(t)$, shown in Figure 5.16(d), is expanded which leads to a corresponding contraction of its sinc function. In the limit, the sinc

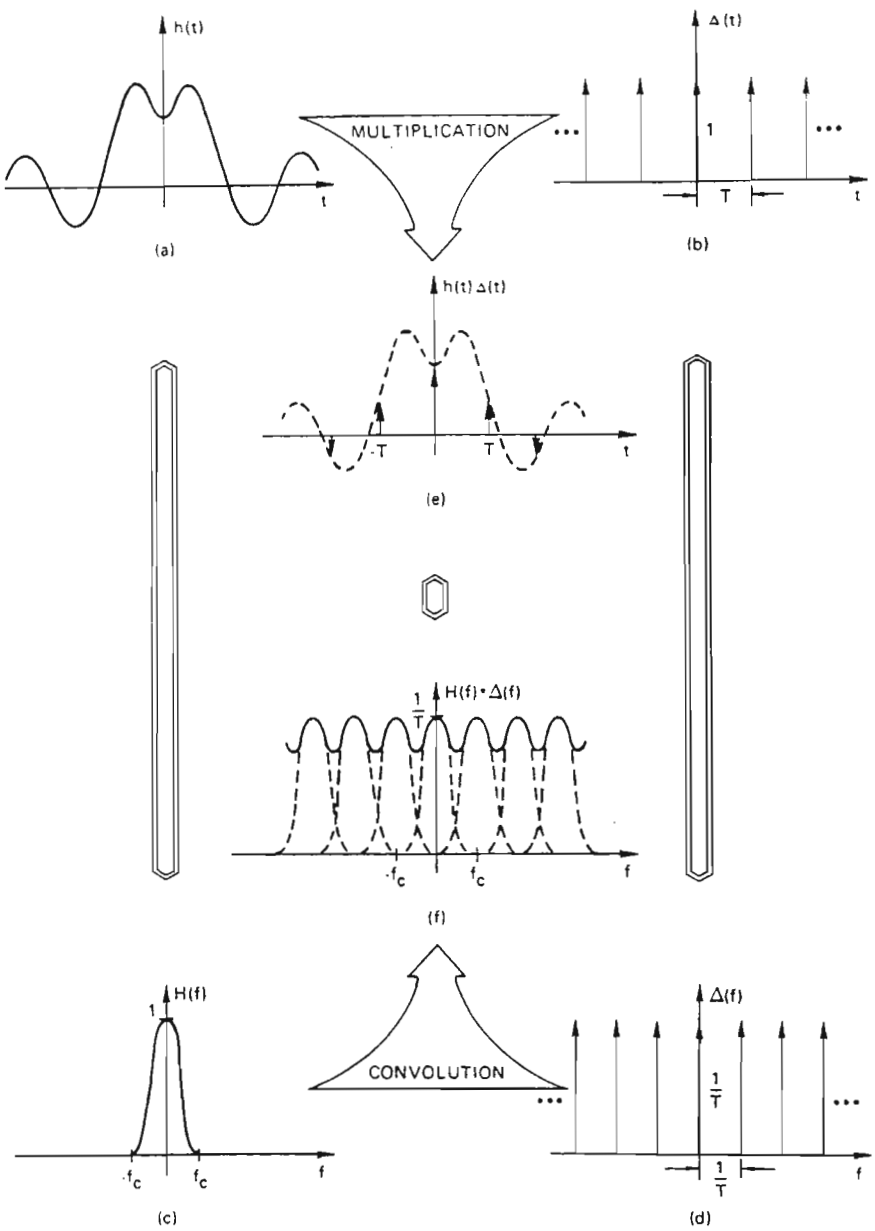

Figure 5.15: The aliased Fourier transform that results from insufficient sampling of a waveform. From *Brigham (1974)*.

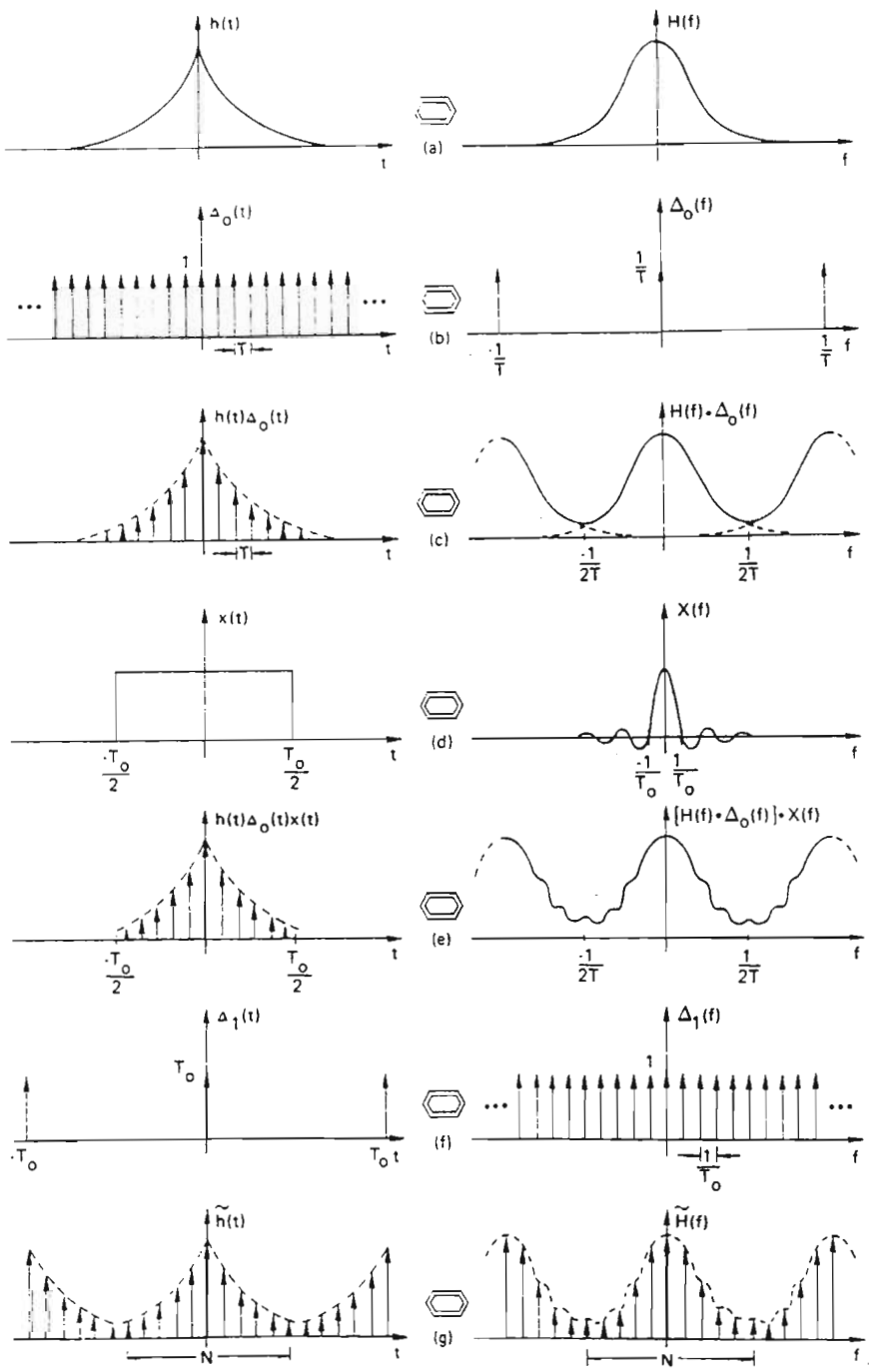

Figure 5.16: Graphical development of the discrete Fourier transform (DFT). From *Brigham (1974)*.

function will approximate an impulse function, with minimal side lobes resulting in nominal ripple, or error, on the spectrum of h(t). In applications, it is therefore desirable to choose the widest truncation function possible. The modified transform pair depicted in Figure 5.16(e) is still not acceptable since the frequency transform is continuous. In computer applications, only sample values of the frequency transform can be computed. Modification of the frequency transform is therefore necessary and is achieved by multiplying it with the frequency sampling function $\Delta_1(f)$ which has a frequency sampling interval of $\frac{1}{T_0}$. The discrete frequency domain implies a periodic time domain. The original time function $\tilde{h}(t)$ and the original Fourier transform $\tilde{H}(f)$ are both approximated by N samples, as illustrated in Figure 5.16(g). These N samples define the discrete Fourier transform pair.

The mathematical version of the DFT can be intuitively defined from the Fourier integration applied to a continuous function. For example, the $\int$ sign in equation (5.1) can exchanged for $\sum$ sign. In addition, the continuous function h(t) in equation (5.1) can be replaced by a series of points $x_0(k)$. The Discrete Fourier Transform (DFT) is then be defined as (*Brigham, 1974*)

$$X(n) = \sum_{k=0}^{N-1} x_0(k)\, W^{nk}, \tag{5.32}$$

where

$$n = 0, 1, 2, ...N - 1$$

and W^{nk} is the complex number

$$W = e^{-\frac{i2\pi}{N}}.$$

The mathematical derivation of these equations is given in *Brigham (1974)*.

In general, $x_0(k)$ may be a complex number i.e.

$$x_0(k) = r_0(k) + ii_0(k).$$

However, the Fourier analysis performed in this thesis is applied to a function that is purely real. There is, of course, an associated reverse DFT, not shown here as it is obsolete for the 'one way' application in the work outlined later in this chapter.

5.2.7 The Fast Fourier Transform (FFT)

The number of calculations in a DFT govern the amount of time required by a computer to determine a Fourier transform. Since a DFT applied to a data set of reasonable length requires many, many computations it is clear that an impracticable amount of

computer time would be required. For years this was the stumbling block for the numerical calculation of transforms. A more efficient method was required, motivating *Cooley and Tukey (1965)* to develop the Fast Fourier Transform. The procedure employed that can reduce the number of calculations, is outlined below by an example for which the number of samples (N) is 4. A rigorous proof of the FFT algorithm can be found on page 176 of *Brigham (1974)*.

The DFT equation (5.32) may be written in the alternate form

$$X(0) = x_0(0)W^0 + x_0(1)W^0 + x_0(2)W^0 + x_0(3)W^0 \tag{5.33}$$

$$X(1) = x_0(0)W^0 + x_0(1)W^1 + x_0(2)W^2 + x_0(3)W^3 \tag{5.34}$$

$$X(2) = x_0(0)W^0 + x_0(1)W^2 + x_0(2)W^4 + x_0(3)W^6 \tag{5.35}$$

$$X(3) = x_0(0)W^0 + x_0(1)W^3 + x_0(2)W^6 + x_0(3)W^9. \tag{5.36}$$

This set of equations may expressed as a set of matrices

$$\begin{bmatrix} X(0) \\ X(1) \\ X(2) \\ X(3) \end{bmatrix} = \begin{bmatrix} W^0 & W^0 & W^0 & W^0 \\ W^0 & W^1 & W^2 & W^3 \\ W^0 & W^2 & W^4 & W^6 \\ W^0 & W^3 & W^6 & W^9 \end{bmatrix} \begin{bmatrix} x_0(0) \\ x_0(1) \\ x_0(2) \\ x_0(3) \end{bmatrix} \tag{5.37}$$

or, by a compact matrix notation

$$\mathbf{X(n)} = \mathbf{W^{nk}} \cdot \mathbf{x_0(k)}. \tag{5.38}$$

Examination of equation (5.37) reveals that since W, and in some instances $x_0(k)$, are complex, N^2 complex multiplications followed by N(N-1) complex additions are required to calculate the transform. Since the number multiplications (N^2), are largely responsible for the amount of computation time required, a DFT will, obviously, require an enormous amount of computer time.

The first step, leading to significant reduction in the number of multiplications, is the choice of the number (N) of sample points $x_0(k)$ according to the relation

$$N = 2^{\gamma},$$

where γ is a positive integer. This DFT example has $\gamma = 2$, and so $N = 4$. Equation (5.37) can be re-written as

$$\begin{bmatrix} X(0) \\ X(1) \\ X(2) \\ X(3) \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & W^1 & W^2 & W^3 \\ 1 & W^2 & W^0 & W^2 \\ 1 & W^3 & W^2 & W^1 \end{bmatrix} \begin{bmatrix} x_0(0) \\ x_0(1) \\ x_0(2) \\ x_0(3) \end{bmatrix} \tag{5.39}$$

since

$$W^{nk} = W^{nk \bmod (N)}. \tag{5.40}$$

where [nk mod (N)] is the remainder upon division of nk by N. W^0 is not reduced to unity in order to develop a generalized result. The matrix in equation (5.39) is factorized so that

$$\begin{bmatrix} X(0) \\ X(2) \\ X(1) \\ X(3) \end{bmatrix} = \begin{bmatrix} 1 & W^0 & 0 & 0 \\ 1 & W^2 & 0 & 0 \\ 0 & 0 & 1 & W^1 \\ 0 & 0 & 1 & W^3 \end{bmatrix} \begin{bmatrix} 1 & 0 & W^0 & 0 \\ 0 & 1 & 0 & W^0 \\ 1 & 0 & W^2 & 0 \\ 0 & 1 & 0 & W^2 \end{bmatrix} \begin{bmatrix} x_0(0) \\ x_0(1) \\ x_0(2) \\ x_0(3) \end{bmatrix}. \tag{5.41}$$

Each row in a matrix that contains W elements, has only two non-zero elements. The reader's attention is drawn to the interchanged rows in $\mathbf{X(n)}$, apparent in the leftmost matrix of equation (5.41). This 'scrambled' vector may be attributed to the factorization process and is denoted by

$$\overline{\mathbf{X(n)}} = \begin{bmatrix} X(0) \\ X(2) \\ X(1) \\ X(3) \end{bmatrix}. \tag{5.42}$$

The calculation of $\overline{\mathbf{X(n)}}$ by equation (5.41) requires substantially less calculations. This is illustrated below. By application of the rightmost W matrix in equation (5.41) to $\mathbf{x_0(k)}$ an intermediate matrix $\mathbf{x_1(k)}$ is formed

$$\begin{bmatrix} x_1(0) \\ x_1(1) \\ x_1(2) \\ x_1(3) \end{bmatrix} = \begin{bmatrix} 1 & 0 & W^0 & 0 \\ 0 & 1 & 0 & W^0 \\ 1 & 0 & W^2 & 0 \\ 0 & 1 & 0 & W^2 \end{bmatrix} \begin{bmatrix} x_0(0) \\ x_0(1) \\ x_0(2) \\ x_0(3) \end{bmatrix}. \tag{5.43}$$

Element $x_1(0)$ is computed by one complex multiplication and one complex addition

$$x_1(0) = x_0(0) + W^0 x_0(2). \tag{5.44}$$

Since $W^0 = -W^2$, element $x_1(1)$ can likewise be determined by one complex multiplication and one complex addition

$$x_1(2) = x_0(0) + W^2 x_0(2) \tag{5.45}$$

$$x_1(2) = x_0(0) - W^0 x_0(2). \tag{5.46}$$

The complex multiplication has already been preformed in determining $x_1(0)$ above and so only one complex addition is required to compute $x_1(2)$. By the same reasoning, $x_1(3)$ can also be computed from only one complex addition. Therefore, the intermediate

vector $\mathbf{x_1(k)}$ is determined by four complex additions and two complex multiplications. The leftmost W matrix in equation (5.41) is applied to $\mathbf{x_1(k)}$

$$\begin{bmatrix} X(0) \\ X(2) \\ X(1) \\ X(3) \end{bmatrix} = \begin{bmatrix} x_2(0) \\ x_2(1) \\ x_2(2) \\ x_2(3) \end{bmatrix} = \begin{bmatrix} 1 & W^0 & 0 & 0 \\ 1 & W^2 & 0 & 0 \\ 0 & 0 & 1 & W^1 \\ 0 & 0 & 1 & W^3 \end{bmatrix} \begin{bmatrix} x_1(0) \\ x_1(1) \\ x_1(2) \\ x_1(3) \end{bmatrix}. \tag{5.47}$$

The term $x_2(0)$ is determined by one complex multiplication and one complex addition viz.

$$x_2(0) = x_1(0) + W^0 x_1(1). \tag{5.48}$$

The determination of $x_2(2)$ also requires one complex multiplication and one complex addition. Element $x_2(1)$ is computed by one addition since $W^0 = -W^2$ and, by similar reasoning, $x_2(3)$ can be determined by one complex addition.

Therefore, computation of $\overline{\mathbf{X(n)}}$ with N=4 by means of equation (5.41) requires only four complex multiplications and eight complex additions. Whereas, computation of $\mathbf{X(n)}$ by equation (5.39) requires sixteen complex multiplications and twelve complex additions. The matrix factorization process introduces zeros into the factored matrices and, as a result, reduces the required number of operations.

In summary, by letting the number of samples $N = 2^\gamma$ the FFT algorithm can then factorize the N × N matrix into γ matrices, each N × N, such that the factorized matrices have the property of reducing the amount of mathematical operations to be preformed. The FFT computes $\frac{N\gamma}{2}$ complex multiplications and N γ complex additions, whereas the unfactorized method requires N^2 complex multiplications and N(N-1) complex additions. If we assume that computing time is proportional to the number of multiplications, then the approximate ratio of direct to FFT computing time is

$$\frac{N^2}{\frac{N\gamma}{2}} = \frac{2N}{\gamma}. \tag{5.49}$$

The comparison of the number of calculations, and so the computing time, as a function of the number of sample points is graphed in Figure 5.17. This graph illustrates the efficiency of the FFT algorithm. It is particularly effective for a large numbers of sample points. For the case $N = 10^6$ samples, an FFT can be preformed in roughly 30 seconds of CPU time on a 1 MHz computer whereas a DFT would take approximately two weeks!

