Spring 1998

Inversion and Analysis of Remotely Sensed Atmospheric Water Vapor Measurements at 940nm

Jack Cutler Larsen
Old Dominion University

INVERSION AND ANALYSIS OF REMOTELY SENSED ATMOSPHERIC WATER VAPOR MEASUREMENTS AT 940nm

by

Jack Cutler Larsen
B.S., March 1975, The Ohio State University
M.S., September 1977, George Washington University

A Dissertation submitted to the Faculty of
Old Dominion University in Partial Fulfillment of the
Requirements for the Degree of

DOCTOR OF PHILOSOPHY

MECHANICAL ENGINEERING

OLD DOMINION UNIVERSITY
May 1998

Approved by:

Surendra N. Tiwari(Director)

Robert L. Ash(Member)

William P. Chu (Member)

Gary E. Copeland(Member)

A. Sidney Roberts, Jr.(Member)

ABSTRACT

INVERSION AND ANALYSIS OF REMOTELY SENSED
ATMOSPHERIC WATER VAPOR MEASUREMENTS AT 940nm.

Jack Cutler Larsen
Old Dominion University, 1998
Director: Dr. Surendra N. Tiwari

The understanding and acceptance of remotely sensed atmospheric data depends strongly on the steps taken to characterize experiment error and validate observations through comparisons to other independent measurements. A formal error analysis of the Stratospheric Aerosol and Gas Experiment II (SAGE II) water vapor operational inversion algorithm is performed and compared to previous results. Random measurement errors were characterized by segmented least-squares profile fitting of the slant path absorption which found the error to be uncorrelated in the stratosphere with estimated variances significantly smaller than expected from 18-30 km. Estimates of null space error were developed from radiosonde hygrometers in the troposphere and from SAGE II in the stratosphere. Systematic model bias errors are significant only in the troposphere where they reach 20% at the surface. Systematic errors associated with radiative transfer modeling are similar to previous analyses.

A comparative error budget study between the operational inversion algorithm and several other algorithms was conducted with formal error analysis and by examining the error characteristics of two years' data inverted with each algorithm. Four other algorithms were considered; onion peel, Mill-Drayson, Mill-Drayson with stratospheric profile smoothing, and a sparse grid non-linear least-squares fitting method. Stratospheric random errors were largest for the onion peel due to the lack of stratospheric profile smoothing while the Mill-Drayson with smoothing was identical to the operational. The Mill-Drayson algorithm exhibited random error

reduction greater than expected from the form of the contribution function with stratospheric random errors approaching operational levels. The sparse grid contribution function was found to be relatively insensitive to grid point density and computationally intensive.

SAGE II upper tropospheric observations are compared to radiosonde climatologies and *in situ* radiosonde reports. SAGE II clear sky climatologies are shown to be half the level of the clear/cloudy sky radiosonde climatologies while correlative comparisons display nearly the same amount of bias. Much of the bias is attributed to the least sensitive hygrometers with SAGE II agreeing quite well with the most accurate and responsive hygrometer. Incorporating isentropic trajectories into the pair matching process greatly increases the number of correlative points but does not materially affect the comparisons.

ACKNOWLEDGEMENTS

I gratefully acknowledge the support, guidance, and encouragement of my advisor, Dr. Tiwari, throughout the course of this work. I also thank Dr. Chu for taking the time from his busy schedule to serve on my committee and for his continuing interest and many helpful suggestions. I would also like to thank the other members of my guidance committee, Drs. R.L. Ash, A.S. Roberts, and G.E. Copeland, for their help and advice. I also thank Dr. C.R. Trepte for generously providing the isentropic trajectory code and Dr. R.E. Veiga for critically reviewing portions of the manuscript. This work could not have been performed without the support of the Aerosol Research Branch at NASA Langley Research Center which supplied financial support and access to branch resources.

This work was supported by the NASA Langley Research Center through grant NAG-1-363 entitled "Institute for Computational and Applied Mechanics(ICAM)" and through a Graduate Student Researchers Program grant, NGT-51164, entitled "Inversion and Analysis of Remotely Sensed Water Vapor Data".

TABLE OF CONTENTS

Section Page

LIST OF TABLES

LIST OF FIGURES

NOMENCLATURE

Latin symbols

$\mathbf{b}$	model parameter vector
c	speed of light, cm/sec
$\mathbf{c}$	inverse model parameter vector
e	partial pressure of water vapor, mb
f	line shape function
g	gravitational acceleration, cm/s^2
k	Boltzmann's constant
k_λ, k_ω	absorption cross section, cm^2/molecule
l	path length
m	measure of radiatively active matter
m^g	mass of gas g
n	refractive index
q	specific humidity, grams/kilogram
r	aerosol particle radius, μm
r'	mass concentration
r_g	aerosol mode radius, μm
r^g	mixing ratio of gas g
t	time, seconds
$\mathbf{x}$	atmospheric parameter vector
$\mathbf{y}^m$	measurement vector
$\mathbf{y}^{mi}$	intermediate measurement vector
z	altitude, km
z_{th}	tangent height, km
A_λ	absorptance

$\mathbf{A_b}$	sensitivity of $\mathcal{T}$ to $\mathbf{b}$
$\mathbf{A_x}$	averaging kernel, sensitivity of $\mathcal{T}$ to $\mathbf{x}$
C	number density, molecules/cm^3
$\mathbf{C}$	correlation matrix
$\mathbf{D_c}$	sensitivity of I to $\mathbf{c}$
$\mathbf{D_b}$	sensitivity of I to $\mathbf{b}$
$\mathbf{D_y}$	contribution function, sensitivity of I to $\mathbf{y}$
E'	lower state energy, cm^{-1}
F	forward model
FF	channel filter function, μm^{-1} or cm
H^i_λ	measured irradiance, W/cm^2
I	inverse model
IA	water vapor inverse algorithm model
$\mathbf{I}$	identity matrix
$\mathbf{K_b}$	sensitivity of F to $\mathbf{b}$
$\mathbf{K_x}$	weighting function, sensitivity of F to $\mathbf{x}$
$\mathbf{L}$	pathlength matrix, km
M	molecular weight
MMR	mass mixing ratio
N	particle size distribution, number/(cm^3-μm)
P	pressure, mb
Q_λ	Mie extinction cross section, cm^2/particle
R	specific gas constant for air
R_λ	radiance, W/(cm^2-μm-sr)
RE	relative error
RH	relative humidity, per cent
S	separation model

S_λ	solar radiance, W/(cm^2-μm-sr)
$S^{g,m}$	line strength, cm^{-1}/(molecule $-$ cm^{-2})
$\mathbf{S}$	covariance matrix
SQ	sum of squared residuals
$\mathcal{T}$	transfer function
T	temperature, K
T_λ, T_i	atmospheric transmittance
U	layer number density, molecules/cm^2
V	molar volume
VMR	volume mixing ratio
V^{H_2O}	water vapor volume mixing ratio
W_λ	radiometer field of view function
Greek symbols	
α_D^g	Doppler halfwidth(HWHM), cm^{-1}
$\alpha_L^{g,m}$	Lorentzian halfwidth(HWHM), cm^{-1}
$\alpha_V^{g,m}$	Voigt halfwidth(HWHM), cm^{-1}
β	vertical temperature gradient, K/km
γ	profile curvature, (mixing ratio/km)^2
δ_λ	optical depth
$\epsilon_{\mathbf{y}}$	measurement noise vector
ζ^g	exponent, $S^{g,m}$ temperature dependence
θ	view angle
Θ	potential temperature, K
λ	wavelength, μm
$\nu^{g,m}$	exponent, $\alpha_L^{g,m}$ temperature dependence
ρ	density, g/cm^3
ρ^g	density of gas g, g/cm^3

σ	standard deviation
σ_λ	total extinction coefficient, km^{-1}
σ_g	aerosol distribution spread
ϕ	azimuthal angle
ω	wavenumber, cm^{-1}
Ω	solid angle

SECTION 1

INTRODUCTION

The state of the terrestrial climate system is determined by the transfer of energy, momentum, and mass across the boundaries of five subsystems; the atmosphere, hydrosphere, cryosphere, lithosphere and biosphere. The hydrological cycle, which describes the reservoirs of water and exchange rates amongst reservoirs, links these subsystems through a myriad of complex physical processes occurring on a broad range of time and space scales. For instance, water evaporating from the surface of a body of water moves from the hydrosphere to the atmosphere. As water vapor it will experience dynamical transport before condensation processes return it as precipitation to the hydrosphere or transfer it to the lithosphere. In the lithosphere water may contribute to the chemical weathering of rock, flow to a surface body of water thereby returning to the hydrosphere or evaporate again to the atmosphere. The reservoir of atmospheric water is a miniscule 0.001% of the global total. However, the high latent heat of vaporization enables efficient energy transport and elevates the role of atmospheric water not only within the hydrological cycle but also within the climate system where it has significant influence on subsystem energy budgets.

Within the atmospheric subsystem, water vapor is classified as a trace gas with fractional concentrations ranging from $\approx 10^{-6}$ in the stratosphere to $\approx 10^{-2}$ in the troposphere. Globally averaged water vapor, expressed as total precipitable water, is estimated to be 2.57 cm (Peixoto and Oort, 1983)[†], with about 75% of this residing in the planetary boundary layer. The primary sources of atmospheric water vapor are evaporation from surface bodies of water such as oceans, lakes and rivers, the soil

[†]The journal model followed is a compromise between the *Journal of Geophysical Research* and the *Journal of Heat Transfer*.

and vegetative transpiration. The dominant sink is precipitation. Over long time periods (decades), globally averaged evaporation balances precipitation so that the total amount of atmospheric water may be considered constant. For shorter time periods and less than global scales, imbalances between these two processes are found that provide driving forces for many interesting phenomena. Water flux across the atmospheric subsystem boundary replaces the total amount of tropospheric water in only 10 days. It is the rapid evaporative/precipitative processes and associated phase changes that directly couple the atmospheric subsystem to terrestrial subsytems.

Phase changes enable not only the transfer of water across the boundaries of the atmospheric subsystem but are also important for many processes within the atmospheric subsystem. Evaporation of water from the surface carries energy from the surface to the atmosphere in the form of latent heat. Latent heat release to the atmosphere, due to the condensation of water vapor in cloud formation, is one of the major energy sources for atmospheric circulation. Relative humidity controls the formation of clouds which in turn modulate the radiation field at both long and short wavelengths and affects the global energy budget.

Water vapor interacts directly with the atmospheric radiation field through a number of vibrational-rotational bands. The visible bands of water vapor are weak (Goody, 1964) and thus transparent to solar radiation for the most part. Somewhat stronger bands ($\rho, \sigma, \tau, \phi, \psi$, and Ω) are located between 1 and 2μm that do absorb some solar radiation. The ν_2 fundamental at 6.27μm is the most important water vapor band because it absorbs infrared radiation emitted by the surface. The ν_1 and ν_3 fundamentals, at 2.74μm and 2.66μm respectively, are of minor importance to both solar and terrestrial radiation. Because the water vapor spectrum is transparent to solar radiation and traps or absorbs terrestrial emissions, it is considered one of the primary greenhouse gases and plays a central role in maintaining Earth's climate system.

The importance of water vapor to the climate system has become apparent as concern about anthropogenic increases in greenhouse gases, such as CO_2, has grown. In a review of geophysical and biogeochemical feedbacks Lashof(1989) shows that the water vapor feedback dominates all other feedback mechanisms. The water vapor feedback mechanism is thought to work as follows. Anthropogenic increases in greenhouse gases reduce the amount of infrared radiation escaping to space. The absorbed radiation is radiated towards the surface producing a warmer climate in the troposphere. The higher surface temperature enhances evaporation from the oceans while the higher atmospheric temperatures, through the Clausius-Clapeyron equation, increase the capacity of the atmosphere to hold water. The higher levels of water in turn absorb more infrared radiation leading to even greater heating at new equilibrium levels. This positive water vapor feedback has been confirmed in computer simulations of warmer climates. In one such simulation (Del Genio et al., 1991), the absolute humidity increased 25% at 950 mb and 60% at 300 mb in response to a 4°C global increase in sea surface temperature. The strong increase in atmospheric water content due to a relatively small increase in surface temperature suggests that monitoring water vapor may be just as effective as monitoring temperature for detecting climate change (Barnett et al., 1991).

Recognition that climate change may be more detectable in water vapor than temperature has lead to examination of historical data bases for water vapor trends. The largest tropospheric database, in terms of global and seasonal coverage, is available from the radiosonde network. An early attempt (Angell et al., 1984) to extract trend information from the historical recorded at two stations in the U.S. experienced difficulty separating water vapor changes from changes in instrumentation and data reduction procedures. In another study, Hense et al.(1988) found an increase in precipitable water and water vapor in the 500/700 mb layer at equatorial Pacific stations between 1965 and 1984. Elliot et al.(1991) reported significant upward trends

in precipitable water (surface to 500 mb) for the 1973-88 time period in the region north of the equator and western North America. Gaffen et al.(1991) found a shift to higher specific humidity values in the lower troposphere, occurring during the winter of 1976/77 from analysis of 118 stations. Ross and Elliot(1996) examined 21 years of radiosonde data over North America for trends and found south of 45° N precipitable water (surface to 500 mb) increased at a rate of 0.3-0.7% yr^{-1} with smaller trends of varying sign north of this latitude circle. Lower stratospheric trends using data from a balloon-borne frost-point instrument(Oltmans and Hofmann, 1995) at Boulder, Colorado ranged from 0.34-1.03% yr^{-1} over 10-26 km. Clearly, establishing the existence of trends of 0.3-1% yr^{-1} with a high level of confidence places strong demands on observing systems in terms of stability, precision, accuracy, analysis algorithms, and consistency in data analysis procedures.

The Stratospheric Aerosol and Gas Experiment II (SAGE II) is a multiwavelength spectrometer which infers the vertical distribution of aerosols, ozone, nitrogen dioxide, and water vapor. Global observations from cloud top into the stratosphere have been routinely obtained since launch aboard the Earth Radiation Budget Satellite in October 1984. The water vapor channel is approximately 20 nm wide and centered at 935.5 nm. SAGE II is probably the largest data base of stratospheric and upper tropospheric water vapor currently available. It is ideally suited for climate change studies because of its global extent, high vertical resolution, internal self consistency and self calibrating nature.

The proposed investigation includes a quantitative characterization of the SAGE II water vapor measurements with an emphasis on retrieval algorithm refinements and error analysis followed by intercomparisons to other observational data sets. These studies will result in an improved understanding of the SAGE II water vapor data set and enhance its usefulness for global change studies.

For the past two decades the inversion algorithm most often applied to limb viewing experiments has been some variation of the onion peel algorithm. The onion peel has been applied to both emission (McKee et al., 1969; Russell and Gordley, 1979; Russell et al., 1984) and absorption (Chu and McCormick, 1979; Chu et al., 1993) experiments. In one variation of the onion peel the iterative process starts with the bottom most tangent layer and proceeds upwards (Mill and Drayson, 1978). As part of the initial validation process this algorithm was applied to the SAGE II water vapor measurements (Pruvost et al., 1993) and yielded profiles that agreed within the estimated uncertainties to profiles from the operational algorithms. Analysis of high spectral resolution limb observations from interferometric instruments is best performed with least squares methods. Early applications were limited to spectral fitting of a single tangent layer (Rinsland et al., 1983) but with todays computer capabilities simultaneous spectral fitting over multiple tangent layers are now possible (Carlotti,1988; Rinsland et al., 1992).

Five inversion algorithms suitable for inverting the SAGE II water vapor channel data are studied in this work. A basic form of the onion peel algorithm provides a comparison baseline for the other four algorithms. To understand the error characteristics of the SAGE II archived water vapor data, the modified onion peel of Chu et al.(1993) employed in the operational data processing is a requisite choice for study. The Mill and Drayson(1978) algorithm (or Mill-Drayson) was selected for its ability to reduce random measurement errors without smoothing the measurements. The fourth algorithm is a variation of the Mill-Drayson that incorporates the stratospheric mixing ratio smoothing used by Chu et al.(1993). The last algorithm studied produces stratospheric smoothing by solving an over-determined equation set with a non-linear least squares solution method.

The SAGE II water vapor error analysis (Chu et al., 1993) was developed as an extension of the technique applied to the ozone and aerosol measurements.

Systematic errors were addressed through a sensitivity analysis of each error term. Random errors are estimated from the standard error equation applied to the occultation geometry with adjustments for the effects of vertical smoothing. Concurrent with this work, a new approach to error analysis was under development. The formal error analysis of Rodgers(1990) provides a structured approach to error analysis applicable to all remote sensing experiments. The analysis accounts for interlevel error correlations with error covariance matrices, provides for estimates of null space error, and accounts for differences between the forward and inverse model. Since its introduction, the formal error analysis has become the standard technique for remote sensing error studies and widely used in inversion algorithm comparisons.

This study follows a sequence typical of most remote sensing projects: algorithm development, application, error analysis, and data validation. Section 2 presents much of the introductory material related to atmospheric remote sensing, the atmosphere and water vapor in particular, and the formal error analysis to be applied to the measurements. Section 3 contains background information on the SAGE II experiment, outlines the inversion procedures applied to the water free channels to yield vertical profiles of aerosol, ozone and nitrogen dioxide, and presents sample "raw" data in the form of slant path optical depths. In Sec. 4 the formal error analysis is applied to each of the inversion algorithms to develop an estimate of the total experimental uncertainty and then each algorithm is used to invert two years of data (1987,1990). The character of the inverted profiles is analyzed with respect to that expected from the error analysis and from inversion of synthetic observations with known errors. Section 5 is concerned with validating the inversion results of Sec. 4. The SAGE II upper tropospheric observations are compared to radiosonde hygrometer measurements. A technique to increase the number and quality of correlative profile pairs by incorporating information from air mass isentropic trajectories is

developed and applied to correlative radiosonde observations. Section 6 summarizes the results and conclusions of this work.

SECTION 2

ATMOSPHERIC REMOTE SENSING

Atmospheric remote sensing is based on the idea that radiatively active constituents produce distinctive spectral signatures that are functions of constituent distributions. To invert measured signatures into constituent distributions requires an accurate physical model of atmospheric radiative processes and a solution technique or inversion algorithm. This section provides background information on atmospheric remote sensing, characteristics of the water vapor distribution, and general features of the atmosphere.

Section 2.1 presents the radiative transfer governing equations describing the absorption and scattering of solar radiation by the optically active atmospheric constituents at visible and near infrared wavelengths. The solar occultation measurement geometry is introduced in Sec. 2.2 and a simple onion peel inversion algorithm that is the starting point for several of the methods studied in Sec. 4 is examined in order to illustrate some basic remote sensing concepts and definitions. In Sec. 2.3 a brief description of atmospheric thermal structure and chemical composition is given along with commonly used terms and definitions. Next, background information on the distribution of water vapor is presented in Sec. 2.4 for the stratosphere and upper troposphere along with a discussion of chemical and dynamical time scales. The remainder of the section, Sec. 2.5, is concerned with the derivation and philosophy, in general terms, of the formal error analysis to be applied to the SAGE II observations.

2.1 Radiative Transfer in Earth's Atmosphere

Radiation in the Earth's atmosphere interacts with molecules and aerosol particles through the processes of absorption, emission, and scattering. For solar occultation experiments operating at visible and near infrared wavelengths atmospheric emission is negligible relative to the transmitted solar radiation. This is easily shown with Planck's law for blackbody radiation where the ratio of solar(6000 K) to atmospheric(250 K) emitted radiation is $\approx 9.8 \cdot 10^{23}$ at 1000 nm and increases significantly at shorter wavelengths. Under conditions of severe atmospheric attenuation, due perhaps to optically thick clouds, heavy aerosol loading from a recent volcanic eruption or an extremely long slant path at grazing incidence, the direct solar beam might become comparable to the thermally emitted radiation. For all practical purposes however, under these conditions a measured signal would likely have fallen to levels below the instrument noise long before this condition was reached for an instrument designed to respond to the full solar radiance.

Radiation scattered into a direction along the direct solar beam from the ambient radiation is also assumed to be negligible compared to the direct transmitted solar radiation. Multiple scattering in a spherical atmosphere at grazing angles of incidence is not a trivial problem and has been attempted by only a few workers(for a review of earlier methods see Lenoble(1985) or Herman et al.(1995) for a more recent approach). Given these two simplifying assumptions related to thermal emission and multiple scattering, it is now possible to describe the atmospheric attenuation of solar radiation with the Bougeut-Lambert extinction law.

The Bougeut-Lambert law states that the incremental change of radiance (specific intensity) due to extinction processes along an infinitesimal path is linearly proportional to the radiance and incremental amount of matter along the path, *i.e.*,

$$dR_\lambda = -\sigma_\lambda R_\lambda dm \tag{2.1}$$

where R_λ is the radiance (specific intensity) at wavelength λ, σ_λ is the total extinction coefficient and dm is the amount of matter encountered by the radiance. The variable σ_λ and dm may be combined into a non dimensional quantity referred to as the incremental optical depth

$$d\delta_\lambda = \sigma_\lambda dm \tag{2.2}$$

and the Bougeut-Lambert law becomes

$$\frac{dR_\lambda}{d\delta_\lambda} = -R_\lambda \tag{2.3}$$

which is the equation of transfer without the source function and is consistent with the simplifying assumptions. Note that in this form R_λ is no longer dependent on specific physical dimensions but is dependent on the amount of interaction between the radiation and medium as expressed by the optical depth. Several choices are possible for the units of σ_λ and dm. Choosing the path length dl to represent dm makes σ_λ the volume extinction coefficient with units of inverse length. Another common choice of units is Cdl for dm where C is the number density and σ_λ becomes the molecular extinction coefficient with units of length squared.

The solution to the equation of transfer is

$$R_\lambda(l) = R_\lambda(0)e^{-\delta_\lambda(l)} \tag{2.4}$$

with the optical depth given by

$$\delta_\lambda(l) = \int_0^l \sigma_\lambda(l')dl' \tag{2.5}$$

Radiance $R_\lambda(l)$ is the portion of the initial radiance $R_\lambda(0)$ that reaches distance l without interacting with the medium and is a simple exponentially decreasing function of the optical depth. The fraction of unattenuated radiance to initial radiance is defined as the transmittance, $T_\lambda(l)$

$$T_\lambda(l) = e^{-\delta_\lambda(l)} \tag{2.6}$$

Another commonly used quantity based on T_λ is the absorptance

$$A_\lambda(l) = 1 - T_\lambda(l) \tag{2.7}$$

Absorptance is the fraction of attenuated(removed) radiance to the initial radiance.

In the visible and near infrared wavelength regions, extinction may be caused by aerosol scattering, molecular scattering (or Rayleigh scattering as it is commonly known), and molecular absorption. Thus, the total extinction coefficient is expressible as

$$\sigma_\lambda(l) = \sigma_\lambda^{\mathrm{Aer}}(l) + \sigma_\lambda^{\mathrm{Ray}}(l) + \sum_g \sigma_\lambda^g(l) \tag{2.8}$$

where $\sigma_\lambda^{\mathrm{Aer}}$ is the aerosol extinction coefficient, $\sigma_\lambda^{\mathrm{Ray}}(l)$ accounts for Rayleigh extinction, and $\sum_g \sigma_\lambda^g$ is the total extinction from all atmospheric molecules g. The Rayleigh and gaseous extinctions are given by

$$\sigma_\lambda^{\mathrm{Ray}}(l) = C(l)\sigma_\lambda^{\mathrm{Mol}}(l) \tag{2.9}$$

$$\sigma_\lambda^g(l) = C^g(l)k_\lambda^g(l) \tag{2.10}$$

where C is the number density of air, $\sigma_\lambda^{\mathrm{Mol}}(l)$ is the scattering cross-section of air, and C^g is the number density of gas g with absorption cross-section k_λ^g. Aerosol and molecular scattering coefficients are smoothly varying functions of wavelength throughout this region while molecular absorption is limited to particular absorption bands exhibiting a range of strengths. The summation in the equation above accounts for overlapping absorption features produced by multiple atmospheric species. Of the three processes contributing to the total extinction, molecular scattering is perhaps the most self-contained in its final form as it can be calculated to a high degree of accuracy after specifying the number density of air. The aerosol scattering term requires knowledge of aerosol composition, size distribution, and physical structure while molecular absorption accuracy depends for the most part on the ancillary line parameter database quality.

Molecular scattering describes the interaction of radiation and matter for the case when the size of the scattering particle is much smaller than the wavelength of radiation. From the viewpoint of classical electromagnetic theory, an incident electromagnetic wave impinging on a particle with properties different than the surrounding medium suffers a loss of amplitude or energy which is reradiated into an outward traveling spherical wave at the same wavelength. The Rayleigh scattering cross-section, developed for the primary scattering of solar radiation by assuming natural light (Goody, 1964), is expressed as

$$\sigma_{\lambda}^{\mathrm{Mol}}(l) = \frac{32\pi^3(n(l)-1)^2}{3\lambda_0^4 C^2(l)} \frac{(6+3\triangle)}{(6-7\triangle)} \tag{2.11}$$

where λ_0 is the vacuum wavelength, n is the refractive index of air and $\triangle$ is the depolarization factor due to non-spherical molecules which has an effective mean of 0.035 for dry air. The scattering cross-section is inversely proportional to both the fourth power of wavelength and the square of the number density C. Since the refractive index term, $n-1$, is proportional to the number density, C, the cross-section is independent of temperature or pressure and is a strong function of only wavelength in the atmosphere.

The standard optical model of aerosol extinction is based on the Mie theory of scattering by homogeneous spheres. Full development of Mie theory may be found in Van De Hulst (1957) and Bohren and Huffman (1983) and enters the expression for $\sigma_{\lambda}^{\mathrm{Aer}}$

$$\sigma_{\lambda}^{\mathrm{Aer}}(l) = \int_0^{\infty} Q_{\lambda}(n,r) N(r,l) dr \tag{2.12}$$

through $Q_{\lambda}(n,r)$, the Mie extinction cross-section. The aerosol particle size distribution, N, describes the particle number density as a function of particle radius r and n is the aerosol refractive index. Software packages for calculating Q_{λ} have been available (Wiscombe, 1979; Dave, 1968) to the scientific community for some time.

Aerosol particles are found globally throughout the troposphere and stratosphere. The stratospheric aerosol layer, first observed by Junge et al.(1961), is composed for the most part of sulfuric acid, water droplets (75% H_2SO_4, 25% H_2O by weight) for background conditions. The composition and size distributions are controlled by a number of processes; formation by nucleation, growth through condensation and coagulation, shrinkage by evaporation, gravitational sedimentation and dynamical transport(Turco et al., 1982). Background aerosol levels may be enhanced by volcanic injection of SO_2 and H_2S into the stratosphere that rapidly convert to sulfuric acid (Mckeen et al., 1984) followed by particle formation. The volcanic aerosols disperse globally and sedimentation from the stratosphere into the troposphere produces a net loss of particles (Junge et al., 1961; Kasten, 1968). The rate at which enhanced aerosol levels return to background conditions is often described in terms of optical depth or mass loading e-folding times (an exponential decay time) and has been observed to be about one year following the October 1974 eruption of Volcán de Fuego (McCormick et al., 1978) and 14-16 months in the northern hemisphere three years after the El Chichon eruption(Yue et al., 1991). Unlike stratospheric aerosols, tropospheric aerosols are composed of a variety of substances such as organics, sea salt, sulfates and soil dust (Toon and Pollack, 1976) and exhibit a large degree of nonuniformity in composition and size distribution. Above the planetary boundary layer a reasonable optical model includes a water soluble component representing ammonium and calcium sulfates and organic compounds and a soil dust component with proportions of 70/30% (Shettle and Fenn, 1979) or 60/40% (Toon and Pollack, 1976). Even though the soil dust particles in this model are irregularly shaped and do not satisfy the Mie spherical requirement in the strictest sense, they are normally assumed to be spherical for computational purposes. In most situations this may be an acceptable assumption, especially given the lack of better alternatives, since many of the reported particle sizes are based on measurements of scattered light.

The reported particle sizes are viewed as "effective spherical particles" in the sense that they produce scattering equivalent to irregular particles (Pinnick et al., 1973).

The particle size distribution depends on the source and evolution of individual particles within the atmosphere. A number of analytical distributions based on aerosol measurements are found in the literature (Russell et al., 1981). One distribution commonly used in the stratosphere and occasionally in the troposphere is the log-normal distribution

$$N(r) = \frac{N_0}{\sqrt{2\pi} r ln\sigma_g} e^{-\left(\frac{ln(r/r_g)}{\sqrt{2}ln\sigma_g}\right)^2} \tag{2.13}$$

The total particle number density is designated by N_0, r is the particle radius, r_g defines the mode radius, and σ_g is representative of the distribution spread. For background or nonvolcanic stratospheric conditions these parameters take the values $N_0 = 10\text{cm}^{-3}$, $r_g = 0.0725\mu\text{m}$, and $\sigma_g = 1.86$ (Pinnick et al., 1976). Segmented power law model distributions developed for the troposphere (Russell et al., 1981) break the range of particle radii into separate regions for fitting purposes. In the upper troposphere, for example, the distribution model is expressed as (Toon and Pollack, 1976)

$$N(r) = Ar^{-\eta} \tag{2.14}$$

$$\eta = 0.0, \quad r < 0.045\mu\text{m}$$

$$\eta = 3.6, \quad 0.045\mu\text{m} \leq r \leq 5\mu\text{m}$$

$$\eta = 5.6, \quad 5\mu\text{m} \leq r \leq 30\mu\text{m}$$

$$\eta = \infty, \quad r > 30\mu\text{m}$$

where A is an adjustable constant. Figure 2.1 shows the variation of $N(r)$ in the stratosphere and troposphere as a function of particle radius for the log-normal and segmented power law models. Although a large fraction of the particles have radii less than $0.1\mu\text{m}$, they are rendered optically inefficient by the Mie cross-section weighting

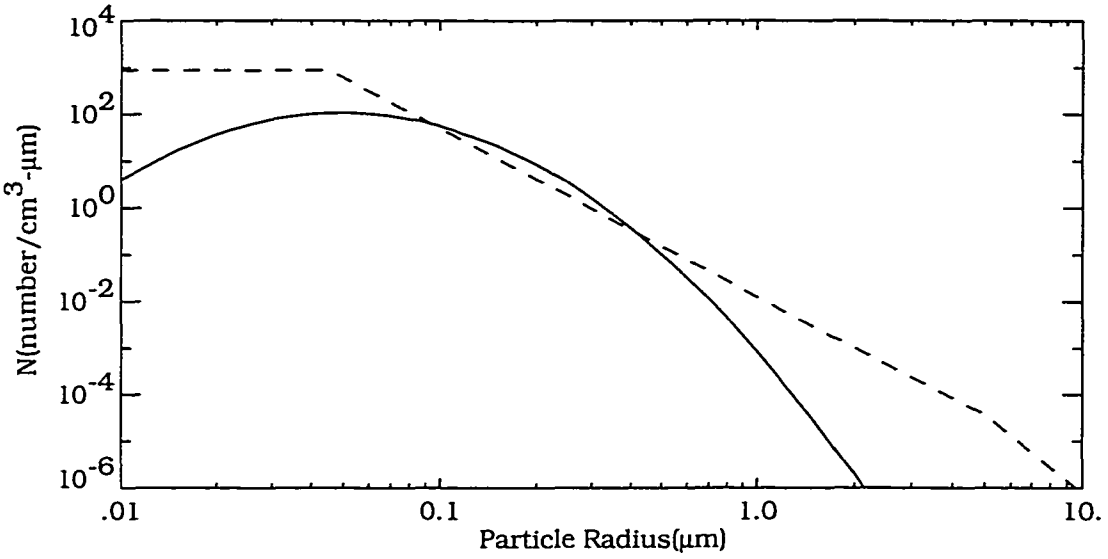


Figure 2.1: Particle size distributions representative of the nonvolcanic stratosphere obtained with the log-normal model (solid line) and upper troposphere segmented power law (dashed line). Model parameters are listed in the text. There are a total of 10 and 500 particles/cm^3 in each distribution, respectively.

in Eq. (2.12) at these radii. In fact, Russell et al.(1981) show that 80% of the 1μm extinction is produced by particles with radii between 0.1 and 0.7μm for their stratospheric log-normal distribution. Large particles contribute little because of their small number. For the upper troposphere power law distribution the radii range from approximately 0.25 to 2.5μm and the range shift relative to the stratospheric distribution reflects both the somewhat smaller number of particles from 0.1 to 0.4μm radii and a much greater number of large radii particles.

Molecular absorption describes the transfer of energy from the radiation field to the molecule whereby the internal energy is increased in either quantized steps (bound-bound transitions) or unquantized steps (bound-free transitions). The bound-bound transitions generally involve simultaneous changes of electronic, rotational, and vibrational energy that follow particular transition rules for each molecule and

appear as a loss of radiative energy at frequencies determined by the changes in molecular energy between the lower and upper states. Because energy loss occurs over a discrete range of frequencies, an absorption line of finite width results. In some situations, such as the Chappuis ozone band, many overlapping individual lines will give the appearance of continuum type absorption. The bound-free transitions produce continuum absorption since all frequencies with energy greater than the ionization potential may ionize the molecule. Both types of transitions are found in the absorption spectrum of atmospheric gases at visible and near infrared wavelengths and in some cases overlap of the two types may occur.

Continuum type absorption displays a slowly varying function of wavelength. It is typically measured in the laboratory over a range of discrete pressures(P_t) and temperatures(T_t) and reported as tabulated data, $K_\lambda(P_t, T_t)$. The absorption cross-section becomes simply

$$k_\lambda^g(l) = k_\lambda^g(P(l), T(l)) \tag{2.15}$$

for gas g, where P is atmospheric pressure and T is atmospheric temperature. First order interpolation from $K_\lambda(P_t, T_t)$ to $K_\lambda(P, T)$ is normally adequate since K_λ is a weak function of pressure and temperature. Uncertainties in K_λ follow directly from the reported measurement uncertainties. In some cases, additional uncertainties may arise when pressure and temperature dependencies are suspected but they are unverified with laboratory measurements.

Modeling absorption of bound-bound transitions is considerably more complicated than that for continua. The natural broadening due to the quantum uncertainty principle is overshadowed relative to Doppler broadening caused by molecular thermal motions and pressure or collisional broadening. For calculating atmospheric absorption, three line models are normally employed; Doppler for low pressure situations, Lorentz for high pressure, and Voigt (Tiwari, 1978; Goody, 1964; Armstrong,

1967). It can be shown that the Voigt line model reduces to the Lorentz and Doppler line models given appropriate limiting conditions and thus can be used at all pressure levels within the atmosphere.

The equations governing absorption by a single line m of gas g based on the Voigt line shape model are

$$k_{\omega}^{g,m}(P,T) = \frac{S^{g,m}(T)}{\alpha_D^g(T)} \left[\frac{ln2}{\pi}\right]^{1/2} \frac{Y}{\pi} \int_{-\infty}^{\infty} \frac{e^{-v^2}}{Y^2 + (X-v)^2} dv \tag{2.16}$$

Here, the absorption cross-section is expressed in terms of wavenumber (ω) instead of wavelength in keeping with standard spectroscopic practices. The conversion is $\omega[\mathrm{cm}^{-1}]=10^4/\lambda[\mu\mathrm{m}]$. The line strength $S^{g,m}(T)$ is calculated from a reference strength $S^{g,m}(T_0)$ at temperature T_0 through the expression

$$S^{g,m}(T) = S^{g,m}(T_0) \left[\frac{T_0}{T}\right]^{\zeta^g} exp\left\{-\frac{E'}{k}\left[\frac{1}{T} - \frac{1}{T_0}\right]\right\} \tag{2.17}$$

where k is Boltzmann's constant. The lower state energy E' is also line dependent but for clarity the g, m superscripts are not shown. The exponent ζ^g of the temperature ratio is $\approx$1 for linear molecules and $\approx$1.5 for asymmetric molecules. The Doppler halfwidth

$$\alpha_D^g(T) = \left[\frac{2kTln2}{M^g c^2}\right]^{1/2} \omega \tag{2.18}$$

is proportional to wavenumber and inversely proportional to molecular weight M^g. The quantity c is the speed of light. The expressions for X and Y are defined by

$$X = \frac{|\omega - \omega_0| ln2^{1/2}}{\alpha_D^g(T)} \tag{2.19}$$

$$Y = \frac{\alpha_L^{g,m}(P,T) ln2^{1/2}}{\alpha_D^g(T)} \tag{2.20}$$

The term X can be viewed as controlling the wavenumber dependence of the profile shape through the term $|\omega - \omega_0|$, where ω_0 is the line center wavenumber of line m, while Y controls profile shape through the ratio of Lorentz and Doppler halfwidths.

The Lorentz or pressure broadened halfwidth $\alpha_L^{g,m}(P,T)$ is given by

$$\alpha_L^{g,m}(P,T) = \alpha_L^{g,m}(P_0,T_0)\left[\frac{P}{P_0}\right]\left[\frac{T_0}{T}\right]^{\nu^{g,m}} \tag{2.21}$$

where $\alpha_L^{g,m}(P_0,T_0)$ is a reference halfwidth corresponding to pressure P_0 and temperature T_0. The exponent $\nu^{g,m}$ is 0.5 from classical theory but laboratory measurements have shown this to be line dependent.

The absorption cross-section defined by Eq. (2.16) is formulated such that

$$\int_{-\infty}^{\infty} k_\omega^{g,m}(P,T)d\omega = S^{g,m}(T) \tag{2.22}$$

Sometimes Eq. (2.16) is presented in terms of a line shape function f

$$k_\omega^{g,m}(P,T) = S^{g,m}(T) f_\omega^{g,m}(\alpha_V^{g,m}) \tag{2.23}$$

where $\alpha_V^{g,m}(P,T)$, the Voigt function halfwidth, may be approximated by

$$\alpha_V(P,T) = \frac{\alpha_L}{2} + \left[\frac{\alpha_L^2}{4} + \alpha_D^2\right]^{1/2} \tag{2.24}$$

It follows from the property of Eq. (2.22) that the line shape function is normalized to unity.

It is instructive to examine the variation of these halfwidths throughout the atmosphere. For an isothermal atmosphere at 250 K, α_D=0.0133 cm^{-1} at 1μm for water vapor. The reference halfwidth $\alpha_L(P_0,T_0)$ for a typical water vapor line in this region is $\approx$0.07 cm^{-1}, so at the surface $\alpha_L\approx$0.07 cm^{-1} (ignoring the temperature correction) while at 40 km($\approx$2.87 mb) $\alpha_L\approx 2\cdot 10^{-4}$ cm^{-1}. Corresponding values of α_V are 0.072 and 0.0134 cm^{-1}, respectively. Thus, at the surface collisional broadening dominates and the Voigt profile shape becomes Lorentzian. The situation reverses at 40 km where thermal broadening controls the line shape. Monochromatic absorption calculations resolving individual lines typically use a wavenumber increment determined by some fraction of the Voigt halfwidth. For a fixed wavenumber range,

the number of monochromatic points at which the line shape is evaluated increases as the Voigt halfwidth decreases. Since much of the total computational time is expended on repetitive line shape evaluation, monochromatic calculations become increasingly more time consuming when going from the surface to higher altitudes.

Practical applications of the Voigt line shape model depends on two issues; fast and accurate evaluation of the integral in Eq. (2.16), the Voigt line shape function, and accurate specification of the molecular line parameters in Eqs. (2.17) to (2.21).

Methods of computing the Voigt function are based for the most part on approximate expressions applicable to specific regions of the $X-Y$ computational plane. In addition to those methods cited in Tiwari(1978) may be added Drayson(1976), Hui et al.(1978), Pierluissi et al.(1977), and Humlicek(1979) among others. As noted in Drayson(1976), a Voigt algorithm providing 1 part in 10^6 accuracy may be overly accurate when one considers the accuracy of subsequent quadrature manipulations required for integrating transmittance over frequency and of even greater impact, when one considers the accuracy of the molecular line parameters. Computational speed and accuracy comparisons (Twitty et al., 1980; Klim, 1981) found that speed and accuracy depended on location within the $X-Y$ domain which prevented selection of an "optimal" algorithm applicable to all situations. One algorithm (Hui et al., 1978) provides 1 part per 10^5 accuracy with nearly uniform speed throughout the $X-Y$ plane that is, at worst, twice as slow (for moderate and large X, Y) as that of Drayson(1976) which has an accuracy of 1 part per 10^4. A more recent study incorporating several newer algorithms (Schreier, 1992) found that algorithm vectorization decreased execution time by two orders of magnitude on a vector machine for the technique of Hui et al.(1978) along with several others. Another approach, that does not directly evaluate the Voigt line shape, decomposes the Lorentz profile into four precalculated subfunctions (Clough et al., 1979) and combines them with the Doppler profile shape in an approximation to the Voigt profile (Clough et al., 1981).

The maximum line shape error in this technique is 3% for $\alpha_L/(\alpha_L + \alpha_V) = 0.05$ and is about 1% when $\alpha_L = \alpha_V$. While not as accurate as some of the previously mentioned functions it is much faster and the error in band averaged transmittances or absorptances will be considerably smaller than that quoted above.

The molecular line parameters that must be specified for a single transition are the line strength $S^{g,m}(T_0)$, the lower state energy E', the line center ω_0, the air broadened halfwidth $\alpha_L^{g,m}(P_0, T_0)$, and the exponent $\nu^{g,m}$ controlling the halfwidth temperature dependence. While numbering only five parameters, the task of compiling these parameters for all atmospheric gas transitions is daunting not to mention time consuming. The HITRAN database of molecular line parameters developed by the U.S. Air Force is the most comprehensive list currently available for atmospheric radiative transfer calculations. For the most part it is composed of data derived from laboratory observations augmented with some theoretical analyses. It has undergone a series of revisions as new information enters the scientific literature; the 1992 edition (Rothman et al., 1992), 1986 edition (Rothman et al., 1987), 1982 edition (Rothman et al., 1983a; Rothman et al., 1983b), and the 1980 edition (Rothman et al., 1981). With each revision, new transitions are incorporated and old data are replaced with updated values of higher accuracy. The current edition contains over 709,000 transitions between 0 and 20,000 cm^{-1} for 31 gaseous species and also includes cross-sectional data for selected heavy molecules of atmospheric interest with currently unresolvable transitions.

The absorption cross-section for gas g at wavenumber ω includes contributions from nearby transitions having non-negligible cross-sections at ω

$$k_\omega^g(l) = k_\omega^g(P(l), T(l)) = \sum_m k_\omega^{g,m}(P, T) \tag{2.25}$$

The point at which the cross-section becomes negligible varies from line to line and depends on the line strength, the Voigt halfwidth and $|\omega - \omega_0|$. Although a variety

of cutoff criteria based on these parameters have been used in the past, todays computer capabilities do not severely penalize overcalculations and a simple criteria based on $|\omega - \omega_0|$ is normally employed.

2.2 Measurement Geometry and Basic Inversion

In a solar occultation experiment the sun is viewed through the limb of the atmosphere during spacecraft sunset or sunrise, the measurement geometry is shown in Fig. 2.2 for a single line of sight or tangent ray between the sun and instrument. The atmospheric optical depth for tangent height z_{th} is given by a Volterra equation

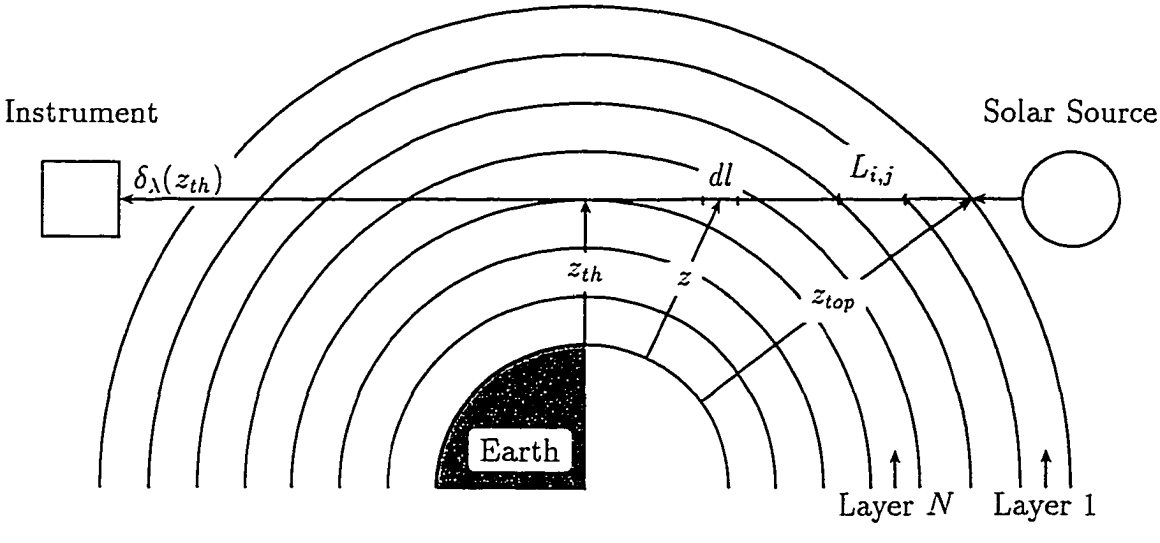


Figure 2.2: The solar occultation measurement geometry.

of the first kind

$$\delta_\lambda(z_{th}) = 2\int_{z_{th}}^{z_{top}} \sigma_\lambda(z)dl(z) \tag{2.26}$$

where dl is the incremental path length as a function of z, the radial distance from Earth's center and σ_λ is defined by Eq. (2.8). The topmost atmospheric layer,

above which σ_λ may be considered negligible, is designated as z_{top}, and z_{th}, the tangent height or tangent altitude, is defined as the altitude along the ray where z and the line of sight ray are normal to each other. It is also the point of closest approach of the line of sight ray to the surface. The assumption of symmetry in atmospheric properties relative to z_{th} accounts for the factor of 2 in Eq. (2.26). The line of sight, shown in Fig. 2.2 as a straight line, experiences refractive curvature in layers close to the surface that depends on the refractive index profile. Application of Snell's law to spherical atmospheres is relatively straight forward (Chu, 1983) and in this situation simply increases the pathlength through a layer. Refraction induced path lengthening becomes significant when the tangent ray passes through the troposphere, that is, when $z_{th} < 15$ km at mid-latitudes.

Integration of Eq. (2.26) is achieved by partitioning the atmosphere into a series of concentric spherical layers or shells as shown in Fig. 2.2. The slant path optical depth may be expressed as a sum of layer optical depths

$$\delta_\lambda(z_{th}) = 2\sum_{j=1}^{N(i)} \Delta\delta_\lambda(i,j) \tag{2.27}$$

where $\Delta\delta_\lambda(i,j)$ is the optical depth for layer j of the particular slant path specified by tangent height i. In this and following sections the index i and z_{th} will be used interchangeably to designate tangent height. The summation runs from the top layer ($j = 1$) to the tangent layer($j = N$) where the total number of layers N is a function of the tangent height. For shallow layers mean layer quantities may be used to accurately model the layer optical depth as follows

$$\Delta\delta_\lambda(i,j) = \overline{\sigma}_\lambda(i,j)L_{i,j} \tag{2.28}$$

where $\overline{\sigma}_\lambda(i,j)$ is the mean extinction cross-section and $L_{i,j}$, the pathlength for layer j, is given by

$$L_{i,j} = \int_{z_{j-1}}^{z_j} dl(z) \tag{2.29}$$

The integration is performed from the lower boundary level z_{j-1}, of layer j, to the upper boundary level z_j. Level z_0 corresponds to z_{top} and $z_{N(i)}$ corresponds to z_{th}. It follows that the mean layer extinction cross-section is given by

$$\overline{\sigma}_\lambda(i,j) = \frac{1}{L_{i,j}} \int_{z_{j-1}}^{z_j} \sigma_\lambda(z) dl(z) \tag{2.30}$$

Each occultation event produces a set of measured slant path optical depths that may be described by the following set of equations

$$\begin{aligned}
\delta_\lambda(1) &= 2\overline{\sigma}_\lambda(1,1)L_{1,1} \\
\delta_\lambda(2) &= 2\overline{\sigma}_\lambda(2,1)L_{2,1} + 2\overline{\sigma}_\lambda(2,2)L_{2,2} \\
&\vdots \\
\delta_\lambda(N^T) &= 2\overline{\sigma}_\lambda(N^T,1)L_{N^T,1} + 2\overline{\sigma}_\lambda(N^T,2)L_{N^T,2} + \cdots + 2\overline{\sigma}_\lambda(N^T,N^T)L_{N^T,N^T}
\end{aligned} \tag{2.31}$$

Equation (2.31) constitutes a set of N^T equations for N^T measurements of δ_λ. It is desired to invert this set of equations and solve for $\overline{\sigma}_\lambda(i,j)$ given the pathlengths $L_{i,j}$. As written, this system of equations is under-determined with $1/2(N^T)^2$ unknowns and only N^T equations. The last major assumption needed for closure recognizes the fact that for shallow layers of $\approx$1 km, the path length weighting of Eq. (2.30) is similar for all tangent heights and the mean extinction cross-section can be expressed as

$$\overline{\sigma}_\lambda(i,j) = \overline{\sigma}_\lambda(j) \tag{2.32}$$

In matrix form Eq. (2.31) then becomes

$$\begin{vmatrix} \delta_\lambda(1) \\ \delta_\lambda(2) \\ \vdots \\ \delta_\lambda(N^T) \end{vmatrix} = 2 \begin{vmatrix} L_{1,1} & 0 & \cdots & 0 \\ L_{2,1} & L_{2,2} & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ L_{N^T,1} & L_{N^T,2} & \cdots & L_{N^T,N^T} \end{vmatrix} \begin{vmatrix} \overline{\sigma}_\lambda(1) \\ \overline{\sigma}_\lambda(2) \\ \vdots \\ \overline{\sigma}_\lambda(N^T) \end{vmatrix} \tag{2.33}$$

These are the basic governing equations for a solar occultation absorption experiment describing radiative processes in an inhomogeneous atmosphere assuming spherical

symmetry and layer averaged atmospheric properties. They may be viewed as applicable to an ideal instrument, that is, monochromatic, with an infinitesimally narrow field of view and noise free. Realistic instrument effects are considered in later sections. Alternate equivalent formulations cast in terms of slant path transmission (Eq. (2.6)) or slant path absorption (Eq. (2.7)) can be derived from Eq. (2.33) but Eq. (2.33) is the simplest formulation and best suited for illustrating the onion-peel inversion concept.

The mathematical solution of Eq. (2.33) for $\overline{\sigma}_\lambda(j)$ is readily obtained through repetitive back substitution since the path length matrix $L_{i,j}$ is lower triangular. The physical interpretation of this inversion process is somewhat more interesting. Starting with the topmost slant path optical depth $\delta_\lambda(1)$, one solves for the unknown $\overline{\sigma}_\lambda(1)$, which then determines the optical depth contribution of this shell

$$\Delta\delta_\lambda(i,1) = 2\overline{\sigma}_\lambda(i,1)L_{i,1} \quad i = 2,3,\ldots N^T \tag{2.34}$$

to the slant path optical depths of all lower tangent heights. Subtracting $\Delta\delta_\lambda(i,1)$ from both sides of Eq. (2.33) leads to a reduced set of N^T-1 equations and unknowns which are

$$\begin{vmatrix} \delta_\lambda(2) - \Delta\delta_\lambda(2,1) \\ \delta_\lambda(3) - \Delta\delta_\lambda(3,1) \\ \vdots \\ \delta_\lambda(N^T) - \Delta\delta_\lambda(N^T,1) \end{vmatrix} = 2 \begin{vmatrix} L_{2,2} & 0 & \cdots & 0 \\ L_{3,2} & L_{3,3} & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ L_{N^T,2} & L_{N^T,3} & \cdots & L_{N^T,N^T} \end{vmatrix} \begin{vmatrix} \overline{\sigma}_\lambda(2) \\ \overline{\sigma}_\lambda(3) \\ \vdots \\ \overline{\sigma}_\lambda(N^T) \end{vmatrix} \tag{2.35}$$

the governing equations describing the original atmosphere but without the topmost shell. The process is repeated until a full solution for $\overline{\sigma}_\lambda$ is reached. Successive removal of the "known" optical depths from the equation set "peels" corresponding layers from the original atmosphere, hence the descriptive name, onion-peel inversion.

The solution to Eq. (2.33) may be written as

$$\sigma_\lambda = \frac{1}{2}\mathbf{L}^{-1}\delta_\lambda \tag{2.36}$$

where σ and δ are now vectors and $\mathbf{L}$[1] is the pathlength matrix. The pathlength matrix can be inverted provided it is nonsingular or $\det(\mathbf{L})\neq 0$. Since the determinant can be expressed as an eigenvalue product, nonsingularity requires nonzero eigenvalues. The eigenvalues of $\mathbf{L}$, which for an upper or lower triangular matrix are given by the diagonal elements, correspond to the tangent shell path lengths. The tangent shell path lengths are nonzero and vary slowly with tangent altitude, thus the inversion will be highly stable.

In this simple example it was possible to solve directly for σ_λ. As the governing equations become more complex and increasingly non-linear from incorporating instrument effects, measurement noise, and inversion effects, a direct solution is no longer possible and an iterative approach to the solution must be taken instead.

2.3 Physical Properties of the Atmosphere

A model of atmospheric temperature, pressure, and number density as a function of altitude is necessary for the evaluation of total extinction cross-section and pathlength of Eq. (2.26). A horizontally stratified atmosphere in hydrostatic equilibrium will follow the hydrostatic equation (U.S. Standard Atmosphere, 1976)

$$\frac{dP}{dz} = -g\rho \tag{2.37}$$

which expresses the vertical differential change in pressure, dP, to the differential change in geometric altitude, dz. Here g is the height dependent gravitational acceleration and ρ is the air density. With the help of the perfect gas law

$$P = \rho RT \tag{2.38}$$

[1]Boldface lower case letter denotes a vector, boldface upper case letter a matrix.

the hydrostatic equation takes the form

$$\frac{dP}{dz} = -\frac{Pg}{RT} \tag{2.39}$$

where R is the gas constant for air, and T is temperature. The formal solution to the equation of hydrostatic equilibrium may be stated as

$$P(z) = P(z_0)exp\left\{-\frac{1}{R}\int_{z_0}^{z}\frac{g(z')}{T(z')}dz'\right\} \tag{2.40}$$

For an isothermal layer over which $g(z)$ can be considered constant, the pressure varies exponentially with altitude.

In practical usage, one is normally concerned with expressing the variation of pressure, temperature, altitude and density or number density between profile levels in a manner consistent with the hydrostatic equation. Given profiles of $P(z_i)$ and $T(z_i)$, it is assumed following Chu(1983), that temperature varies linearly through each layer with the form

$$T(z) = T(z_i) + \beta_i(z - z_i) \quad z_i \leq z \leq z_{i+1} \tag{2.41}$$

where

$$\beta_i = \frac{T(z_{i+1}) - T(z_i)}{z_{i+1} - z_i} \tag{2.42}$$

is the vertical temperature gradient for the ith layer. Recognizing that $dT = \beta_i dz$, Eq. (2.40) for pressure becomes

$$P(z) = \begin{cases} P(z_i)\left(\frac{T(z)}{T(z_i)}\right)^{-\frac{g}{R\beta_i}} & \beta_i \neq 0 \\ P(z_i)exp\left(-\frac{g(z-z_i)}{RT(z_i)}\right) & \beta_i = 0 \end{cases} \tag{2.43}$$

Although g is assumed to be constant within the layer, it is not necessary for g to be constant for all layers. Since the initial $P(z_i), T(z_i)$ pairs are known for all i, another coefficient that accounts for the layer variation of g can be defined

$$G_i = \frac{g}{R} = \begin{cases} \frac{1}{\beta_i}\frac{ln[P(z_{i+1})/P(z_i)]}{ln[T(z_{i+1})/T(z_i)]} & \beta_i \neq 0 \\ \frac{-ln[P(z_{i+1})/P(z_i)]}{T(z_i)/(z_{i+1}-z_i)} & \beta_i = 0 \end{cases} \tag{2.44}$$

This not only simplifies computational aspects of the hydrostatic interpolation but also accommodates those situations where the given profile may not strictly follow the hydrostatic equation. Relationships for density or number density follow directly from application of the perfect gas law.

All of the radiatively and chemically important gaseous species have extremely low concentrations and it is usually more convenient to express their abundances in fractional terms relative to air. From Kondrat'ev (1973), the mixing ratio r^g of gas g can be written as the ratio of mass, m^g, or density, ρ^g of gas g, to the mass or density of air without that component

$$r^g = \frac{m^g}{m - m^g} = \frac{\rho^g}{\rho - \rho^g} \tag{2.45}$$

The mass concentration is the ratio of the mass or density of gas g to that of air

$$r' = \frac{m^g}{m} = \frac{\rho^g}{\rho} \tag{2.46}$$

For $\rho^g \ll \rho$, there is no practical distinction between r and r' and the two definitions have blurred into a generic mass mixing ratio over time. The mass mixing ratio MMR of gas g is defined as the ratio of ρ^g, the density of gas g, to that of air

$$\text{MMR} = \frac{m^g}{m} = \frac{\rho^g}{\rho} \tag{2.47}$$

while the volume mixing ratio, VMR, is defined by

$$\text{VMR} = \frac{C^g}{C} = \frac{P^g}{P} = \frac{V^g}{V} \tag{2.48}$$

where P^g is the partial pressure, and V^g is the molar volume of gas g. Both mixing ratios may be scaled by factors of $10^3, 10^6$, and 10^9 as the situation warrants. The traditional nomenclature for VMR leads to values reported in parts per thousand by volume(pptv), parts per million(ppmv), and parts per billion(ppbv). Conversion between volume and mass mixing ratio is accomplished with the perfect gas law

$$\text{MMR} = \text{VMR}\frac{M^g}{M} \tag{2.49}$$

M is the mean molecular weight of air and M^g is the molecular weight of gas g. In addition to these commonly used measures of abundance that are applicable to all species, water vapor has its own units of measure that are partly based on phase changes and partly due to the historical evolution of water vapor instrumentation.

Following the Smithsonian Meteorological Tables(1951), the mixing ratio of Eq. (2.45) applied to water vapor becomes the ratio of water mass to the mass of dry air

$$r = \frac{m^{H_2O}}{m - m^{H_2O}} = \frac{\rho^{H_2O}}{\rho - \rho^{H_2O}} \tag{2.50}$$

with units of grams/kilograms.

The mass concentration of Eq. (2.46), known as the specific humidity q, is normally expressed in grams/kilogram, and represents the ratio of water vapor mass to the mass of moist air.

$$q = r' = \frac{m^{H_2O}}{m} = \frac{\rho^{H_2O}}{\rho} = \text{MMR} = \text{VMR}\frac{M^{H_2O}}{M} \tag{2.51}$$

The water vapor partial pressure, e, can be related to total pressure P and mixing ratio r with the perfect gas law applied to a binary mixture of dry air and water vapor

$$e = \frac{r}{0.62197 + r}P \tag{2.52}$$

and from Eq. (2.48) the ratio e/P is the volume mixing ratio of water. The constant 0.62197 represents M^{H_2O}/M.

Absolute humidity is simply the water vapor density ρ^{H_2O} expressed in units of grams/meter^3. Relative humidity, RH, is expressed as the ratio of partial pressure e at a given pressure P and temperature T to the saturation partial pressure e_s at the same conditions

$$RH = \frac{e}{e_s(T)} = \frac{e_s(T_d)}{e_s(T)} = \frac{e_s(T_f)}{e_s(T)} \tag{2.53}$$

Normally, RH is stated in terms of a percentage. The saturation partial pressure e_s represents saturation with respect to either water or ice, depending on the temperature, and is primarily a function of temperature in both cases. The dew-point temperature T_d is the temperature to which moist air at constant pressure P and mixing ratio r(or partial pressure e) must be cooled so that it is saturated with respect to water. The frost-point temperature T_f is defined in a similar manner for saturation over ice. The dew point depression is defined as the difference between T and T_d.

Pressure, temperature, and number density profiles of air for a model atmosphere (U.S. Standard Atmosphere, 1976) representative of middle latitude annual mean conditions are shown in Fig. 2.3. Initially, temperature decreases upwards from the surface with a negative gradient and then reverses at higher altitudes. The tropopause, thermally defined as the altitude where the temperature ceases to decrease, separates the atmosphere into two regimes, the troposphere(below) where rapid mixing occurs and the dynamically quiescent stratosphere(above). The tropopause varies from about 15 km at the equator to 10 km at high latitudes. Both pressure and number density vary nearly exponentially over three decades through the troposphere and stratosphere. A model profile for water vapor volume mixing ratio is also shown in the lower right panel. Mixing ratio decreases rapidly with height throughout the troposphere and is almost constant in the stratosphere. The dashed line in the temperature panel, corresponding to a combined dew/frost point (weighted transition from $\approx$ 258 K to 238 K) temperature profile, tracks the air temperature throughout the troposphere and then diverges above the tropopause. The water vapor partial pressure, e, given by Eq. (2.48), ranges over six decades with a stratospheric gradient matching that of air while the tropospheric gradient is determined more by the mixing ratio gradient.