A more concise way of representing the mathematical operations of a FFT, is the signal flow graph, illustrated in Figure 5.18 for N = 4 samples. The data vector $\mathbf{x_0(k)}$ is represented by the vertical column of nodes on the left side in Figure 5.18. Each computational array corresponds to a factorized matrix so that γ vertical arrays of

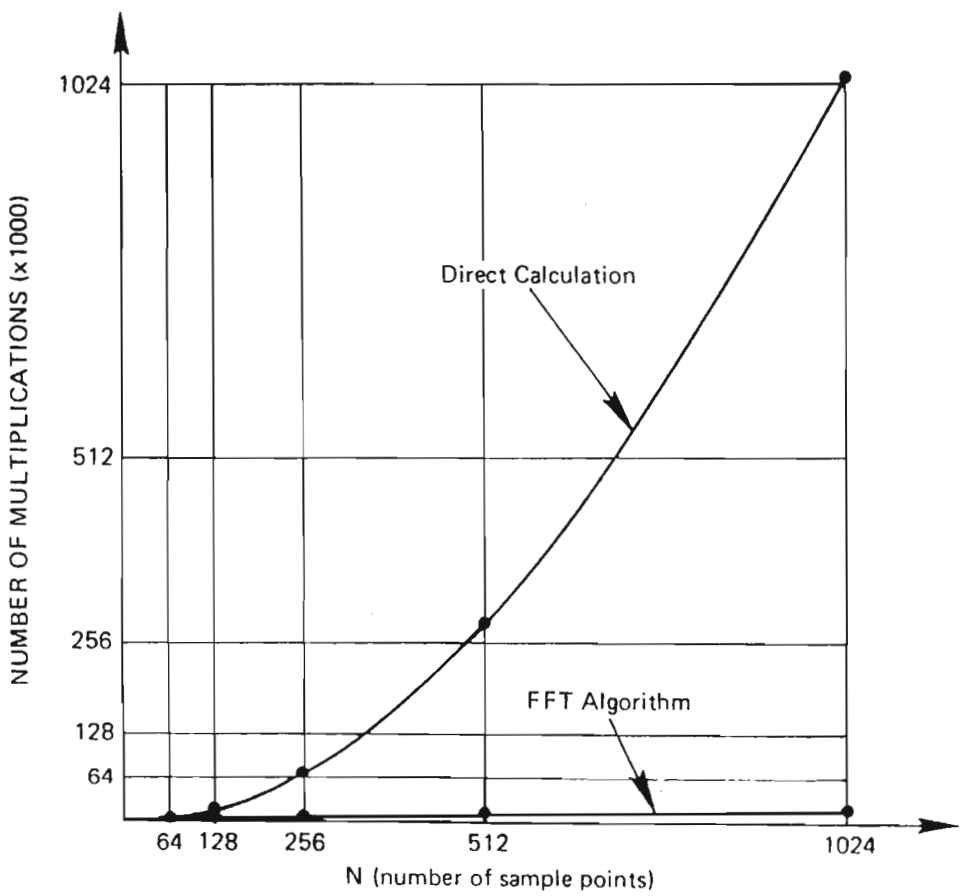

Figure 5.17: Comparison of the multiplications required by direct calculation and the FFT algorithm. From *Brigham (1974)*.

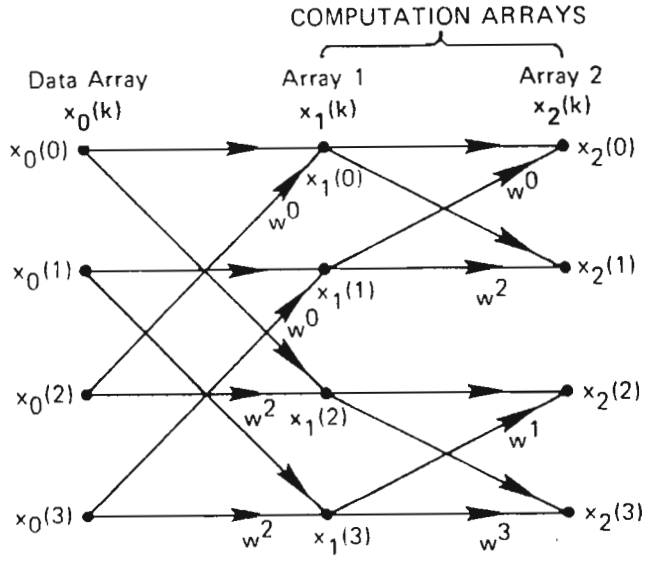

Figure 5.18: FFT signal flow diagram with 4 data points. From *Brigham (1974)*.

N points are required. The middle vertical array is the vector $\mathbf{x}_1(\mathbf{k})$, computed in equation (5.43). The rightmost computational array corresponds to the vector $\mathbf{x}_2(\mathbf{k}) = \overline{\mathbf{X}(\mathbf{n})}$. The signal flow graph may be interpreted in the following manner. Each node is entered by two solid lines, or transmission paths, from previous nodes. The path transmits or carries a quantity from a node in one array, multiples the quantity by W^p, and inputs the result into the node of the next array. The factor W^p appears near the arrowhead of the transmission path, its absence implies p = 0. Results entering a node from two transmission paths are added together. To illustrate this interpretation, consider the node $x_1(1)$ in the middle computation array of Figure 5.18. The rules outlined above imply that

$$x_1(1) = x_0(1) + W^0 x_0(3).$$

The necessity of these flow graphs, which allow clear illustration of the computational process, becomes obvious when considering large numbers of sampling points. Figure 5.19 is the signal flow graph for N = 16. Inspection of Figure 5.19 reveals that in each array there are always two nodes whose input transmission paths stem from the same pair of nodes in the previous array, i.e. the array to the left. Two such nodes are referred to as a dual node pair, some examples of which are labelled in Figure 5.19. Since the determination of any dual node pair is independent of all the other nodes, it is possible to perform *in-place* computation. For example, in Figure 5.19, $x_1(0)$ and $x_1(8)$ can be simultaneously computed from $x_0(0)$ and $x_0(8)$ and the results can be returned to the storage locations previously occupied by $x_0(0)$ and $x_0(8)$. Computer memory requirements can therefore be limited to the storage space of the input arrays.

The form of the Fourier algorithm discussed here implies that the input data is in sequential order, while the output data, which is written over the input data, in not ordered. To *unscramble* the output data the technique of bit reversal is applied. The output number n is converted to a binary number and the order of the bits (1's and 0's) of the binary number is then reversed. This is illustrated in Figure 5.20 where N, number of samples, is 16. The thirteen output, 1101, in $\overline{\mathbf{X}(\mathbf{n})}$, left hand column is converted to 1011 which is the eleventh output of $\mathbf{X}(\mathbf{n})$. By this method the output is sequentially ordered. It is completely legitimate to rearrange the signal flow chart in Figure 5.20 so that the input data is scrambled, or bit reversed, resulting in an output that is sequential.

Although the Cooley-Tukey form of the FFT has been outlined in this thesis, there exist many variations of the FFT algorithm which are, in a sense, canonic. Each particular algorithm variation is formulated to exploit a particular property of the data being transformed and the computer being used to perform the computation. Most of these variants are based on the Cooley-Tukey or the Sande-Tukey algorithms. The FFT algorithm used in this analysis of long term variations in total column ozone can be found on page 394 of *Press et al. (1987)*.

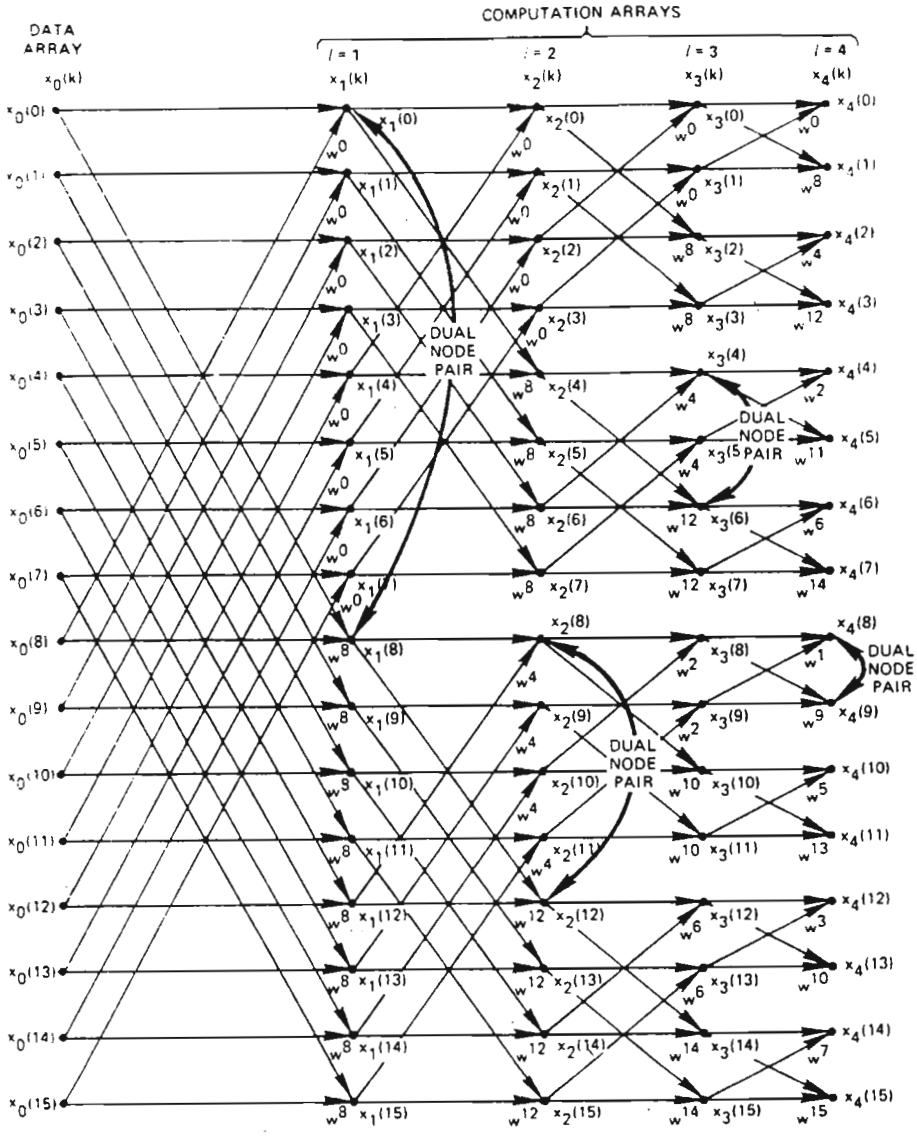

Figure 5.19: FFT signal flow chart for 16 samples. Dual nodes are indicated. From *Brigham (1974)*.

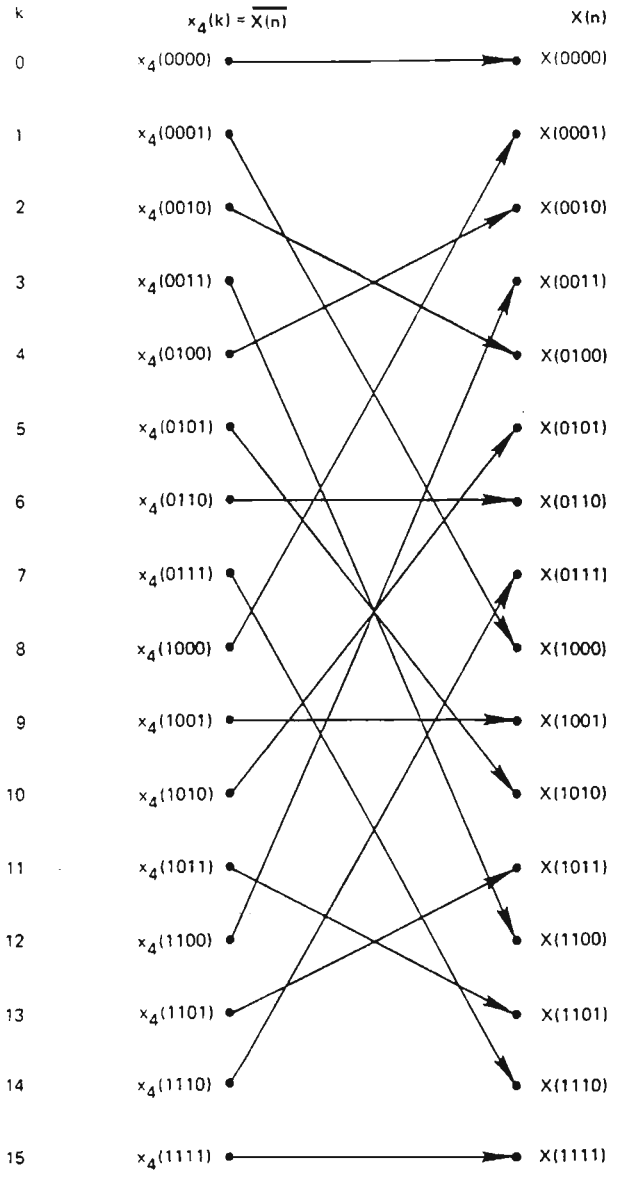

Figure 5.20: Illustration of the *bit reversal* technique for 16 data points. From *Brigham (1974)*.

Fourier analysis is a very powerful tool however, unless careful attention is paid to correct data preparation and interpretation of the resulting spectrum, erroneous results are obtained. For this reason, we have outlined the necessary Fourier theory that enables the reader to gain an appreciation of the many pitfalls that must be carefully negotiated.

5.3 Data preparation and subsequent Fourier analysis

Daily total column ozone mass values, of the region 90-70°S, were calculated from the TOMS total column ozone data for each day in the eleven year interval 1981 to 1991. The technique employed to determine ozone mass values is described in section 4.2. A measurement of the ozone mass over the polar cap every day provides a means by which influences on an extensive area of ozone can be assessed. The inability of the TOMS instrument to monitor ozone poleward of the day-night terminator, due to lack of irradiance, results in ozone data gaps over the polar regions during the winter season. These annual data gaps were deemed to have serious consequences on the spectrum resulting from a FFT of the data and so, the missing data were estimated. Two methods of data approximation were employed. If the terminator was within the region of ozone mass determination ($\geq$70°S) then, the zonally averaged ozone values poleward of the day-night terminator were assumed to be equal to the average value of those nearest to the terminator. This technique has been applied by *Schoeberl et al. (1989)* to estimate polar ozone data during the dark winter period. When no data were available in the latitude zone 90-70°, the data gap in each year was estimated by a high order polynomial fit. A least squares regression algorithm from a software graphical analysis package, *Graftool*, was used to provide a best curve fit. Figure 5.21 illustrates the application of this curve fitting algorithm to the 1987 ozone mass data. The theoretically determined polynomial, used to estimate the missing data between day numbers 131 and 213, is without micro-structure that corresponds to short term variations of polar ozone mass. The approximated ozone mass values compare favourably with those derived from ground-based measurements within the polar cap. For example, ozonesonde measurements during winter at Halley, Antarctica (76°S;27°W) (*Gardiner and Farman, 1988*) translate to polar ozone masses that are never more than 5% higher then the estimated values. Of course, the assumption here is that the Halley total ozone data is representative of the entire polar cap.

Eleven years of daily data corresponds to

$$(9\,\text{years} \times 365\,\text{days}) + (2\,\text{years} \times 366\,\text{days}) = 4017\,\text{data records}. \tag{5.50}$$

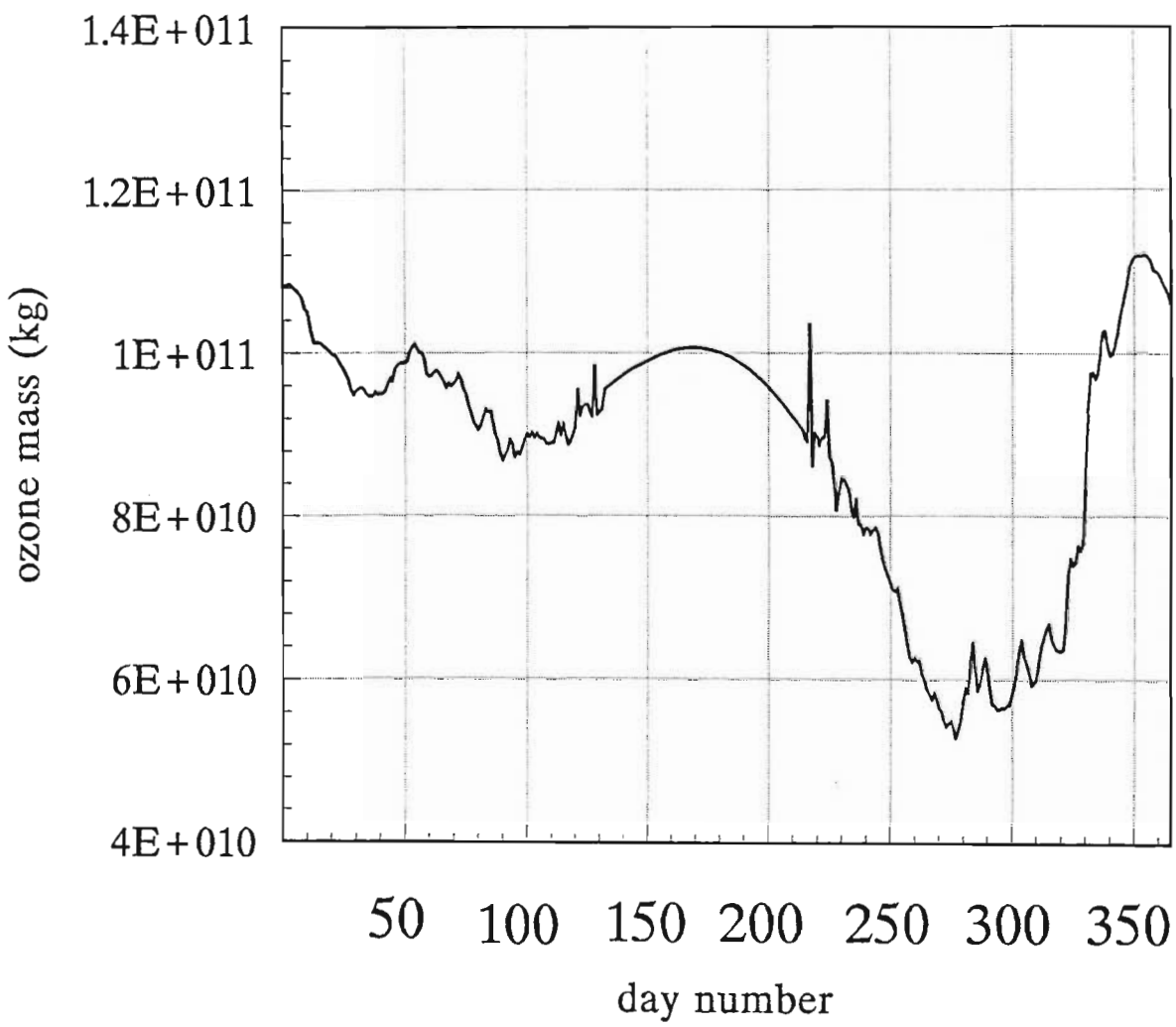

Figure 5.21: An illustration of the application of the curve fitting algorithm in order to estimate the ozone mass (kg) over Antarctica during the polar night period in 1987.

The necessary condition for the application of a FFT is that the data series must have N data records, or points, such that

$$2^{\gamma} = \mathrm{N}. \tag{5.51}$$

Therefore, the data sequence must be padded with zero values to nearest value of N that satisfies the equation above, which in this case is

$$2^{12} = 4096. \tag{5.52}$$

Therefore, only 79 zero values must be appended to the data sequence. Zero padding is equivalent to multiplying the sampled ozone mass function by a top hat function. In this case, very little zero padding was necessary, implying a very wide top hat function. Figure 5.4 illustrates that by expanding the width of the top hat function, the sinc transform contracts, and so approximates a δ function with nominal high frequency side lobes. Consequently, an almost 'ripple-free' spectrum is obtained. Although thirteen years of TOMS data were available, which corresponds to 4748 data records, this would require 3444 zero padding records to achieve the nearest power of two i.e.

$$2^{13} = 8192. \tag{5.53}$$

The resultant narrow top hat function would represent considerable ripple on the spectrum. **For this reason eleven, and not thirteen, years of data were chosen as the interval on which to perform a FFT.**

Ozone mass values, tailored to FFT requirements, are plotted in Figure 5.22. Each year of ozone mass data is contained within one grid width. Data for 1981 is plotted on the left side of the graph, followed by subsequent years, with 1991 data plotted on the extreme right side. On the extreme right, values drop off the scale, indicating the amount of zero padding. Inspection of Figure 5.22 reveals some defined periodicities. In addition, the seasonal dip to extremely low ozone mass values, representing the chlorine-catalyzed Antarctic ozone hole, is clearly visible throughout this eleven year period.

A time series of daily readings of solar radio flux (10.7 cm) at Ottawa, over the same interval from 1981 to 1991, was also prepared. Solar radio flux values were read from a compact disc, prepared by NOAA, on which many solar parameters over this extended period are recorded. There were no missing solar radio flux data in this eleven year period. A detailed discussion of the format of these data appears in section 3.5. Once again, 4017 data records were padded with zero values to 4096 data points. The time series of these data is shown in Figure 5.23. Conventionally, 90 $\times 10^{-22}\,\mathrm{W\,m^{-2}\,Hz^{-1}}$ defines a quiet sun characteristic of the solar minimum years 1985, 1986 and 1987, a value of 150 $\times 10^{-22}$ corresponds to moderate activity and values higher than 220 $\times 10^{-22}$ represent active conditions. The solar cycle signature is obvious in Figure

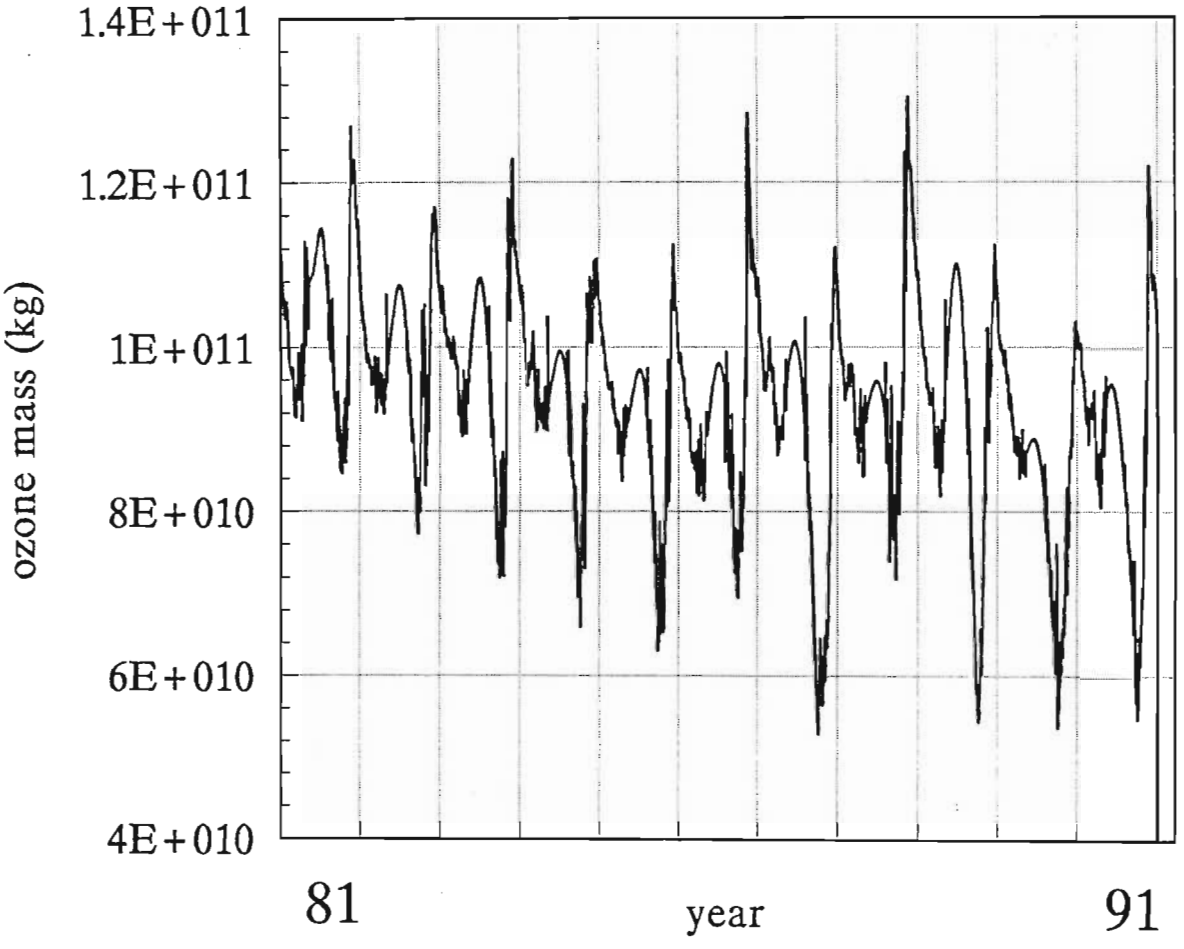

Figure 5.22: The diurnal polar ozone mass, calculated from TOMS total ozone measurements over 90°S to 70°S latitude, for the eleven year period 1981 to 1991, inclusively. Data gaps over the winter pole are evident as 'smooth' curves.

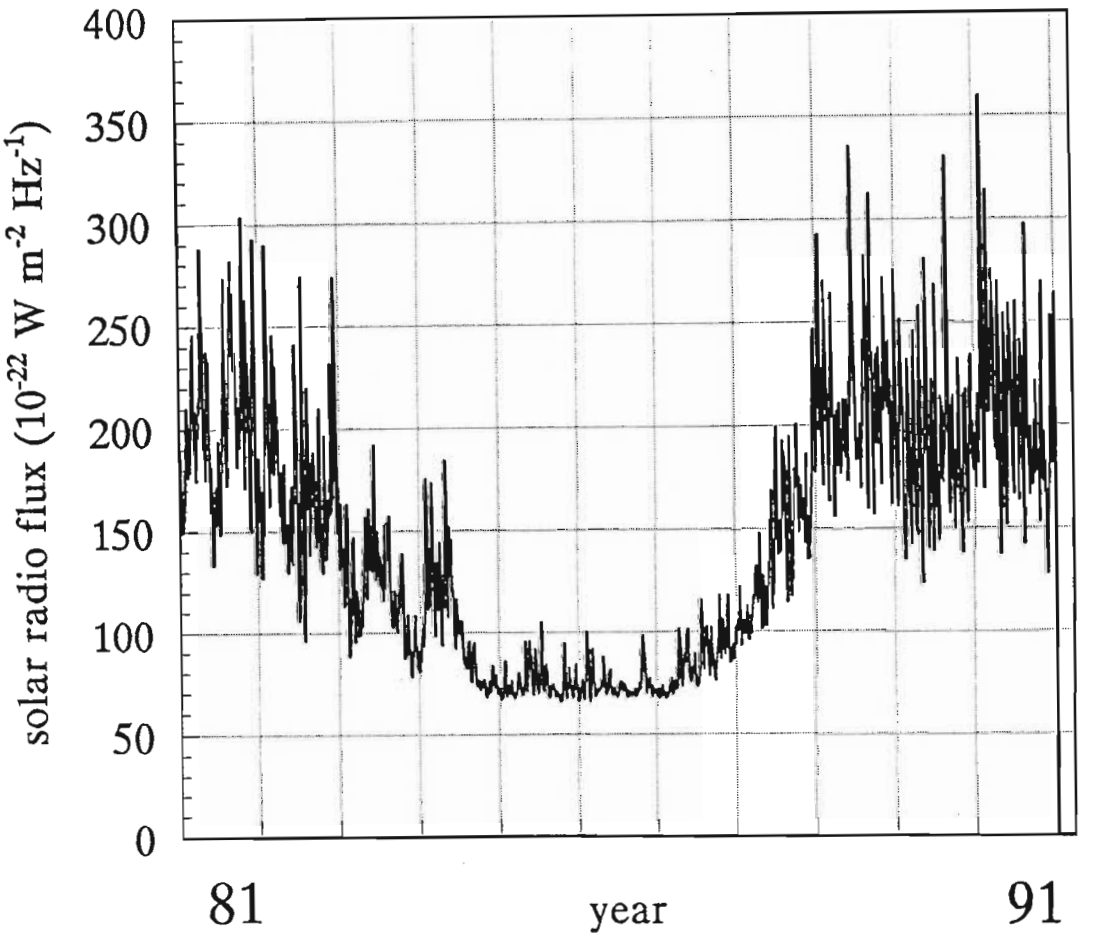

Figure 5.23: The 10.7 cm solar radio flux (10^{-22} W m^{-2} Hz^{-1}) values measured at Ottawa, for the eleven year period 1981 to 1991, inclusively. Large variations in solar radio flux during solar maxima are evident, corresponding to the occurrence of solar flares.

5.23, the maximum of solar cycle 21 occurring over the early 1980's and that of solar cycle 22 beginning in the late 1980's. Figure 5.23 reveals that relatively little variation in the solar flux data occurs during solar minima, whereas considerable fluctuation, corresponding to the presence of large solar flares, are characteristic of years during solar maximum periods.

All programs to prepare the data for the FFT, to call the FFT algorithm and to provide plot ready data files were written, in Fortran for the Hewlett-Packard 9000 mainframe computer, by the author. A schematic representation of the procedure for obtaining the spectrum of the daily ozone mass and solar flux data is given in Figure 5.24. Program names are boxed. Firstly, an array of 4096 real values (ozone mass or solar radio flux) with zero padding already appended, is the input to the program **dcout.f**. This program removes the 'dc', or component of zero frequency, from the data. If not removed, the Fourier transform would have a very broad 'dc' spike that would interfere with other low frequency peaks. Therefore, the 'dc' value of the time sampled data is calculated and subtracted, carefully avoiding the zero padding records. This program returns the data, minus the zero frequency component, to the input array. This data array is then passed to the program **drivefft.f** which creates another array, again of length 4096, with zero values in each element. This second array represents the imaginary component of the time sampled data. The two arrays are combined by dovetailing their values so that a new array, complex(8192), is created whose first value is the real part of the first day in 1981 data and, whose second value represents the imaginary part of the first day in 1981 data. The value stored in complex(3) is the real part of the second days data and so on. The program **drivefft.f** then calls the procedure **four1.f**. This procedure, from *Press et al. (1987)*, is a Fortran version of a Fast Fourier Transform algorithm which scrambles the order of the input data by bit reversal, described in sub-section 5.2.7, before computing the FFT. The frequency data points of the calculated Fourier spectrum are returned to the input array, complex(8192), i.e. it performs *in place* calculation. The real component of the f=0 frequency spike is stored in complex(1), while the imaginary part of the f=0 frequency spike is stored in complex(2). As the 'dc' was removed from the data by **dcout.f**, the values stored in complex(1) and (2) are very nearly zero. Hence, the real value of the nth frequency component is stored in the (n+1) position of the output array. Imaginary and real parts of the Fourier transform are then an input for two programs. The program **magn.f** determines the amplitude of the Fourier transform given by equation (5.3), while the program **pha.f** calculates the phase as given by equation (5.4). Finally, in order to provide an easy method of determining large periodicities in the Fourier spectrum, a program **spike.f** was written by the author whereby peaks are sought that exceed a user defined threshold. The reader is reminded that when providing the intuitive development of the DFT, Figure 5.16, a sampled time function implied a periodic frequency spectrum. This characteristic of

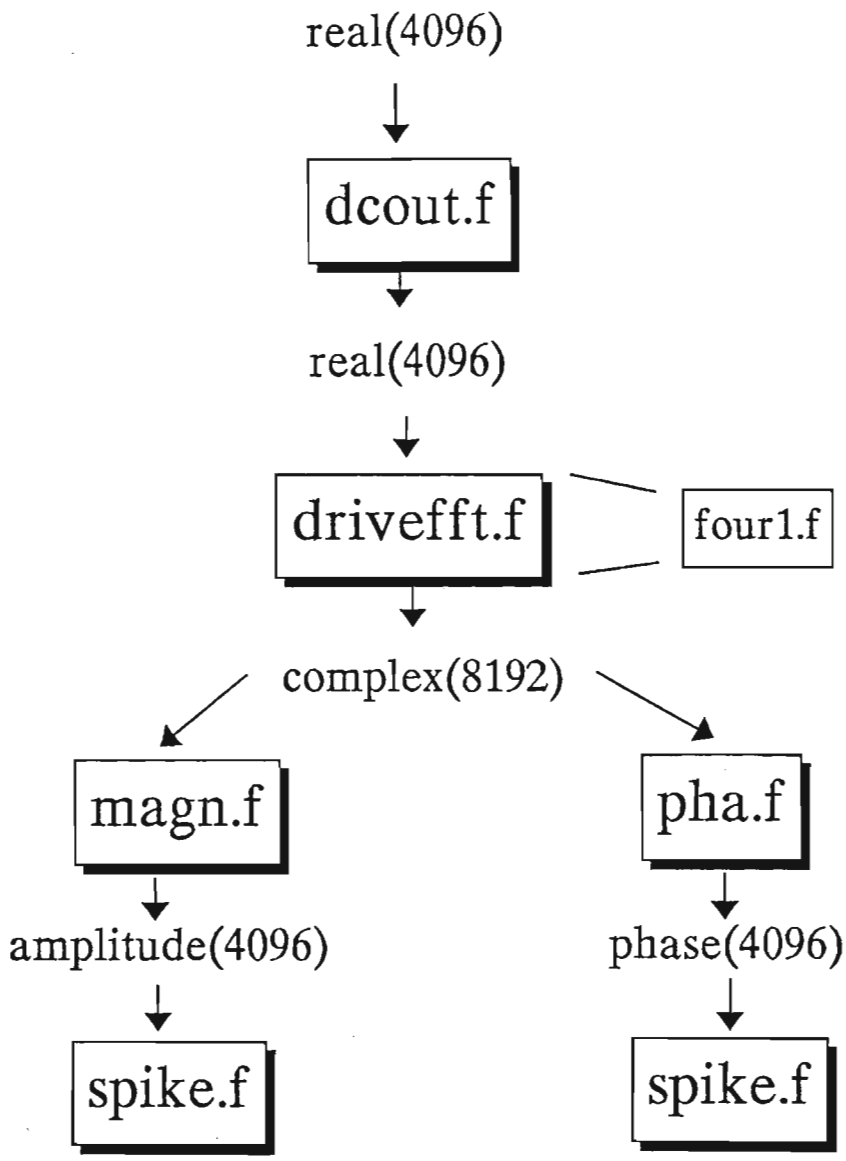

Figure 5.24: A schematic diagram outlining the computational procedure required to determine the Fourier transform of a real function of eleven years of diurnal data.

the DFT is illustrated in panel (c) of this figure. The frequency spectrum is folded about $\frac{1}{2T}$, where T is the sampling period. Therefore, it is only necessary for **spike.f** to search and identify peaks with a frequency record number < 2048. This implies that the smallest periodicity that can be identified is one cycle every two days. Therefore a diurnal peak cannot be identified.