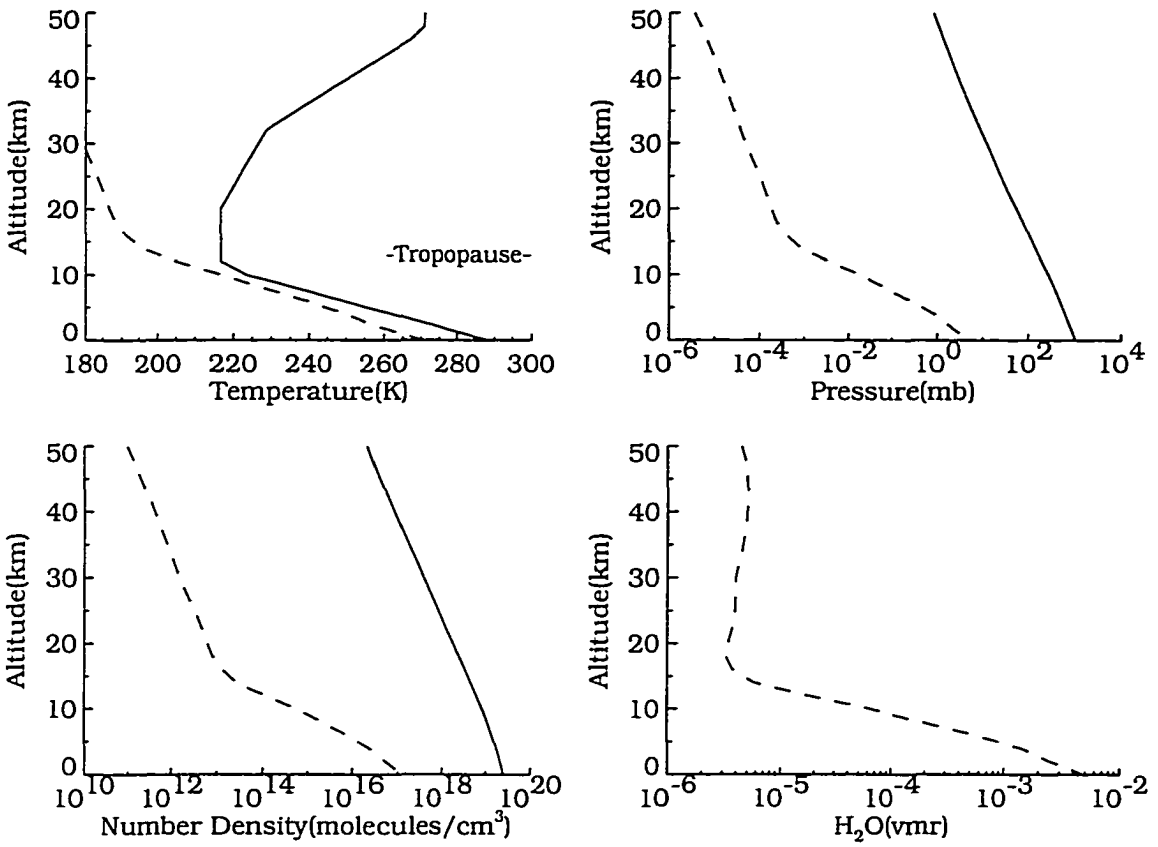


Figure 2.3: Illustrative model atmosphere. Solid lines indicate temperature, pressure and number density as specified by the 1976 Standard Atmosphere. Dashed lines indicate a combined frost/dew point temperature, partial pressure and number density for the water vapor profile shown in the lower right panel.

2.4 Remote Sensing Error Analysis

In the past, a wide variety of approaches to error analysis of remotely sensed atmospheric data have been taken that were highly instrument dependent and best described as ad hoc. In the simplest and most general approach, the governing equations, describing atmospheric physics and instrument effects to an appropriate level of detail, are combined with the best estimates of ancillary information to construct synthetic measurements followed by inversion to produce a reference profile. Systematic errors are assessed by perturbing each of the systematic ancillary parameters by

its estimated uncertainty in the inversion model, inverting, and then comparing the resulting profile to the reference. To assess random errors, a Monte Carlo approach is usually taken whereby the synthetic measurements or ancillary parameters are perturbed randomly and then inverted. This must be done for a number (25-30) of independent sets of random numbers before statistics can be derived from the inverted profiles. The error analysis becomes more complicated when considering interlevel error correlations or the importance of *a priori* profile information to the inverted profile. In addition, any comparison or combination of profile data from different instruments requires an understanding of experimental resolution. Weighting functions provide a measure of resolution but since they are usually presented neglecting instrument effects or resolution degradation in the inversion, they are of limited use.

The formal error analysis developed by Rodgers(1990) provides a rigorous approach that is independent of observing system or inversion method and addresses many of the issues mentioned above. The development of this approach was triggered by observational evidence from satellites and ground based networks that levels of ozone were decreasing in the stratosphere and in a total column sense from 1979 through the mid 1980's. A search for similar trends in other ozone databases and subsequent evaluation of trend results through instrument intercomparisons was hampered by the different approaches taken to define vertical resolution and error propagation. As a result of this situation, data from nine ozone experiments (including SAGE II) was completely re-analyzed, each with the best available inversion method appropriate to that experiment and then uniformly evaluated with Rodgers's error analysis (WMO Report No. 18, 1988). The analysis is also useful when comparing inversion algorithms, such as was done for the Improved Stratospheric and Mesospheric Sounder ozone channel (Connor and Rodgers, 1989) and for a broad band limb scanning radiometer also measuring ozone (Gille and Bailey, 1989).

The formal error analysis is based on the linearization of the forward and inverse models describing the remote sensing experiment at a reference state such that uncertainties appear in the forward rather than the inverse model. The advantages of this approach are several. Interlevel error correlations are easily accommodated with error covariance matrices. Overall experimental resolution is defined by the averaging kernel which gives a truer measure of profile resolution than the weighting functions. For under-determined systems that require *a priori* data to close the system of equations, the method provides a way to determine how much profile information is related to the measurements and not the *a priori* data. Our systems of equations are well posed, so *a priori* will not be a concern in this work.

Since Rodgers's error analysis will be used extensively in Sec. 4 a discussion of the general methodology will be presented here using the terminology and symbols found in Rodgers(1990) for the most part. Specific details will be developed in Sec. 4 when the method is applied to the SAGE II water vapor observations.

The remote sensing observational scenario is shown schematically in Fig. 2.4. The atmospheric spectral signature is a function, in some complicated way, of atmospheric composition and thermodynamic state. The atmospheric parameters, viewed as the unknowns of the experiment for which information is desired, are denoted by $\mathbf{x}$. In the following development the superscript t denotes the true value of the parameter, $\hat{}$ denotes a best estimate and an unscripted parameter may take on both meanings. The instrument and inversion method together compose the observing system whose final product is the retrieved atmospheric parameter(s) $\mathbf{x}^r$. Typically, $\mathbf{x}^r$ is thought of as a vector representing a profile of some atmospheric quantity (gas concentration or mixing ratio, temperature, pressure) but it may also represent a set of coefficients that describe these quantities. For the sake of discussion, $\mathbf{x}^r$ will be treated as a profile with n elements located at levels sufficiently close together to adequately resolve significant profile variations.

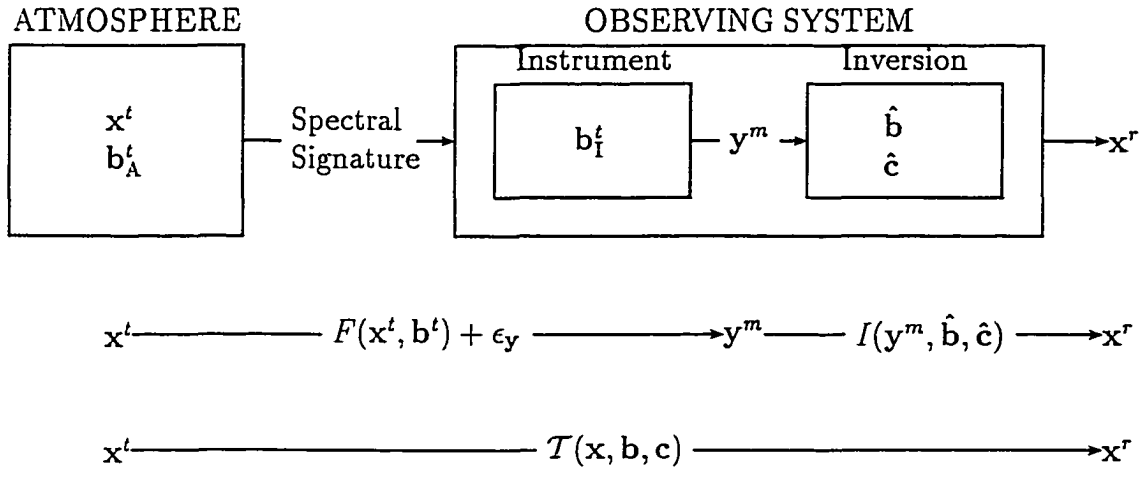


Figure 2.4: Remote sensing elements.

The measurement vector $\mathbf{y}^m$ is a profile of m quantities that include measurement error $\epsilon_\mathbf{y}$. The error $\epsilon_\mathbf{y}$ is assumed to be a random variable with a normal distribution, zero mean, and covariance $\mathbf{S}_{\epsilon\mathbf{y}}$. When the number of measurements, m, is less than the number, n, of unknowns, the inversion problem is under-determined and additional constraints must be included as part of the inversion methodology. Fortunately, the occultation measurement geometry produces well posed inversions with $n \leq m$.

Atmospheric parameters (other than the unknown $\mathbf{x}^t$) that materially affect the spectral signature are indicated by $\mathbf{b}_A^t$ and $\mathbf{b}_I^t$ denotes instrument parameters that influence $\mathbf{y}^m$. Together, $\mathbf{b}_A^t$ and $\mathbf{b}_I^t$ form the model parameter vector $\mathbf{b}^t$ with n_b elements. The atmospheric and model parameters thus represent all significant experiment variables. Since $\mathbf{b}^t$ is unknown but must be specified, a best estimate $\hat{\mathbf{b}}$ is used in the inversion that may contain both systematic and random errors.

Parameters that do not influence $\mathbf{y}^m$ but are part of the inversion method are grouped in vector $\hat{\mathbf{c}}$ composed of n_c elements. These inversion model parameters might include *a priori* data, convergence acceleration parameters, convergence

criteria, and iterative step sizes for example. Since many of these parameters are algorithmic in nature, with a permissible range of values of which one must be selected, the meaning of true and best estimate does not strictly hold. In this case they should be viewed as bracketing a range of values for sensitivity studies.

2.4.1 Forward Model

The forward model F is defined through the following relationship.

$$\mathbf{y}^m = F(\mathbf{x}, \mathbf{b}) + \epsilon_{\mathbf{y}} \tag{2.54}$$

The forward model combines the atmospheric physics needed to calculate a spectral signature with the observational characteristics of the instrument and measurement geometry. The model parameter vector $\mathbf{b}$ describes all quantities, other than the unknown, $\mathbf{x}$, that must be specified prior to calculating $\mathbf{y}^m$. Provided F accurately describes the physics of the measurement, errors in $\mathbf{y}^m$ due to F are attributable directly to $\mathbf{b}$. In general, the true values of $\mathbf{b}$ are not known and best estimates of $\mathbf{b}$ may introduce systematic and random errors in $\mathbf{y}^m$.

A Taylor series expansion of $\mathbf{y}^m$ with respect to $\mathbf{x}$ and $\mathbf{b}$ at the reference state $\overline{\mathbf{x}}, \overline{\mathbf{b}}$ becomes (after dropping second order terms)

$$\mathbf{y}^m = F(\overline{\mathbf{x}}, \overline{\mathbf{b}}) + \frac{\partial F}{\partial \mathbf{x}}(\mathbf{x} - \overline{\mathbf{x}}) + \frac{\partial F}{\partial \mathbf{b}}(\mathbf{b} - \overline{\mathbf{b}}) + \epsilon_{\mathbf{y}} \tag{2.55}$$

By defining

$$\mathbf{K}_{\mathbf{x}} = \frac{\partial F}{\partial \mathbf{x}}, \quad \mathbf{K}_{\mathbf{b}} = \frac{\partial F}{\partial \mathbf{b}} \tag{2.56}$$

Eq. (2.55) becomes

$$\mathbf{y}^m = F(\overline{\mathbf{x}}, \overline{\mathbf{b}}) + \mathbf{K}_{\mathbf{x}}(\mathbf{x} - \overline{\mathbf{x}}) + \mathbf{K}_{\mathbf{b}}(\mathbf{b} - \overline{\mathbf{b}}) + \epsilon_{\mathbf{y}} \tag{2.57}$$

For m measurements, n unknowns, and n_b model parameters the $\mathbf{K}_{\mathbf{x}}$ matrix will be $m \times n$ and $\mathbf{K}_{\mathbf{b}}$ is $m \times n_b$. Matrices $\mathbf{K}_{\mathbf{x}}$ and $\mathbf{K}_{\mathbf{b}}$ define the sensitivity of the forward

model to the atmospheric and model parameters, respectively. The matrix $\mathbf{K_x}$ is interpreted as a weighting function since the rows of $\mathbf{K_x}$ describe the weight given to information from each level of $\mathbf{x}$. The row weighting functions tend to peak over a range of levels and are indicative of the vertical resolution of the experiment. The interpretation of $\mathbf{K_b}$ is straightforward, that is, individual model parameters may potentially influence all values of $\mathbf{y}^m$ depending on their role in F.

Analytical expressions describing the forward model are quite complicated except for the simplest remote sensing experiments and expressions for $\mathbf{K_x}$ and $\mathbf{K_b}$ are difficult to obtain. In most cases it is easier to build up the $\mathbf{K}$ matrices through a series of perturbation runs using a computer model of F. To do this, let $\epsilon_\mathbf{y} = 0$, $\mathbf{b} = \overline{\mathbf{b}}$, and for a perturbation Δx_k in $\mathbf{x}$ at level k, solve Eq. (2.57) for the k^{th} column vector of $\mathbf{K_x}$ as

$$\mathbf{K_x}|_k = \frac{\mathbf{y}^m - F(\overline{\mathbf{x}}, \overline{\mathbf{b}})}{\Delta x_k} \tag{2.58}$$

Similarly, let $\epsilon_\mathbf{y} = 0$, $\mathbf{x} = \overline{\mathbf{x}}$, and the k^{th} column vector of $\mathbf{K_b}$ may be calculated for a perturbation Δb_k from

$$\mathbf{K_b}|_k = \frac{\mathbf{y}^m - F(\overline{\mathbf{x}}, \overline{\mathbf{b}})}{\Delta b_k} \tag{2.59}$$

2.4.2 Inverse Model

The retrieved atmospheric parameter vector $\mathbf{x}^r$ depends on $\mathbf{y}^m, \mathbf{b}, \mathbf{c}$ and the physics of the inversion or retrieval model as follows

$$\mathbf{x}^r = I(\mathbf{y}^m, \mathbf{b}, \mathbf{c}) \tag{2.60}$$

Expanding as before in a Taylor series at the reference state $\overline{\mathbf{y}}^m, \overline{\mathbf{b}}, \overline{\mathbf{c}}$ and neglecting second order terms yields

$$\mathbf{x}^r = I(\mathbf{y}^m, \mathbf{b}, \mathbf{c}) = I(\overline{\mathbf{y}}^m, \overline{\mathbf{b}}, \overline{\mathbf{c}}) + \mathbf{D_y}(\mathbf{y}^m - \overline{\mathbf{y}}^m) + \mathbf{D_b}(\mathbf{b} - \overline{\mathbf{b}}) + \mathbf{D_c}(\mathbf{c} - \overline{\mathbf{c}}) \tag{2.61}$$

where

$$\mathbf{D_y} = \frac{\partial I}{\partial \mathbf{y}^m}, \quad \mathbf{D_b} = \frac{\partial I}{\partial \mathbf{b}}, \quad \mathbf{D_c} = \frac{\partial I}{\partial \mathbf{c}} \tag{2.62}$$

The matrix $\mathbf{D_y}$ is $n \times m$ while $\mathbf{D_b}$ is $n \times n_b$ and $\mathbf{D_c}$ is $n \times n_c$. Matrices $\mathbf{D_b}$ and $\mathbf{D_c}$ simply describe the sensitivity of the inversion to the model and inversion parameters. Matrix $\mathbf{D_y}$ is referred to as the contribution function with the interpretation that for a unit change in $\mathbf{y}^m$ at level j, the j^{th} column of $\mathbf{D_y}$ describes the contribution to $\mathbf{x}^r$ for all levels.

The reference states $\overline{\mathbf{x}}, \overline{\mathbf{b}}, \overline{\mathbf{c}}$, and $\overline{\mathbf{y}}^m$ referred to in this and the previous section may assume a number of different meanings. While $\overline{\mathbf{b}}$ and $\overline{\mathbf{c}}$ may take on values that correspond to their true or best estimates, $\overline{\mathbf{y}}^m$ is an intermediate product that is eliminated in the end-to-end analysis that follows. The reference parameter vector $\overline{\mathbf{x}}$ can correspond to the true or retrieved values or even be specified from other sources such as climatological data, atmospheric reference models and photochemical model calculations.

2.4.3 Transfer Function and Error Analysis

In some instances it is desirable to express the relationship between $\mathbf{x}$ and $\mathbf{x}^r$ in terms of a transfer function $\mathcal{T}(\mathbf{x}, \mathbf{b}, \mathbf{c})$. The transfer function will be developed by combining the forward and inverse models assuming that the forward and reverse models are linearized at the same state $(\overline{\mathbf{x}}, \overline{\mathbf{b}}, \overline{\mathbf{c}})$. Then, with $\overline{\mathbf{y}}^m = F(\overline{\mathbf{x}}, \overline{\mathbf{b}})$ and $\mathbf{y}^m$ given by Eq. (2.55), Equation (2.61) becomes

$$\begin{aligned} \mathbf{x}^r &= I(F(\mathbf{x}_F, \mathbf{b}_F) + \epsilon_\mathbf{y}, \mathbf{b}_I, \mathbf{c}_I) \\ &= I(F(\overline{\mathbf{x}}, \overline{\mathbf{b}}), \overline{\mathbf{b}}, \overline{\mathbf{c}}) + \mathbf{D_y}[\epsilon_\mathbf{y} + \mathbf{K_x}(\mathbf{x}_F - \overline{\mathbf{x}}) + \mathbf{K_b}(\mathbf{b}_F - \overline{\mathbf{b}})] \\ &\quad + \mathbf{D_b}(\mathbf{b}_I - \overline{\mathbf{b}}) + \mathbf{D_c}(\mathbf{c}_I - \overline{\mathbf{c}}) \end{aligned} \tag{2.63}$$

Since the forward model will be evaluated using "true" parameters while the inverse model uses "estimated" parameters in subsequent sections, the parameters in

Eq. (2.63) have been temporarily tagged with subscripts F and I to clearly distinguish where each should be used. Thus, for the moment, it would be inappropriate to combine terms in the above equation for the model parameters $\mathbf{b}$. This then, is the most general equation describing the linearization of the atmosphere and observing system. Although it contains a number of terms, many of these will be eliminated after the linearization and evaluation states are specified.

In an experiment with no measurement noise ($\epsilon_y = 0.0$), Eq. (2.63) becomes

$$\begin{aligned} \mathbf{x}^r &= I(F(\mathbf{x}_F, \mathbf{b}_F), \mathbf{b}_I, \mathbf{c}_I) \\ &= I(F(\overline{\mathbf{x}}, \overline{\mathbf{b}}), \overline{\mathbf{b}}, \overline{\mathbf{c}}) + \mathbf{D_y}[\mathbf{K_x}(\mathbf{x}_F - \overline{\mathbf{x}}) + \mathbf{K_b}(\mathbf{b}_F - \overline{\mathbf{b}})] \\ &\quad + \mathbf{D_b}(\mathbf{b}_I - \overline{\mathbf{b}}) + \mathbf{D_c}(\mathbf{c}_I - \overline{\mathbf{c}}) \end{aligned} \tag{2.64}$$

The expression above suggests a general definition of $\mathcal{T}$ as

$$\mathbf{x}^r = I(\mathbf{y}^m, \mathbf{b}, \mathbf{c}) = I(F(\mathbf{x}, \mathbf{b}), \mathbf{b}, \mathbf{c}) = \mathcal{T}(\mathbf{x}, \mathbf{b}, \mathbf{c}) \tag{2.65}$$

where the forward and inverse models are required to use the same atmospheric, model and inversion parameters. Since only the model parameters, $\mathbf{b}$, are found in both F and I this is analogous to requiring $\mathbf{b}_F = \mathbf{b}_I$. Then Eq. (2.64) becomes

$$\mathbf{x}^r = \mathcal{T}(\mathbf{x}, \mathbf{b}, \mathbf{c}) = \mathcal{T}(\overline{\mathbf{x}}, \overline{\mathbf{b}}, \overline{\mathbf{c}}) + \mathbf{D_y}\mathbf{K_x}(\mathbf{x} - \overline{\mathbf{x}}) + [\mathbf{D_y}\mathbf{K_b} + \mathbf{D_b}](\mathbf{b} - \overline{\mathbf{b}}) + \mathbf{D_c}(\mathbf{c} - \overline{\mathbf{c}}) \tag{2.66}$$

This defines $\mathcal{T}$, and the linearization of $\mathcal{T}$ at the state $(\overline{\mathbf{x}}, \overline{\mathbf{b}}, \overline{\mathbf{c}})$ in terms of F and I for a noise free experiment provided F and I are evaluated at identical states. Note that $\mathbf{x}^t$ may not be recovered even when the true values of all the experiment parameters are known, that is

$$\mathbf{x}^r = \mathcal{T}(\mathbf{x}^t, \mathbf{b}^t, \mathbf{c}^t) \neq \mathbf{x}^t \tag{2.67}$$

To develop an expression for $\mathbf{x}^r$ using the transfer function suitable for an error analysis of a real experiment we examine two approaches. The first recognizes that

only the best estimates $\hat{\mathbf{b}}$ and $\hat{\mathbf{c}}$ are available to the experimenter for computational purposes. This suggests linearization at the state $(\overline{\mathbf{x}}, \hat{\mathbf{b}}, \hat{\mathbf{c}}, \overline{\epsilon}_\mathbf{y})$, where the reference state $\overline{\mathbf{x}}$ stands in place of the unknown $\mathbf{x}^t$. This is the approach taken by Rodgers. The second and more traditional approach simply uses the state $(\mathbf{x}^t, \mathbf{b}^t, \mathbf{c}^t, \overline{\epsilon}_\mathbf{y})$ as the linearization point. Since all of these parameters are unknown, one is forced to use the best estimates in place of the true values for numerical computations.

In both approaches it will be necessary to evaluate the inverse model using best estimates of the model and inversion parameters, $\hat{\mathbf{b}}$ and $\hat{\mathbf{c}}$, while the forward model uses the true values $\mathbf{x}^t$ and $\mathbf{b}^t$. This is shown pictorially in Fig. 2.4.

Linearization at $(\overline{\mathbf{x}}, \hat{\mathbf{b}}, \hat{\mathbf{c}})$

Equation (2.63) defines the linearization of $\mathbf{x}^r$ at the reference state $(\overline{\mathbf{x}}, \overline{\mathbf{b}}, \overline{\mathbf{c}})$. If, instead, the equation is linearized at $\overline{\mathbf{x}}, \hat{\mathbf{b}}, \hat{\mathbf{c}}(\overline{\mathbf{x}} = \overline{\mathbf{x}}, \overline{\mathbf{b}} = \hat{\mathbf{b}}, \overline{\mathbf{c}} = \hat{\mathbf{c}})$ and evaluated at $\hat{\mathbf{b}}, \hat{\mathbf{c}}(\mathbf{b} = \hat{\mathbf{b}}, \mathbf{c} = \hat{\mathbf{c}})$, for terms involving the inverse model and forward model terms at $\mathbf{x}^t, \mathbf{b}^t(\mathbf{x} = \mathbf{x}^t, \mathbf{b} = \mathbf{b}^t)$, it reduces to

$$\begin{aligned} \mathbf{x}^r &= I(F(\mathbf{x}^t, \mathbf{b}^t) + \epsilon_\mathbf{y}, \hat{\mathbf{b}}, \hat{\mathbf{c}}) \\ &= \mathcal{T}(\overline{\mathbf{x}}, \hat{\mathbf{b}}, \hat{\mathbf{c}}) + \mathbf{D}_\mathbf{y}\mathbf{K}_\mathbf{x}(\mathbf{x}^t - \overline{\mathbf{x}}) + \mathbf{D}_\mathbf{y}\mathbf{K}_\mathbf{b}(\mathbf{b}^t - \hat{\mathbf{b}}) + \mathbf{D}_\mathbf{y}\epsilon_\mathbf{y} \end{aligned} \tag{2.68}$$

after introducing the transfer function. The retrieved vector $\mathbf{x}^r$ is determined by the transfer function evaluated at the linearization state $(\overline{\mathbf{x}}, \hat{\mathbf{b}}, \hat{\mathbf{c}})$ with perturbation terms in $\Delta\mathbf{x}$ and $\Delta\mathbf{b}$ acted on by both the forward and inverse models and a measurement error term that is acted on by the inverse model. With this choice of linearization, uncertainties in $\mathbf{x}$ and $\mathbf{b}$ translate into uncertainties in $\mathbf{y}^m$ that are then acted on by I. Alternately, Eq. (2.68) implies the linearization of $\mathcal{T}$ at the state $(\overline{\mathbf{x}}, \hat{\mathbf{b}}, \hat{\mathbf{c}})$ and evaluation at $(\mathbf{x}^t, \mathbf{b}^t)$ or

$$\begin{aligned} \mathbf{x}^r &= \mathcal{T}(\overline{\mathbf{x}}, \hat{\mathbf{b}}, \hat{\mathbf{c}}) + \mathbf{A}_\mathbf{x}(\mathbf{x}^t - \overline{\mathbf{x}}) + \mathbf{A}_\mathbf{b}(\mathbf{b}^t - \hat{\mathbf{b}}) + \mathbf{D}_\mathbf{y}\epsilon_\mathbf{y} \\ &= \mathcal{T}(\mathbf{x}^t, \mathbf{b}^t, \hat{\mathbf{c}}) + \mathbf{D}_\mathbf{y}\epsilon_\mathbf{y} \end{aligned} \tag{2.69}$$

with $\mathbf{A_x} = \frac{\partial \mathcal{T}}{\partial \mathbf{x}} = \mathbf{D_y K_x}$ and $\mathbf{A_b} = \frac{\partial \mathcal{T}}{\partial \mathbf{b}} = \mathbf{D_y K_b}$. The measurement error term arises because the transfer function was defined with $\epsilon_\mathbf{y} = 0$. Note that this form is a simplified version of Eq. (2.66) resulting from the elimination of some terms.

The square matrix, $\mathbf{A_x}$, smooths variations of $\mathbf{x}^t$ from the reference state $\overline{\mathbf{x}}$ and is more commonly given as

$$\mathbf{A} = \mathbf{A_x} = \mathbf{D_y K_x} \tag{2.70}$$

The rows of $\mathbf{A}$ are referred as the "averaging kernel". The j^{th} row weights the difference profile $\Delta\mathbf{x} = \mathbf{x}^t - \overline{\mathbf{x}}$ for a contribution to $\mathbf{x}^r$ at level j. The columns define the contribution to $\mathbf{x}^r$ for a δ-function perturbation in $\Delta\mathbf{x}$. The rows and columns are typically smoothly varying functions with peaks that are indicative of experimental resolution. Functions with narrow peaks are more desirable than those with broad peaks as this limits the influence of a single level of $\Delta\mathbf{x}$ to adjacent levels. A "perfect" experiment would have an $\mathbf{A}$ matrix of unity. Nadir viewing experiments generally exhibit broad functions while occultation experiments tend to be much narrower.

Linearization at $(\mathbf{x}^t, \mathbf{b}^t, \mathbf{c}^t)$

Repeating the analysis but linearizing at $(\mathbf{x}^t, \mathbf{b}^t, \mathbf{c}^t)$ and evaluating at $\hat{\mathbf{b}}, \hat{\mathbf{c}}$ in the inverse model and $\mathbf{x}^t, \mathbf{b}^t, \epsilon_\mathbf{y}$ in the forward model leads to

$$\begin{aligned} \mathbf{x}^r &= I(F(\mathbf{x}^t, \mathbf{b}^t) + \epsilon_\mathbf{y}, \hat{\mathbf{b}}, \hat{\mathbf{c}}) \\ &= I(F(\mathbf{x}^t, \mathbf{b}^t), \mathbf{b}^t, \mathbf{c}^t) + \mathbf{D_b}(\hat{\mathbf{b}} - \mathbf{b}^t) + \mathbf{D_c}(\hat{\mathbf{c}} - \mathbf{c}^t) + \mathbf{D_y}\epsilon_\mathbf{y} \\ &= \mathcal{T}(\mathbf{x}^t, \mathbf{b}^t, \mathbf{c}^t) + \mathbf{D_b}(\hat{\mathbf{b}} - \mathbf{b}^t) + \mathbf{D_c}(\hat{\mathbf{c}} - \mathbf{c}^t) + \mathbf{D_y}\epsilon_\mathbf{y} \end{aligned} \tag{2.71}$$

In this form all information about the forward model is implicitly contained in the transfer function. Uncertainties in $\mathbf{b}$ and $\mathbf{c}$ are acted on by only the inverse model and perturbation terms containing $\mathbf{x}$ are no longer present. In the past it

has been this form favored in error analyses. One advantage to this formulation for error studies is that only the contribution functions must be calculated. Even though $\mathbf{x}^t, \mathbf{b}^t, \mathbf{c}^t$ are not known, the experimenter proceeded computationally instead with $\overline{\mathbf{x}}, \hat{\mathbf{b}}, \hat{\mathbf{c}}$ or even $\mathbf{x}^r, \hat{\mathbf{b}}, \hat{\mathbf{c}}$ (in a post retrieval analysis from the same experiment) to calculate F or $\mathcal{T}$ and the contribution functions. In the development that follows, only linearization at $\overline{\mathbf{x}}, \hat{\mathbf{b}}, \hat{\mathbf{c}}$ is considered.

Error Analysis

The total error for the experiment, $\mathbf{x}^r - \mathbf{x}^t$, can now be derived. Returning to the preferred linearization form, $(\overline{\mathbf{x}}, \hat{\mathbf{b}}, \hat{\mathbf{c}})$, Eq. (2.68) may be written as

$$\mathbf{x}^r - \mathbf{x}^t = [\mathcal{T}(\overline{\mathbf{x}}, \hat{\mathbf{b}}, \hat{\mathbf{c}}) - \overline{\mathbf{x}}] + (\mathbf{A} - \mathbf{I})(\mathbf{x}^t - \overline{\mathbf{x}}) + \mathbf{D_y K_b}(\mathbf{b}^t - \hat{\mathbf{b}}) + \mathbf{D_y}\epsilon_{\mathbf{y}} \qquad (2.72)$$

after adding $\mathbf{x}^t - \mathbf{x}^t + \overline{\mathbf{x}} - \overline{\mathbf{x}}$ to the right hand side and rearranging. The identity matrix is denoted by $\mathbf{I}$. A full description of the total error $\mathbf{x}^r - \mathbf{x}^t$ is composed of both systematic and random components while the error terms on the right hand side may be solely systematic, solely random or a combination of both.