5.4 The Fourier spectrum of polar ozone mass

The amplitude of the Fourier transform, as determined by **magn.f**, of the eleven year data set of polar ozone mass is plotted in Figure 5.25. The folded nature of the spectrum, described in the previous section, is evident. Ground-based observations of total column ozone reveal that, on some occasions, ozone variations associated with weather systems occur within half a day (*Krueger, 1989*). Since the sampling period of the ozone mass data is once per day, there will be some, small amount of aliasing at higher frequencies. However, inspection of Figure 5.25 reveals that the large, and therefore important, peaks in the spectrum are lower than the frequency record number 500 and **therefore any distortion by aliasing can be effectively discounted.** This analysis is confined to long term effects on polar ozone and is therefore not affected by aliasing. A low frequency window ($n \leq 100$) is plotted in Figure 5.26 (see below for the meaning of frequency record number n). The program **spike.f** identified the peaks that are labelled in Figure 5.26 as being over an amplitude threshold of 3×10^{12} magnitude. By inspection of Figure 5.26 these peaks are very much larger than any that might be attributed to noise so that statistical analysis of the peaks is not necessary here. The peaks are denoted by letters, a to m, the significance of each peak will be discussed in the following section.

The reader is reminded of the discrete nature of the spectra which implies that only Fourier spectrum amplitudes corresponding to integral frequency record numbers (n) may be considered. The frequency record number (n), plotted against peak magnitude in Figure 5.26, is interpreted in the following manner,

$$\frac{n}{2^{12}} = x \text{ cycles day}^{-1} \tag{5.54}$$

and so,

$$\frac{n \times 365.25 \text{ days}}{2^{12}} = x \text{ cycles year}^{-1}. \tag{5.55}$$

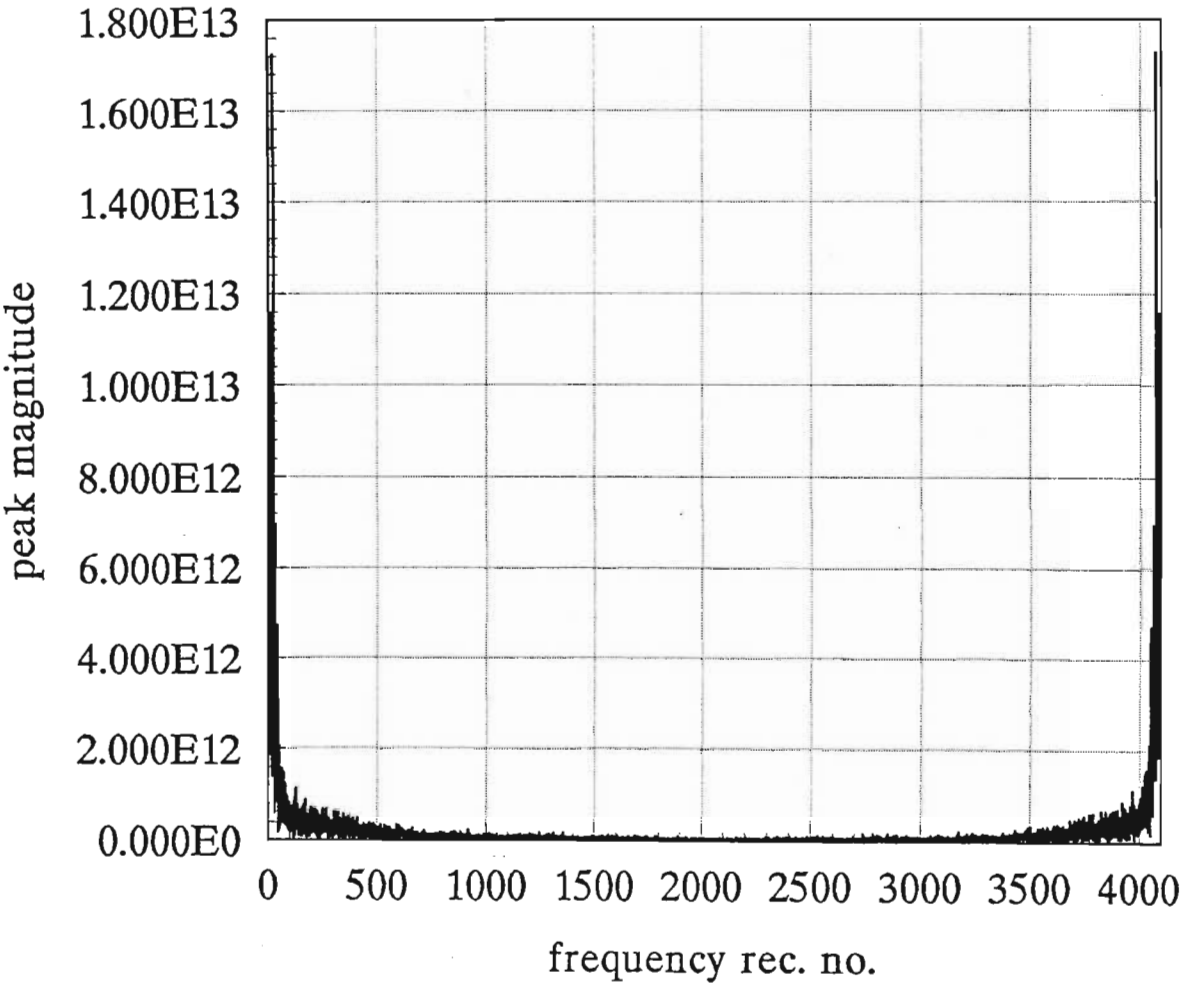

Figure 5.25: The Fourier spectrum of eleven years of diurnal polar ozone mass, plotted as amplitude versus frequency record number (n). The spectrum is folded about the n=2000 record number.

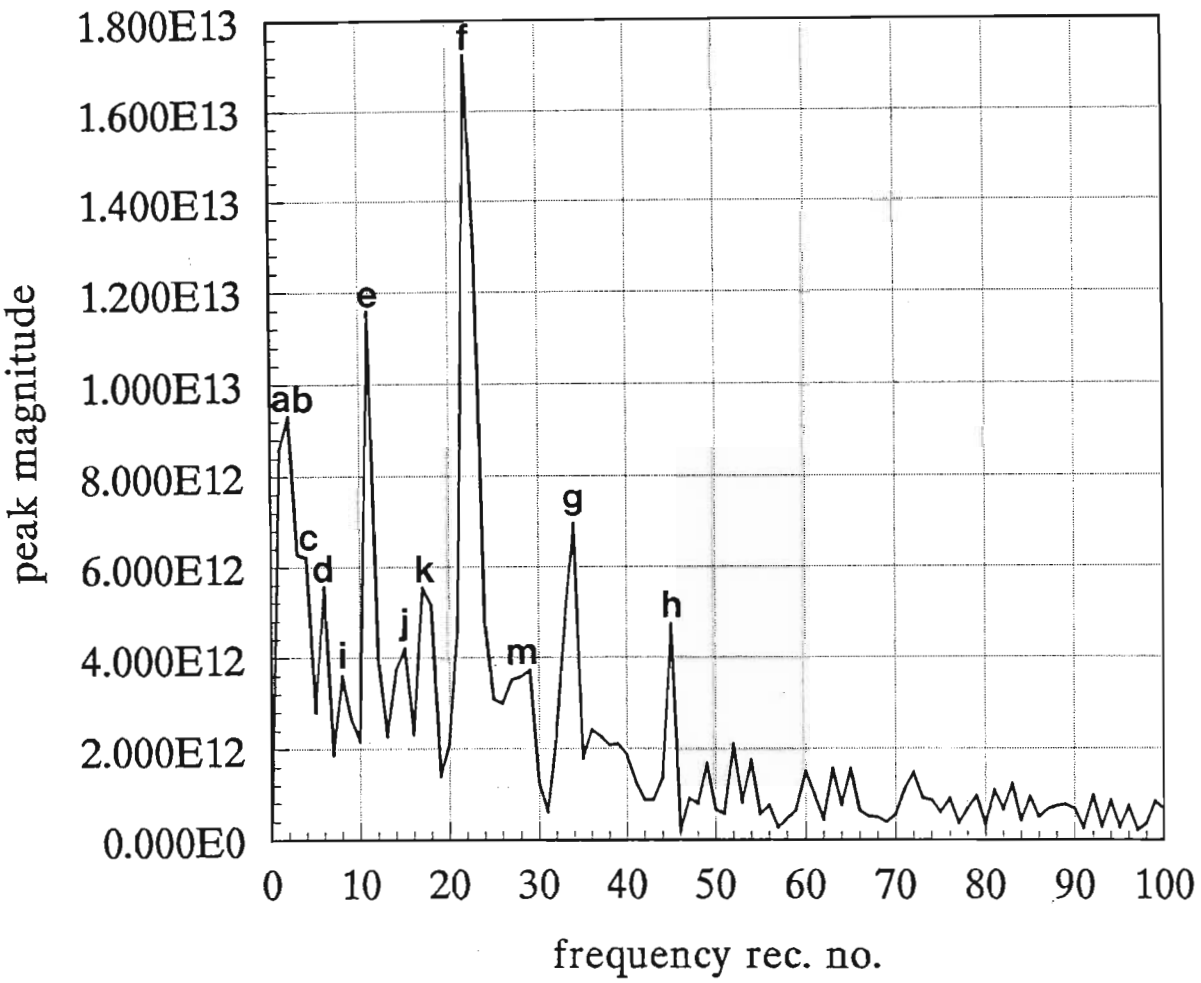

Figure 5.26: A low frequency window (n$\leq$100) of the Fourier spectrum of eleven years of diurnal polar ozone mass. Peaks are labelled that exceed 3 $\times 10^{12}$ magnitude and may be associated with a plausible physical mechanism.

5.4.1 Interpretation and discussion of significant peaks

The amplitude of the Fourier transform of austral polar ozone mass (Figure 5.26) reveals that long term oscillations are primarily responsible for determining the amount of ozone over the southern polar cap at any particular time. The current understanding of the physical mechanisms, behind each of these long term oscillations, is highlighted below. However, a more comprehensive review of them can be found in chapter 2 of this thesis. Once again, we stress that the interpretation of this spectrum applies only to the polar ozone distribution during the interval 1981 to 1991. Important factors in the interannual variability of ozone at high latitudes remain unexplained (*Zerefos et al., 1992*) and therefore peaks (amplitudes $> 3\times10^{12}$) will only be cited, and labelled in Figure 5.26, for which a plausible agency can be identified. Other influences e.g. volcanic emissions have not been considered here.

Interannual peaks

- **Peak a**, frequency record number n = 1, corresponds to 0.09 cycles year^{-1} (equation 5.55), or, one cycle every 11 years. This peak may be interpreted to be the effect of the solar cycle on polar ozone.
- **Peak b**, frequency record number n = 2, corresponds to 0.18 cycles year^{-1}, or, one cycle every 5.6 years which is the period of the first harmonic of solar cycle.

Modern photochemical theory provides three possible mechanisms by which interannual solar variability can affect ozone. These are changes in solar ultraviolet irradiance, proton precipitation and electron precipitation. Each of these aspects of solar variability is discussed below.

Large variations are observed in the ultraviolet and extreme ultraviolet portions of the solar spectrum. These wavelengths control the stratospheric and mesospheric temperature (λ <310 nm), and the photochemistry of the middle atmosphere. In the middle atmosphere, the variability of the emission of electromagnetic radiation from the sun can have two direct consequences. Firstly, this variation can produce a change in the atmospheric heating rate by molecular oxygen and ozone. The resulting temperature variation is fairly small for altitudes below 90 km, where the major heating is provided by ozone absorption in the Hartley band (near 250 nm), since the amplitude of the solar flux variation at these wavelengths is quite small. The change of temperature is also dependent on the ozone-temperature feedback discussed in sub-section 2.8.3. Model calculations e.g. *Garcia et al. (1984)* indicate a maximum stratospheric change of 2 to 3 K near the stratopause. Temperature variations can, in turn, produce perturbations in dynamics.

The second effect that must be considered is the modification to the rates of photodissociation and photoionization of atmospheric constituents, particularly in the upper atmosphere, where the short wavelength radiation exhibiting the largest solar variability can be found. The variability of the shorter wavelengths of extreme ultraviolet radiation (including the Lyman α line) is about a factor of 2 over the course of the eleven year solar cycle. The Lyman α penetrates deeply into the upper mesosphere, and is largely responsible for the photodissociation of water vapour and methane at those altitudes, producing short lived odd hydrogen radicals that destroy odd oxygen. Lyman α also controls the rate of photoionization of NO in the D-region. Extreme ultraviolet radiation also produces nitric oxide in the thermosphere, which can be transported down into the mesosphere and stratosphere. Therefore, a change in the chemistry of the thermosphere and upper mesosphere in connection with the eleven year solar cycle is predicted (*Garcia et al., 1984*). The possible propagation effects into the lower atmosphere will be examined in more detail later in this sub-section.

The radiation in the Schumann-Runge bands can penetrate to the stratopause, and its variability can therefore influence the entire middle atmosphere. The rate of photodissociation of molecular oxygen in the γ(0-0) and γ(1-0) bands can vary by 15-20% over the solar cycle. The variability of the Hertzberg continuum, which penetrates into the stratosphere, is weaker. Since an important part of the photodissociation in this spectral region comes from the atmospheric window near 200 nm, an upper limit of 15% variation of the photodissociation rate is adopted in the Hertzberg continuum. This number applies both to the photodissociation of molecular oxygen and to source gases, N_2O, $CFCl_3$, etc., thus controlling the variability of odd oxygen, odd nitrogen and odd chlorine production. The concentration of ozone is a function of the ratio of two photolysis rates, J_{O_2} and J_{O_3}. Both of these tend to increase along with solar activity, thus tending to cancel one another. However, J_{O_3} depends primarily on solar radiation at wavelengths near 250 nm, where variability is much smaller, while J_{O_2} is determined by wavelengths less than 240 nm, whose variation with solar activity is much larger. The net effect should be a small increase in ozone, although temperature and dynamical feedbacks should also be considered. These increases apply only to low and middle latitudes where chemical production is a dominant process. *Brasseur et al. (1988)* calculated that a solar cycle total ozone change of $\sim$1.6% for a solar ultraviolet flux change of 9% near 205 nm.

High solar activity levels are also accompanied by increased injection of energetic particles, during SPE's and auroral activity. These events increase atmospheric ionization and produce Joule heating and chemical perturbations. SPE's can influence the chemistry of the whole of the middle atmosphere whereas the effect of auroral particles is limited to the thermosphere. However, there is a possibility to influence lower levels by long range transport (*Solomon and Garcia, 1984*). The long term effects of parti-

cle precipitation on the budget of atmospheric NO_x must be considered. *Crutzen et al. (1975)* suggested that the production of NO_x in particle precipitation events could be an important part of the budget of odd nitrogen in the middle atmosphere, since the odd nitrogen input provided by a single SPE can be as large as that produced by oxidation of atmospheric N_2O in an entire year for latitude poleward of 50°. These concepts were explored in a budget analysis in Figure 2.28 by *Jackman et al. (1990)*. *Orsini and Frederick (1982)* presented a one-dimensional model study of the impact of the SPE related NO_x production on NO densities in the mesosphere, and found it to be an important source. *Thorne (1980)* determined that relativistic electrons were also important in NO_x production in the mesosphere. *Solomon et al. (1982)* found that the production of NO_x in aurorae near 100 km could also provide an important source in the mesosphere and stratosphere as a result of large scale transport.

The mechanisms outlined above are addressed in more detail in sub-section 2.8.3 of this thesis.

In order to estimate the potential response of ozone to solar variations, photochemical model simulations are performed. The results of the two-dimensional model of *Garcia et al. (1984)* are given in Figures 2.25, 2.26 and 2.27. This figures illustrate the predicted perturbation of temperature, NO_x and ozone respectively in the middle atmosphere at all latitudes. The model predicts increased production of odd nitrogen at polar latitudes over the solar maximum. In addition, substantial increases are predicted at low altitudes here as a result of downward transport during the polar night. As a result of the NO_x enhancements, in a region which is largely controlled by chemical destruction and dynamics, calculated ozone reductions of up to 40% are predicted at an altitude of 40 km during springtime over the poles. Other aspects of these figures are addressed in detail in sub-section 2.8.3.

- **Peak c**, frequency record number n = 3, corresponds to 0.27 cycles year^{-1}, or, one cycle every 3.7 years. This peak represents the quasi-four year period of ENSO.

The ENSO phenomenon is a quasi-four year oscillation in the tropical troposphere, although its extremes can be detected in the extratropics and the stratosphere. The primary manifestation of the southern oscillation is a seesaw in atmospheric pressure between the south-east Pacific and the Indian Ocean between Australia and Africa. *Shiotani (1992)* has shown that mean zonal ozone values are strongly anti-correlated with the east-west gradient and therefore exhibit a quasi-four year oscillation. *Zerefos et al. (1992)* demonstrated with data from 28 Dobson stations, that large ENSO events are followed by low total ozone at middle, and even high, latitudes.

- **Peak d** frequency record number n = 5, corresponds to 0.45 cycles year^{-1}, or, one cycle every 2.2 years. This interval corresponds to the average period of the oscillation in the 50 mb zonal wind in the tropics, referred to as the QBO.

The 26-30 month (*Zerefos et al., 1992*) oscillation of zonal winds in the equatorial stratosphere affects zonal mean ozone values at tropical latitudes (<15°) resulting in one hemisphere that is positively correlated with the QBO coefficient, while the other is negatively correlated. The low latitude total ozone QBO is driven by vertical advection of ozone by mean meridional circulation induced by the zonal wind QBO. During intervals of westerly wind shear with increasing altitude at the equator, downwelling tends to occur bringing ozone from the middle stratosphere to the lower stratosphere where ozone lifetime is longer. The reverse occurs at adjacent low latitudes where upwelling dominates.

Dunkerton et al. (1988) have noted that mid-winter major warmings appear to be modulated by the QBO. Their study indicates that the polar QBO arises indirectly from changes in eddy activity. They suggest that the link is the increase in eddy activity, which increases the temperature within the polar vortex through the transport of heat from mid-latitudes, during the east phase of the QBO. In addition, strong eddy mixing disturbs the integrity of the polar air mass through injection of high NO_x mid-latitude air, reducing the level of active chlorine. Both the depth of the ozone hole and the rate of the September decline show an apparent QBO signal (*Garcia and Solomon, 1987; Lait et al., 1989*). For example, in 1985 and 1987 (westerly phase QBO years) a strong polar vortex and cold lower stratosphere persisted well into the spring. Both of these years showed a rapid September decline of total ozone which resulted in extreme low October total ozone amounts. In contrast, during 1986 and 1988 (easterly phase QBO years) the lower stratosphere was significantly more disturbed by planetary waves. The result was a warmer, weaker polar vortex and a smaller ozone depletion. Finally, a study of the QBO modulation of Antarctic ozone depletion has been performed by *Lait et al. (1989)* using a 9 year TOMS data set. Their results confirm the findings of *Garcia and Solomon (1987)*.

To conclude the discussion on interannual variations in ozone, related work by other authors is summarized below. Interannual periodicities in the TOMS data set have been estimated by regression analysis (*Stolarski et al., 1991*). The regression model has the form, with constants,

$$O_3(t) = \mu + \alpha \cdot \text{trend} + \beta \cdot \text{QBO} + \gamma \cdot \text{solar} + \text{noise}. \tag{5.56}$$

Fitting the statistical model of equation (5.56) to a period of TOMS data extending from November 1978 to May 1990 between 65°S and 65°N it was estimated that the influence of the solar cycle was 3.7 ± 0.6 DU per 100 units of solar radio flux. Since

the solar radio flux typically varies by more than 130 units from solar maximum to minimum, the implied total ozone change is 1.5 to 1.8%. This is in approximate accord with model predictions e.g. *Brasseur et al. (1988)*. This figure is a global average and the solar coefficient has different values over different latitude ranges. The QBO influence was quantified as -0.4 $\pm$ 0.1 DU per 10 knots of the 30 mb Singapore wind.

A similar regressional model has been applied to 13.2 year of TOMS data by *Hood and McCormack (1992)*. The existence of a solar cycle component is indicated by empirical studies of model residuals and by the approximate agreement of the derived global mean solar coefficient amplitude with photochemical calculations. Initial estimates for the latitude dependence of the solar coefficient (γ in equation (5.56)) suggest higher amplitudes with increasing latitude. The southern polar cap in spring was found to have the largest solar cycle component. If the solar variability is omitted in the regression model, the residual exhibits a variation that approximately follows the solar cycle.

Shiotani (1992) has analyzed 11 years of TOMS at tropical latitudes from which he determined annual, QBO and ENSO signals in the equatorial ozone field.

From analysis of total ozone from a network of 28 Dobson stations, during the interval 1957 to 1991, *Zerefos et al. (1992)* determined that the quasi-four year oscillation associated with the ENSO phenomenon can be detected at middle and high latitudes.

Seasonal peaks

- **Peak e**, frequency record number n = 11, corresponds to 0.98 cycles year^{-1}. This peak represents the annual variation in ozone over Antarctica.
- **Peak f**, frequency record number n = 22, corresponds to 1.9 cycles year^{-1}. This may be interpreted as the first harmonic of annual variation, or semi-annual variation.
- **Peak g**, frequency number n = 33, corresponds to 2.9 cycles year^{-1} which is the second harmonic of annual variation.
- **Peak h**, frequency record number n = 44, corresponds to 3.9 cycles year^{-1} which is the third harmonic of the annual variation.

Perliski and London (1989) have observed annual and semi-annual oscillations in a nine year vertical profile ozone data set from the SBUV instrument, aboard Nimbus 7. At high latitudes three main regions of maximum annual amplitude were located near 40, 7 and 1.5 mb, each separated by a region of minimum amplitude called a transition zone. In the transition zones, where the phase of the annual variation changes most

rapidly with height, the semi-annual oscillation accounts for most of the observed long-term ozone variance. The largest maximum in the semi-annual amplitude is located near 2 mb in the high latitude stratosphere. *Perliski et al. (1989)* have compared these measurements with a two-dimensional photochemical model. The observed SBUV data and modelled predictions for the amplitudes of the annual and semi-annual variations are depicted in Figures 2.2 and 2.3 respectively. The reader is reminded that no polar ozone (>65°) observations were available. Model predictions, however, cover all latitudes. This model simulates the processes believed to be responsible for the seasonal variations of ozone. These processes are seasonal changes in odd oxygen production rates, temperature dependent ozone destruction rates, and transport. At high latitudes and low altitudes, modelled ozone abundances increase in the winter due to poleward transport and decrease in the summer due to chemical destruction. However, at higher altitudes the annual ozone variation is largely due to the annual variation in the odd oxygen production rate. The seasonal variation of ozone as a function of latitude is plotted in Figure 1.25, which shows clearly that seasonal variations are largest in the polar regions. Polar semi-annual ozone oscillations are due to the modulation of the radiatively driven odd oxygen production rate by the temperature dependent chemical destruction processes (*Perliski et al., 1989*). The amplitudes of the modelled annual variation (Figures 2.2) is almost twice that of the semi-annual variation (Figure 2.3) at polar latitudes. However, the amplitude of the peak corresponding to the semi-annual oscillation (**peak f** in Figure 5.26) in the Fourier spectrum of the austral polar ozone mass is one and a half times larger than that of the annual variation (**peak e** in Figure 5.26). The semi-annual amplitude increment is a direct consequence of the presence of the Antarctic ozone hole at a time in the year when ozone values would have previously maximized. This enhanced semi-annual variation over the austral pole has also been found by regression analysis of the TOMS data (private communication, *G. E. Bodeker*), and so it is not peculiar to this analysis method. In addition, the slight over-estimation of ozone mass values during the polar night (section 5.3) could also contribute in strengthening the semi-annual signal.

Other authors e.g. *Varotsos et al. (1992)* have determined annual and semi-annual waves in global ozone from SBUV vertical ozone profiles. *Stolarski et al. (1991)* applied a model to eleven years of TOMS data to remove seasonal variations in order to detect global ozone loss from anthropogenic sources. After the evaluation of the statistical structure of the noise on the data, a regression analysis estimated the annual variation and its three harmonics, 6 month, 4 month and 3 month. In addition, the solar cycle, QBO and linear trend signals were determined.

Modulated peaks

When a slow varying function, h(t), is modulated onto a higher carrier frequency, f_0, the resulting transform is

$$h(t)\cos(2\pi f_0 t) \Leftrightarrow \frac{1}{2} H(f - f_0) + \frac{1}{2} H(f + f_0). \tag{5.57}$$

Modulation is a special case of the frequency convolution theorem which states that the spectrum of two time domain functions which are multiplied together, can be determined by convolving the individual spectra. This process results in two peaks, one at f_0 + f and another at f_0 - f. For more detail on the modulation process, see sub-section 5.2.4.

By inspection of Figure 5.26 it is clear that the annual peak (marked **peak e**) is flanked, at an equal distance, on either side by **peaks i** and **j**. **Peak j** is at frequency record number n = 14, or n = 11 (annual peak) + 3 (ENSO peak). Similarly, **peak i** which is at frequency record number 8 which is n = 11 - 3. Therefore the ENSO signal in polar ozone is modulated onto the annual variation. Similarly, **Peaks d** and **k** correspond to the modulation of the QBO (n = 5) onto the annual variation. The width of **peak e**, corresponding to the annual oscillation, prevents the detection of the presence of a modulated solar cycle on the seasonal variation.

Furthermore, the semi-annual variation (**peak f**) is clearly modulated by the QBO, represented by **peaks k** and **m**. The presence of the Antarctic ozone hole is directly responsible for enhancing the magnitude of the semi-annual variation of austral polar ozone, therefore modulation of **peak f** suggests that the QBO modulates the chlorine catalyzed ozone depletion over Antarctica. Unfortunately, both the modulation of ENSO and the solar cycle on the semi-annual oscillation cannot be resolved.

It has been proposed by other authors that the QBO may be modified by either the solar cycle (*Labitzke and van Loon, 1988; Bojkov et al., 1990; Angell, 1989*) and/or a quasi-four year oscillation (*Hasebe, 1983; Schuster et al., 1989*). However, as is the case with the analysis outlined in this chapter, definitive proof requires a much longer data set before the modulation can be resolved.

Finally, the modulation of the QBO on seasonal variations at equatorial latitudes has been detected in the TOMS data by *Shiotani (1992)*. *Hood and McCormack (1992)* noted that the QBO regression coefficient of total ozone at high latitudes has a seasonal dependence. The coefficient is very much enhanced over the austral pole during the months of September, October and November.

Other peaks

Variations in the ultraviolet region (120 to 210 nm) of the solar spectrum over the 27 day solar rotation period have been observed (*Rottman, 1983*), and so a corresponding peak in the Fourier polar ozone mass spectrum was sought. The 27 day periodicity corresponds to frequency record number

$$\frac{n \times 365.25}{2^{12}} = 13.53\,\text{cycles year}^{-1} \tag{5.58}$$

$$n = 152. \tag{5.59}$$

Figure 5.27 is a plot of the frequency window extending from frequency record number 140 to 160. The peak magnitudes in Figure 5.27 are two magnitudes smaller than those corresponding to the interannual periodicities in Figure 5.26 which reflects the weak influences on polar ozone at higher frequencies. The peak corresponding to n = 152 is not statistically significant. Ozone variations with a 27 day period have never been detected in total ozone measurements, although they are presents in vertical profile ozone observations (*Brasseur et al., 1987*). This result is therefore consistent with observations.

5.5 The Fourier spectrum of solar radio flux

The 10.7 cm solar radio flux has been found to be closely related to solar ultraviolet variations at wavelengths that are photochemically important for ozone and odd nitrogen production in the upper stratosphere (*Donnelly, 1991*). Therefore, the Ottawa 10.7 cm radio flux was used as a proxy to solar activity during the eleven year interval from 1981 to 1991. This implies that solar cycle forcing occurs primarily in the form of ultraviolet variations and so the possible middle atmospheric chemical effects of charged particle precipitation are excluded in this particular analysis.

The first one hundred frequency record numbers of the Fourier spectrum of the eleven years of solar radio flux data are plotted in Figure 5.28. The only prominent peak in the spectrum is at frequency record number n=1 corresponding to the solar cycle, which is marked as **peak s** in Figure 5.28.

5.5.1 Relationship between the two spectra

Cross-correlation between the solar radio flux spectrum and the ozone mass produces a spectrum with one large peak at frequency record number n=1, corresponding to the

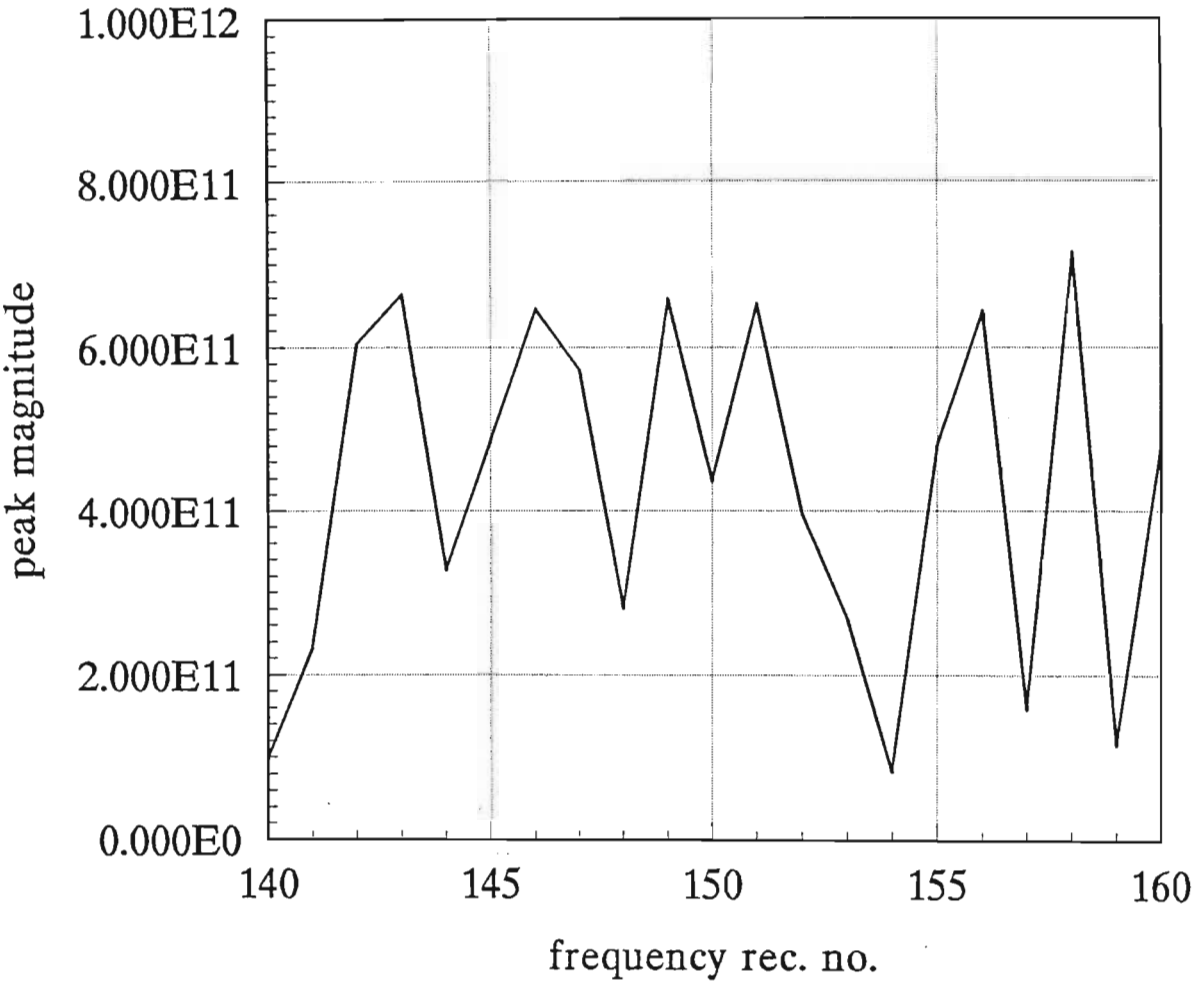

Figure 5.27: A frequency window (n=140 to n=160) of the polar ozone mass Fourier spectrum. A peak (n=152) corresponding to the 27 day solar rotation can not be identified.