Error propagation for a linear system of equations, or for a linearized system such as Eq. (2.72), must include the covariance of the random variable. If the random dependent variables $\mathbf{u} = (u_1, u_2, \ldots, u_m)$ are functions of the random independent variables $\mathbf{v} = (v_1, v_2, \ldots, v_n)$ through

$$\mathbf{u} = G(\mathbf{v}) \qquad (2.73)$$

and this system is linearized to give

$$\mathbf{u} = \mathbf{u}(E\{\mathbf{v}\}) + \mathbf{H}(\mathbf{v} - E\{\mathbf{v}\}), \qquad H_{ij} = \frac{\partial u_i}{\partial v_j} \qquad (2.74)$$

then the covariance matrix of $\mathbf{u}$, $\mathbf{S^u}$, will be

$$\mathbf{S^u} = \mathbf{H}\mathbf{S^v}\mathbf{H^T} \qquad (2.75)$$

The expected value or mean of vector $\mathbf{v}$ is $E\{\mathbf{v}\}$, $\mathbf{S^v}$ is the covariance matrix of $\mathbf{v}$, and the superscript T denotes the matrix transpose operation. The symmetric covariance matrix $\mathbf{S^v}$ may be written in matrix notation as

$$\mathbf{S^v} = \mathbf{E}\{(\mathbf{v} - \mathbf{E}\{\mathbf{v}\})(\mathbf{v} - \mathbf{E}\{\mathbf{v}\})^{\mathbf{T}}\} \tag{2.76}$$

or for element i, j

$$S_{ij}^{\mathbf{v}} = E\{(\mathbf{v}_i - E\{\mathbf{v}_i\})(\mathbf{v}_j - E\{\mathbf{v}_j\})\} \tag{2.77}$$

As defined, the covariance matrix diagonal elements are the variances of $\mathbf{v}$ and the off-diagonal elements, the covariances, are a measure of the dependence between the elements of $\mathbf{v}$. When the elements of $\mathbf{v}$ are independent, $\mathbf{S^v}$ is a diagonal matrix. The uncertainty associated with $\mathbf{u}$, the variances of $\mathbf{u}$ found on the diagonal of $\mathbf{S^u}$, include contributions from the off-diagonal covariant elements of $\mathbf{S^v}$.

Another measure of the dependence of two elements v_i, v_j is the correlation coefficient

$$C_{ij} = \frac{S_{ij}}{\sigma_i \sigma_j} \qquad -1. \leq C_{ij} \leq 1. \tag{2.78}$$

where the standard deviation, σ_i, is the square root of the variance

$$\sigma_i^2 = E\{(\mathbf{v}_i - E\{\mathbf{v}_i\})^2\} \tag{2.79}$$

and a similar equation holds for σ_j. The covariance or correlation coefficient is positive if $\mathbf{v}_i$ and $\mathbf{v}_j$ together preferentially fall above or below their means and is negative if one is preferentially above its mean while the other is below. For $S_{ij} = C_{ij} = 0$, the variables are uncorrelated but not necessarily independent unless $\mathbf{v}$ is Gaussian.

The physical basis of each error term will be presented in the following sections along with the corresponding error covariances.

Measurement error $\epsilon_{\mathbf{y}}$. The measurement error term appears as $\mathbf{D}_{\mathbf{y}}\epsilon_{\mathbf{y}}$. The quantity $\epsilon_{\mathbf{y}}$ has covariance $\mathbf{S}_{\epsilon_{\mathbf{y}}}$ and the error term will have covariance $\mathbf{S}_E$. Thus,

$$\mathbf{S}_E = \mathbf{D}_{\mathbf{y}}\mathbf{S}_{\epsilon_{\mathbf{y}}}\mathbf{D}_{\mathbf{y}}^T \tag{2.80}$$

The error covariance $\mathbf{S}_{\epsilon_{\mathbf{y}}}$ may be derived from laboratory or operational calibrations, or analytically based design expectations.

Model parameter errors $\epsilon_{\mathbf{b}}$. The model parameter error $\epsilon_{\mathbf{b}}$ is defined as $\mathbf{b}^t - \hat{\mathbf{b}}$ and may contain both systematic and random error components depending on the particular parameter. The random component has error covariance $\mathbf{S}_{\epsilon_{\mathbf{b}}}$. The retrieval error covariance for this term is

$$\mathbf{S}_B = (\mathbf{D}_{\mathbf{y}}\mathbf{K}_{\mathbf{b}})\mathbf{S}_{\epsilon_{\mathbf{b}}}(\mathbf{D}_{\mathbf{y}}\mathbf{K}_{\mathbf{b}})^T \tag{2.81}$$

Model bias error. The term $[\mathcal{T}(\overline{\mathbf{x}}, \hat{\mathbf{b}}, \hat{\mathbf{c}}) - \overline{\mathbf{x}}]$ describes systematic errors resulting from differences between the forward and inverse models. *A priori*, the experimenter cannot be certain that the inverse model contains all the significant physical processes of the forward model, that approximations to certain physical processes are sufficiently accurate, or even that the operational inversion is error free. While some gross errors will be immediately obvious, evidence for subtle model biases may not appear until the validation phase of the experiment when $\mathbf{x}^r$ is compared to other experimental observations. In either case, if the modeling defect cannot be identified, the bias can be absorbed in the inversion model as a correction term.

Null space error. The null space error term, $(\mathbf{A} - \mathbf{I})(\mathbf{x}^t - \overline{\mathbf{x}})$ describes how the overall experimental resolution, as defined by $\mathbf{A}$, affects retrieval accuracy. Null space exists whenever components of $\mathbf{x}^t$ are orthogonal to $\mathbf{K}_{\mathbf{x}}$ in the forward model or components of $\mathbf{y}^m$ are orthogonal to $\mathbf{D}_{\mathbf{y}}$ in the inverse model. Since $\mathbf{K}_{\mathbf{x}}$ for the

occultation measurement geometry is square ($m = n$) with a nonzero determinant, null space in $\mathbf{A}$ is due entirely to the form of $\mathbf{D_y}$. When the sensitivity of the inverse model matches the forward model, or $\mathbf{D_y} = \mathbf{K_x^{-1}}$, then $\mathbf{A} = \mathbf{I}$, and the null space error vanishes.

The systematic null space error is $(\mathbf{A} - \mathbf{I})(\mathbf{x}^t - \overline{\mathbf{x}})$ while the random component is given by

$$\mathbf{S}_N = (\mathbf{A} - \mathbf{I})\mathbf{S}_{\mathbf{x}^t}(\mathbf{A} - \mathbf{I})^T \tag{2.82}$$

where $\mathbf{S}_{\mathbf{x}^t}$ is the covariance of $\mathbf{x}^t - \overline{\mathbf{x}}$, or $\mathbf{x}^t$ since $\overline{\mathbf{x}}$ is constant. To evaluate the systematic and random error terms requires knowledge of $\mathbf{x}^t$, our unknown. Estimates of the null space errors are possible if independent estimates of $\mathbf{x}^t$ are available, preferably from other experiments with superior resolution.

The role of the model bias error and null space error seem to overlap since both involve terms in $\mathbf{x}$. Rewriting the error equation for terms in $\mathbf{x}$ (Eq. (2.72), terms in $\mathbf{b}$ and $\mathbf{c}$ discarded) as

$$\mathbf{e}(\mathbf{x}^t) = \mathbf{x}^r - \mathbf{x}^t = [\mathcal{T}(\overline{\mathbf{x}}, \hat{\mathbf{b}}, \hat{\mathbf{c}}) - \overline{\mathbf{x}}] + (\mathbf{A} - \mathbf{I})(\mathbf{x}^t - \overline{\mathbf{x}}) \tag{2.83}$$

$$= \mathbf{e}(\overline{\mathbf{x}}) + \frac{\partial \mathbf{e}}{\partial \mathbf{x}}(\mathbf{x}^t - \overline{\mathbf{x}}) \tag{2.84}$$

we see that the model bias error $\mathbf{e}(\overline{\mathbf{x}})$, representing the error at the reference state $\overline{\mathbf{x}}$, accounts for the bulk of the error caused by differences between the inverse and forward models while the null space term represents a linear adjustment to $\mathbf{e}(\overline{\mathbf{x}})$ that depends on inverse and forward model sensitivity differences ($\mathbf{A}$) at $\overline{\mathbf{x}}$. Thus, even if $\mathbf{A} = \mathbf{I}$ at $\overline{\mathbf{x}}$, $\mathbf{e}(\overline{\mathbf{x}})$ may not vanish. Also, we note that choosing $\overline{\mathbf{x}}$ close to $\mathbf{x}^t$ can reduce the null space systematic error but will not affect $\mathbf{S}_{\mathbf{x}^t}$ or $\mathbf{S}_N$.

The linearized equations for the forward model, inverse model, and transfer function governing $\mathbf{y}^m$ and $\mathbf{x}^r$ are given in Appendix A along with the covariances for both.

SECTION 3

THE SAGE II EXPERIMENT

This section contains background information on the SAGE II instrument and observational characteristics. Section 3.1 traces the historical development of the SAGE series of instruments, instrument operational characteristics, and measurement coverage. Because the water vapor inversion requires the simultaneous ozone and aerosol profile data obtainable from the other channels, the two step procedure that first separates the slant path optical depths by species and then inverts to yield vertical profiles of aerosol, ozone and nitrogen dioxide is briefly described in Sec. 3.2. The section closes in Sec. 3.3 with a presentation of typical slant path optical depths measured by SAGE II.

3.1 Experiment Description

SAGE II was launched as part of the Earth Radiation Budget Satellite(ERBS) on October 5, 1984 during shuttle mission 41-G. Scientific data acquisition began October 24 and remains operational at the time of this writing. SAGE II is a third generation satellite experiment that originated with SAM II, a single channel instrument that measured aerosol extinction at 1000 nm primarily in the polar regions (McCormick et al., 1979). The second instrument, SAGE, expanded on this concept with additional channels in the visible wavelength region to detect gaseous O_3 and NO_2 (McCormick et al., 1979). The SAGE II design incorporates additional aerosol channels to improve the aerosol wavelength characterization and provide size distribution information. The NO_2 absorption is covered with two narrow band channels instead of a single broadband channel to provide a low noise differential signal. The O_3 channel remained the same and a new channel for water vapor was added. The

full complement of SAGE II channels is listed in Table 3.1 along with the target and interfering constituents. A complete description of the SAGE II instrument may be found in McMaster(1986) and Mauldin et al.(1985).

Table 3.1: The SAGE II Channel Centers and Bandwidths

Channel	Center(Bandwidth) λ (nm)	Target Constituent	Interfering Constituents
1	1020 (20)	Aerosol	
2	940 (15)	H_2O	Aerosol, O_3
3	600 (15)	O_3	Aerosol
4	525 (15)	Aerosol	
5,6	453,448 (2,3)	NO_2	Aerosol
7	385 (20)	Aerosol	NO_2

SAGE II scans the solar disk during spacecraft sunrise and sunset events. Scans through the Earth's limb are normalized to exoatmospheric scans (above 100 km), this removes solar limb darkening and instrumental effects while converting the scan data to transmittances without having to perform a rigorous radiometric calibration. The exoatmospheric normalization performed for each measurement event makes the experiment essentially self-calibrating since instrument changes over periods greater than the event duration (30 to 90 seconds) are automatically removed. Tangent heights are calculated for each transmittance and the transmittances are placed in 1 km bins over the event altitude range. Mean values and standard deviations of the binned data define the slant path atmospheric transmittance and associated uncertainty versus altitude. To improve the noise characteristics of some channels, ratioed measurements are created by ratioing channel transmittances before the altitude binning is performed.

The solar occultation technique used by SAGE II produces 15 sunrise and sunset measurements each day evenly spaced over longitude (by 24° between adjacent events) along two separate parallels of latitude. Latitudinal measurement coverage

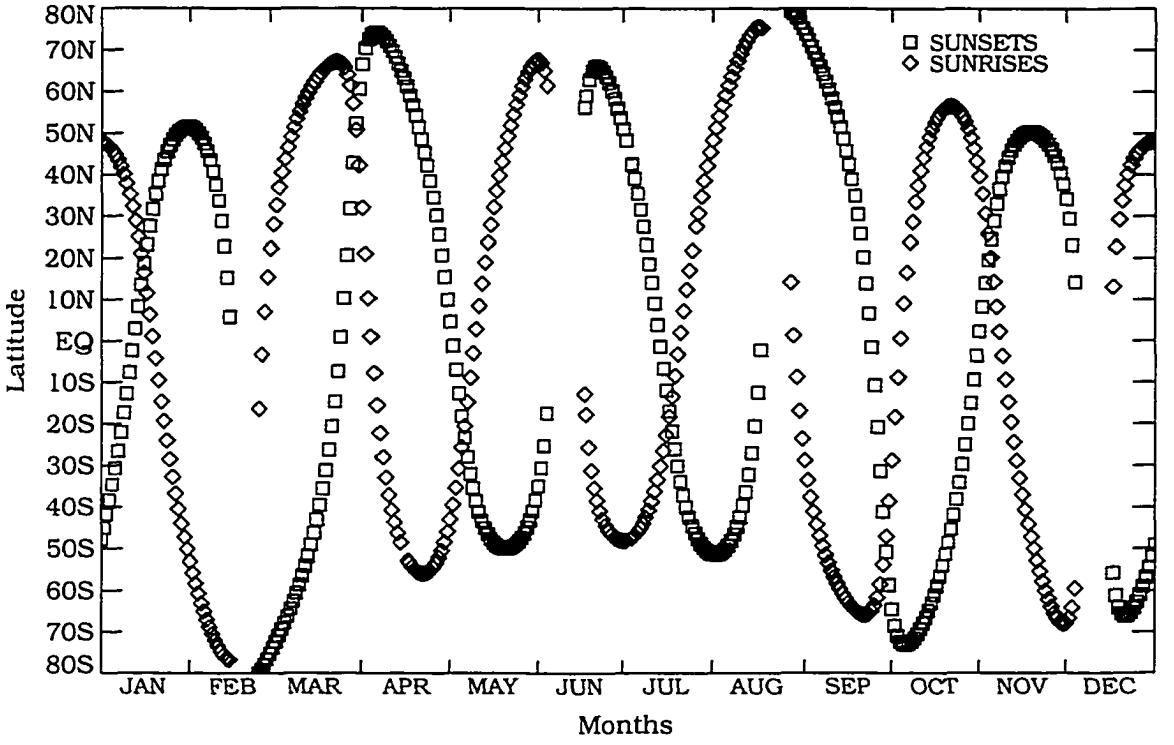


Figure 3.1: Typical SAGE II annual latitudinal measurement coverage. For clarity, only the first event from each day is plotted. Diamonds denote sunrise events and squares denote sunset events. Measurement patterns for specific years will be similar to this but slightly shifted in time.

for a typical year is shown in Fig. 3.1. Only the first sunrise or sunset event of each day is plotted. Sunrise events are indicated with diamonds while sunset events are indicated with squares. The measurement parallels sweep north or south to provide global coverage ranging from 80° S to 80° N at different times of the year. The tropics and middle latitudes are measured every two to three weeks while the high latitudes are measured less often. The measurement parallels sweep through the equatorial region rather quickly, at the rate of $\approx 9°$ per day and tend to dwell at the highest latitudes during sweep reversal. Evident in Fig. 3.1 are periods of no measurements, such as June 5 to June 16, when the sun-instrument geometry is such that the instrument remains in view of the sun at all times and occultation events

do not occur. Figure 3.2 displays the global measurement coverage for March 1987. At the beginning of March, sunset events are located at 73° and move northward throughout the month to 65° N. Sunrise events initially located in the northern extratropics move northward to ≈ 65° N and then south to the northern mid-latitudes.

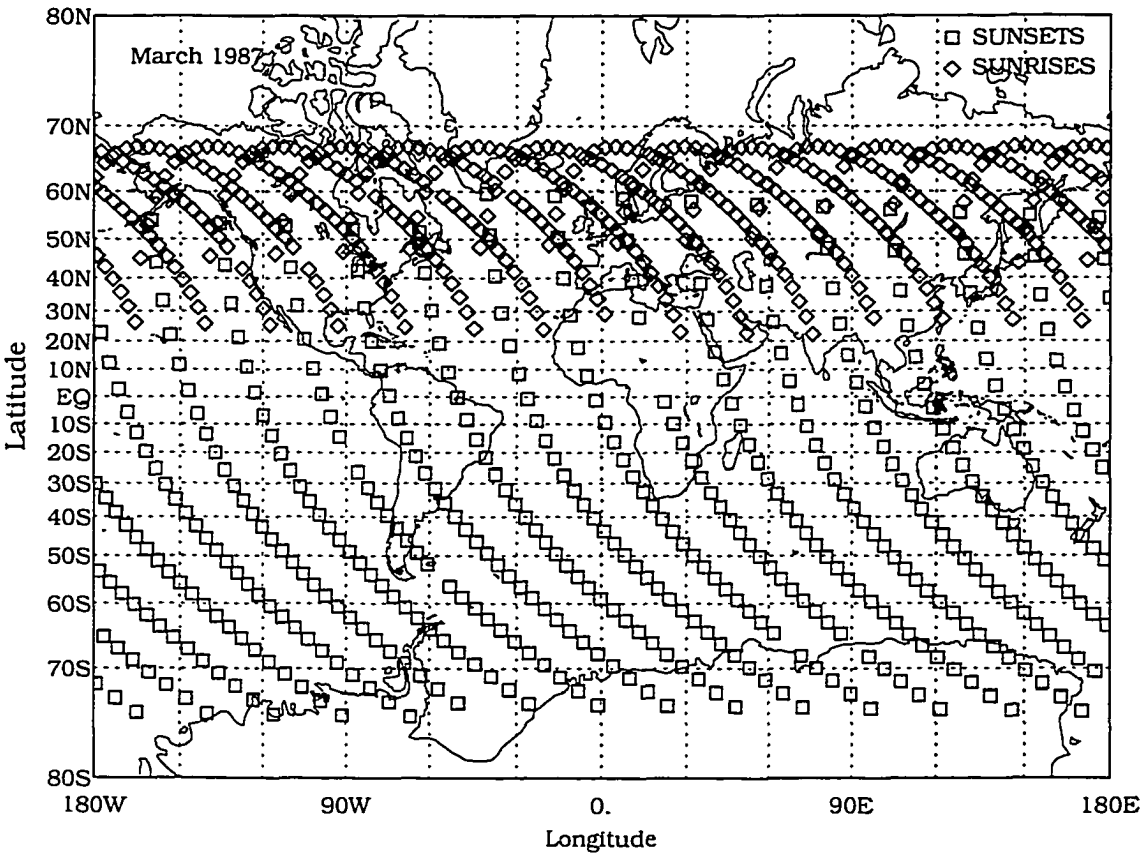


Figure 3.2: An example of global measurement coverage achievable for a single month. March 1987 is shown. Diamonds denote sunrise events and squares denote sunset events.

The spectral region covered by the water vapor channel includes a number of optically active constituents. Rayleigh or molecular scattering, not included in Table 3.1 since it can be calculated with reasonable accuracy given the appropriate temperature profile, produces slowly varying extinction proportional to λ^{-4} not only

throughout the water vapor channel but also in all other channels. Aerosol extinction, due to particulate scattering, is also found in all channels. Some attenuation due to ozone, resulting from the slow decay of the Chappuis band at 600 nm, is present in the water vapor channel. The water vapor band, while one of the strongest for the visible and near-ir, produces a relatively weak signal in the stratosphere that does however increase rapidly in the troposphere. Except for NO_2, all constituents contribute directly to the total signal in the water vapor channel.

Horizontal resolution, along the instrument line of sight, is approximately 200 km for an atmosphere discretized in 1 km thick shells. The instantaneous field of view gives a 0.5 km vertical resolution at the tangent height for each of the channels. The vertical resolution is degraded to 1-5 km by vertical smoothing implemented during the operational data processing. The strength of the smoothing is determined by the noise characteristics of each channel and may vary with altitude. For instance, the water vapor channel smoothing produces 3 km vertical resolution above 20 km and 1 km resolution below 20 km. The ozone channel vertical resolution is degraded to 5 km in the upper stratosphere because of the need to reduce measurement noise levels. Other channels are smoothed to 1 km. Vertical coverage varies from event to event and from channel to channel. Since the slant path optical depth increases downwards the upper limit is determined when the signal, in this case a change in transmission from unity, is discernible from the measurement noise. The lower limit is fixed when the atmospheric transmission drops below the noise level at some altitude in the troposphere. This can be caused by either increased opacity due to the atmospheric constituents or the presence of clouds. For water vapor the upper limit is $\approx$ 45 km and almost all profiles extend down to the tropopause. Penetration to levels below this depends on the distribution of clouds as shown in Table 1 of Rind et al.(1992). Typically, the upper troposphere is sampled 50% of the time and the middle troposphere is sampled 25% of the time relative to the number of

stratospheric measurements. Because the water vapor inversion must be terminated when clouds are detected in the field of view, moist atmospheric states, associated with convective activity and cloud formation, will be undersampled.

3.2 Inversion of Aerosol, Ozone and Nitrogen Dioxide

As shown in Table 3.1, the SAGE II water vapor channel contains spectral signatures from aerosol and ozone. To remove these, information from the other six channels is required. Operationally, this is achieved by first inverting the other six channels to produce profiles of ozone, nitrogen dioxide and aerosol extinction as will be described briefly in the next section.

3.2.1 Forward Models

Details of the instrument configuration, operation and data acquisition may be found in Chu and McCormick(1979) and the description of the SAGE II inversion algorithm is contained in Chu et al.(1989). Development of the governing equations here follows that of Chu et al.(1989). The experimental geometry is shown in Fig. 3.3.

SAGE II, located on the ERB Satellite moving with velocity $\vec{V}_{\text{Sat}}$ in a circular orbit, repeatedly scans the solar disk vertically during each solar occultation event. The instantaneous irradiance H_i for channel i at time t is defined as

$$H_i(t) = \int_{\Delta\lambda}\int_{\Delta\omega} W_\lambda(\theta,\phi)S_\lambda(\theta,\phi,t)T_\lambda(\theta)d\Omega d\lambda \tag{3.1}$$

where W is the radiometer field of view function, S is the exoatmospheric solar radiance at wavelength λ, T is the atmospheric transmittance, ϕ is the azimuthal angle, and Ω is the solid angle. A one to one correspondence exists between θ, the view angle and z_{th}, the tangent height.

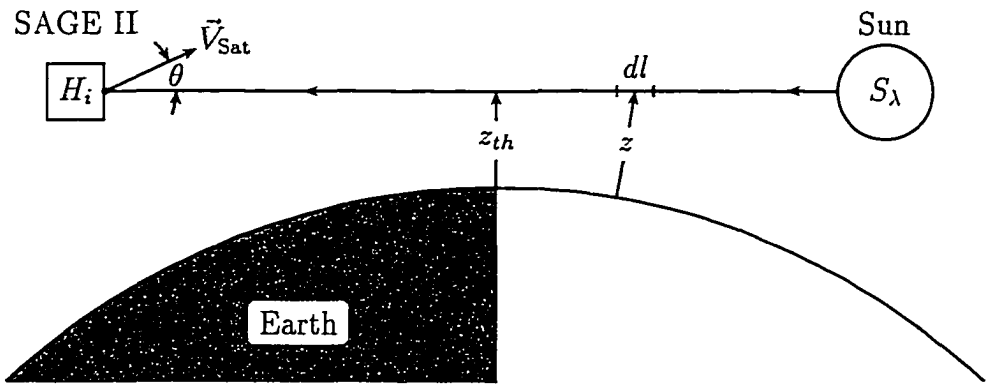


Figure 3.3: Occultation geometry.

At tangent heights located above the radiatively active portion of the atmosphere (T_λ=1) an exoatmospheric irradiance may be defined as

$$H_i^{exo} = \int_{\Delta\lambda}\int_{\Delta\omega} W_\lambda(\theta,\phi)S_\lambda(\theta,\phi,t)d\Omega d\lambda \tag{3.2}$$

When converted from view angle to tangent altitude and expressed as the ratio of the atmospheric irradiance to exoatmospheric irradiance the expression for atmospheric transmittance becomes

$$\overline{T}_i(z_{th}) = H_i(z_{th})/H_i^{exo} \tag{3.3}$$

This defines the mean atmospheric transmittance $\overline{T}_i(z_{th})$ over the spectral bandwidth of channel i and instrumental field of view function W as a function of z_{th}. With the help of Eqs. (2.6) and (2.26), the mean transmittance $\overline{T}_i(z_{th})$ can be expressed in terms of atmospheric properties as

$$\overline{T}_i(z_{th}) = \exp[-\delta_i(z_{th})] = \exp[-2\int_{z_{th}}^{z_{top}} \sigma_i(z)dl_i(z)] \tag{3.4}$$

The total extinction coefficient (Eq. (2.8)), as a function of altitude z, is a summation of the extinction due to each of the atmospheric species that interacts with

the radiation field within the wavelength range of channel i

$$\sigma_i(z) = \sigma_i^{\mathrm{Ray}}(z) + \sigma_i^{\mathrm{O_3}}(z) + \sigma_i^{\mathrm{NO_2}}(z) + \sigma_i^{\mathrm{Aer}}(z) \tag{3.5}$$

where the superscripts $^{\mathrm{Ray}}$,$^{\mathrm{O_3}}$,$^{\mathrm{NO_2}}$,$^{\mathrm{Aer}}$ denote extinction due to Rayleigh scattering, ozone absorption, nitrogen dioxide absorption, and aerosol scattering, respectively. The absorption cross sections of ozone and nitrogen dioxide exhibit little pressure or temperature dependence at the SAGE II wavelengths so the extinctions are simply the product of the band averaged absorption cross section with the species number density or

$$\begin{aligned} \sigma_i^{\mathrm{O_3}}(z) &= k_i^{\mathrm{O_3}} C^{\mathrm{O_3}}(z) \\ \sigma_i^{\mathrm{NO_2}}(z) &= k_i^{\mathrm{NO_2}} C^{\mathrm{NO_2}}(z) \end{aligned} \tag{3.6}$$

with k_i the channel averaged absorption cross section and $C(z)$ is the respective gaseous number density profile. Aerosol is somewhat more complicated because it is a function of aerosol size distribution, shape, and index of refraction (which is determined by the aerosol composition) as presented in Eq. (2.12) for a spherical homogeneous aerosol. Aerosol size distributions are not calculated explicitly as part of the operational inversion because of uncertainties related to composition, sphericity, and homogeneity. Instead, aerosol extinction coefficients are reported.

3.2.2 Inversion Algorithm

The inversion algorithm converts the slant path mean atmospheric transmission (or equivalently, optical depth) of the six channels into profiles describing the vertical distribution of ozone, nitrogen dioxide, and aerosol extinction. Details of the water vapor inversion are deferred to Sec. 4. The slant path optical depth may also be expressed as the summation of optical depths produced by each of the optically active species in channel i. Analogous to Eq. (3.5), the optical depth components

are

$$\delta_i(z_{th}) = \delta_i^{\mathrm{Ray}}(z_{th}) + \delta_i^{\mathrm{O_3}}(z_{th}) + \delta_i^{\mathrm{NO_2}}(z_{th}) + \delta_i^{\mathrm{Aer}}(z_{th}) \tag{3.7}$$

In the development that follows, the explicit dependence of δ_i on z_{th} will not be shown since the steps taken to reduce the equation set require manipulation of terms only within the specified tangent height. While Rayleigh and aerosol scattering is present in all six channels, ozone is not found in channels 1 or 7, and nitrogen dioxide is absent from channel 1. The first step in the inversion is the calculation and removal of the Rayleigh component. Rayleigh slant path optical depths are calculated for each channel at z_{th} using an atmospheric temperature profile provided by the National Weather Service corresponding to the observation time and location. After subtracting the Rayleigh component and dropping terms that do not contribute, the set of six equations describing the SAGE II measurements become

$$\delta_1' = \delta_1 - \delta_1^{\mathrm{Ray}} = \delta_1^{\mathrm{Aer}} \tag{3.8}$$

$$\delta_3' = \delta_3 - \delta_3^{\mathrm{Ray}} = \delta_3^{\mathrm{O_3}} + b_3\delta_6^{\mathrm{NO_2}} + \delta_3^{\mathrm{Aer}} \tag{3.9}$$

$$\delta_4' = \delta_4 - \delta_4^{\mathrm{Ray}} = a_4\delta_3^{\mathrm{O_3}} + b_4\delta_6^{\mathrm{NO_2}} + \delta_4^{\mathrm{Aer}} \tag{3.10}$$

$$\delta_5' = \delta_5 - \delta_5^{\mathrm{Ray}} = a_5\delta_3^{\mathrm{O_3}} + b_5\delta_6^{\mathrm{NO_2}} + \delta_5^{\mathrm{Aer}} \tag{3.11}$$

$$\delta_6' = \delta_6 - \delta_6^{\mathrm{Ray}} = a_6\delta_3^{\mathrm{O_3}} + b_6\delta_6^{\mathrm{NO_2}} + \delta_6^{\mathrm{Aer}} \tag{3.12}$$

$$\delta_7' = \delta_7 - \delta_7^{\mathrm{Ray}} = b_7\delta_6^{\mathrm{NO_2}} + \delta_7^{\mathrm{Aer}} \tag{3.13}$$

The slant path optical depth minus the Rayleigh component is denoted by δ_i'. The contribution of ozone in each of the channels has been referenced to $\delta_3^{\mathrm{O_3}}$. Likewise, the nitrogen dioxide contribution is referenced to $\delta_6^{\mathrm{NO_2}}$. Thus, the a_i's and b_i's are defined by

$$a_i = \delta_i^{\mathrm{O_3}}/\delta_3^{\mathrm{O_3}} \tag{3.14}$$

and

$$b_i = \delta_i^{NO_2}/\delta_6^{NO_2} \tag{3.15}$$

The a_i's and b_i's are tabulated in Table 1 of Chu et al.(1989) and were calculated by convolving laboratory absorption spectra with the measured instrument spectral response function.

The six equations for δ_i' contain contributions from each of the species and the next step is to separate the species to the greatest extent possible. The nitrogen dioxide contribution in channel 6, $\delta_6^{NO_2}$, is separated from the set of equations by taking the difference between channels 5 and 6 to yield

$$\delta_D^{NO_2} = \delta_6' - \delta_5' = \delta_6^{Aer} - \delta_5^{Aer} + (a_6 - a_5)\delta_3^{O_3} + (b_6 - b_5)\delta_6^{NO_2} \tag{3.16}$$

Since channels 5 and 6 have a small spectral separation it is reasonable to assume that $\delta_6^{Aer} - \delta_5^{Aer} \approx 0.$ and if one also assumes that ozone is the predominant term in channel 3 at the nitrogen dioxide inversion altitudes then $\delta_3^{O_3} = \delta_3$. With these assumptions Eq. (3.16) provides an estimate for $\delta_6^{NO_2}$ as

$$\delta_6^{NO_2} = \delta_D^{NO_2} - (a_6 - a_5)\delta_3/(b_6 - b_5) \tag{3.17}$$

Inserting the estimated $\delta_6^{NO_2}$ into the set of equations leads to

$$\delta_1'' \quad = \delta_1' \quad = \delta_1^{Aer} \tag{3.18}$$

$$\delta_3'' = \delta_3' - b_3\delta_6^{NO_2} = \delta_3^{O_3} + \delta_3^{Aer} \tag{3.19}$$

$$\delta_4'' = \delta_4' - b_4\delta_6^{NO_2} = a_4\delta_3^{O_3} + \delta_4^{Aer} \tag{3.20}$$

$$\delta_5'' = \delta_5' - b_5\delta_6^{NO_2} = a_5\delta_3^{O_3} + \delta_5^{Aer} \tag{3.21}$$

$$\delta_7'' = \delta_7' - b_7\delta_6^{NO_2} = \delta_7^{Aer} \tag{3.22}$$

The equations defining the slant path optical depth δ_i'' after removal of the Rayleigh and nitrogen dioxide components still contain overlap between ozone and

aerosol. Each channel contains an aerosol term and several channels contain ozone for a total of six unknowns in the five equations. To solve this set of equations in a unique way, additional information concerning the spectral variation of the slant path optical depth is incorporated. Equation (2.12) is recast to express δ_i^{Aer} as a function of the extinction cross section and integrated slant path aerosol size distribution. Through a series of steps described in detail in Chu et al.(1989), one solves implicitly for the aerosol size distribution using the slant path optical depths $\delta_1^{\mathrm{Aer}}, \delta_4^{\mathrm{Aer}}, \delta_5^{\mathrm{Aer}}$, and δ_7^{Aer} and then uses the size distribution to calculate δ_3^{Aer}. Subtracting δ_3^{Aer} from Eq. (3.18) yields $\delta_3^{\mathrm{O_3}}$. The separation of species, starting with the nitrogen dioxide estimate given by Eq. (3.17), is repeated once more using updated ozone and aerosol values. Because nitrogen dioxide contributes in a minor way to the channels determining ozone and aerosol there is little change in these values on the second iteration. The two step species separation is carried out for each z_{th} to build up vertical distributions of $\delta_3^{\mathrm{O_3}}, \delta_6^{\mathrm{NO_2}}, \delta_1^{\mathrm{Aer}}, \delta_4^{\mathrm{Aer}}, \delta_5^{\mathrm{Aer}}, \delta_7^{\mathrm{Aer}}$ which are inverted to yield vertical profiles of ozone, nitrogen dioxide and aerosol extinction for channels 1, 4, 5, and 7. Details of the inversion process are not presented here since only the slant path optical depths are needed to perform the water vapor inversion.