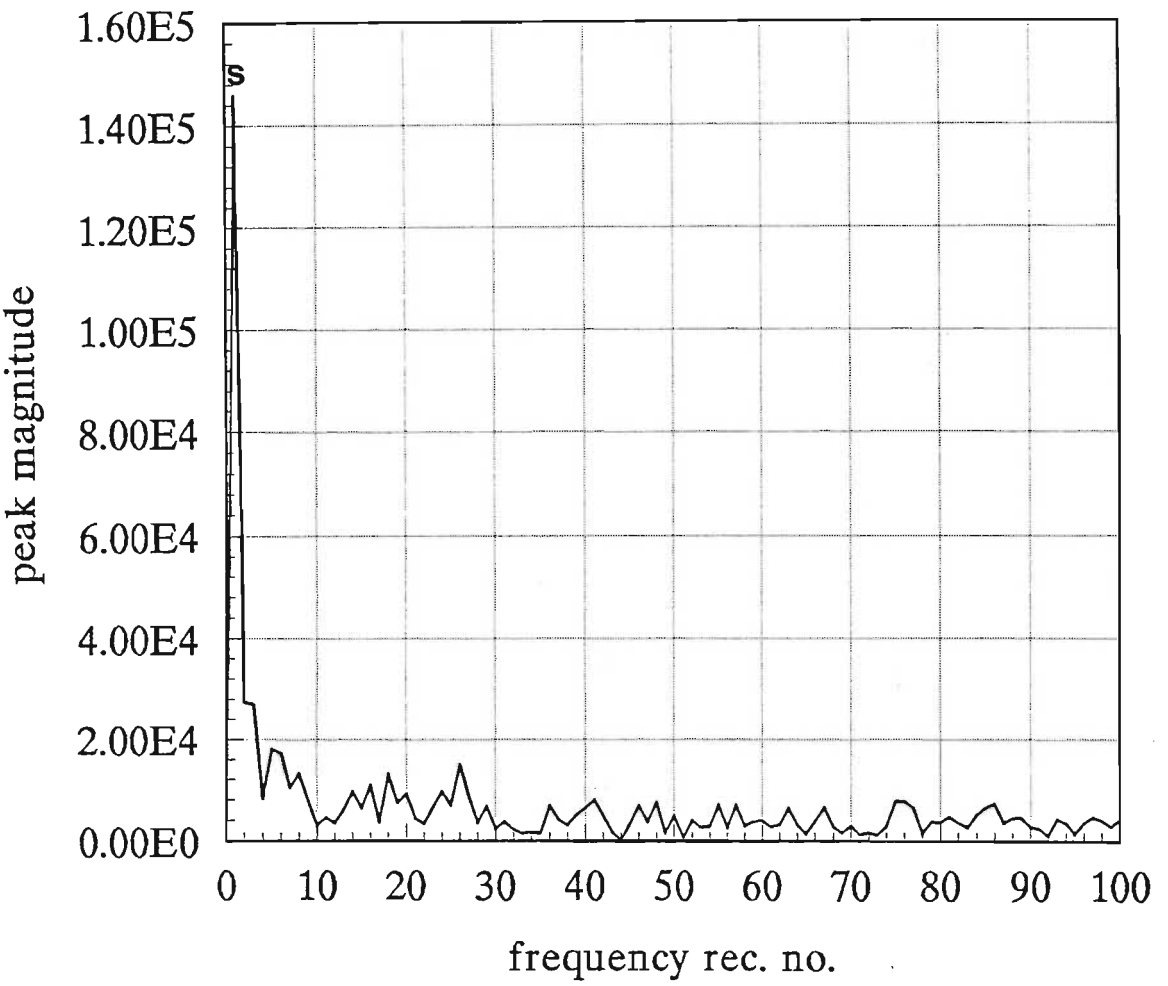

Figure 5.28: The Fourier spectrum of eleven years of diurnal solar radio flux values. The lowest 100 frequency record numbers are plotted. The peak corresponding to the solar cycle periodicity is marked 's'.

solar cycle. The phase of each solar cycle peak (n = 1 in each spectrum) was calculated by the Fortran program **pha.f** using equation (5.4). The phase of the solar cycle peak in the ozone mass spectrum was 67° while its phase in the solar radio flux spectrum was -25°. The phase difference between the solar cycle peak in the solar radio spectrum and the polar ozone mass spectrum was therefore 92°. In other words, the ozone minimum lags, as predicted by the two-dimensional model of *Garcia et al. (1984)*, the solar radio flux maximum by 92°. This phase difference corresponds to a time lag in polar ozone of ~2.8 years, for this eleven year period. This is the first time, to our knowledge, that the solar cycle phase lag of polar ozone has been determined by Fourier analysis. The maximum of solar cycle 22 was achieved in the first months of 1990 (*Solar geophysical data, 1990, a*), and so a a polar ozone minimum would be predicted in 1987. *Garcia et al. (1984)* predict that the solar cycle effect is most pronounced in spring, after the additional NO_x has been transported down to altitudes where ozone is most abundant. The lowest total ozone levels achieved in this eleven year period over Antarctica were recorded in 1987 (*Stolarski et al., 1990*). This result is evident in Figure 5.22 were austral polar ozone mass values are plotted for 1981 through to 1991. The seasonal ozone decline corresponding to the springtime ozone hole achieves lower and lower values until 1987, after which there is a marginal improvement.

The analysis of ground-based total ozone and 100 mb temperature data over a 30 year interval has yielded some provisional evidence for a solar modulation of the Antarctic ozone in the southern hemisphere spring (*Angell, 1988*).

Lag correlation coefficients have been determined from ground-based measurements by other authors. For example, a lag of 3 years was determined from Arosa data by *London and Oltmans (1973)*.

5.6 Sources of error

The TOMS algorithm has been shown to be sensitive to profile shape at large solar zenith angles where penetration to the ground is not good (*Klenk et al., 1982*). Therefore, the TOMS measurements have their greatest error at zenith angles greater than 80° i.e. data taken near the terminator. *Stolarski et al. (1990)* put an upper limit of 12% error in version 5 of the TOMS observations. The version 6 data, used in this analysis, are considered to be more accurate. A more detailed discussion on the accuracy of the TOMS instrument can be found in section 3.3.

The TOMS instrument determines ozone amounts by the backscattered ultraviolet method. Therefore, no measurements can be made during the polar winters. The data during this period were estimated by methods outlined in sub-section 5.3. Ground-

based total ozone measurements at Halley, over the same interval, suggest that the polar ozone mass values were over-estimated by ~5%. Errors in the winter polar ozone mass would certainly affect the relative amplitudes of the seasonal peaks in the ozone spectrum.

Although double precision definitions were used when computing the FFT, rounding errors by the Hewlett-Packard mainframe cannot be discounted.

As previously indicated, the frequency spectrum computations are limited to integer multiples of the frequency sampling interval. The 'picket fence' effect prevents the determination of frequency spikes that occur between the multiple frequency intervals. For example, we cannot determine the spike in the ozone mass spectrum that corresponds exactly to 1 cycle year^{-1} since this corresponds to frequency record number

$$n = \frac{1 \times 2^{12}}{365.25} \tag{5.60}$$

$$n = 11.2. \tag{5.61}$$

An effective solution is to vary the total number of time samples, by appending zero values to the front, or end, of the time sampled array. This process artificially alters the frequency multiple size so that peaks corresponding to 'new' periodicities can be viewed, effectively allowing one to 'look in between' the pickets.

Some aliasing in the spectra was known to occur. However, this has no effect on the low frequency end of the spectra from which interesting peaks were sited and discussed.

The leakage is small as very little zero padding was required. Therefore, the ripple effect on the spectra was negligible.

5.7 Summary

There are only a few analyses of the large scale structure of total ozone to determine long term oscillations. This may be attributed to the unavailability of appropriate data. The results of the Fourier analysis of diurnal ozone mass during the eleven year period between 1981 and 1991 over Antarctica show:

- peaks associated with the solar cycle, ENSO, QBO, annual and semi-annual oscillations. In addition, the second and third harmonics of the annual signal.

- modulation of interannual variations on annual and semi-annual peaks is evident.

- comparison of the phase of the solar cycle peak in the solar radio flux spectrum with that of the polar ozone mass spectrum reveals, for this eleven year period, that the solar cycle periodicity in austral polar ozone lags solar cycle forcing by ~2.8 years.

- the polar ozone spectrum indicates that complex relationships exist between solar and dynamics effects.

5.8 Proposals for future work

To conclude this final chapter, suggestions for further work with this technique are proposed.

1. The polar ozone phase lag following strong ENSO events could be determined by comparison of the polar ozone spectrum with a FFT of eleven years of diurnal 30 mb zonal wind values at Singapore.

2. A Fourier analysis, identical to that performed in this work, of northern polar ozone would be valuable. The comparison of the peak amplitudes in the two polar ozone spectra would enable interhemispheric differences to be quantified. This result would reveal the relative importance of the different long term processes governing polar ozone over each polar cap. In particular, this study could give further, and verify existing, evidence as to why chlorine catalyzed depletion is less effective over northern polar regions.

3. The analysis described in this chapter could provide a valuable baseline. Comparison with similar analyses performed in the future could provide clues as to which processes are most responsible for the rapidly changing polar ozone distribution.

References

Aikin, A. C., Spring polar ozone behaviour, *Planet. Space Sci.*, **40**, 7, 1992.

Allen, J., L. Frank, H. Sauer and P. Reiff, Effects of the March 1989 solar activity, *Eos*, **70**, 1479, 1989.

Anderson, J. G., Free radicals in the earth's atmosphere: measurement and interpretation in *Ozone Depletion, Greenhouse Gases and Climate Change*, Proceedings of a Joint Symposium by the Board on Atmospheric Sciences and Climate and the Committee on Global Change, National Academy Press, Washington D. C., U.S.A., 1988.

Anderson, J. G., W. H. Burne and M. J. Proffitt, Ozone destruction by chlorine radicals within the Antarctic vortex: the spatial and temporal evolution of ClO - O_3 anticorrelation based on *in situ* ER-2 data, *J. Geophys. Res.*, **94**, 11465, 1989 a.

Anderson, J. G., W. H. Burne, S. A. Lloyd, D. W. Toohey, S. P. Sander, W. L. Starr, M. Loewenstein and J. R. Podolske, Kinetics of O_3 destruction by ClO and BrO within the Antarctic vortex: an analysis based on *in situ* ER-2 data, *J. Geophys. Res.*, **94**, 11480, 1989 b.

Angell, J. K. and J. Korshover, Quasi-biennial and long term fluctuations in total ozone, *Mon. Weather Rev.*, **101**, 476, 1973.

Angell, J. K. and J. Korshover, Global ozone variations: an update into 1976, *Mon. Weather Rev.*, **106**, 725, 1978.

Angell, J. K., Relation of Antarctic 100 mb temperature and total ozone to equatorial QBO, equatorial SST, and sunspot number, *Geophys. Res. Lett.*, **15**, 915, 1988.

Angell, J. K., On the relation between atmospheric ozone and sunspot number, *J. Clim.*, **2**, 1404, 1989.

Arnold, F. Paper presented during AEP spring meeting of German Physical Society, Munich, 1978.

Atkins, P. W., *Physical chemistry*, 2nd edition, Oxford University Press, Oxford, United Kingdom, 1982.

Austin, J. and N. Butchart, A three-dimensional modeling study of the influence of planetary wave dynamics on polar ozone photochemistry, *J. Geophys. Res.*, **97**, 10165, 1992.

Austin, J., N. Butchart and K. P. Shine, Possibility of an Arctic ozone hole in a doubled-CO_2 climate, *Nature*, **360**, 221, 1992.

Barth, C. A., D. W. Rusch, R. J. Thomas, G. H. Mount, G. J. Rottman, G. E. Thomas, R. W. Saunders and G. M. Lawrence, Solar Mesosphere Explorer: scientific objectives and results, *Geophys. Res. Lett.*, **10**, 237, 1983.

Berggren, R. and Labitzke, The distribution of ozone on pressure surfaces, *Tellus*, **20**, 88, 1968.

Bethe, H. A. and J. Ashkin, in *Experimental nuclear physics, Volume I*, John Wiley and Sons, New York, U.S.A, 1953.

Bhartia, P. K., K. F. Klenk, C. K. Wong, D. Gordon and A. J. Fleig, Intercomparison of the Nimbus 7 SBUV/TOMS total ozone data sets with Dobson and M83 results, *J. Geophys. Res.*, **89**, 5239, 1984.

Blackshear, W. T. and R. H. Tolson, High correlations between variations in monthly averages of solar activity and total atmospheric ozone, *Geophys. Res. Lett.*, **5**, 921, 1978.

Bodeker, G. E., The dynamics of the ozone layer between South Africa and Antarctica, *MSc Thesis*, University of Natal, Durban, South Africa, 1992.

Bojkov, R. D., The 1983 and 1985 anomalies in ozone distribution in perspective, *Mon. Weather Rev.*, **115**, 2187, 1987.

Bojkov, R. D., L. Bishop, W. J. Hill, G. C. Reinsel and G. C. Tiao, A statistical analysis of revised Dobson total ozone data over the northern hemisphere, *J. Geophys. Res.*, **95**, 9785, 1990.

Bracewell, R. N., *The fourier transform and its applications*, McGraw-Hill Kogakusha, Tokyo, Japan, 1978.

Brasseur, G. *Physique et chimie de l'atmosphère moyenne*, Masson, Paris, France, 1982.

Brasseur, G., P. De Baets and A. De Rudder, Solar variability and minor constituents in the lower thermosphere and in the mesosphere, *Space Sci. Rev.*, **34**, 377, 1983.

Brasseur, G. and S. Solomon, *Aeronomy of the middle atmosphere*, 2nd edition, D. Reidel, Dordrecht, Holland, 1986.

Brasseur, G., A. De Rudder, G. M. Keating and M. C. Pitts, Response of the middle atmosphere to short term solar ultraviolet variations: 2. Theory, *J. Geophys. Res.*, **92**, 903, 1987.

Brasseur, G., M. H. Hitchman, P. C. Simon and A. De Rudder, Ozone reduction in the 1980's: a model simulation of anthropogenic and solar perturbations, *Geophys. Res. Lett.*, **15**, 1361, 1988.

Brewer, A. W., Evidence for a world circulation provided by measurements of helium and water vapour distribution in the stratosphere, *Q. J. R. Meteorol. Soc.*, **75**, 351, 1949.

Brigham, E. Oran, *The fast fourier transform*, Prentice-Hall, New York, U.S.A., 1974.

Browell, E. V., C. F. Butler, S. Ismail, P. A. Robinette, A. F. Carter, N. S. Higdon, O. B. Toon, M. R. Schoeberl and A. F. Tuck, Airborne lidar observations in the wintertime Arctic stratosphere: polar stratospheric clouds, *Geophys. Res. Lett.*, **17**, 385, 1990.

Burne, B., D. W. Toohey, J. G. Anderson and K. R. Chan, *In situ* observations of ClO in the Arctic stratosphere: ER-2 aircraft results from 59°N to 80°N latitude, *Geophys. Res. Lett.*, **17**, 505, 1990.

Callis, L. B. and M. Natarajan, The Antarctic ozone minimum: relationship to odd nitrogen, odd chlorine, the final warming, and the 11-year solar cycle, *J. Geophys. Res.*, **91**, 10771, 1986.

Chandra, S. and R. D. McPeters, Some observations on the role of planetary waves in determining the springtime ozone distribution in the Antarctic, *Geophys. Res. Lett.*, **13**, 1224, 1986.

Charney, J. C. and P. G. Drazin, Propagation of planetary-scale disturbances from the lower to the upper atmosphere, *J. Geophys. Res.*, **66**, 83, 1961.

Chatfield, R. B. and P. J. Crutzen, Sulfur dioxide in remote oceanic air: cloud transport of reactive precursors, *J. Geophys. Res.*, **89**, 7111, 1984.

Chemical Rubber Company handbook for chemistry and physics, edited by R. C. Weast, CRC Press Inc., Boca Raton, Florida, U.S.A., 1984.

Coffey M. T., W. G. Mankin and A. Goldman, Simultaneous spectroscopic determination of the latitudinal, seasonal and diurnal variability of stratosphere N_2O, NO, NO_2 and HNO_3, *J. Geophys. Res.*, **86**, 7731, 1981.

Cooley, J. W. and J. W. Tukey, An algorithm for the machine calculation of complex Fourier series, *Mathematics of Computation*, **19**, 297, 1965.

Covington, A. E., Solar radio emission at 10.7 cm, *J. Royal Astron. Soc., Canada*, **63**, 125, 1969.

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

Crutzen, P. J., I. S. A. Isaksen and G. C. Reid, Solar Proton Events: stratospheric sources of nitric oxide, *Science*, **189**, 457, 1975.

Crutzen, P. J. and F. Arnold, Nitric acid cloud formation in the cold Antarctic stratosphere: a major cause for the springtime 'ozone hole', *Nature*, **324**, 651, 1986.

Cunnold, D. M., F. Alyea, N. Phillips and R. G. Prinn, A three-dimensional dynamical-chemical model of atmospheric ozone, *J. Atmos. Sci.*, **32**, 170, 1975.

Dalgarno, A., Atmospheric reactions with energetic particles, *Space Research*, **7**, 849, 1971.

Dave, J. V. and C. L. Mateer, A preliminary study of the possibility of estimating total atmospheric ozone from satellite measurements, *J. Atmos. Sci.*, **24**, 414, 1967.

Dobson, G. M. B., D. C. Harrison and J. Lawrence, Measurement of the amount of ozone in the earth's atmosphere and its relation to other geophysical conditions. Part III, *R. Soc. (Lond.) Proc. ser. A.*, **122**, 456, 1929.

Dobson, G. M. B., Annual variation of ozone in Antarctica, *Q. J. R. Meteorol. Soc.*, **92**, 549, 1966.

Donnelly, R. F., Solar UV spectral irradiance variations, *J. Geomagn. Geoelect.*, **43**, 835, 1991.

Dubach, J. and W. A. Barker, Charged particle induced ionization rates in planetary atmospheres, *J. Atmos. Terr. Phys.*, **33**, 1287, 1971.

Dunkerton, T., On the mean meridional mass motions of the stratosphere and mesosphere, *J. Atmos. Sci.*, **35**, 2325, 1978.

Dunkerton, T., C. -P. F. Hsu and M. E. McIntyre, Some Eulerian and Lagrangian diagnostics for a model stratospheric warming, *J. Atmos. Sci.*, **38**, 819, 1981.

Dunkerton, T. J., D. P. Delisi and M. P. Baldwin, Distribution of major stratospheric warmings in relation to the quasi-biennial oscillation, *Geophys. Res. Lett.*, **15**, 136, 1988.

Dütsch, H. U., The search for solar cycle-ozone relationships, *J. Atmos. Terr. Phys.*, **41**, 771, 1979.

Eisberg, R. and R. Resnick, *Quantum physics of atoms, molecules, solids, nuclei and particles*, 2nd edition, John Wiley and Sons Pub. Co., New York, U.S.A., 1985.

Fabian, P., J. A. Pyle and R. J. Wells, The August 1972 solar proton event and the atmospheric ozone layer, *Nature*, **277**, 458, 1979.

Fabry, C. and H. Buisson, Étude de l'extrémité ultraviolette du spectre solaire, *J. de Physique (Paris) Serie 6*, **2**, 197, 1921.

Fahey, D. W, D. M. Murphy, K. K. Kelly, M. K. W. Ko, M. H. Proffitt, C. S. Eubank, G. V. Ferry, M. Loewenstein and K. R. Chan, Measurement of nitric oxide and total reactive nitrogen in the Antarctic stratosphere: observations and chemical implications, *J. Geophys. Res.*, **94**, 16665, 1989.

Fahey, D. W., S. R. Kawa and K. R. Chan, Nitric oxide measurements in the Arctic winter stratosphere, *Geophys. Res. Lett.*, **17**, 489, 1990.

Farman, J. C., B. G. Gardiner and J. D. Shanklin, Large losses of total ozone in Antarctica reveal seasonal ClO_x/NO_x interaction, *Nature*, **315**, 201, 1985.

Fleig, A. J., K. F. Klenk, P. K. Bhartia, D. Gordon and W. H. Schneider, Users' guide for the Solar Backscatter Ultraviolet (SBUV) instrument first year ozone-S data set, *NASA-RP*, **1095**, 1982.

Fleig, A .J., P. K. Bhartia, C. G. Wellemeyer and D. S. Silberstein, Seven years of total ozone from the TOMS instrument - a report on data quality, *Geophys. Res. Lett.*, **13**, 1355, 1986.

Forbush, S. E., Three unusual cosmic ray intensity increases due to charged particles from the sun, *Phys. Rev.*, **70**, 771, 1946.

Frederick, J. E. and N. Orsini, The distribution and variability of mesospheric odd nitrogen: a theoretical investigation, *J. Atmos. Terr. Phys.*, **44**, 4798, 1982.

Freier, P. S. and W. R. Webber, Exponential rigidity spectrums for solar flare cosmic rays, *J. Geophys. Res.*, **68**, 1605, 1963.

Garcia, R. R. and S. Solomon, A numerical model of the zonally averaged dynamical and chemical structure of the middle atmosphere, *J. Geophys. Res.*, **88**, 1379, 1983.

Garcia R., S. Solomon, R. G. Roble and D. W. Rusch, A numerical study of the response of the middle atmosphere to the 11 year solar cycle, *Planet. Space Sci.*, **32**, 411, 1984.

Garcia, R. R. and S. Solomon, A possible relationship between interannual variability in Antarctic ozone and the quasi-biennial oscillation, *Geophys. Res. Lett.*, **14**, 848, 1987.

Gardiner, B. G. and J. C. Farman, *Results of the 1987 ozonesonde programme at Halley Bay, Antarctica*, British Antarctic Survey, Cambridge, United Kingdom, 1988.

Geller, M. A. and J. C. Alpert, Planetary wave coupling between the troposphere and the middle atmosphere as a possible sun-weather mechanism, *J. Atmos. Sci.*, **37**, 1197, 1980.

Gille, J. C., C. M. Smythe and D. F. Heath, Observed ozone response to variations in solar ultraviolet radiation, *Science*, **225**, 315, 1984.

Gordley, L. L. and J. M. Russell III, Rapid inversion of limb radiance data using an emissivity growth approximation, *Appl. Optics*, **20**, 807, 1981.

Gorney, D. J., Solar cycle effects on the near-earth space environment, *Rev. Geophys.*, **28**, 315, 1990.

Gribbin, J., *The hole in the sky*, Corgi, London, United Kingdom, 1988.

Guthrie, P. D., C. H. Jackman, J. R. Herman and C. J. McQuillian, A diabatic circulation experiment in a two-dimensional photochemical model, *J. Geophys. Res.*, **89**, 9589, 1984.

Hamill, P., O. B. Toon and R. P. Turco, Characteristics of polar stratospheric clouds during the formation of the Antarctic ozone hole, *Geophys. Res. Lett.*, **13**, 1288, 1986.

Hare, F. K. and B. W. Boville, *The polar circulation*, Technical Note 70, World meteorological organization, Geneva, Switzerland, 1965.

Hargreaves, J. K., *The upper atmosphere and solar-terrestrial relations*, Van Nostrand Reinhold, Wokingham, United Kingdom, 1979.

Hartmann, D. L., K. R. Chan, B. L. Gary, M. R. Schoeberl, P. A. Newman, R. L. Martin, M. Loewenstein, J. R. Podolske and S. E. Strahan, Potential vorticity and mixing in the south polar vortex during spring, *J. Geophys. Res.*, **94**, 11625, 1989.

Harwood, R. S. and J. A. Pyle, A two-dimensional mean circulation model for the atmosphere below 80 km, *Q. J. R. Meteorol. Soc.*, **101**, 723, 1975.

Harwood, R. S. and J. A. Pyle, Studies of the ozone budget using a zonal mean circulation model and linearized photochemistry, *Q. J. R. Meteorol. Soc.*, **103**, 319, 1977.

Hasebe, F., Interannual variations of global total ozone revealed from Nimbus 4 BUV and ground-based observations, *J. Geophys. Res.*, **88**, 6819, 1983.

Heath, D. F., A. J. Krueger, H. A. Roeder and B. D. Henderson, The solar backscatter ultraviolet and total ozone mapping spectrometer (SBUV/TOMS) for Nimbus G, *Opt. Eng.*, **14**, 323, 1975.

Heath, D. F. and M. P. Thekaekara, The solar spectrum between 1200 and 3000 Åin *The solar output and its variation*, Colorado University Press, Colorado, U.S.A., 1977.

Heath, D. F., A. J. Krueger and P. J. Crutzen, Solar proton event: influence on stratospheric ozone, *Science*, **197**, 886, 1977.

Heath, D. F., Non-seasonal changes in total column ozone from satellite observations, *Nature*, **332**, 219, 1988.

Herman, J. R., R. Hudson, R. McPeters, R. Stolarski, Z. Ahmad, X-Y. Gu S. Taylor and C. Wellemeyer, A new self-calibration method applied to TOMS/SBUV backscattered ultraviolet data to determine long term global ozone change, *J. Geophys. Res.*, **96**, 7531, 1991.

Hilsenrath E. and B. M. Schlesinger, Total ozone seasonal and interannual variations derived from the 7 year Nimbus 4 BUV data set, *J. Geophys. Res.*, **86**, 12087, 1981.

Hines, C. O., A possible mechanism for the production of sun-weather correlations, *J. Atmos. Sci.*, **31**, 589, 1974.

Hirman, J. W., G. R. Heckman, M. S. Greer and J. B. Smith, Solar and geomagnetic activity during cycle 21 and implications for cycle 22, *Eos Trans. AGU*, **69**, 962, 1988.

Hofmann, D. J., J. W. Harder, J. M. Rosen, J. V. Hereford and J. R. Carpenter, Ozone profile measurements at McMurdo station, Antarctica, during the spring of 1987, *J. Geophys. Res.*, **94**, 16527, 1989.

Holton, J. R. and R. S. Lindzen, An updated theory for the quasi-biennial cycle of the tropical stratosphere, *J. Atmos. Sci.*, **29**, 1076, 1972.

Holton, J. R., *An introduction to dynamic meteorology*, Harcourt Brace Jovanovich, London, United Knigdom, 1992.

Hood, L. L. and J. L. Jirikowic, Stratospheric dynamical effects of solar ultraviolet variations: evidence from zonal mean ozone and temperature data, *J. Geophys. Res.*, **96**, 7565, 1991.

Hood L. L. and J. P. McCormack, Components of interannual ozone change based on Nimbus 7 TOMS data, *Geophys. Res. Lett.*, **19**, 2309, 1992.

Humble, J. E., M. L. Duldig, D. F. Smart and M. A. Shea, Detection of 0.5-15 GeV solar protons on 29 September 1989 at Australian stations, *Geophys. Res. Lett.*, **18**, 737, 1991.

Isaksen, I. S. A., B. Rognerad, F. Stordal, M. T. Coffey and W. G. Mankin, Studies of Arctic stratospheric ozone in a 2-D model including some effects of zonal asymmetries, *Geophys. Res. Lett.*, **17**, 557, 1990.

Jackman, C, H. and R. D. McPeters, The response of ozone to solar proton events during solar cycle 21: a theoretical interpretation, *J. Geophys. Res.*, **90**, 7955, 1985.

Jackman, C. H., P. D. Guthrie and J. A. Kaye, An intercomparison of nitrogen containing species in Nimbus 7 LIMS and SAMS data, *J. Geophys. Res.*, **92**, 995, 1987.

Jackman, C. H. and P. E. Meade, Effect of solar proton events in 1978 and 1979 on the odd nitrogen abundance in the middle atmosphere, *J. Geophys. Res.*, **93**, 7084, 1988.

Jackman, C. H., A. R. Douglass, R. B. Rood, R. D. McPeters and P. E. Meade, Effect of solar proton events on the middle atmosphere during the past two solar cycles as computed using a two-dimensional model, *J. Geophys. Res.*, **95**, 7417, 1990.

Jackman, C. H., Response of the middle atmosphere to solar proton events, *XX General Assembly of IUGG, IAGA symposium GAM 2.8*, Vienna, Austria, 1991.

Jackman, C. H., J. E. Nielsen, D. J. Allen, M. C. Cerniglia, R. D. McPeters, A. R. Douglass and R. B. Rood, The effects of the October 1989 solar proton events on the stratosphere as computed using a three-dimensional model, *Geophys. Res. Lett.*, **20**, 459, 1993.

Johnston, H. S., Reduction of stratospheric ozone by nitrogen oxide catalysts from SST exhaust, *Science*, **173**, 517, 1971.

Kanzawa, H. and Kawaguchi, S., Large stratospheric sudden warming in Antarctic late winter and shallow ozone hole in 1988, *Geophys. Res. Lett.*, **17**, 77, 1990.

Keckhut, P. and M. L. Chanin, Middle atmosphere response to the 27-day solar rotation as observed by lidar, *Geophys. Res. Lett.*, **19**, 809, 1992.

Keys, J. G. and P. V. Johnston, Stratospheric NO_2 and O_3 in Antarctica: dynamic and chemically controlled variations, *Geophys. Res. Lett.*, **13**, 1260, 1986.

King, J. H., Solar proton fluences for 1977-1983 space missions, *J. Spacecr. Rockets*, **11**, 401, 1974.

Klenk, K. F., P. K. Bhartia, A. J. Fleig, V. G. Kaveeshvar, R. D. McPeters and P. M. Smith, Total ozone determination from the Backscatter UV experiment (BUV), *J. Appl. Meteorol.*, **21**, 1672, 1982.

Ko, M. K. W., K. K. Tung, D. K. Weisenstein and N. D. Sze, A zonal mean model of stratospheric tracer transport in isentropic coordinates: numerical simulations for nitrous oxide and nitric acid, *J. Geophys. Res.*, **90**, 2313, 1985.

Kodama, M., T. Kohno and H. Kanzawa, Stratospheric sudden cooling after solar proton event over Syowa station, Antarctica, *J. Geomag. Geoelectr.*, **44**, 361, 1992.

Krueger, A. J., The global distribution of total ozone: TOMS satellite measurements, *Planet. Space Sci.*, **37**, 1555, 1989.

Krüger, B. C., Observations of polar stratospheric clouds in the Arctic winter 1989 at 79°N, *Geophys. Res. Lett.*, **17**, 365, 1990.

Labitzke, K. and H. van Loon, Associations between the 11 year solar cycle, the QBO and the atmosphere. Part I: the troposphere and stratosphere in the northern hemisphere winter, *J. Atmos. Terr. Phys.*, **50**, 197, 1988.

Lait, L. R., M. R. Schoeberl and P. A. Newman, Quasi-biennial modulation of the Antarctic ozone depletion, *J. Geophys. Res.*, **94**, 11559, 1989.

Lenoble, J., *Standard procedures to compute atmospheric radiative transfer in a scattering atmosphere*, I.A.M.A.P., NCAR, Boulder, U.S.A, 1977.

Lindzen, R. S. and J. R. Holton, A theory of the quasi-biennial oscillation, *J. Atmos. Sci.*, **22**, 341, 1968.

London, J. and S. Oltmans, Further studies of ozone and sunspots, *Pure and Appl. Geophys.*, **106-108**, 1302, 1973.

MacDowall, J., Distribution of atmospheric ozone: a preliminary analysis of some International Geophysical Year observations, *Nature*, **187**, 382, 1960.

Maeda, K. and D. F. Heath, Stratospheric ozone response to a solar proton event: hemispheric asymmetries, *Pure Appl. Geophys.*, **119**, 1, 1980/81.

Maeda, K., D. F. Heath and T. Aruga, North-south asymmetries of solar particle events in upper stratospheric ozone, *Planet. Space Sci.*, **32**, 857, 1984.

Mahlman, J. D., H. Levy II and W. J. Moxim, Three-dimensional tracer structure and behaviour as simulated in two ozone precursor experiments, *J. Atmos. Sci.*, **37**, 655, 1980.

Mahlman, J. D. and S. B. Fels, Antarctic ozone decreases: a dynamical cause?, *Geophys. Res. Lett.*, **13**, 1316, 1986.

Mateer, C. L., D. F. Heath and A. J. Krueger, Estimation of total ozone from satellite measurements of backscattered ultraviolet earth radiance, *J. Atmos. Sci.*, **28**, 1307, 1971.

Matsuno, T. and K. Nakamura, The Eulerian and Lagrangian mean meridional circulations in the stratosphere at the time of a sudden warming, *J. Atmos. Sci.*, **36**, 640, 1979.

McCormick, M. P., P. Hamill, T. J. Pepin, W. P. Chu, T. J. Swissler and L. R. McMaster, Satellite studies of the stratospheric aerosol, *Bull. Am. met. Soc.*, **60**, 1038, 1979.

McCormick, M. P. and C. R. Trepte, SAM II measurements of Antarctic PSC's and aerosols, *Geophys. Res. Lett.*, **13**, 1276, 1986.

McCormick, M. P., J. M. Zawodny, R. E. Veiga, J. C. Larsen and P. H. Wang, An overview of SAGE I and II ozone measurements, *Planet. Space Sci.*, **37**, 1567, 1989.

McElroy, M. B., R. J. Salawitch, S. C. Wofsy and J. A. Logan, Reduction of Antarctic ozone due to synergistic interactions of chlorine and bromine, *Nature*, **321**, 759, 1986.

McPeters, R. D., C. H. Jackman and E. G. Stassinopoulos, Observations of ozone depletion associated with solar proton events, *J. Geophys. Res.*, **86**, 12071, 1981.

McPeters, R. D., D. F. Heath and P. K. Bhartia, Average ozone profiles for 1979 from the Nimbus 7 SBUV instrument, *J. Geophys. Res.*, **89**, 5199, 1984.

McPeters, R. D. and C. H. Jackman, The response of ozone to solar proton events during solar cycle 21: the observations, *J. Geophys. Res.*, **90**, 7945, 1985.

McPeters, R. D., A nitric oxide increase observed following the July 1982 solar proton event, *Geophys. Res. Lett.*, **13**, 667, 1986.

McPeters, R. D. and W. D. Komhyr, Long-term changes in SBUV/TOMS relative to World Primary Standard Dobson Spectrometer 83, *J. Geophys. Res.*, **96**, 2987, 1991.

Meier, R. R., D. E. Anderson Jr. and M. Nicolet, Radiation field in the troposphere and stratosphere from 240 to 1000 nm. I. General analysis, *Planet. Space Sci.*, **30**, 923, 1982.

Miller, A. J., A review of satellite observations of atmospheric ozone, *Planet. Space Sci.*, **37**, 1539, 1989.

Molina, L. T. and M. J. Molina, Production Cl_2O_2 from the self-reaction of the ClO radical, *J. Phys. Chem.*, **91**, 433, 1987.

Molina, M. J. and F. S. Rowland, Stratospheric sink for chlorofluoromethanes: chlorine atom catalysed destruction of ozone, *Nature*, **249**, 810, 1974.

Molina, M. J., T. -L. Tso, L. T. Molina and F. C. -Y. Wang, Antarctic stratospheric chemistry of chlorine nitrate, hydrogen chloride, and ice: release of active chlorine, *Science*, **238**, 1253, 1987.