3.3 Slant Path Optical Depth Characteristics

Typical signal levels of the water vapor channel and adjacent channels making the greatest information contribution to the water vapor inversion are shown in Fig. 3.4. Channel 3 provides information on ozone and channels 1 and 4 help define the aerosol spectral variations. The solid line indicates the monthly zonal mean of the slant path optical depth using all observations in a 10° latitude band centered at 40° N. The dashed line is the standard deviation of the slant path optical depth and can be considered a combination of the geophysical variability of the signal around

the globe within the latitude band and experimental error. The dotted line shows the typical estimated error for a single observation and the dot-dash lines bracketing this line indicate its range($\pm\sigma$). The estimated error includes instrument error, altitude binning errors, and geophysical variability (or vertical variation) within the 1 km vertical altitude bin. The average estimated error should always be smaller than the standard deviation of the slant path optical depth.

The slant path optical depth varies over four orders of magnitude in each of the channels from the surface to 60 km. Due to the Rayleigh and aerosol extinction spectral variation the channels display increasing optical depths towards the shorter wavelengths. The standard deviation of the mean optical depth is greatest in the troposphere and decreases upwards. The estimated error also follows this trend. In channels 1 and 2 the presence of aerosol between 15 and 30 km is manifested not only in the zonal mean but also in the standard deviation and estimated error. In these two channels the estimated error and standard deviation of the mean are approximately the same in the upper stratosphere indicating either little global geophysical variability at these levels or an overestimate of the experimental error. It will be shown in Sec. 4 that the estimated error for channel 2 in the stratosphere is greater than the actual error.

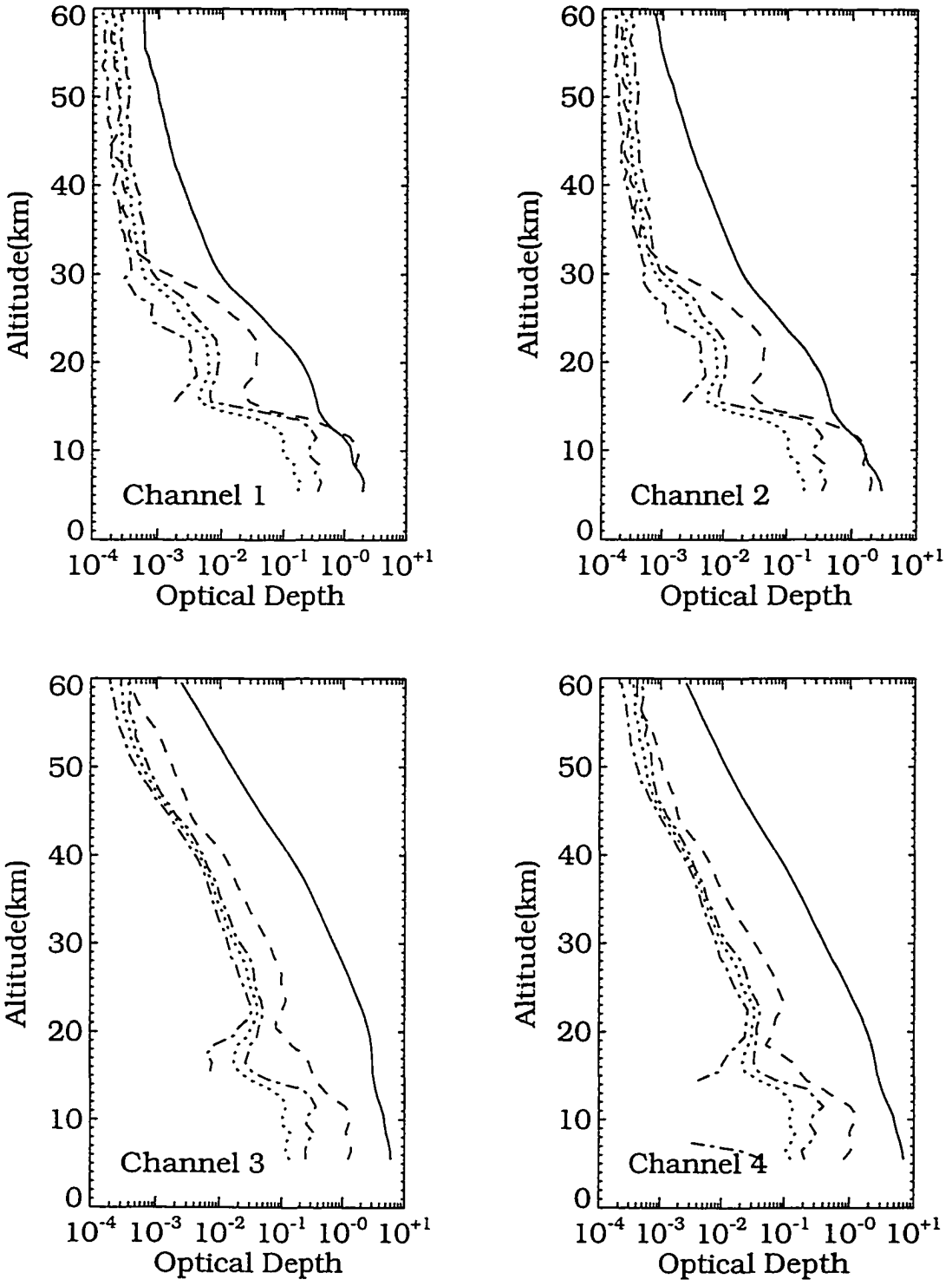


Figure 3.4: Average slant path optical depth(solid line), standard deviation(dashed line), average estimated error(dotted line), estimated error standard deviation(dash-dot line) for the 40° N latitude band, February 1987.

SECTION 4

INVERSION ALGORITHM COMPARISON AND ERROR ANALYSIS

Extraction of useful information from indirect observations requires a full understanding of the physical processes linking the desired information to the observations and a fast, accurate, and stable solution method supported with accurate error estimates. In this section four inversion algorithms are investigated as possible alternatives to the existing operational algorithm (Chu et al.,1993) with the aid of a formal error analysis.

Sections 4.1 and 4.2 begin the section with the development of the SAGE II water vapor forward model and inverse model for each of the inversion algorithms. The significant experimental parameters are identified and quantified in Sec. 4.3. Next, in Sec. 4.4, the formal error analysis is applied to each of the algorithms to develop an estimate of the total experimental uncertainty and then each algorithm is used to invert two years of data (1987,1990). The character of the inverted profiles is analyzed with respect to that expected from the formal error analysis and from inversion of synthetic observations with known errors in Sec. 4.5. The results of this section are summarized in Sec. 4.6 with algorithm comparisons and a discussion of the advantages and disadvantages of each method.

4.1 SAGE II Forward Water Vapor Model

Unlike ozone and nitrogen dioxide, which exhibit smooth spectrally varying absorption cross-sections in their primary channels, water vapor absorption in this channel, and throughout the visible region, is caused by vibrational-rotational transitions that produce cross-sections with strong spectral variations. These must be

accounted for by incorporating the channel spectral response in the forward model. The development of the water vapor model is analogous to that found in Sec. 3.2.1 for the other channels. The instantaneous irradiance for channel 2 is given by Eq. (3.1) modified to include the channel spectral response explicitly through the filter function FF, *i.e.*,

$$H_2(t) = \int_{\Delta\lambda}\int_{\Delta\omega} W_\lambda(\theta,\phi)S_\lambda(\theta,\phi,t)FF(\lambda)T_\lambda(\theta)d\Omega d\lambda \tag{4.1}$$

The filter function $FF(\lambda)$ was obtained from preflight laboratory measurements and is shown in Fig. 4.1.

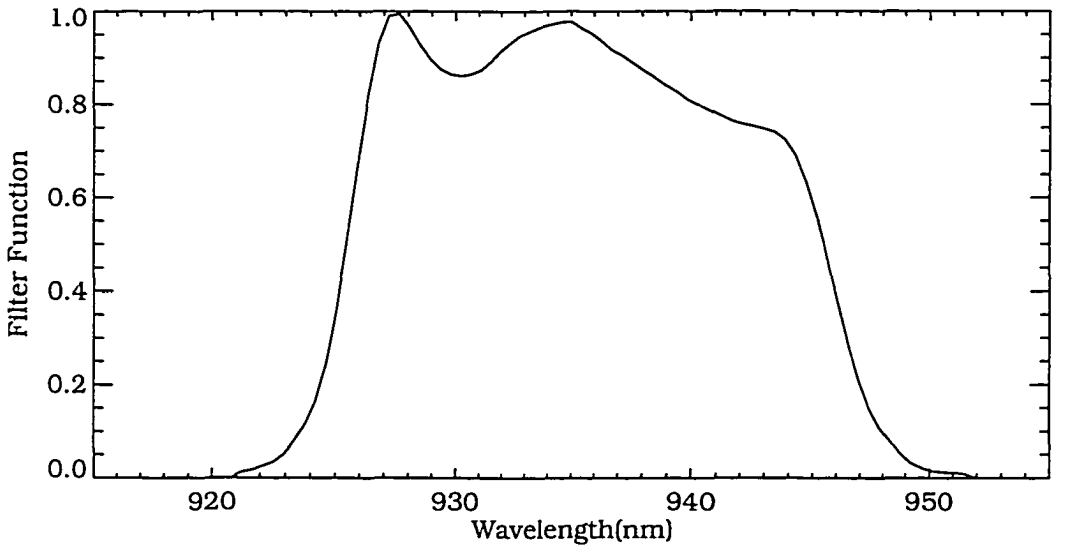


Figure 4.1: The SAGE II water vapor channel filter function.

The exoatmospheric irradiance is

$$H_2^{exo} = \int_{\Delta\lambda}\int_{\Delta\omega} W_\lambda(\theta,\phi)S_\lambda(\theta,\phi,t)FF(\lambda)d\Omega d\lambda \tag{4.2}$$

Again expressing the mean transmittance (in terms of z_{th}) as a ratio of the instantaneous and exoatmospheric irradiance one has

$$\overline{T}_2(z_{th}) = H_2(z_{th})/H_2 = \frac{\int_{\Delta\lambda}\int_{\Delta\omega} W_\lambda(\theta,\phi)S_\lambda(\theta,\phi,t)FF(\lambda)T_\lambda(\theta)d\Omega d\lambda}{\int_{\Delta\lambda}\int_{\Delta\omega} W_\lambda(\theta,\phi)S_\lambda(\theta,\phi,t)FF(\lambda)d\Omega d\lambda} \tag{4.3}$$

Provided that most of the wavelength dependence is contained in $FF(\lambda)$ and T_λ, and that $T_\lambda(\theta)$ has minimal θ dependence over $d\Omega$, the terms $FF(\lambda)$ and T_λ may be removed from the solid angles integral, and the solid angle and wavelength integrations separate. If W_λ is constant during the measurement event and the portion of the solar disk viewed through the atmosphere is properly located in the exoatmospheric solar limb scan, such that $S_\lambda(z_{th}) \approx S_\lambda(exo)$, then the solid angle integrals cancel. This leaves

$$\overline{T}_2(z_{th}) = \frac{\int_{\Delta\lambda} FF(\lambda)T_\lambda(z_{th})d\lambda}{\int_{\Delta\lambda} FF(\lambda)d\lambda} \tag{4.4}$$

Defining $\int_{\Delta\lambda} FF(\lambda)d\lambda = 1$, then the above expression becomes

$$\overline{T}_2(z_{th}) = \int_{\Delta\lambda} FF(\lambda)T_\lambda(z_{th})d\lambda \tag{4.5}$$

The transmittance $\overline{T}_2(z_{th})$ can be separated into spectrally slow (Other) and fast (H_2O) components as follows

$$\overline{T}_2(z_{th}) = \overline{T}_{\overline{\lambda}}^{Other}(z_{th}) \int_{\Delta\lambda} FF(\lambda)T_\lambda^{H_2O}(z_{th})d\lambda \tag{4.6}$$

where $\overline{\lambda}$ indicates channel mean values. The spectrally slow transmission component (Eqs. (3.4), (3.5) and (3.7) applied to channel 2) is defined by

$$\begin{aligned} \overline{T}_{\overline{\lambda}}^{Other}(z_{th}) &= \exp[-\delta_2^{Other}(z_{th})] = \exp[-2\int_{z_{th}}^{z_{top}} \sigma_2^{Other}(z)dl(z)] \\ \delta_2^{Other} &= \delta^{Ray}(\lambda_2) + \delta^{O_3}(\lambda_2) + \delta^{Aer}(\lambda_2) \\ \sigma_2^{Other} &= \sigma^{Ray}(\lambda_2) + \sigma^{O_3}(\lambda_2) + \sigma^{Aer}(\lambda_2) \end{aligned} \tag{4.7}$$

The spectrally fast component, due to water vapor, defines a channel mean water vapor transmission weighted by the filter function as

$$\overline{T}^{H_2O}(z_{th}) = \int_{\Delta\lambda} FF(\lambda)T_\lambda^{H_2O}(z_{th})d\lambda \tag{4.8}$$

The monochromatic water vapor transmission $T_\lambda^{H_2O}(z_{th})$ is defined by Eq. (2.10) with $\sigma_\lambda^g = \sigma_\lambda^{H_2O}$

$$T_\lambda^{H_2O}(z_{th}) = \exp[-2\int_{z_{th}}^{z_{top}} k_\lambda(z)V^{H_2O}(z)C^{Air}(z)dl(z)] \tag{4.9}$$

The absorption cross-section for water, $k_\lambda(z)$, is given by Eq. (2.25). The water vapor mixing ratio, $V^{H_2O}(z)$, is the unknown quantity to be inferred from the transmission measurements and $C^{Air}(z)$ is the number density of air.

A significant reduction in measurement noise can be achieved by manipulating channels 1 and 2 to form a transmission ratio of channel 2 to channel 1. As it turns out, this ratio exhibits a higher signal to noise ratio than that of channel 2 alone, some of the improvement appears to be due to cancellation of correlated errors in the channels. This transmission ratio is given by

$$T_{2/1}(z_{th}) = \frac{\overline{T}_{\overline{2}}^{Other}(z_{th})}{\overline{T}_{\overline{1}}^{Other}(z_{th})} \overline{T}^{H_2O}(z_{th}) \tag{4.10}$$

Equation (4.10) is obtained from Eqs. (4.6) and (4.7) and is performed prior to altitude binning. Further consequences of this approach are that while the water vapor signal remains unchanged, the transmission ratio of the other species is smaller than $\overline{T}_{\overline{2}}^{Other}$. Thus, the water vapor signal is a greater fraction of the measured signal in this case. From Eq. (4.7) the transmission ratio of the other species may be written as

$$\begin{aligned} \frac{\overline{T}_{\overline{2}}^{Other}(z_{th})}{\overline{T}_{\overline{1}}^{Other}(z_{th})} &= \exp[-\Delta\delta_{2-1}^{Other}(z_{th})] = \exp[-2\int_{z_{th}}^{z_{top}} \Delta\sigma_{2-1}^{Other}(z)dl(z)] \qquad (4.11) \\ \Delta\delta_{2-1}^{Other} &= \delta^{Ray}(\lambda_2) - \delta^{Ray}(\lambda_1) + \delta^{O_3}(\lambda_2) + \delta^{Aer}(\lambda_2) - \delta^{Aer}(\lambda_1) \\ \Delta\sigma_{2-1}^{Other} &= \sigma^{Ray}(\lambda_2) - \sigma^{Ray}(\lambda_1) + \sigma^{O_3}(\lambda_2) + \sigma^{Aer}(\lambda_2) - \sigma^{Aer}(\lambda_1) \end{aligned}$$

Note that the ratioing process has not introduced additional unknowns since the channel 1 variables are already available following the inversion of the other channels as described in Sec. 3.2.2. The channel 2 unknowns must be specified in some manner before Eq. (4.10) can be solved for $\overline{T}^{H_2O}(z_{th})$. Details of this process will be presented in a later section. Each of the variables in $\Delta\delta_{2-1}^{Other}$ or $\Delta\sigma_{2-1}^{Other}$ transfers its random noise and systematic biases to $\overline{T}^{H_2O}(z_{th})$ via Eq. (4.10). Of course, the $T_{2/1}(z_{th})$ measurements also contain random and systematic errors, and all of these must be considered when developing the water vapor error budget.

4.2 SAGE II Inverse Model

The overall inversion process can be stated as follows. First, calculate $\frac{\overline{T}_2^{Other}(z_{th})}{\overline{T}_1^{Other}(z_{th})}$ and solve for $\overline{T}^{H_2O}(z_{th})$ from Eq. (4.10). Next, invert $\overline{T}^{H_2O}(z_{th})$ for $V^{H_2O}(z)$. Since these two steps are independent of one another, we have subdivided the inverse model into sub-models as shown diagrammatically in Fig. 4.2. The separation model

$$\mathbf{y}^m \longrightarrow I(\mathbf{y}^m, \hat{\mathbf{b}}, \hat{\mathbf{c}}) \longrightarrow \mathbf{x}^r$$

$$\mathbf{y}^m \longrightarrow S(\mathbf{y}^m, \hat{\mathbf{b}}, \hat{\mathbf{c}}) \longrightarrow \mathbf{y}^{mi} \longrightarrow IA(\mathbf{y}^{mi}, \hat{\mathbf{b}}, \hat{\mathbf{c}}) \longrightarrow \mathbf{x}^r$$

Figure 4.2: Division of the inverse model (I) into two sub-models representing the removal of interfering species signatures (S) and water vapor inversion algorithm(IA).

S removes the Rayleigh, ozone and aerosol signatures, as defined by $\frac{\overline{T}_2^{Other}(z_{th})}{\overline{T}_1^{Other}(z_{th})}$ from the measurement $T_{2/1}(z_{th})$ ($\mathbf{y}^m$) to give an intermediate measurement $\overline{T}^{H_2O}(z_{th})$ ($\mathbf{y}^{mi}$) containing random error designated as ϵ_s. The inversion algorithm model, IA, performs the inversion. The accuracy and precision of the final water vapor data product thus depends not only on the quality of the "raw" channel 2 data, but also on the quality of the measurement data from the other channels.

Breaking up the inverse model into two sub-models introduces equations for $\mathbf{y}^{mi}$ in addition to those found in Table A.1. They are

$$\begin{aligned} \mathbf{y}^{mi} &= S(\mathbf{y}^m, \hat{\mathbf{b}}, \hat{\mathbf{c}}) \\ &= S(\overline{\mathbf{y}}^m, \hat{\mathbf{b}}, \hat{\mathbf{c}}) + \frac{\partial S}{\partial \mathbf{y}^m}(\mathbf{y}^m - \overline{\mathbf{y}}^m) \qquad (4.12) \\ \mathbf{S}_{\mathbf{y}^{mi}} &= (\frac{\partial S}{\partial \mathbf{y}^m})\mathbf{S}_{\mathbf{y}^m}(\frac{\partial S}{\partial \mathbf{y}^m})^T \\ &= (\frac{\partial S}{\partial \mathbf{y}^m})[\mathbf{K}_\mathbf{x}\mathbf{S}_{\mathbf{x}^t}\mathbf{K}_\mathbf{x}^T + \mathbf{K}_\mathbf{b}\mathbf{S}_{\epsilon_\mathbf{b}}\mathbf{K}_\mathbf{b}^T + \mathbf{S}_{\epsilon_\mathbf{y}}](\frac{\partial S}{\partial \mathbf{y}^m})^T \qquad (4.13) \end{aligned}$$

The covariance of the random error ϵ_s is given by

$$\mathbf{S}_{\epsilon_S} = (\frac{\partial S}{\partial \mathbf{y}^m})\mathbf{S}_{\epsilon_Y}(\frac{\partial S}{\partial \mathbf{y}^m})^T \tag{4.14}$$

and

$$\begin{aligned} \mathbf{x}^r &= IA(\mathbf{y}^{mi}, \hat{\mathbf{b}}, \hat{\mathbf{c}}) \\ &= IA(\overline{\mathbf{y}}^{mi}, \hat{\mathbf{b}}, \hat{\mathbf{c}}) + \frac{\partial IA}{\partial \mathbf{y}^{mi}}(\mathbf{y}^{mi} - \overline{\mathbf{y}}^{mi}) \qquad (4.15) \\ \mathbf{S}_{\mathbf{x}^r} &= (\frac{\partial IA}{\partial \mathbf{y}^{mi}})\mathbf{S}_{\mathbf{y}^{mi}}(\frac{\partial IA}{\partial \mathbf{y}^{mi}})^T \\ &= (\frac{\partial IA}{\partial \mathbf{y}^{mi}})(\frac{\partial S}{\partial \mathbf{y}^m})[\mathbf{K}_\mathbf{x}\mathbf{S}_{\mathbf{x}^t}\mathbf{K}_\mathbf{x}^T + \mathbf{K}_\mathbf{b}\mathbf{S}_{\epsilon_\mathbf{b}}\mathbf{K}_\mathbf{b}^T + \mathbf{S}_{\epsilon_\mathbf{y}}](\frac{\partial S}{\partial \mathbf{y}^m})^T(\frac{\partial IA}{\partial \mathbf{y}^{mi}})^T \qquad (4.16) \end{aligned}$$

Since $\mathbf{D}_\mathbf{y} = \frac{\partial IA}{\partial \mathbf{y}^{mi}}\frac{\partial S}{\partial \mathbf{y}^m}$, equations for $\mathbf{x}^r - \mathbf{x}^t$ still hold and we obtain the previous expression for $\mathbf{S}_E$

$$\mathbf{S}_E = (\frac{\partial IA}{\partial \mathbf{y}^{mi}})\mathbf{S}_{\epsilon_S}(\frac{\partial IA}{\partial \mathbf{y}^{mi}})^T = \mathbf{D}_\mathbf{y}\mathbf{S}_{\epsilon_\mathbf{y}}\mathbf{D}_\mathbf{y}^T \tag{4.17}$$

Specifics of the two sub-models will now be presented.

4.2.1 SAGE II Separation Model

The separation model S is given as

$$\overline{T}^{\mathrm{H_2O}} = T_{2/1}\frac{\overline{T}_{\overline{1}}^{Other}}{\overline{T}_{\overline{2}}^{Other}} \tag{4.18}$$

or

$$-\ln(\overline{T}^{\mathrm{H_2O}}) = -\ln(T_{2/1}) - [\delta^{\mathrm{Ray}}(\lambda_2) - \delta^{\mathrm{Ray}}(\lambda_1)] - \delta^{\mathrm{O_3}}(\lambda_2) - [\delta^{\mathrm{Aer}}(\lambda_2) - \delta^{\mathrm{Aer}}(\lambda_1)] \tag{4.19}$$

when expressed in terms of optical depths for a single z_{th}. The Rayleigh components are calculated as described in Sec. 3.2.2 and the ozone signature in channel 2, $\delta^{\mathrm{O_3}}(\lambda_2)$ is modeled with $a_2\delta^{\mathrm{O_3}}(\lambda_3)$. The ozone contribution coefficient a_2, calculated from the optical depth measurements of Anderson et al.(1991), was found to be 0.0039. The aerosol contribution is modeled following Eq. (2.12) with a segmented power law particle size distribution (Thomason, 1991).

4.2.2 SAGE II Inversion Algorithm Models

Solving Fredholm equations of the first kind for an under-determined system of equations with overlapping kernels is difficult and instability can be a concern (Twomey, 1977). Kernels associated with Volterra equations of the first kind, produced by the occultation geometry, generally exhibit less overlap and greater independence. When Volterra equations are transformed into Fredholm equations (Linz, 1985), the reduced overlap appears as truncated kernels which lead to upper or lower triangular matrices in the discretized form. For a linear system of equations direct solution is acceptable but a non-linear equation set is best approached iteratively. Since the equations relating water vapor mixing ratio to absorption are nearly linear in the stratosphere and non-linear in the troposphere we will focus on iterative solution techniques.

Four other inversion algorithms were selected for comparison to the operational algorithm: onion peel, Mill-Drayson, smoothed Mill-Drayson, and sparse grid least squares. The onion peel inversion algorithm can be viewed as a simplified form of the operational algorithm that represents the simplest possible solution method. The Mill-Drayson algorithm works much like the onion peel except that multiple passes through the atmosphere from bottom to top are made. The smoothed Mill-Drayson includes stratospheric smoothing identical to the operational to see if the Mill-Drayson approach offers any advantages relative to the onion peel when combined with strong smoothing. In the last technique we attempt to mimic stratospheric smoothing by taking advantage of the water vapor profile smoothness to create an over-determined set of equations that is solved with a non-linear least squares technique.

In the description that follows, $\overline{T}(z_{th})$ or $\overline{A}(z_{th})$ will be understood to represent the mean channel transmission or absorption due solely to water vapor as described

by Eq. (4.8). The assumptions and simplifications developed in Sec. 2.2, used to model the inhomogeneous atmosphere with a series of homogeneous layers, are applicable to Eq. (4.9) and combined with Eq. (4.8) lead to

$$\overline{A}(z_{th}) = 1 - \overline{T}(z_{th}) = 1 - \int_{\Delta\lambda} FF(\lambda) \exp[-2 \sum_{j=1}^{N(i)} \overline{k}_\lambda(j) \overline{V}^{H_2O}(j) \overline{C}^{Air}(j) L_{i,j}] d\lambda \quad (4.20)$$

where

$$\overline{k}_\lambda(j) \overline{V}^{H_2O}(j) \overline{C}^{Air}(j) = \frac{1}{L_{i,j}} \int_{z_{j-1}}^{z_j} k_\lambda(z) V^{H_2O}(z) C^{Air}(z) dl(z) \quad (4.21)$$

The variables $\overline{k}_\lambda(j)$, $\overline{V}^{H_2O}(j)$, and $\overline{C}^{Air}(j)$ are evaluated at the tangent layer($i = j$) and it is assumed that the product of the means closely approximates the mean of the product for these variables.

ALGORITHM COMMONALITY

Before presenting each of the inversion techniques in detail, operations common to all are described.

The radiative transfer calculations are performed with the "emissivity growth approximation" (EGA) (Gordley and Russell, 1981) reformulated in terms of absorption. The absorption produced by an individual homogeneous layer, j, can be written as

$$A_j = A(T_j, P_j, U_j) \quad (4.22)$$

where T_j, P_j are layer averaged temperatures and pressures. The total number density, U_j, through the layer j, is given by

$$U_j = \overline{V}^{H_2O}(j) \overline{C}^{Air}(j) L_{i,j} \quad (4.23)$$

Total number density may also be referred to as layer number density or absorber amount. Defining A^j as the cumulative absorption up through layer j and expressing the absorption as a function of U with

$$A_j(U) = A_j(T_j, P_j, U) \quad (4.24)$$

the technique may be mathematically stated as

$$A^j = A_j(U_j + U_j^e) \tag{4.25}$$

The cumulative absorption through layers $1 \ldots j$ is expressed as a function of the sum of U_j, the actual total number density for layer j, and U_j^e, an effective total number density that accounts for the absorption due to all previous homogeneous layers $1 \ldots j-1$ and is defined by

$$A_j(U_j^e) = A^{j-1} \tag{4.26}$$

The process starts at the first level, $j = 1$, by calculating A^1 with $U_j^e = 0.0$ in Eq. (4.25). For each new layer j, one first computes $A_j(U)$ for layer j, solves Eq. (4.26) for U_j^e by interpolation, and evaluates Eq. (4.25) for A^j. The process continues until $j = N(i)$ and $A^{N(i)}$ is the slant path absorption for tangent height i. As noted in Gordley and Russell(1981), the technique is exact for the monochromatic case and becomes an approximation when applied to a channel averaged absorption as defined by Eq. (4.20). A onetime core database of $A(U)$ for a set of standard pressures and temperatures encompassing expected atmospheric extremes is constructed from which $A_j(T_j, P_j, U)$ for each homogeneous layer is derived, again by interpolation. The EGA technique replaces time consuming line-by-line monochromatic radiative transfer computations with a table lookup solution based on repeated interpolation, after the core database has been constructed.

A coarse first guess water vapor profile is constructed for each measurement that is used by the inversion algorithms as an iterative starting point. Above 10.5 km the profile is set to 4 ppmv everywhere. Strong latitudinal gradients in water vapor below this made it necessary to use three profiles representative of the equatorial, middle, and high latitudes to achieve convergence in a reasonable number of iterations.

Preconditioning of the slant path absorption profiles in the form of vertical smoothing with the purpose of reducing random noise is performed in each of the

algorithms. In the stratosphere the vertical smoothing is given by

$$\overline{A}(i) = exp[\{log(\overline{A}(i-1)) + log(\overline{A}(i)) + log(\overline{A}(i+1))\}/3] \tag{4.27}$$

for all i with z_{th} greater than 18.5 km. Somewhat weaker smoothing is applied in the troposphere since the signal-to-noise ratios and profile gradients are greater there. Below 18.5 km the vertical smoothing becomes

$$\overline{A}(i) = exp[.25log(\overline{A}(i-1)) + .5log(\overline{A}(i)) + .25log(\overline{A}(i+1))] \tag{4.28}$$

The vertical smoothing proceeds from profile top downwards.

Onion Peel Inversion

The inversion of the set of N^T equations given by Eq. (4.20) for N^T unknown $\overline{V}^{H_2O}(j)$ is performed following the sequence of back substitution steps outlined in Sec. 2.2. An iterative solution technique based on the secant method is used here, instead of the direct solution that was possible with the simplified equation set of Sec. 2.2, to accommodate the spectral integration over the channel bandpass. The inversion is performed in a single pass from top to bottom, at each tangent height a solution to Eq. (4.20) is obtained by iteratively adjusting the tangent layer U with the secant method. The secant method has been used previously in conceptual measurement feasibility studies (Allario et al, 1980a; Allario et al.,1980b) and was found to converge rapidly.

For a given tangent level, the total number density is updated by applying the secant method to solve

$$RE(U) = 0. \tag{4.29}$$

where RE is the relative error between the measured (A^m) and calculated (A^c) slant path absorption defined by

$$RE(U^n) = \frac{\overline{A}^m - \overline{A}^c(U^n)}{\overline{A}^m} \tag{4.30}$$

The quantity U^n is the tangent layer number density for iteration n and layer number densities above the tangent layer are fixed from previous solutions. The secant method requires two initial number density values before a new number density can be predicted, the first (U^1) is taken from the first guess profile and the second is predicted from $U^2 = U^1\overline{A}^m/\overline{A}^c(U^1)$. The relative absorption error is calculated for each and used to linearly predict U^3 with

$$U^3 = U^2 - RE(U^2)[\frac{U^2 - U^1}{RE(U^2) - RE(U^1)}] \tag{4.31}$$

Relative error $RE(U^3)$ is calculated and the two number densities corresponding to the lowest error are used for the next prediction. Occasionally, negative number densities are predicted and these are replaced with number densities from the previous tangent level. In some cases the denominator in the predictor approaches zero due to computer word size limitations and the secant predictor is replaced by

$$U^{n+1} = U^n\frac{\overline{A}^m}{\overline{A}^c(U^n)} \tag{4.32}$$

for that iteration. Updating continues until $|\overline{A}^m - \overline{A}^c(U^{n+1})| < \Delta_{cl}$ or 20 updates, where Δ_{cl} is the convergence error limit defined as the maximum allowable difference between $\overline{A}^m$ and $\overline{A}^c$ for which the iterative solution is acceptable. The secant method converges to within the error limit rapidly, generally within two to four iterations at all tangent levels. This solution technique is very stable in the presence of noise. Incorrect absorption values deliberately introduced in the slant path absorption profile at a particular tangent altitude were found to cause an error in the inverted number density at the tangent level that propagated downwards in an oscillatory fashion two to three layers before damping out. This is to be expected given Fig. 4.3 which shows that most of the absorption signal results from the tangent layer and roughly the next two layers overhead so number density errors in layers above these will have little affect on the solution. The advantages of this form of the onion peel inversion are each profile inversion is accomplished in a single pass,

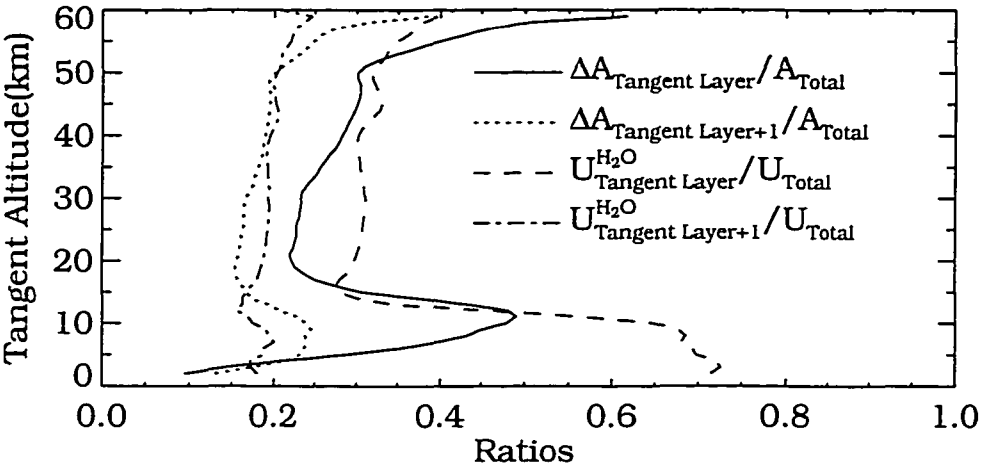


Figure 4.3: Fraction of slant path absorption and number density in the tangent layer and next highest layer relative to slant path total. ΔA is the incremental change in absorption through the layer, $A_j - A_{j-1}$.

inversion stability with noisy data, and rapid convergence. The only disadvantage is the the downward propagation of error caused by limiting error compensation to the tangent level.

Operational Inversion

The algorithm selected by the SAGE II project to invert the water vapor data on an operational basis, described in Chu et al.(1993), is a multiple pass, two step process. The first step consists of a limited onion peel inversion, limited in the sense that only a single update is allowed at each tangent level using Chahine's predictor (Chahine, 1972)

$$U^{n+1/2} = U^n \frac{\overline{A}^m}{\overline{A}^c(U^n)} \tag{4.33}$$

when $|\overline{A}^m - \overline{A}^c(U^n)| > \Delta_{\mathrm{cl}}$ and

$$U^{n+1/2} = U^n \tag{4.34}$$

when $|\overline{A}^m - \overline{A}^c(U^n)| < \Delta_{cl}$. In the second step, vertical smoothing of the number density above 18.5 km is performed with

$$\overline{U}^{n+1}(i) = exp[\{log(\overline{U}^{n+1}(i-1)) + log(\overline{U}^{n+1/2}(i)) + log(\overline{U}^{n+1/2}(i+1))\}/3] \quad (4.35)$$

This counteracts the relatively low signal-to-noise ratios in the stratosphere. Because Chahine's predictor converges slowly and because the vertical smoothing may cause $|\overline{A}^m - \overline{A}^c(U^{n+\frac{1}{2}})| > \Delta_{cl}$ for some tangent layers, the two step process is repeated until all absorption residuals are less than Δ_{cl} or 20 passes. This inversion is also stable in the presence of noise and disperses the effects of measurement errors over several layers in the stratosphere. The biggest drawback to this technique is the multiple pass aspect which requires significantly more radiative transfer computations than the straight onion peel inversion.

Mill-Drayson Inversion

If slant path absorption is viewed as a measure of the information content of a signal, then the greatest information content is found in signals from the lowest measurement tangent altitudes. The inversion algorithm designed by Mill and Drayson (1978) was designed to utilize the highest information content signals first by starting at the lowest tangent layer and working upwards. Instead of updating just the tangent layer U as in the onion peel inversion, U at and above the tangent layer are updated with a constant multiplicative factor that adjusts the absolute number density but preserves the profile shape. Multiple passes from bottom to top are necessary since adjustments in higher layers may change absorption residuals in layers below.

At each tangent layer i, the previously described secant method updates the multiplicative factor r_i and old number densities U^{old}, resulting from adjustments

performed at the previous tangent layer, are scaled by

$$U_{i,j}^{n+1} = U_{i,j}^{old} r_i^{n+1} \quad \text{layers } j = 1, 2, \ldots i \tag{4.36}$$

when $|\overline{A}^m - \overline{A}^c(U^n)| > \Delta_{\text{cl}}$ or

$$U_{i,j}^{n+1} = U_{i,j}^{old} \tag{4.37}$$

when $|\overline{A}^m - \overline{A}^c(U^n)| < \Delta_{\text{cl}}$. Updating continues until $|\overline{A}^m - \overline{A}^c(U^n)| < \Delta_{\text{cl}}$ or 20 updates are made. After the updating process is completed for layer i, the algorithm increments to the layer immediately above $(i-1)$ and repeats the process. Bottom to top passes repeat until the absorption residuals at all tangent layers satisfy the error criterion or 20 passes have been made. The first guess water vapor profile provides the initial $U_{i,j}^{old}$ profile.

The fixed step size search procedure used in the original Mill-Drayson formulation was replaced by the secant method here because the secant method seemed to be both simpler and faster and it also allowed a more direct comparison of error propagation qualities between this and the onion peel inversion.

The advantages of this approach include stability in the presence of noise and limited error propagation. According to Mill-Drayson (1978) the error propagation can be explained as follows. In the idealized situation where the slant path absorption profile contains a single value with measurement noise, the profile fitting of this technique will spread the retrieval errors over all of the layers being fitted (primarily the lowest layers where most of the absorption takes place). The error at the tangent layer will be smaller than the onion peel retrieval error, which focuses the retrieval error in the tangent layer, and since the tangent layer U or $\overline{V}^{H_2O}$ is not part of subsequent profile fittings at higher levels, there is no direct error propagation. The main disadvantage to the technique is the number of radiative transfer calculations needed for the multiple passes.

MILL-DRAYSON(SMOOTHED) INVERSION

This algorithm is based on the operational inversion with the onion peel step replaced by the Mill-Drayson bottom to top profile fitting. Like the operational inversion, it is a multiple pass, two step process. In the first step, a single update of the profile with the multiplicative factor r_i calculated with Chahine's predictor is made for each tangent layer moving from the bottom to top of the atmosphere. The vertical smoothing performed in the second step is unchanged from the operational algorithm. The process repeats until the convergence criterion of the operational algorithm are satisfied. One would expect this inversion to be stable, considering its origin, and combine the error characteristics of both algorithms, error reduction through smoothing in the stratosphere and reduced error propagation.

SPARSE GRID LEAST SQUARES INVERSION

The global fit inversion (Carlotti, 1988) developed for application to limb scanning multi-wavelength interferometric spectra is an extension of earlier techniques whereby an onion peel inversion is performed by updating an experiment parameter (such as number density or other experimental variables) at a given tangent level until spectral residuals between the measured and calculated signals are minimized. The global fit algorithm takes this one step farther by simultaneously minimizing the residuals summed over wavelength and tangent layer. Mathematically, the residual sum is stated as

$$SQ(p_1, p_2 \ldots p_{N^E}) = \sum_{i=1}^{N^T} \sum_{k=1}^{\lambda^T} [A_{i,k}^m - A_{i,k}^c(p_1, p_2 \ldots p_{N^E})]^2 \qquad (4.38)$$

where N^E is the maximum number of experiment parameters, i is the tangent layer index, and k is the wavelength index. The total number of wavelengths being fitted is given by λ^T and p represents the adjustable experiment parameters. For a single wavelength, representing the water vapor channel measurement, $k = 1$, and the

previous equation simplifies to a minimization of the tangent layer residuals

$$SQ(U_1, U_2 \ldots U_{N^T}) = \sum_{i=1}^{N^T} [A_i^m - A_i^c(U_1, U_2 \ldots U_{N^T})]^2 \qquad (4.39)$$

after limiting the unknown experiment parameters to the number densities U. This is a well posed problem with N^T equations for N^T unknowns. One of the advantages to this algorithm is the capability to uncouple the number density profile from the measurement geometry, which will be used here as a means to provide stratospheric smoothing. Instead of solving for U at every layer in the stratosphere, a solution every second or third layer is sought above 17.5 km. The water vapor mixing ratio is a slowly varying function of altitude so linear interpolation between solution layers to fill in skipped levels should be reasonably accurate. Random errors in adjacent layers should cancel to some extent in the minimization process of this over-determined system of equations, producing a reduction in errors similar to the explicit smoothing of the operational algorithm. This variation of the global fit algorithm will be referred to as the sparse grid least squares algorithm (SGLS).

The search for the set of U that minimizes SQ to within some convergence criterion is carried out with the Levenberg-Marquardt non-linear least squares fitting routine (Brown and Dennis, 1972). The technique is computationally intensive, for example, adjustment of U_1 in the topmost layer requires recalculation of A_i^c for all tangent layers before SQ can be recomputed. To improve computational efficiency, advantage was taken of the fact that not every U_i contributes to all A_i^c, for instance, U_{N^T} contributes only to $A_{N^T}^c$. Slant path absorption values ranging over four orders of magnitude produce residuals over a similar broad range and prevents convergence at the higher tangent layer where the smaller residuals have little to no affect on SQ. To equalize the residual contribution from all tangent layers, Eq. (4.39) was recast into

$$SQ(U_1, U_2 \ldots U_{N^T}) = \sum_{i=1}^{N^T} [log(A_i^m) - log(A_i^c(U_1, U_2 \ldots U_{N^T}))]^2 \qquad (4.40)$$

where the logarithm function is used to compress the range of absorption values.

The convergence criteria may take several forms. Iterations may continue until SQ is minimized or some preset value based on the random measurement noise is reached. Another approach is to iterate until the relative change of the free parameters falls below some level or to specify the number of significant digits in the free parameters. For this work, iterations proceed until the solution is unchanged in three significant digits, or better than 0.1%, up to a 100 iteration maximum.

The multi-spectral version of this technique exhibits less error propagation than an onion peel algorithm (Carlotti, 1988) and a similar error dispersal should be apparent in the single wavelength application. This inversion will be particularly easy to apply in a systematic way to the upcoming SAGE III multi-channel observations. The primary disadvantage of the technique is computational, execution times are significant multiples of the other algorithms.

4.3 Experiment Parameters

Table 4.1 summarizes the experiment parameters relevant for the SAGE II water vapor measurements and maps them to the generalized variables employed in the discussion of the formal error analysis found in Sec. 2.4 and shown in Figs. 2.4 and 4.2.

As indicated in the table, model parameters may have both systematic and random error components and are shown grouped by where they are encountered in the experiment. All of the model parameters, except one, have atmospheric origins. The sole instrument model parameter is the reference altitude. The contribution function $\mathbf{D_y}$, defined as the incremental change in the water vapor mixing ratio due to an incremental change in $\mathbf{y}^m$ for channel 2, does not account for incremental changes in the slant path absorption of other channels on the inverted water vapor. One might

Table 4.1: Experiment Parameters for the SAGE II Water Vapor Measurement

$\mathbf{y}^m$	$T_{2/1}(z_{th})$		
$\mathbf{y}^{mi}$	$\overline{T}^{\mathrm{H_2O}}(z_{th}), \overline{A}(z_{th})$		
$\mathbf{x}$	$\overline{V}^{\mathrm{H_2O}}(i)$		
$\mathbf{b}_\mathrm{A}$		Systematic	Random
	T	$\surd$	$\surd$
	Separation Model S		
	$\Delta\delta_{2-1}^{Rayleigh} = \delta^{\mathrm{Ray}}(\lambda_2) - \delta^{\mathrm{Ray}}(\lambda_1)$	$\surd$	
	$\Delta\delta_{2-1}^{Ozone} = \delta^{\mathrm{O_3}}(\lambda_2) = a_2\delta^{\mathrm{O_3}}(\lambda_3)$	$\surd$	$\surd$
	$\Delta\delta_{2-1}^{Aerosol} = \delta^{\mathrm{Aer}}(\lambda_2) - \delta^{\mathrm{Aer}}(\lambda_1)$	$\surd$	$\surd$
	Inversion Algorithm Model IA		
	Molecular line parameters (k_λ)		
	Band strength $\sum S^m$	$\surd$	
	Lorentz half widths α_L	$\surd$	
	Exponent ν	$\surd$	
	Continuum	$\surd$	
	EGA accuracy	$\surd$	
$\mathbf{b}_\mathrm{I}$	Reference altitudes		$\surd$
$\mathbf{c}$	First guess profile	$\surd$	$\surd$
	Convergence criteria Δ_{cl}	$\surd$	$\surd$

define a set of contribution functions as the response of the inverse model to incremental changes in the measured absorption of each of the other channels or simply declare the other six measured absorption profiles to be model parameters. Both approaches are equivalent, computationally intensive, and require estimates of the systematic and random profile uncertainties which effectively forces one to analyze the entire experiment. Instead, coupling to the other channels will be limited to the separation model parameters shown in Table 4.1, $\Delta\delta_{2-1}^{Rayleigh}, \Delta\delta_{2-1}^{Ozone}, \Delta\delta_{2-1}^{Aerosol}$. The estimated systematic and random uncertainties of these parameters account for uncertainties appearing elsewhere in the experiment (except for temperature and reference altitude which are discussed below). For example, a systematic error in an NO_2 cross-section affecting the ozone and aerosol solution and then $\Delta\delta_{2-1}^{Ozone}, \Delta\delta_{2-1}^{Aerosol}$, is contained in the systematic uncertainty of $\Delta\delta_{2-1}^{Ozone}, \Delta\delta_{2-1}^{Aerosol}$.

Two of the model parameters, temperature and reference altitude, are singled out for special attention because perturbations in either one tend to produce simultaneous and similar changes in the measured slant path absorptions of all channels. For these, calculating $\mathbf{A_b}$ directly instead of multiplying $\mathbf{D_y K_b}$ properly takes into account the overall effect of uncertainties in these parameters, balancing sensitivities in channel 2 against those in the other channels that enter the analysis through the separation model.

Applying the formal error analysis to a large number of individual profiles quickly becomes computationally overwhelming because of the perturbation approach taken to develop $\mathbf{K_x}$, $\mathbf{D_y}$, $\mathbf{K_b}$, and $\mathbf{A_b}$. Since the results of the error analysis would most likely be averaged in some manner for presentation, the error analysis will be performed in terms of averaged quantities, monthly zonal means. If 10° latitude zones are used only 16 sets of functions per month are required instead of 900. In the following sections, unless otherwise noted in the figure caption or text, examples of the $\mathbf{K_x}$, $\mathbf{D_y}$, $\mathbf{K_b}$, and $\mathbf{A_b}$ functions were calculated with the 40° N reference

atmosphere described in Sec. 4.3.1. Functions used in the actual error analysis are calculated assuming reference profiles taken from the SAGE II inversion results.

In Sec. 4.3.1 an example of $\mathbf{K_x}$ is presented for a representative model atmosphere. In Sec. 4.3.2 examples of $\mathbf{K_b}$ for each of the model parameters are presented and estimates of the systematic errors are given. Discussion of the model parameter error covariances and measurement errors are found in Sec. 4.3.4.

4.3.1 ATMOSPHERIC PARAMETERS x

The goal of the inversion process is the accurate determination of the unknown atmospheric parameters, defined for this experiment as the layer averaged water vapor mixing ratio $\overline{V}^{H_2O}(i)$.

The sensitivity of the forward model F to the atmospheric parameters $\mathbf{x}$, represented by $\mathbf{K_x}$ and defined by Eq. (2.7), is shown in Fig. 4.4 for selected rows and columns. The columns indicate the response of the forward model, a change in slant path absorption, to a delta function change in water vapor volume mixing ratio Δx. The response peaks at the perturbation layer, falls off below it as the perturbed layer makes a diminishing contribution to the slant path absorption, and is zero above the perturbed layer since the perturbed layer contributes nothing to the absorption at higher tangent layers. The rows of $\mathbf{K_x}$ are viewed as weighting functions applied to the Δx to determine the slant path absorption. The tangent layer Δx receives the greatest weight and layers above receive diminishing weights. Again, because of the occultation geometry, Δx below the tangent layer receives zero weight.

When the row and column profiles are normalized by their maximum values, the profile shapes are found to be quite similar. The strongly peaked narrow profiles are typical of that found for an occultation experiment and in a sense are a restatement of the information found in Fig. 4.3. The figure shows the proportion of water vapor

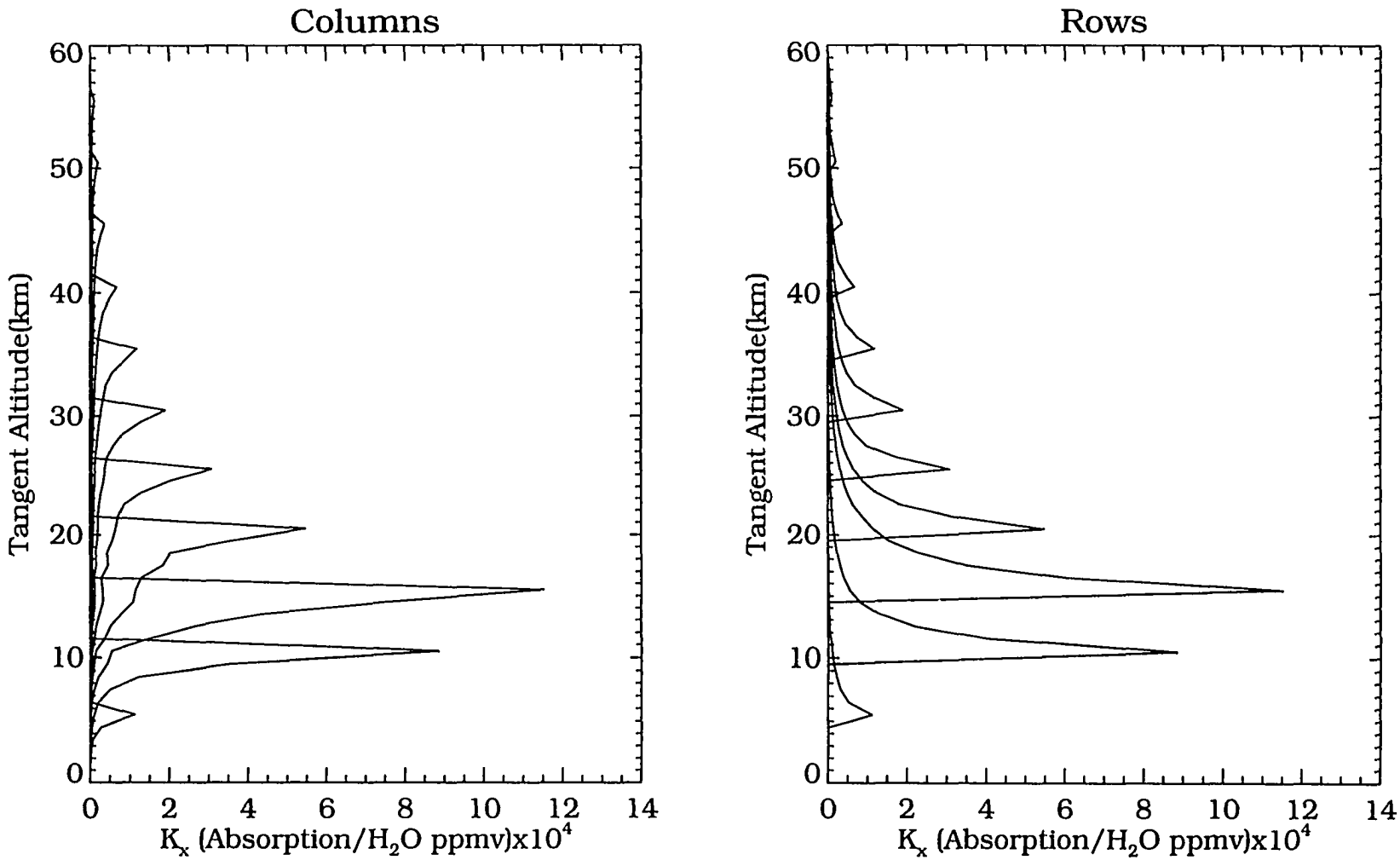


Figure 4.4: Columns and rows of the weighting function matrix $\mathbf{K_x}$, every 5 km starting at 5.5 km.

absorber amount, U, contained in the tangent layer ($j = i$) and the layer above ($j = i - 1$) to the total absorber amount along the slant path. In the stratosphere, about 30% of the total absorber amount is found in the tangent layer, 20% resides in the layer above, and higher layers are responsible for successively smaller portions of total U^{H_2O}. In the troposphere, the two layers combined account for $\approx$ 90% of the total water absorber amount. The incremental change in slant path absorption through these two layers relative to the total slant path absorption is also shown in the figure and reflects the absorber amount layer contributions where the tangent layer provides the largest contribution. The peak response of $\mathbf{K_x}$ at the perturbation level follows from the high fraction of the total absorber amount contained in the tangent layer, the narrowness of the profile follows from the rapidly decreasing contributions in layers above.

The envelope peak in Fig. 4.4 results from expressing $\mathbf{K_x}$ in terms of absorption/ppmv. An incremental change in mixing ratio causes a larger change in U at lower altitudes than at higher altitudes. Below 15 km the peaks fall off as the increased slant path optical depths push the exponential function in Eq. (4.9) out of the linear response region.

The $\mathbf{K_x}$ shown here was calculated using an annualized northern middle latitude profile constructed from two sources. The interim reference model for water (Chiou et al., 1996) was used in the stratosphere above 14 km. The tropospheric data of Oort (1983), decreased approximately by a factor of two to approximate the cloud free and dryer conditions under which SAGE II makes most of its observations, was used from the surface up to 9 km. The troposphere and stratosphere were connected smoothly to produce a reasonable hygropause. Linearizations using equatorial, high latitude, and dehydrated Antarctic springtime profiles showed minor changes, on the order of 10-20%, from Fig. 4.4. Corresponding but opposite shifts in $\mathbf{D_y}$ were observed, making $\mathbf{A}$ nearly independent of the linearization profile.

4.3.2 Model parameters b

In this section the sensitivity of the forward model is investigated by developing $\mathbf{K_b}$ for each of the model parameters **b** listed in Table 4.1. For model parameters containing systematic errors a review of the uncertainty associated with each parameter is made and a reasonable error is chosen for use in the final error budget calculations. It is assumed that the model parameters are independent of one another. The dissimilar units in which **b** is expressed make comparisons amongst $\mathbf{K_b}$ difficult, therefore $\mathbf{K_b}$ is presented as the incremental change in slant path absorption for the specified uncertainty.

Atmospheric Model Parameters b_A

Temperature. Sensitivity of the forward model to uncertainties in temperature provided by the National Weather Service (NWS) analysis is shown in Fig. 4.5 for a 2° C perturbation on each of the standard pressure levels. Temperature is found in both separation and inversion algorithm models and the uncertainty in Fig. 4.5 includes the affects of temperature on shell path lengths, Rayleigh slant path optical depths and water vapor absorption. Peak sensitivity generally occurs at the altitude of the corresponding standard pressure level with an increase in temperature resulting in a decrease in absorption. Since the model temperature profile for the 1 km shells is developed by interpolation from the somewhat coarser standard pressure profile, perturbations extend upwards to the next standard pressure level. Below, not only do perturbations appear in the model profile at the tangent level but perturbations in shells above also contribute to extend the sensitivity across several standard pressure levels. The response of the forward model on the lower three standard pressure levels are slightly different than the other levels. The lowest two levels show an increase in absorption while the third level appears to be a transition

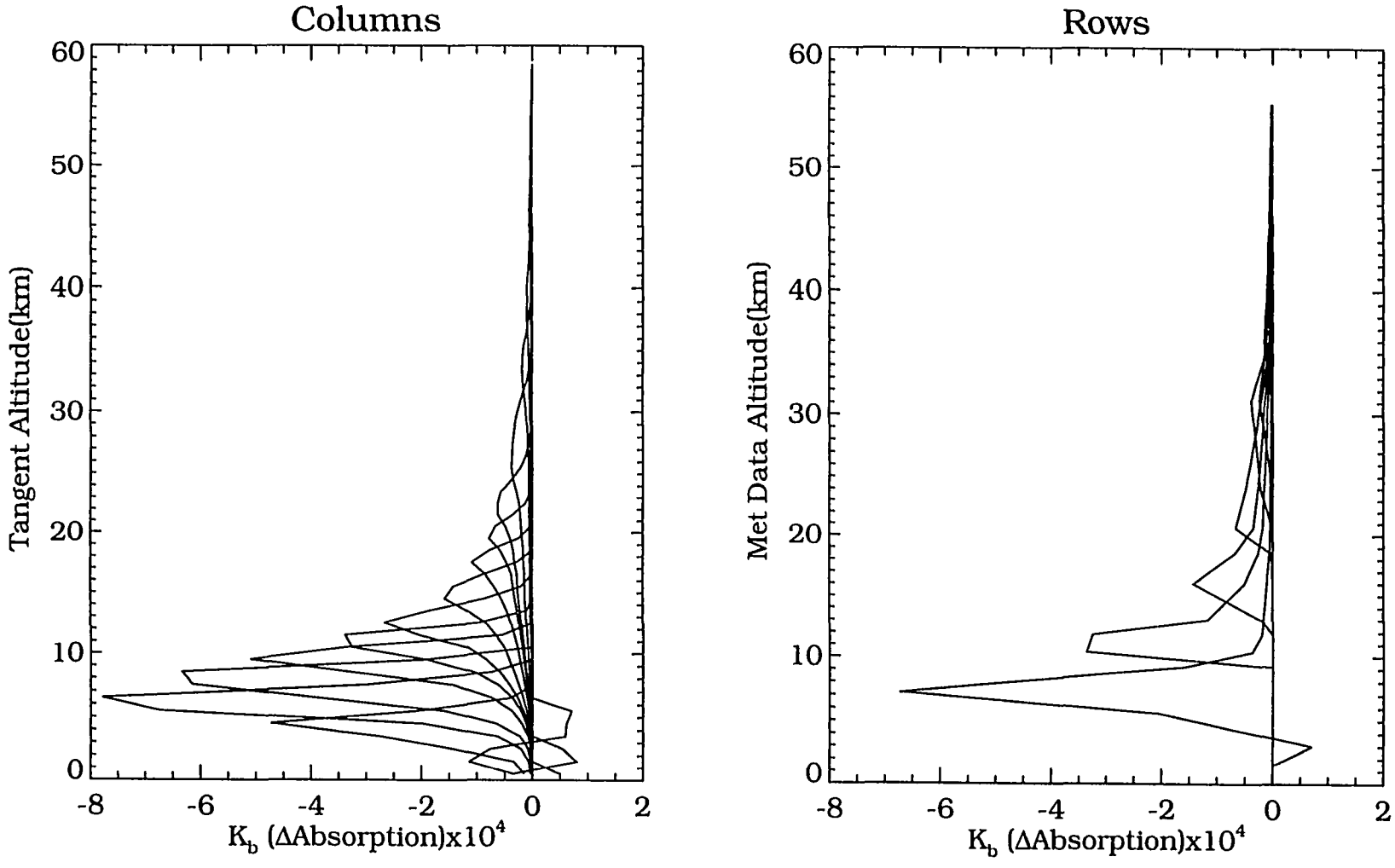


Figure 4.5: Columns of $\mathbf{K_b}$ for a 2° C perturbation in the meteorological temperatures on the standard pressure levels. Rows are shown every 5 km starting at 5.5 km.

between the two regimes. Matrix $\mathbf{K_b}$ in Fig. 4.5 will overestimate the temperature error since $\mathbf{D_y K_b} \neq \mathbf{A_b}$ where $\mathbf{A_b}$ is calculated directly and includes the affects of temperature perturbations in the other channels.

The NWS analysis represents a blend of *in situ* radiosonde observations in the troposphere and lower stratosphere, and satellite observations in the stratosphere. Systematic differences reported for various radiosonde intercomparisons (Hooper, 1975) range from 0.6 to 1.5° C. The NWS satellite upper stratospheric satellite based temperatures are corrected (Geller et al., 1983) with regression coefficients developed from rocketsonde comparisons which makes estimation of systematic errors difficult. We assume a systematic temperature error of 1° C throughout the atmosphere for the error budget calculations but recognize that this may be an underestimate for the upper stratosphere.