Molina, M. J., Heterogeneous Chemical Processes in Ozone Depletion in *Ozone Depletion, Greenhouse Gases and Climate Change*, Proceedings of a Joint Symposium by the Board on Atmospheric Sciences and Climate and the Committee on Global Change, National Academy Press, Washington D. C., U.S.A., 1988.

Murgatroyd, R. J. and F. Singleton, Possible meridional circulations in the stratosphere and mesosphere, *Q. J. R. Meteorol. Soc.*, **87**, 125, 1961.

Nagatani, R. M., A. J. Miller, M. E. Gelman and P. A. Newman, A comparison of Arctic lower stratospheric winter temperatures for 1988-89 with temperatures since 1964, *Geophys. Res. Lett.*, **17**, 333, 1990.

NASA/WMO, Ozone Trends Panel Report, An assessment report, *WMO Rep.* **18**, Vol. I, II, Geneva, Switzerland, 1988.

Neuber, R. and B. C. Krüger, The stratospheric ozone layer above Spitsbergen in winter 1989, *Geophys. Res. Lett.*, **17**, 321, 1990.

Newman, P. A., R. S. Stolarski, M. R. Schoeberl, L. R. Lait and A. J. Krueger, Total ozone during the 88-89 northern hemisphere winter, *Geophys. Res. Lett.*, **17**, 317, 1990.

Nicolet, M., Aeronomic reactions of hydrogen and ozone in *Mesospheric models and related experiments*, D. Reidel, Dordrecht, Holland, 1971.

Nicolet, M. and W. Peetermans, The production of nitric oxide in the stratosphere by oxidation of nitrous oxide, *Annls. Geophys.*, **28**, 751, 1972.

Nicolet, M., On the production pf nitric oxide by cosmic rays in the mesosphere and stratosphere, *Planet. Space Sci.*, **23**, 637, 1975.

Nimbus 7 users' guide, prepared by the Landsat/Nimbus Project, Goddard Space Flight Center, Edited by C. R. Madrid, Beltsville, Maryland, U.S.A., 1978.

Orsini, N. and J. E. Frederick, Solar disturbances and mesospheric odd nitrogen, *J. Atmos. Terr. Phys.*, **44**, 489, 1982.

Paetzold, H. K., The influence of solar activity on the stratospheric ozone layer, *Pure and Appl. Geophys.*, **106-108**, 1308, 1973.

Penner, J. E. and J. S. Chang, The relation between atmospheric trace species and variabilities and solar uv variability, *J. Geophys. Res.*, **85**, 5523, 1980.

Perliski, L. M. and J. London, Satellite observed long-term averaged seasonal and spatial ozone variations in the stratosphere, *Planet. Space Sci.*, **37**, 1509, 1989.

Perliski, L. M., S. Solomon and J. London, On the interpretation of seasonal variations of stratospheric ozone, *Planet. Space Sci.*, **37**, 1527, 1989.

Peter, Th., C. Brühl, C. and P. J. Crutzen, Increase in the PSC-formation probability caused by high flying aircraft, *Geophys. Res. Lett.*, **18**, 1465, 1991.

Planet, W. G., D. S. Crosby, J. H. Leinesch and M. L. Hill, Determination of total ozone amount from TIROS radiance measurements, *J. Climate appl. Meteorol.*, **23**, 308, 1984.

Podolske, J. R., M. Loewenstein, S. E. Strahan and K. R. Chan, Stratospheric nitrous oxide distribution in the southern hemisphere, *J. Geophys. Res.*, **94**, 16767, 1989.

Pommereau, J. P., F. Goutail, H. LeTexier and T. S. Jorgensen, Stratospheric ozone and nitrogen dioxide monitoring at southern and northern high latitudes,

Our changing atmosphere, Proceedings of the 28th Liège International Astrophysical Coloquium, June 26-30, 1989, Ed. by P. Crutzen, J. -C. Gérard and R. Zander, Universite de Liège, Liège, Belgium, 1989.

Poole, L. R., S. Solomon, M. P. McCormick and M. C. Pitts, The interannual variability of polar stratospheric clouds and related parameters in Antarctica during September and October, *Geophys. Res. Lett.*, **16**, 1157, 1989.

Porter, H. S., C. H. Jackman and A. E. S. Green, Efficiencies for production of atomic nitrogen and oxygen by relativistic proton impact in air, *J. Chem. Phys.*, **65**, 154, 1976.

Press, W. H., B. P. Flannery, S. A. Teukolsky and W. T. Vetterling, *Numerical recipes: the art of scientific computing*, Cambridge University Press, Cambridge, United Kingdom, 1987.

Proffitt, M. H., J. A. Powell, A. F. Tuck, D. W. Fahey, K. K. Kelly, A. J. Krueger, M. R. Schoeberl, B. L. Gary, J. J. Margitan, K. R. Chan, M. Loewenstein and J. R. Podolske, A chemical definition of the boundary of the Antarctic ozone hole, *J. Geophys. Res.*, **94**, 11437, 1989 a.

Proffitt, M. H., K. K. Kelly, J. A. Powell, B. L. Gray, M. Loewenstein, J. R. Podolske, S. E. Strahan and K. R. Chan, Evidence for diabatic cooling and poleward transport within and around the 1987 Antarctic ozone hole, *J. Geophys. Res.*, **94**, 16797, 1989 b.

Randel, W., Ideas flow on Antarctic vortex, *Nature*, **364**, 105, 1993.

Ranta, H., A. Ranta, S. M. Yousef, J. Burns and P. Staunings, D-region observations of polar cap absorption events during the EISCAT operation in 1981-1989, *J. Atmos. Terr. Phys.*, **55**, 751, 1993.

Rapp, D., P. Englander-Golden and D. D. Briglia, Cross-sections for dissociative ionization of molecules by electron impact, *J. Chem. Phys.*, **42**, 4081, 1965.

Reagan, J. B., R. E. Meyerott, R. W. Nightingale, R. C. Gunton, R. G. Johnson, J. E. Evans, W. L. Imhof, D. F. Heath and A. J. Krueger, Effects of the August 1972 solar particle events on stratospheric ozone, *J. Geophys. Res.*, **86**, 1473, 1981.

Reed, R. J., W. J. Campbell, L. A. Rasmusson and D. G. Rodgers, Evidence of a downward propagating annual wind reversal in the equatorial stratosphere, *J. Geophys. Res.*, **66**, 813, 1961.

Reeves, G. D., T. E. Clayton, S. P. Gray and R. D. Belian, The great solar energetic particle events of 1989 observed from geosynchronous orbit, *J. Geophys. Res.*, **97**, 6219, 1992.

Reid, G. C., S. Solomon and R. R. Garcia, Response of the middle atmosphere to the solar proton events of August-December 1989, *Geophys. Res. Lett.*, **18**, 1019, 1991.

Reinsel, G. C., G. C. Tiao, S. K. An, M. Pugh, S. Basu, J. J. DeLuisi, C. L. Mateer, A. J. Miller, P. S. Connell and D. J. Wuebbles, An analysis of the 7-year record of SBUV satellite ozone data: global profile features and trends in total ozone, *J. Geophys. Res.*, **93**, 1689, 1988.

Rodgers, C. D., Retrieval of atmospheric temperature and composition and remote measurements of thermal radiation, *Rev. Geophys. Space Phys.*, **14**, 609, 1976.

Ropelewski, C. F. and P. D. Jones, An extension of the Tahiti-Darwin southern oscillation index, *Mon. Weather Rev.*, **115**, 2161, 1987.

Rottman, G. J., 27-day variations observed in solar ultraviolet (120-300 nm), *Planet. Space Sci.*, **31**, 1001, 1983.

Rusch, D. W., J. -C. Gérard, S. Solomon, P. J. Crutzen and G. C. Reid, The effect of particle precipitation events on the neutral and ion chemistry of the middle atmosphere. I. Odd nitrogen, *Planet. Space Sci.*, **29**, 767, 1981.

Rusch, D. W., G. H. Mount, C. A. Barth, R. J. Thomas and M. T. Callan, Solar Mesosphere Explorer ultraviolet spectrometer: measurements of ozone in the 1.0 to 0.1 mb region, *J. Geophys. Res.*, **89**, 11677, 1984.

Schneider, U. and C. D. Schonwiese, Some statistical characteristics of El Niño-southern oscillation and north Atlantic oscillation indices, *Atmosphera*, 1989.

Schneider, W. H., P. K. Bhartia, K. F. Klenk and C. L. Mateer, *An optimum statistical technique for ozone profile retrieval from backscattered ultraviolet radiances*, Paper presented at fourth conference on atmospheric radiation, Am. Met. Soc., Toronto, Canada, 16-18 June 1981.

Schoeberl, M. R., Stratospheric warmings: observations and theory, *Rev. Geophys.*, **16**, 521, 1978.

Schoeberl, M. R. and A. J. Krueger, Medium scale disturbances in total ozone during southern hemisphere summer, *Bull. Am. Meteorol. Soc.*, **64**, 1358, 1983.

Schoeberl, M. R., R. S. Stolarski and A. J. Krueger, The 1988 Antarctic ozone depletion: comparison with previous year depletions, *Geophys. Res. Lett.*, **16**, 377, 1989.

Schoeberl, M. R. and D. L. Hartmann, The dynamics of the stratospheric polar vortex and its relation to springtime ozone depletions, *Science*, **251**, 46, 1991.

Schuster, G. S., R. B. Rood and M. R. Schoeberl, Quasi-biennial and interannual variability in high resolution ozone data (TOMS), in *Ozone in the atmosphere*,

edited by R. D. Bojkov and P. Fabian, A. Deepak Publishing, Hampton, Virginia, U.S.A., 1989.

Schwentek, H., Regular and irregular behaviour of the winter anomaly in ionospheric absorption, *J. Atmos. Terr. Phys.*, **33**, 1647, 1971.

Shine, K. P., On the modelled thermal response of the Antarctic stratosphere to a depletion of ozone, *Geophys. Res. Lett.*, **13**, 1331, 1986.

Shiotani, M., Annual, quasi-biennial and El Niño-southern oscillation (ENSO) time-scale variations in equatorial total ozone, *J. Geophys. Res.*, **97**, 7625, 1992.

Shumilov, O. I., K. Henriksen, O. M. Raspopov and E. A. Kasatkina, Arctic ozone abundance and solar proton events, *Geophys. Res. Lett.*, **19**, 1647, 1992.

Simons, J. W.,, R. J. Paur, H. A. Webster III and E. J. Bair, Ozone ultraviolet photolysis. IV. The ultraviolet spectrum, *J. Chem. Phys.*, **59**, 1203, 1973.

Solar Geophysical Data, comprehensive report, **487, (Part II)**, 29, 1985.

Solar Geophysical Data, explanation of data reports, **515 (supplement)**, 35, 1987.

Solar Geophysical Data comprehensive report, **541 (Part II)**, 11, 1989.

Solar Geophysical Data comprehensive report, **546 (Part II)**, 21, 1990, b.

Solar Geophysical Data comprehensive report, **547 (Part II)**, 140, 1990, a.

Solar Geophysical Data comprehensive report, **548 (Part II)**, 19, 1990, c.

Solar Geophysical Data comprehensive report, **549 (Part II)**, 34, 1990, d.

Solomon, S. and P. J. Crutzen, Analysis of the August 1972 solar proton event including chlorine chemistry, *J. Geophys. Res.*, **86**, 1140, 1981.

Solomon, S., P. J. Crutzen and R. G. Roble, Photochemical coupling between the thermosphere and the lower atmosphere. I. Odd nitrogen from 50 to 120 km, *J. Geophys. Res.*, **87**, 7206, 1982.

Solomon, S. and R. R. Garcia, Simulation of NO_x partitioning along isobaric parcel trajectories, *J. Geophys. Res.*, **88**, 5497, 1983.

Solomon, S. and R. R. Garcia, Transport of thermospheric NO to the upper stratosphere?, *Planet. Space Sci.*, **32**, 399, 1984.

Solomon, S., R. R. Garcia, F. S. Rowland and D. J. Wuebbles, On the depletion of Antarctic ozone, *Nature*, **321**, 755, 1986.

Solomon, S., Progress towards a quantitative understanding of Antarctic ozone depletion, *Nature*, **347**, 347, 1990.

Steele, H. M., P. Hamill, M. P. McCormick and T. J. Swissler, The formation of polar stratospheric clouds, *J. Atmos. Sci.*, **40**, 2055, 1983.

Stoker, P. H. and J. P. van Wyk, Riometer observations at Sanae ($L = 4.0$) related to solar proton events, *J. Geophys. Res.*, **98**, 17429, 1993.

Stolarski, R. S., A. J. Krueger, M. R. Schoeberl, R. D. McPeters, P. A. Newman and J. C. Alpert, Nimbus 7 satellite measurements of the springtime Antarctic decrease, *Nature*, **322**, 808, 1986.

Stolarski, R. S., M. R. Schoeberl, P. A. Newman, R. D. McPeters and A. J. Krueger, The 1989 Antarctic ozone hole as observed by TOMS, *Geophys. Res. Lett.*, **17**, 1267, 1990.

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

Stordal, F., I. S. A. Isaksen and K. Horntveth, A diabatic circulation two dimensional model with photochemistry: simulations of ozone and long-lived tracers with surface sources, *J. Geophys. Res.*, **90**, 5757, 1985.

Swider, W. and T. J. Keneshea, Decrease of ozone and atomic oxygen in the lower mesosphere during a PCA event, *Planet. Space Sci.*, **21**, 1969, 1973.

Thomas, R. J., C. A. Barth, D. W. Rusch and R. W. Saunders, Solar Mesosphere Explorer near-infrared spectrometer: measurements of 1.27 μm radiance and the inference of mesospheric ozone, *J. Geophys. Res.*, **89**, 9569, 1984.

Thorne, R. M., The importance of energetic particle precipitation on the chemical composition of the middle atmosphere, *Pure Appl. Geophys.*, **118**, 128, 1980.

Toon, O. B., P. Hamill, R. P. Turco and J. Pinto, Condensation of HNO_3 and HCl in the winter polar stratosphere, *Geophys. Res. Lett.*, **13**, 1284, 1986.

Toon, G. C., C. B. Framer, L. L. Lowes, P. W. Schaper, J. F. Blavier and R. H. Norton, Infrared aircraft measurements of stratospheric composition over Antarctica during September 1987, *J. Geophys. Res.*, **94**, 16571, 1989.

Toon, O. B. and R. P. Turco, Polar stratospheric clouds and ozone depletion, *Scientific American*, 40, June 1991.

Toumi, R., R. L. Jones and J. A. Pyle, Stratospheric ozone depletion by $ClONO_2$ photolysis, *Nature*, **365**, 37, 1993.

Trenberth, K. E. and W. T. K. Smith, Quasi-biennial fluctuations in sea level pressures over the northern hemisphere, *Mon. Weather Rev.*, **112**, 761, 1984.

Tuck, A. F., A comparison of one-, two-, and three-dimensional model representations of stratospheric gases, *Phil. Trans. R. Soc. Lond. A*, **290**, 9, 1979.

Tuck, A. F., Synoptic and chemical evolution of the Antarctic vortex in late winter and early spring, 1987, *J. Geophys. Res.*, **94**, 11687, 1989.

Varotsos, C., C. Helmis and C. Cartalis, Annual and semi-annual waves in ozone from SBUV vertical global ozone profiles, *Geophys. Res. Lett.*, **19**, 925, 1992.

Venkateswaran, S. V., J. G. Moore and A. J. Krueger, Determination of the vertical distribution of ozone by satellite photometry, *J. Geophys. Res.*, **66**, 1751, 1961.

Walterscheid, R. L., Solar cycle effects on the upper atmosphere: Implication for satellite drag, *J. Spacecr. Rockets*, **26**, 439, 1989.

Weeks, C. H., R. S. Cuikay and J. R. Corbin, Ozone measurements in the atmosphere during the solar proton event of 2 November 1969, *J. Atmos. Sci.*, **29**, 1138, 1972.

Weisenstein, D. K., M. K. W. Ko, J. M. Rodriguez and N. -D. Sze, Impact of heterogeneous chemistry on model calculated ozone change due to high speed civil transport aircraft, *Geophys. Res. Lett.*, **18**, 1991, 1991.

Wellemeyer, C. G., S. L. Taylor, C. J. Seftor and R. D. McPeters, TOMS profile shape error estimates at high latitude, Paper presented at Atmospheric Studies by Optical Methods, Tromsø, Norway, 1993.

Willett, H. C., The relationship of total atmospheric ozone to the sunspot cycle, *J. Geophys. Res.*, **67**, 661, 1962.

Zadorozhny, A. M., G. A. Tuchkov, V. N. Kikhtenko, J. Laštovička, J. Boška and A. Novák, Nitric oxide and lower ionosphere quantities during solar particle events of October 1989 after rocket and ground-based measurements, *J. Atmos. Terr. Phys.*, **54**, 183, 1992.

Zerefos, C. S. and P. J. Crutzen, Stratospheric thickness variations over the northern hemisphere and their possible relation to solar activity, *J. Geophys. Res.*, **80**, 5041, 1975.

Zerefos, C. S., A. F. Bais, I. C. Ziomas and R. D. Bojkov, On the relative importance of the quasi-biennial oscillation and El Niño-southern oscillation in the revised Dobson total ozone records, *J. Geophys. Res.*, **97**, 10135, 1992.