Separation Model S

Rayleigh. The sensitivity of the forward model to uncertainties in the difference of Rayleigh slant path optical depths, $\Delta\delta_{2-1}^{Rayleigh}$, is shown in Fig. 4.6. Matrix $\mathbf{K_b}$ in this case is a diagonal matrix and only the diagonal elements are plotted in the form of a smooth curve. The Rayleigh scattering cross-section, given by Eq. (2.11), is a function of wavelength and refractive index. The refractive index term $(n(l) - 1)$ is expressible as a wavelength dispersion term scaled by the number density, $a(\lambda)C^{\mathrm{Air}}(l)$. Equation (2.11) becomes

$$\sigma_\lambda^{\mathrm{Mol}} = K' \frac{a(\lambda)^2}{\lambda^4} \tag{4.41}$$

where K' collects the constants. Since $\sigma_\lambda^{\mathrm{Mol}}$ is independent of pressure or temperature along the pathlength l, $\Delta\delta_{2-1}^{Rayleigh}$ may be written as

$$\Delta\delta_{2-1}^{Rayleigh} = \delta^{\mathrm{Ray}}(\lambda_2) - \delta^{\mathrm{Ray}}(\lambda_1) = (R_{2/1} - 1.)\delta^{\mathrm{Ray}}(\lambda_1) \tag{4.42}$$

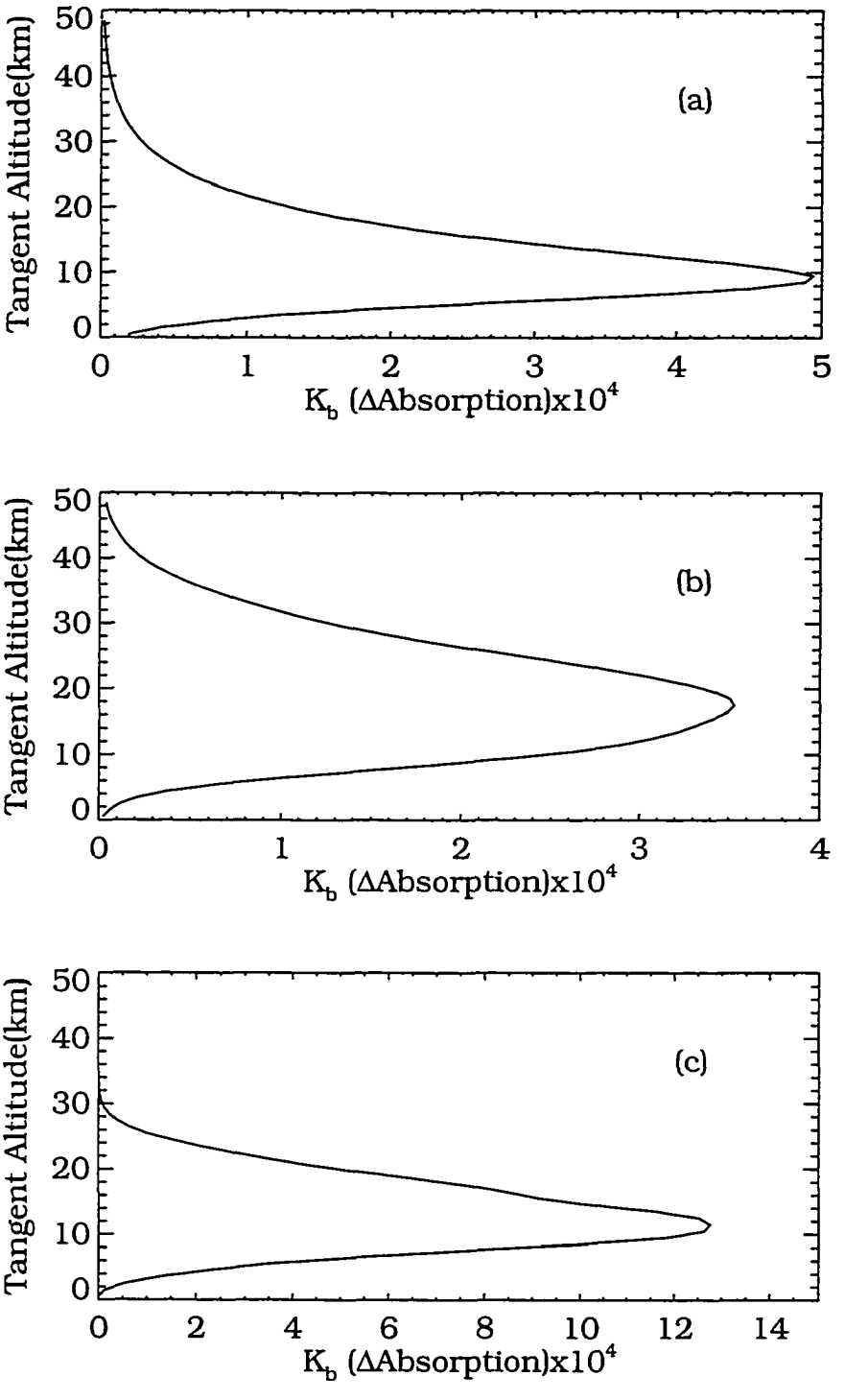


Figure 4.6: Diagonal elements of the matrix $\mathbf{K_b}$ plotted in the form of a smooth curve, for: a) 1% perturbation in $\Delta\delta_{2-1}^{Rayleigh}$, b) 5% perturbation in $\Delta\delta_{2-1}^{Ozone}$, c) 5% perturbation in $\Delta\delta_{2-1}^{Aerosol}$, as defined in Table 4.1. Off diagonal elements are zero.

where $R_{2/1}$, a constant, is the ratio of $\sigma_\lambda^{\text{Mol}}$ calculated at the wavelengths of channel 2 and 1. Uncertainties in the expression for 4.42 result from uncertainties in $R_{2/1}$ and $\delta^{\text{Ray}}(\lambda_1)$, these in turn are determined by uncertainties in $a(\lambda)$ and the channel wavelength λ. The dispersion term, $a(\lambda)$, is known to better than 1 part in 8000 (Edlén, 1966). Error analysis of either $\sigma_\lambda^{\text{Mol}}, R_{2/1}$ or $\delta^{\text{Ray}}(\lambda_1)$ shows that in order for the errors due to wavelength uncertainties to equal those due to $a(\lambda)$ requires a wavelength uncertainty of $6.25 \cdot 10^{-5} \mu$m. This being an order of magnitude smaller than that assumed for the error budget, errors due to uncertainties in $a(\lambda)$ will be ignored.

Note that in this formulation, uncertainties in $\delta^{\text{Ray}}(\lambda_1)$ due to temperature errors have already been addressed in the previous section on temperature sensitivity. Errors in either λ_1 or λ_2 will be systematic. A systematic error of 0.005μm in λ_2 with λ_1 known exactly causes a constant relative error of 0.0071 in $\Delta\delta_{2-1}^{Rayleigh}$, and when λ_1 is equally uncertain the relative error increases to 0.0088.

Ozone. The sensitivity of the forward model to uncertainties in the ozone slant path optical depth, $\Delta\delta_{2-1}^{Ozone} = \delta^{O_3}(\lambda_2) = a_2\delta^{O_3}(\lambda_3)$ is shown in Fig. 4.6. Ozone slant path optical depth increases through the stratosphere down through the peak of the ozone number density and then decreases slightly to a nearly constant value below, giving $\mathbf{K_b}$ its profile shape down to $\approx$15 km. Below this altitude the exponential function in the transmission equation becomes increasingly important in reducing $\mathbf{K_b}$ to 0 at the surface. Slant path optical depth difference, $\Delta\delta_{2-1}^{Ozone}$, has both systematic and random uncertainties due to random errors in $\delta^{O_3}(\lambda_3)$ and systematic errors in a_2 and $\delta^{O_3}(\lambda_3)$.

The ozone contribution coefficient a_2, calculated from the optical depth measurements of Anderson et al.(1991) and found to be 0.0039, is assigned a relative error of 0.10 or 10%. An estimate for a_2 can be developed from the SAGE II observations

by analyzing the slant path optical depths in the stratosphere where the aerosol contributions are low. The Rayleigh component is removed as usual and a water vapor profile from a correlative measurement allows calculation and removal of the water vapor contribution to yield the ozone optical depth. When ratioed to channel 3 an estimate of a_2 is obtained for each slant path. Analysis of several events with this procedure suggests that the uncertainty for a_2 is $\approx$10%. An estimated relative uncertainty of 0.05 is assigned also to $\delta^{O_3}(\lambda_3)$. This estimate may be somewhat high since the determination of $\delta^{O_3}(\lambda_3)$ from $\delta(\lambda_3)$ should be highly accurate because $\delta^{O_3}(\lambda_3)$ is the dominant component of $\delta(\lambda_3)$ in the stratosphere. So, while one estimate may be low, and the other high, the combination of the two estimates to give a systematic relative uncertainty of 0.11 is certainly reasonable for $\Delta\delta_{2-1}^{Ozone}$.

Aerosol. Initial calculations of $\delta^{\mathrm{Aer}}(\lambda_2)$ with the Mie scattering equations using typical aerosol particle size distribution models with adjustable parameters determined by the SAGE II aerosol channels produced unsatisfactory water vapor inversions. An interim solution developed by Chu et al.(1993) is based on expressing $\Delta\delta_{2-1}^{Aerosol}$ as

$$\Delta\delta_{2-1}^{Aerosol} = \delta^{\mathrm{Aer}}(\lambda_2) - \delta^{\mathrm{Aer}}(\lambda_1) = (R_2 - 1.)\delta^{\mathrm{Aer}}(\lambda_1) \tag{4.43}$$

where R_2, the ratio $\frac{\delta^{\mathrm{Aer}}(\lambda_2)}{\delta^{\mathrm{Aer}}(\lambda_1)}$, is related empirically to a third order polynomial in R_1, the ratio $\frac{\delta^{\mathrm{Aer}}(\lambda_4)}{\delta^{\mathrm{Aer}}(\lambda_1)}$. A physical basis for this approach was provided by Thomason(1991) who showed it was possible to model the empirical relationship with a segmented power law particle size distribution.

Systematic uncertainty in $\Delta\delta_{2-1}^{Aerosol}$ is difficult to gage. To do so, we turn to comparisons of the aerosol extinction (km^{-1}) measured by SAGE II in channels 5 and 7 to those calculated with the segmented power law size distribution whose adjustable parameters are determined by aerosol extinction in channels 4 and 1. The mean annual error (Fig. 7, Thomason(1991)) for channel 5 is less than 5%

everywhere except for the equatorial region (± 30°) below 19 km where it reaches 12.5%. The uncertainty in the slant path optical depth will fall somewhat below this due to averaging along the slant path. By assuming that the modeling uncertainty in $\delta^{\mathrm{Aer}}(\lambda_2)$ is comparable to $\delta^{\mathrm{Aer}}(\lambda_5)$, we estimate the systematic error to be 10% for $\Delta\delta_{2-1}^{Aerosol}$. The sensitivity of the forward model to uncertainties in the difference of aerosol slant path optical depths, $\Delta\delta_{2-1}^{Aerosol}$, is shown in Fig. 4.6.

Inversion Algorithm *IA* This section addresses the accuracy of the variables and algorithms underlying the water vapor radiative transfer calculations. This includes accuracy of the molecular line parameters described in Sec. 2.1 that are needed in the calculation of $k_{\omega}^{g,m}(P,T)$ as defined by Eq. (2.16), accuracy of the EGA technique, selection of particular line parameter data sets, and the water vapor continuum. The sensitivity of the forward model to the molecular line parameters and the EGA technique will be developed here and included in the error budget of each inversion algorithm. The effect of data set substitutions and the water vapor continuum on the retrieved water will be studied with the help of the operational inversion and omitted from the formal error budget.

The core EGA database of $A(U)$, which must be recalculated for each molecular line parameter modification, is developed from monochromatic line-by-line calculations performed with FASCOD3 (Clough et al., 1989a; Clough et al., 1989b). The HITRAN 92 line parameter compilation (Rothman et al., 1992) was chosen as the baseline for the studies performed in this section since it contains the most complete and accurate water vapor line parameters. The EGA database utilized operationally for the inversion and archival of the water vapor product is based on the older HITRAN 82 compilation calculated with a Fourier transform line-by-line radiative transfer code. We are primarily interested in the improvements and modifications to the water vapor line parameters in the channel 2 spectral region spanning these

two data sets to understand the uncertainties associated with the operational EGA database.

Molecular line parameters. For a broadband instrument,the molecular line parameters affecting measurement accuracy, in order of importance, are the line strength $S^{g,m}(T_0)$, the air broadened halfwidth $\alpha_L^{g,m}(P_0,T_0)$, and the exponent $\nu^{g,m}$ controlling the halfwidth temperature dependence.

A limited number of spectroscopic experiments have been performed to determine the water vapor line strengths in the spectral region of interest. The first line strength measurements were those of Benedict who determined equivalent widths of atmospheric species observed in the solar spectra of Delbouille and Roland(1963) and analyzed by Swensson et al.(1970). Some time later, measured line strengths of 97 transitions by Giver et al.(1982) found good agreement with Benedicts work for strong lines in the 201-000 band (30 line average of $0.93\pm0.08\sigma$, Benedict higher) and in the 121-000 band (9 line average of $1.0\pm0.03\sigma$). Weaker lines of the 300-000 were found to be much stronger than Benedicts, by an order of magnitude in many cases. In the most recent set of measurements by Chevillard et al.(1989), 718 isolated lines were measured with an average accuracy of 7%, 1695 other lines with accuracies ranging from 6-15% for weakly perturbed lines, 15-25% for perturbed lines and 50% for weak or strongly overlapped lines were obtained with a faster and somewhat less accurate analysis technique. An average uncertainty of 15% is assigned this second group of lines. A subset of 53 isolated lines compared to the work of Giver et al.(1982) found an average difference of $0.96\pm0.077\sigma$, with the strengths of Chevillard et al.(1989) lower. While the quoted uncertainties of individual transitions may seem high, they are typical for transitions in the visible and near infrared. Fortunately, these uncertainties are reduced in a broadband experiment where uncertainty in the line strength total within the pass band is a

better error descriptor. The HITRAN 82 line strengths are based on Benedicts analysis while HITRAN 92 uses that of Chevillard et al.(1989). Using Giver et al.(1982) as a comparison standard, even though the subsets of lines being compared are different, suggests that the HITRAN 92 total line strengths should be $\approx$11% lower than HITRAN 82. In fact, a straight summation of all water vapor lines within the channel 2 passband yields a 10.5% difference. Allowing for improvements in the more recent spectroscopic measurements, we assign a total line strength error of 10% to HITRAN 82 and 5% to HITRAN 92. However, for the error budget calculations, we will assume that these uncertainties can be reduced by half following successful validation activities.

The HITRAN 82 air-broadened halfwidths, based on the theoretical calculations of Davies and Oli(1978), were compared by Giver et al.(1982) to their measurements for 71 lines. They found their measurements were on average $1.08\pm0.07\sigma$ larger than HITRAN 82. The HITRAN 92 air-broadened halfwidths are a mix of experimental values, theoretical calculations, and predefined default values (Rothman et al., 1992). Recent air-broadened halfwidth calculations by Adler-Golden et al.(1992) at 820 nm for 8 lines of the (211) vibrational band found excellent agreement with the HITRAN 92 halfwidths, with an average difference of $1.00\pm0.04\sigma$. We also compared the Giver halfwidth measurements to HITRAN 92 for 70 lines and found an average difference of 1.038. We therefore assign a conservative passband uncertainty of 10% to the HITRAN 82 air-broadened halfwidths and 5% to HITRAN 92.

The air-broadened halfwidth temperature dependence follows Eq. (2.21) and is determined by the coefficient $\nu^{g,m}$. Although it has been known for some time that $\nu^{g,m}$ varies strongly from line to line from theoretical calculations (Benedict and Kaplan, 1959), experimental measurements of this parameter were few and most radiative transfer codes used a gas averaged value ν^{g}. An early estimate for ν^{H_2O} is 0.62 (Benedict and Kaplan, 1959) for the HITRAN 82 era. More recent theoretical

calculations find ν^{H_2O} =0.68 (Gamache and Rothman, 1988). The HITRAN 92 compilation incorporates transition specific values from Gamache and Rothman(1988) and uses 0.68 for transitions not covered by the theoretical calculations. Due to the lack of comparable experimental results it is difficult to place an error estimate on $\nu^{H_2O,m}$. We assign an average passband uncertainty of 10% to ν^{H_2O} used with HITRAN 82 and 5% to the HITRAN 92 values.

To illustrate the relative importance of line strengths, air-broadened halfwidths, and associated temperature coefficients, $\mathbf{K_b}$ was calculated for a 5% average change in each. Figure 4.7 shows that the forward model is more sensitive to uncertainties in line strengths than halfwidths and ten times more sensitive than temperature coefficients. Because $\mathbf{K_b}$ is expressed in terms of absorption, all show peak sensitivity in the same general altitude range. This is somewhat misleading since systematic biases in line strengths affect $\mathbf{x}^r$ at all altitudes while halfwidth biases will be limited to altitudes where pressure broadening becomes significant.

Major updates in line parameter data sets, such as from HITRAN 82 to HITRAN 92, entail typically simultaneous parameter replacements/adjustments and the overall change in a radiative transfer computation may be difficult to predict. For example, the impact of increases/decreases in line strengths on $\mathbf{x}^r$ may be offset by opposite shifts in halfwidths. The combined effects, along with other assumptions related to the radiative transfer modeling, are best studied by the simulation and inversion of synthetic measurements.

For instance, Fig. 4.8a displays the percent change in $\mathbf{x}^r$ when the forward model employs the HITRAN 82 data base to simulate slant path absorptances that are then inverted with HITRAN 92. Although the total line strengths differ by $\approx$ 10.5% between the two data sets, strength partitioning between the strong and weak lines differ significantly. Grouping all lines by orders of magnitude, one finds that the total

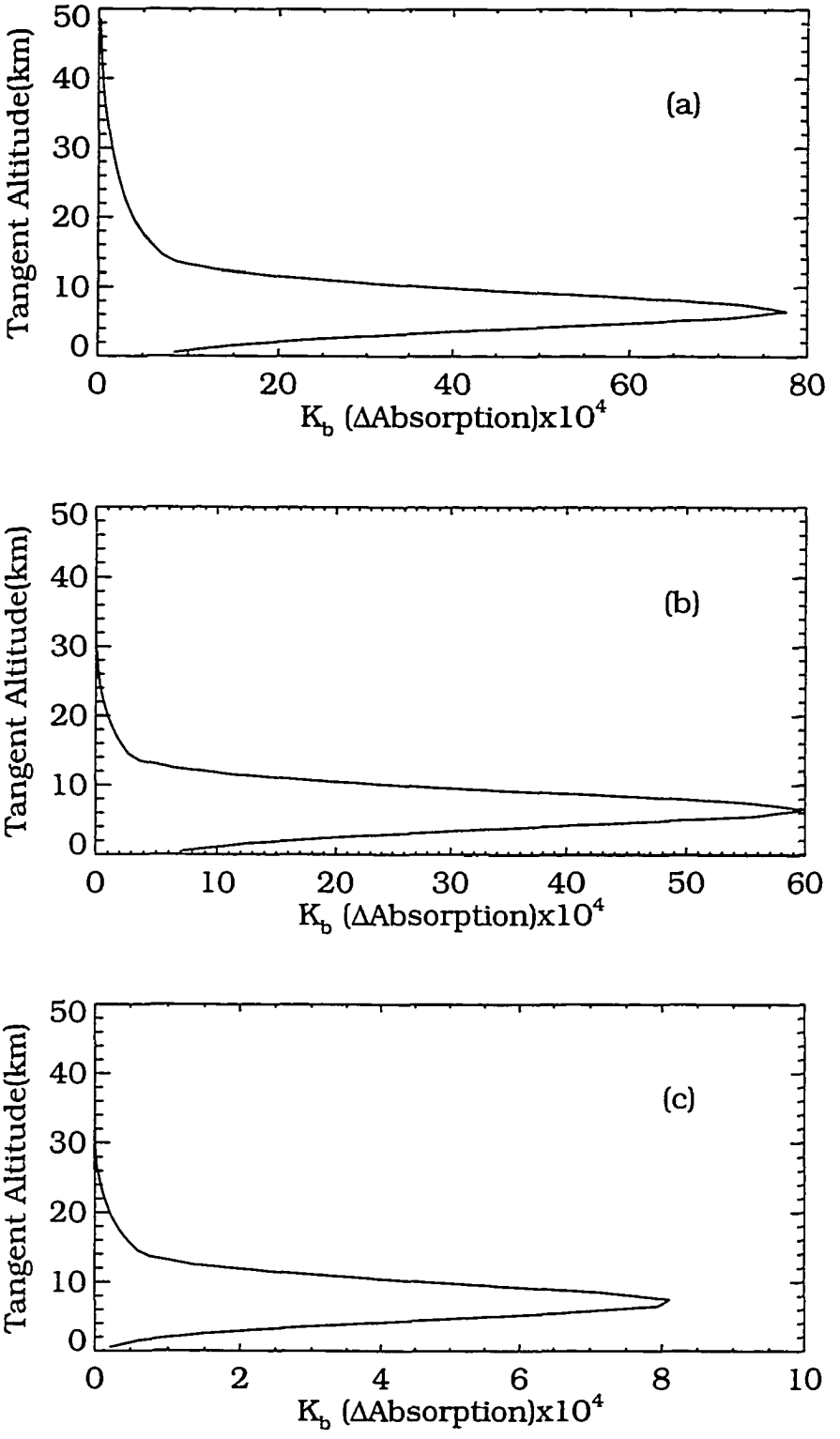


Figure 4.7: $\mathbf{K_b}$ for a 5% perturbation in a)line strengths, b)air-broadened halfwidths, c)air-broadened halfwidths temperature coefficient.

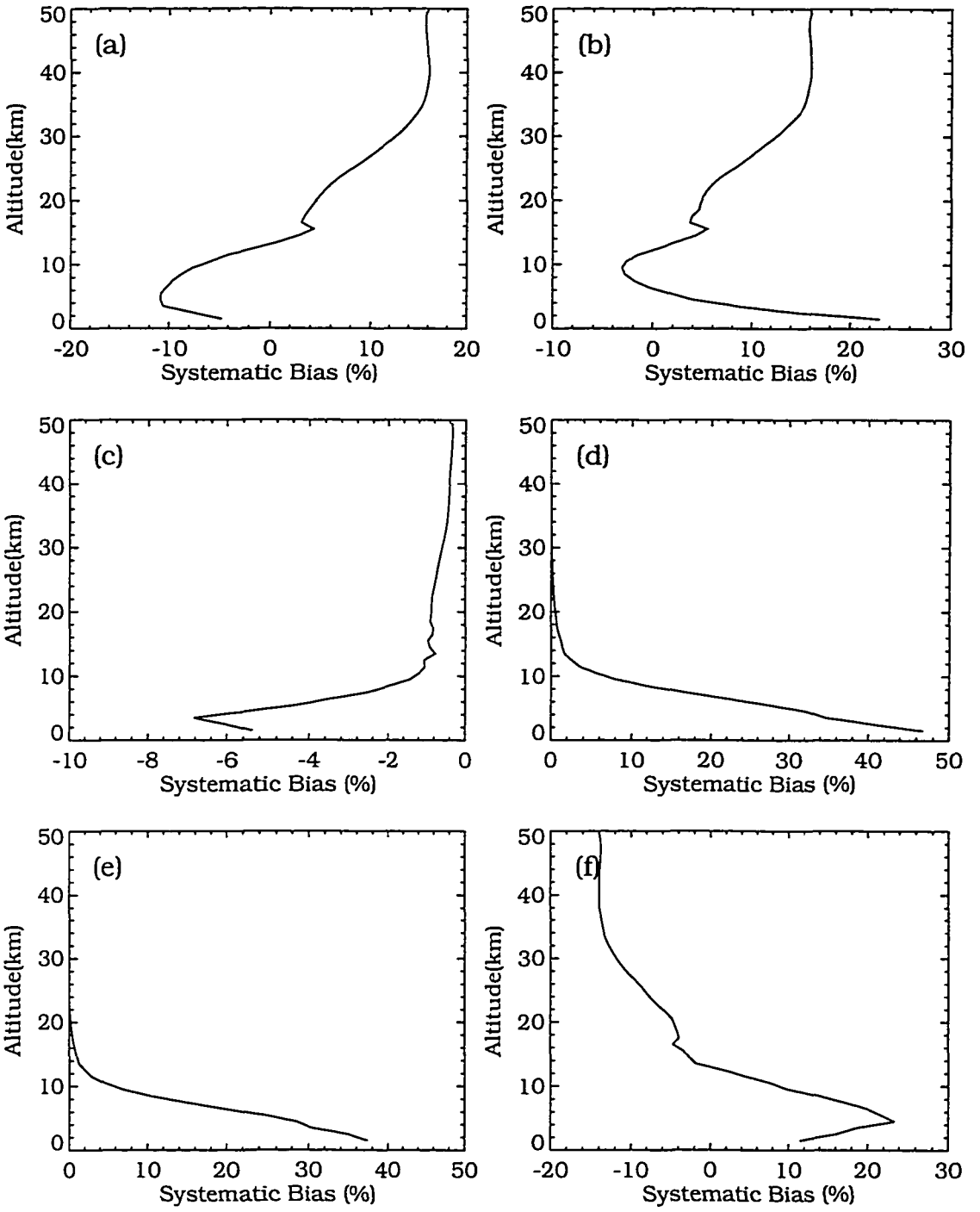


Figure 4.8: Percent change in $\mathbf{x}^r$ when the forward model uses atmospheric model parameters, $\mathbf{b}_A$, as follows: a)HITRAN 82, b)operational EGA database, c)unidentified transitions deleted, d)0.94μm channel incorporates water vapor continuum, e)1.02., 0.94, and 0.6μm channels incorporate water vapor transitions and continuum. Operational inversion uses the HITRAN 92 transitions in a)-e). f)Forward model uses HITRAN 92 transitions and continuum in channels 1-3 and the operational EGA database in the operational inversion.

line strength in the top decade decreases by ≈ 20% in HITRAN 92 but increases by ≈24% and ≈16% in the next two decades. Therefore, in the stratosphere, where the strong lines produce most of the absorption, simulated absorption increases when calculated with HITRAN 82 and the systematic bias in $\mathbf{x}^r$ is 14-15%. This decreases downwards and reverses sign to -12% as the weaker lines contribute more to the absorption signal. The operational EGA data base, which is based on HITRAN 82, was calculated with a Fourier transform radiative transfer code that utilized halfwidth binning. The systematic bias introduced in $\mathbf{x}^r$ when the forward model employs the operational EGA data base while the inverse model uses HITRAN 92 is shown in Fig. 4.8b. In the stratosphere and down to 12 km the bias is identical to Fig. 4.8a but at 10 km the bias reaches -4% and then increases to 10% at 1.5 km and 23% at 0.5 km. Slight variations in halfwidth treatment between the two radiative transfer codes likely explain the difference in the tropospheric bias, between panels a and b, particularly with the high mass paths encountered near the surface where pressure broadening is significant.

Although the quality and completeness of HITRAN 92 is considerably improved over HITRAN 82, there are still transitions lacking quantum identifications that are tagged in HITRAN 92 by setting the lower state energy to a negative value. To make HITRAN 92 a little more realistic, and allow these transitions to be included in the computations, these lower state energies were assigned a nominal value of 300 cm^{-1}, making the transitions largely temperature insensitive. Figure 4.8c shows the change in $\mathbf{x}^r$ when these transitions are deleted from the line data base, a worst case situation but is in fact the standard FASCODE treatment for these transitions. In the stratosphere the bias falls below -1% and decreases to -8% near the surface. It is unlikely however, that a future HITRAN line list with complete quantum identifications would produce as much bias as shown in panel c since changing the lower

state energy from 300 cm^{-1} to their actual values should have a smaller effect than removing the lines entirely.

In the radiative transfer calculations performed so far for k_ω^g in Eq. (2.25) we have assumed that the Voigt line shape is an adequate absorption model and for $|\omega - \omega_0|$ <25cm^{-1} the contribution to k_ω^g is negligible. The wavenumber cutoff assumption, while somewhat arbitrary, was chosen taking into consideration the tradeoff between computational time and accuracy for typical atmospheric conditions. For most trace gases this modeling framework is consistent with laboratory and atmospheric spectral observations. It fails, however, for water vapor, in that it is unable to properly account for the presence of

"anomalous" water vapor absorption in the 8 to 12μm region that appears there as an underlying continuum. In a recent review, Varanasi (1994) traces experimental and theoretical developments as interest in the water vapor continuum moved beyond the spectroscopic community to climate modelers, where continuum absorption in the atmospheric "window region" directly affects heating rates and impairs the predictive accuracy of the models to radiative forcing by greenhouse gases. The 8 to 12μm region is bracketed by the ν_2 fundamental band at 6.25μm and the pure rotational band in the far-infrared. In a comparison of experimental and theoretical spectra, Varanasi shows that applying the 25 cm^{-1} frequency cutoff in the theoretical calculations underestimates the measured continuum absorption but removing the cutoff entirely overestimates the absorption due to the cumulative effect of the far wings from lines in the two bands bracketing this region. Approached from this point of view, explaining the water vapor continuum seems to require a modification to the far wing line shape. Indeed, numerous theories along these and other lines have been advanced with some success, as pointed out by Varanasi in his review. The issue of concern for this work may be stated as follows: If the model of molecular absorption must be modified or new physical processes invoked (such as dimer

formation) to account for "anomalous" absorption in the window region, how does this affect water vapor absorption at 0.94μm?

The importance of continuum absorption at 0.94μm can be estimated with FASCOD3 which employs the line shape formulation developed by Clough et al. (1989c). The Lorentz far wing line shape is modified to account for collision duration effects through a semi-empirical adjustment designed to match experimental measurements in the infrared. Extrapolation of this formulation to 0.94μm may not be entirely accurate but atmospheric validation in the window region and at 4.75μm and 3.2μm found excellent agreement between atmospheric measurements and theoretical calculations (Clough et al., 1989c). Figure 4.8d displays the systematic bias produced by including the continuum in the forward model. The bias, greatest in the troposphere where collisional broadening dominates the line shape, is 10% at 8 km and increases to 46% at 0.5 km. This is a significant bias that shows the continuum must be included for realistic radiative transfer calculations.

We will now examine the tacit assumption that significant water vapor absorption occurs only in channel 2 and may be ignored for other channels. Water absorption in channels 3 and 4 could potentially affect inversion of the other species and then feedback into channel 2 through the aerosol and ozone modeling while water absorption in channel 1 would directly affect the ratio of channel 2 to channel 1. Total line strengths from HITRAN 92 are some 1500, 150, and 15,000 times smaller in channels 1, 3 and 4, respectively, than channel 2, and these ratios are reflected in the simulated stratospheric optical depths after taking into account channel bandwidths. In these channels water absorption makes a negligible contribution to the total signal and may be safely ignored above 10 km. Below however, this is not necessarily true and the situation must be examined in detail for channels 1 and 3. The sparseness and low strength of the water lines in channel 4 are insufficient to produce a significant signal in either the troposphere or stratosphere. In both

channels 1 and 3, the simulated slant path optical depths for water increase at a faster rate than in channel 2 between 10 km and the surface. At 0.5 km the optical depths (including continuum) are $\approx$ 0.122, 4.65, and 0.333 for channels 1-3, respectively. Of somewhat more importance is the relative water contribution to the total signal after Rayleigh removal. In channel 1, water is approximately 12% of the total signal (aerosol and water) between 2 and 6 km while in channel 3 it is insignificant above 7 km and increases to 15% of the total (aerosol, water, and ozone) at 0.5 km.

Figure 4.8e shows the systematic inversion bias produced when the forward model includes water vapor absorption (HITRAN 92, lines and continuum) in channels 1,2, and 3 while the inverse model assumes water absorption (HITRAN 92, lines) is found only in channel 2. Most of the bias in Fig. 4.8e is still caused by the channel 2 continuum. The bias is not quite as great as in panel d because the absorption in channel 1 has reduced $T_{2/1}(z_{th})$, counteracting the continuum increase in channel 2. The bias due to channel 1 (lines only) alone is less than 0.5% in the stratosphere and increases to 7% at 0.5 km. This accounts for most of the bias change between panels d and e, which is roughly 10% at 0.5 km. The remainder is due to the channel 1 continuum and induced channel 2 aerosol modeling errors. Errors in ozone modeling do not contribute since ozone makes an insignificant contribution to the total signal in the troposphere. Channel 2 aerosol modeling errors do not appear to be significant even though aerosol is 80% of the total signal at the top of the troposphere and 20% at 5 km. A 12% error in a 20% contribution produces only a 2% modeling error. Other results (not shown) show that channel 3 water vapor absorption plays a small role in the bias shown in panel e and may also be neglected in the troposphere. From panels d and e we conclude channel 1 water vapor absorption should be included when inverting the $T_{2/1}$ signal and that the continuum must be included in both channels. Aerosol and ozone modeling errors are found to be small and not requiring

correction for an accurate water vapor inversion. Of course, the inverted aerosol and ozone profiles will still contain errors and corrections may be required for them.

Figure 4.8f displays the bias that may now exist in the inverted data product. Here, the forward model has been made as physically realistic as possible by using HITRAN 92 and including the continuum to calculate absorption in channels 1-3. The inverse model uses the operational EGA data base in the operational inversion. Note that this is equivalent to subtracting panel b from panel e. The bias in the stratosphere is the mirror image of that found in panels a and b resulting from the decrease in total channel 2 line strength. Compared to panel e, the tropospheric bias is reduced somewhat and never exceeds 25%. Thus, if the forward atmospheric parameters replace the operational EGA database in the inverse model, $\mathbf{x}^r$ will increase in the stratosphere and decrease in the troposphere.

EGA accuracy. A two step approach was taken to investigate accuracy of the EGA technique, the first examines the accuracy of the core data base by comparing homogeneous path absorptances while the second investigates the overall accuracy of the technique when applied to nonhomogeneous atmospheres. In the first step, homogeneous path absorptances calculated with a set of number densities, U, ranging from $6\text{x}10^{18}$ to $4\text{x}10^{25}$ (molecules/cm^2) at four pressures (0.22, 55.2, 265, and 1013 mb) and a set of temperatures using the core database were compared to FASCOD3 computations. Since the chosen pressures correspond to the data base pressures, only the accuracy of the computations required to turn line-by-line monochromatic transmittances into the database and the temperature coefficient fitting are tested. Most of the comparisons yielded absorption errors $< 0.01\%$ except at 55.2 mb where the error increased in some cases to $\approx 0.7\%$. The second step includes pressure interpolation errors and errors from using an effective absorber amount when combining homogeneous paths to model a nonhomogeneous

atmosphere. Atmospheric absorptances calculated with the EGA technique and then with FASCOD3 for the same middle latitude atmosphere model found differences (EGA-FASCOD3) of -0.05, 0.0, 0.11, 0.29, 0.75, and 0.57% at tangent altitudes of 35, 30, 25, 20, 15, and 10 km, respectively. We assume a systematic error of 0.1% down to 25 km, a linear increase to 0.4% at 20 km, 0.8% at 15 km, a decrease to 0.6% at 10 km, and 0.2% at the surface. The corresponding $\mathbf{K_b}$ is shown in Fig. 4.9. While somewhat conservative, this error is probably a reasonable upper limit that encompasses radiative transfer errors for other conditions encountered with other atmospheres.

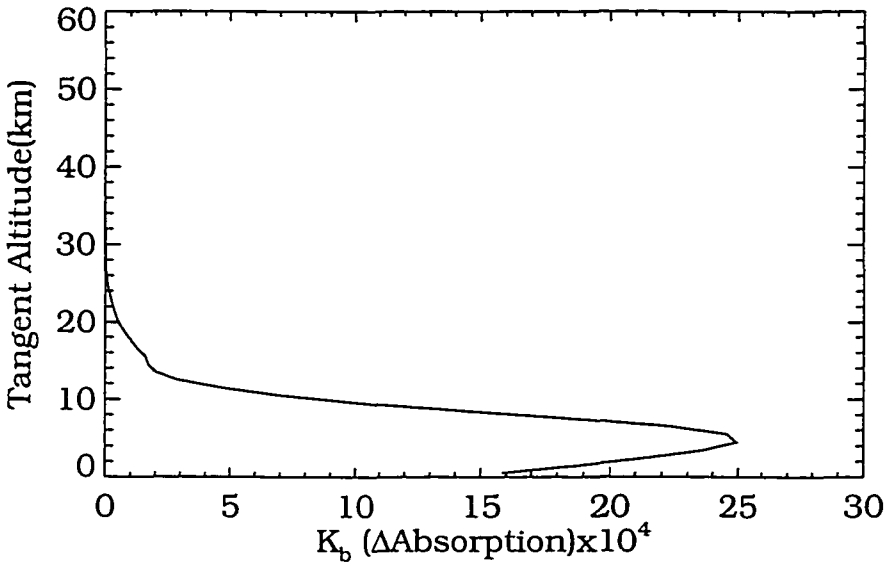


Figure 4.9: $\mathbf{K_b}$ corresponding to a 1% uncertainty in EGA radiative transfer modeling.

INSTRUMENT MODEL PARAMETERS $\mathbf{b_I}$

The sensitivity of the forward model to reference altitude error is shown in Fig. 4.10. A positive perturbation in tangent altitude leads to a decrease in absorption at the perturbation level (due to a thinner shell) while the reverse occurs in the level below. The reference altitude error is assumed to be entirely random. Matrix $\mathbf{K_b}$ in Fig. 4.5 will overestimate the reference altitude error since $\mathbf{D_y K_b} \neq \mathbf{A_b}$ where $\mathbf{A_b}$ is calculated directly and includes the effects of reference altitude perturbations in the other channels.

4.3.3 INVERSION MODEL PARAMETERS c

An iterative solution is also an approximate solution and both systematic and random errors may result from the approximation. The solution process at each tangent level requires a starting point (provided by the first guess profile) and an ending point, given by the convergence criteria. Thus, there are two parameters that fall in the inversion model parameter category, the solution starting point (or first guess profile) and the convergence error limit, $\Delta_{\mathbf{cl}}$. While the iterative solution process determines the statistical relationship between the measured and calculated slant path absorption (designated by $\epsilon_{\mathbf{cl}}$), it is the relationship between the true and calculated absorption (designated as $\epsilon_{\mathbf{tc}}$) that defines the solution error appearing in $\mathbf{x}^r$. As it turns out, the systematic error appearing in $\epsilon_{\mathbf{tc}}$ is primarily a function of $\epsilon_{\mathbf{cl}}$ while random errors in $\epsilon_{\mathbf{tc}}$ may result from both $\epsilon_{\mathbf{cl}}$ and $\Delta_{\mathbf{cl}}$. In the following sections, random errors in $\epsilon_{\mathbf{tc}}$ are addressed first followed by the systematic errors.

Controlled by the experimenter, $\Delta_{\mathbf{cl}}$ is defined as the maximum allowable difference between the measured and calculated slant path absorption for which an iterative solution is acceptable. The difference between measured and calculated

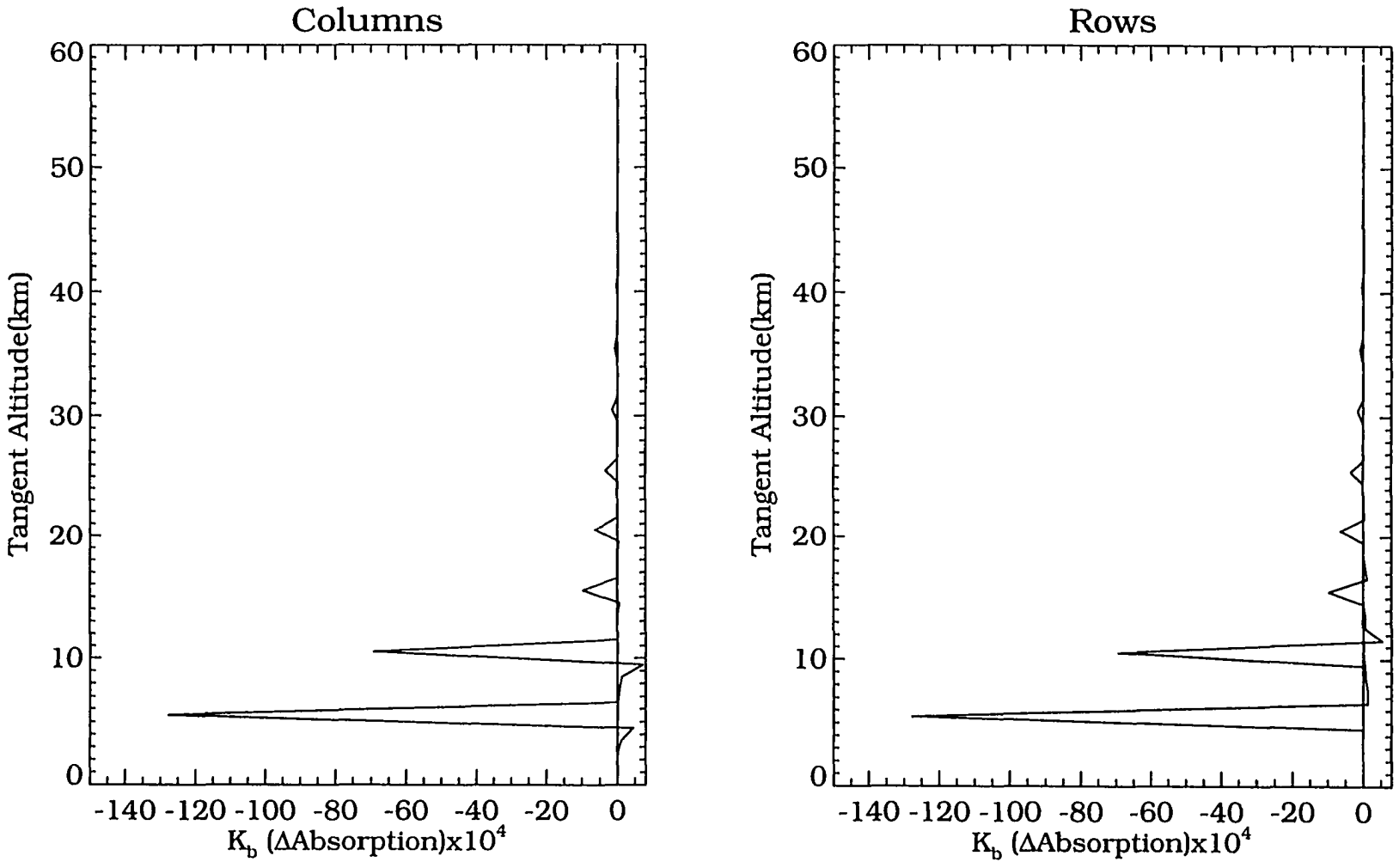


Figure 4.10: Columns and rows of $\mathbf{K_b}$, every 5 km starting at 5.5 km, for a 100 meter perturbation in tangent altitude.

slant path absorption, ϵ_{cl}, is determined by the characteristics of the iterative solution method (secant or Chahine), profile smoothing of U and the upper limit Δ_{cl}.

To better understand how ϵ_{cl} interacts with the measurement uncertainty ϵ_s, where both appear in the analysis at the same point, we refer to Fig. 4.11. $\overline{A}^t, \overline{A}^m$,

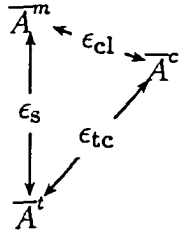


Figure 4.11: Graphical depiction of the relationship between the random measurement error,ϵ_s, the random error introduced by the solution method, ϵ_{cl}, and the resulting random error, ϵ_{tc}.

and $\overline{A}^c$ are the mean slant path absorptances corresponding to the true, measured and calculated values, respectively. The relationship between $\overline{A}^t$ and $\overline{A}^m$ is given by ϵ_s, and between $\overline{A}^m$ and $\overline{A}^c$ by ϵ_{cl}. The uncertainty between $\overline{A}^t$ and $\overline{A}^c$, represented by ϵ_{tc}, is the total uncertainty that must appear in the error equation as follows

$$\mathbf{x}^r - \mathbf{x}^t \propto \frac{\partial IA}{\partial \mathbf{y}^{mi}}\epsilon_s + \frac{\partial IA}{\partial \mathbf{c}}(\hat{\mathbf{c}} - \hat{\hat{\mathbf{c}}}) \tag{4.44}$$

where terms of $\mathbf{x}^r - \mathbf{x}^t$ not related to this discussion are not shown and $\frac{\partial IA}{\partial \mathbf{c}} = \frac{\partial IA}{\partial \mathbf{y}^{mi}}$ for the inversion model parameter ϵ_{cl}, identical to ϵ_s. Instead of using $\mathbf{c}^t$ and $\hat{\mathbf{c}}$ as the true and best estimates, $\hat{\mathbf{c}}$ and $\hat{\hat{\mathbf{c}}}$ are simply two reasonable selections for $\mathbf{c}$ by the experimenter to represent the evaluation and linearization states respectively. Selecting $\epsilon_{cl} = \hat{\mathbf{c}} - \hat{\hat{\mathbf{c}}}$ guarantees that the solution error vanishes for an exact solution ($\epsilon_{cl} = 0$). We then have

$$\mathbf{x}^r - \mathbf{x}^t \propto \frac{\partial IA}{\partial \mathbf{y}^{mi}}\epsilon_{tc} = \frac{\partial IA}{\partial \mathbf{y}^{mi}}(\epsilon_s + \epsilon_{cl}) \tag{4.45}$$

A good choice for Δ_{cl} is the magnitude of the estimated random measurement error ϵ_s (ignoring random errors introduced by the separation model) but it may

also be adjusted upwards or downwards. For an exact solution containing random measurement noise, $\Delta_{cl} \rightarrow 0, \epsilon_{cl} \rightarrow 0, \overline{A}^c \rightarrow \overline{A}^m$, and $\epsilon_{tc} \rightarrow \epsilon_s$ while for a noise free approximate solution, $\epsilon_s \rightarrow 0, \overline{A}^m \rightarrow \overline{A}^t$, and $\epsilon_{tc} \rightarrow \epsilon_{cl}$.

Unlike the other model parameters, ϵ_s and ϵ_{cl} may not be independent since the final choice of $\overline{A}^c$ depends not only on $\overline{A}^m$ at the tangent altitude of interest but other levels also. In this situation the error covariances become

$$\mathbf{S}_{\mathbf{x}^r-\mathbf{x}^t} \propto (\frac{\partial IA}{\partial \mathbf{y}^{mi}})\mathbf{S}_{\epsilon_{tc}}(\frac{\partial IA}{\partial \mathbf{y}^{mi}})^T = (\frac{\partial IA}{\partial \mathbf{y}^{mi}})(\mathbf{S}_{\epsilon_s}+\mathbf{S}_{\epsilon_{cl}}+\mathbf{S}_{\epsilon_s,\epsilon_{cl}}+\mathbf{S}_{\epsilon_{cl},\epsilon_s})(\frac{\partial IA}{\partial \mathbf{y}^{mi}})^T \quad (4.46)$$

with $\mathbf{S}_{\epsilon_s,\epsilon_{cl}}, \mathbf{S}_{\epsilon_{cl},\epsilon_s}$ representing the covariances between ϵ_s and ϵ_{cl}. Neither matrix is symmetric, but since $\mathbf{S}_{\epsilon_s,\epsilon_{cl}} = \mathbf{S}^T_{\epsilon_{cl},\epsilon_s}$, the two matrices when combined $(\mathbf{S}^T_{\epsilon_{cl},\epsilon_s} + \mathbf{S}_{\epsilon_{cl},\epsilon_s})$, make symmetric contributions to $\mathbf{S}_{\mathbf{x}^r-\mathbf{x}^t}$. The error interactions, for a diagonally dominant set of matrix equations, may be loosely interpreted graphically using the the triangle in Fig. 4.11. When ϵ_s and ϵ_{cl} are independent, $\mathbf{S}_{\epsilon_s,\epsilon_{cl}} = \mathbf{S}_{\epsilon_{cl},\epsilon_s} = 0$, and the angle between ϵ_s and ϵ_{cl} is 90°. When ϵ_s and ϵ_{cl} are predominantly correlated negatively, the covariances are negative, and ϵ_{tc} decreases. In the figure, this corresponds to an angle between ϵ_s and ϵ_{cl} <90°. For a positive correlation, the angle is >90° and ϵ_{tc} increases.

To examine how the solution method adjusts ϵ_{cl} to produce ϵ_{tc}, simulated slant path transmittances containing realistic random measurement errors (50 profiles) were inverted using $\Delta_{cl} = |\epsilon_s|$ to create a statistical database suitable for calculating $\mathbf{S}_{\epsilon_{cl}}, \mathbf{S}_{\epsilon_{cl},\epsilon_s}$, and the contribution by each covariance term to $\mathbf{S}_{\mathbf{x}^r-\mathbf{x}^t}$. The uncertainty in $\mathbf{x}^r$, due to the random errors, was compared to that predicted by the formal error analysis $(\mathbf{D}_\mathbf{y}\mathbf{S}_{\epsilon_\mathbf{y}}\mathbf{D}^T_\mathbf{y})$ as a redundant check on the importance of ϵ_{cl}.

In the troposphere it was found that the combined terms $\mathbf{S}_{\epsilon_{cl}} + \mathbf{S}_{\epsilon_s,\epsilon_{cl}} + \mathbf{S}_{\epsilon_{cl},\epsilon_s}$ were small relative to $\mathbf{S}_{\epsilon_s}$ and that $\mathbf{S}_{\epsilon_{tc}} \approx \mathbf{S}_{\epsilon_s}$ for all four inversion algorithms. The uncertainty in $\mathbf{x}^r$ at these levels agreed well with that predicted by the formal error analysis, confirming this result.

In the stratosphere, the response of the inversion algorithms vary. For the onion peel algorithm the combined terms $\mathbf{S}_{\epsilon_{cl}} + \mathbf{S}_{\epsilon_s,\epsilon_{cl}} + \mathbf{S}_{\epsilon_{cl},\epsilon_s}$ remained small compared to $\mathbf{S}_{\epsilon_s}$ up to 35 km. Above 35 km, where lower signal to noise ratios occur, $\mathbf{S}_{\epsilon_{cl}}$ increased to about half of $\mathbf{S}_{\epsilon_s}$ as did the positively correlated $\mathbf{S}_{\epsilon_s,\epsilon_{cl}} + \mathbf{S}_{\epsilon_{cl},\epsilon_s}$, which made the effect of $\mathbf{S}_{\epsilon_{tc}}$ on $\mathbf{S}_{\mathbf{x}^r-\mathbf{x}^t}$ greater than $\mathbf{S}_{\epsilon_s}$ alone at some levels. The uncertainty in $\mathbf{x}^r$ was also greater than predicted with the formal error analysis. However, in several other cases using different solution initializations or first guesses, it was found that $\mathbf{S}_{\epsilon_{tc}} \approx \mathbf{S}_{\epsilon_s}$. It is not surprising that the first guess affects $\mathbf{x}^r$ in low signal to noise ratio situations. We will assume that for a group of profiles there is enough variation within $\mathbf{x}^t$, that when inverted with a fixed first guess, more cases with $\mathbf{S}_{\epsilon_{tc}} \approx \mathbf{S}_{\epsilon_s}$ occur than with $\mathbf{S}_{\epsilon_{tc}}$ greater than $\mathbf{S}_{\epsilon_s}$. Thus, for a zonal mean analysis we assume $\mathbf{S}_{\epsilon_{tc}} \approx \mathbf{S}_{\epsilon_s}$ but recognize that this may not be true for all individual profiles.

The stratospheric vertical smoothing of U that is part of both the operational and Mill-Drayson smoothed algorithms overwhelming controls ϵ_{cl} and ϵ_{tc}. The smoothing pushes $\overline{A}^c$ away from the $\overline{A}^c$ Chahine's predictor would otherwise naturally select, thus increasing $\mathbf{S}_{\epsilon_{cl}}$. The terms $\mathbf{S}_{\epsilon_s,\epsilon_{cl}} + \mathbf{S}_{\epsilon_{cl},\epsilon_s}$ are negatively correlated along the diagonal, more than countering the increase in $\mathbf{S}_{\epsilon_{cl}}$, and in fact, $\mathbf{S}_{\epsilon_{tc}}$ is smaller than $\mathbf{S}_{\epsilon_s}$ by a significant amount. Since the vertical smoothing is an explicit step in both inversion algorithms, the smoothing is automatically included in the formulation of $\mathbf{D_y}$ and the terms $\mathbf{S}_{\epsilon_{cl}} + \mathbf{S}_{\epsilon_s,\epsilon_{cl}} + \mathbf{S}_{\epsilon_{cl},\epsilon_s}$ need not be calculated. The uncertainty in $\mathbf{x}^r$, due to random measurement errors, agreed very well for both algorithms to the errors predicted by the formal error analysis.

The Mill-Drayson algorithm, without explicit smoothing of U, redistributes the measurement uncertainty upwards to higher layers by adjusting the profile of U, at and above the tangent layer, by a multiplicative constant. Profile curvature of $\overline{A}^c$ and $\mathbf{x}^r$ is much less than for the onion peel algorithm, suggesting the Mill-Drayson algorithm provides a form of smoothing that is implicit in nature. Implicit smoothing

is not captured in the formulation of $\mathbf{D_y}$. We will show in a following section that the Mill-Drayson and onion peel $\mathbf{D_y}$ are in fact identical, and because of this the formal error analysis overestimates the actual uncertainty in $\mathbf{x}^r$. To properly account for the effects of the implicit smoothing in the formal error analysis requires a full description of $\mathbf{S}_{\epsilon_{tc}}$, the approach is presented in Sec. 4.4.4. The implicit smoothing acts very much like explicit smoothing, only not quite as effective as the three point smoothing of U.

Inversion random errors are insensitive to reasonable choices of Δ_{cl} for the onion peel, operational, and Mill-Drayson smoothed algorithms. However, for the Mill-Drayson algorithm, selecting Δ_{cl} too small decreases the strength of the implicit smoothing by increasing $\mathbf{S}_{\epsilon_{tc}}$ to $\mathbf{S}_{\epsilon_s}$ and negates the primary advantage of the algorithm over the onion peel. For the operational, smoothed Mill-Drayson, and onion peel algorithms, the effect of $\mathbf{S}_{\epsilon_{cl}}$ on $\mathbf{x}^r$ is negligible and will be dropped from the error analysis. For the Mill-Drayson algorithm the implicit smoothing must be taken into account in the analysis.

From Eq. (4.45) we see that systematic errors in ϵ_{tc} are actually due to systematic biases in ϵ_{cl} since by definition the mean of ϵ_s is zero. Examples of the importance of systematic errors in ϵ_{tc} on $\mathbf{x}^r$ are shown in Fig. 4.12. Fifty simulations containing realistic levels of measurement noise were inverted with two different first guess profiles assuming $\Delta_{cl} = |\epsilon_s|$ as before. The first guess profiles were generated by multiplying a noise free onion peel inversion by 0.5 or 1.5 so that the solution process begins above or below the solution at each tangent level. The greatest uncertainty occurs in the troposphere and in the low signal to noise regions of the upper stratosphere. Using a first guess profile with a profile shape identical to the true profile favors the Mill-Drayson algorithm, but other choices of first guess profile show systematic errors at levels consistent with the other algorithms.

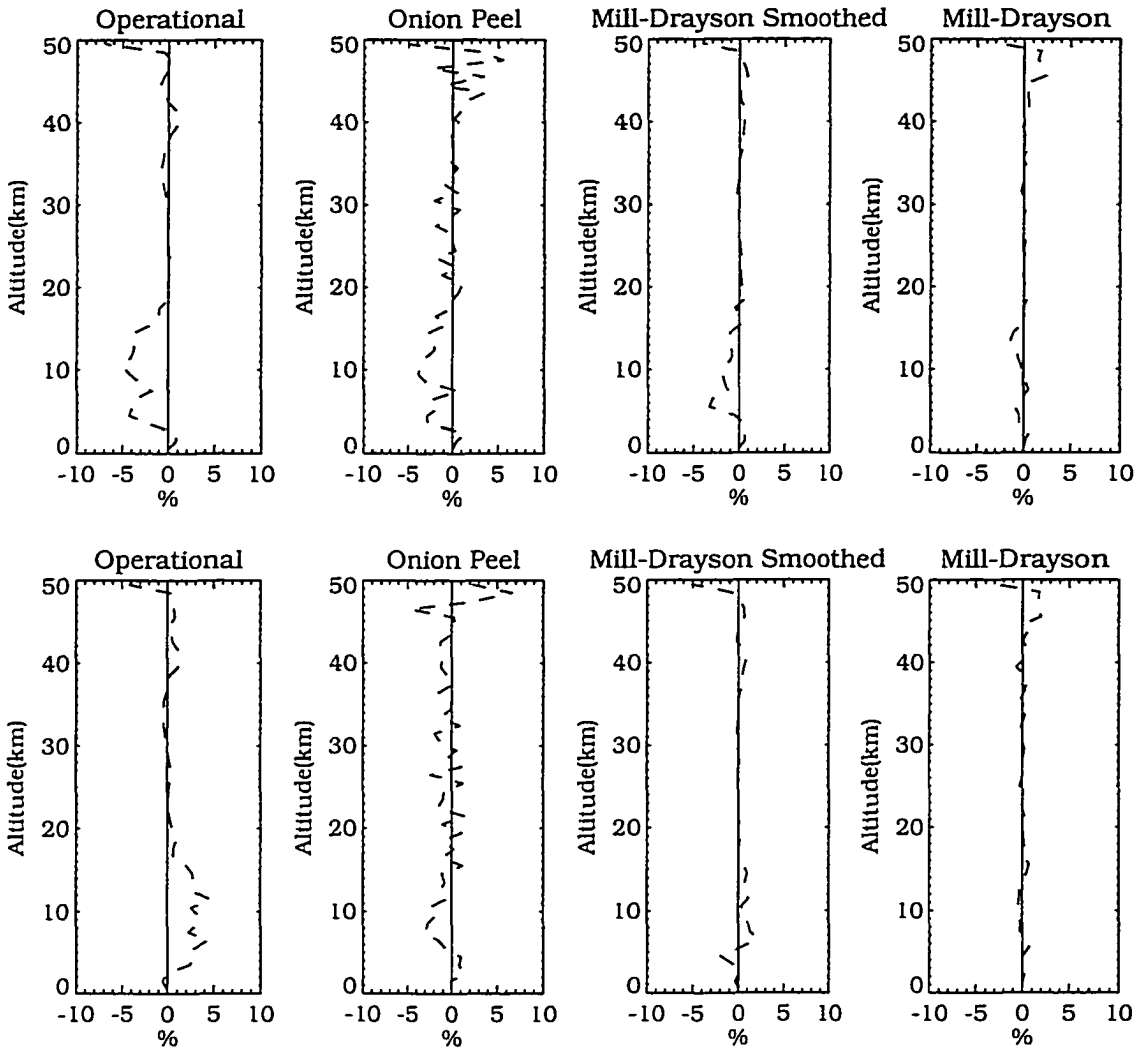


Figure 4.12: Systematic errors in retrieved mixing ratios due to two different first guess profiles. First guess profiles are a noise free onion peel inverted profile multiplied by 0.5 (upper set of panels) and 1.5 (lower set of panels).

For the operational, onion peel, and Mill-Drayson smoothed algorithms, the systematic error in ϵ_{tc} can be reduced by arbitrarily reducing Δ_{cl} without increasing the random errors in ϵ_{tc}. Of course, the number of allowable iterations must be increased for convergence to occur and computational time will show a corresponding increase. The convergence limit Δ_{cl}, for the Mill-Drayson algorithm, cannot be reduced without increasing the random errors significantly. We therefore assume the systematic error in ϵ_{tc} may be ignored for all but the Mill-Drayson algorithm. Systematic errors will be calculated for the Mill-Drayson algorithm using the first guess profiles normally employed in the operational data reduction procedure.

4.3.4 Error Covariances $\mathbf{S}_{\epsilon_b}$,$\mathbf{S}_{\epsilon_y}$

In this section the error covariances $\mathbf{S}_{\epsilon_b}$ for the model parameters listed in Table 4.1 are defined along with $\mathbf{S}_{\epsilon_y}$. The model parameters are assumed to be independent of one another. Unlike the systematic error components, most of the random errors vary with time, measurement location, and altitude. However, in the following development these dependencies are not shown explicitly. The model parameter variances are presented first and then the steps taken to construct the error covariance matrices are presented.

The temperature error variance is

$$\sigma_T^2 = \epsilon_T^2 \tag{4.47}$$

where ϵ_T is the estimated temperature uncertainty specified by the National Weather Service for each temperature profile.

The reference altitude error variance is given by

$$\sigma_{RA}^2 = \epsilon_{RA}^2 \tag{4.48}$$

where ϵ_{RA} is 200 m from Chu et al.(1993).

The estimated measurement error ϵ_y in channel 2 corresponding to each $\mathbf{y}^m(z_{th})$ is used to construct the measurement error variance as follows

$$\sigma^2_{\epsilon_y} = \epsilon^2_y \tag{4.49}$$

Since a_2 is a constant, the ozone error variance is given by

$$\sigma^2_{\Delta\delta^{Ozone}_{2-1}} = (a_2)^2 \sigma^2_{\delta^{O_3}(\lambda_3)} \tag{4.50}$$

where $\sigma^2_{\delta^{O_3}(\lambda_3)}$ is the error variance in the channel 3 ozone slant path optical depth. The approximation $\sigma^2_{\delta^{O_3}(\lambda_3)} \approx \sigma^2_{\delta(\lambda_3)} = \epsilon^2_{\lambda_3}$, where $\epsilon^2_{\lambda_3}$ is the estimated measurement error in channel 3, simplifies the ozone error variance calculations by avoiding a detailed development of an error budget for $\sigma^2_{\delta^{O_3}(\lambda_3)}$.

The variance of the aerosol difference error term is

$$\sigma^2_{\Delta\delta^{Aerosol}_{2-1}} = \sigma^2_{\delta^{Aer}(\lambda_4)}(k_2 + 2k_3R_1 + 3k_4R_1^2)^2 + \sigma^2_{\delta^{Aer}(\lambda_1)}(k_1 - 1 - k_3R_1^2 - 2k_4R_1^3)^2 \tag{4.51}$$

where $\sigma^2_{\delta^{Aer}(\lambda_1)}$ and $\sigma^2_{\delta^{Aer}(\lambda_4)}$ is the error variance in the aerosol slant path optical depth for channels 1 and 4, respectively. The k_j are the coefficients for the empirical expression relating R_2 to R_1 in Thomason(1991). The approximation $\sigma^2_{\delta^{Aer}(\lambda)} \approx \sigma^2_{\delta(\lambda)} = \epsilon^2_\lambda$ is made for both channels 1 and 4.

Construction of the temperature and reference altitude covariance error matrices is straightforward. For both, the error variances define the on-diagonal elements, the off-diagonal covariances are assumed to be zero.

The measurement, ozone, and aerosol error variances are all expressed in terms of the estimated measurement errors in channels 1, 2, 3, and 4. While investigating the error characteristics of channel 2 it was discovered that the error variances were overestimated in the middle atmosphere. This is discussed in Appendix B where a correction procedure is given for selected portions of the atmosphere and is applied here to the variances for all three terms in the construction of the error covariance

matrices. The on-diagonal elements of $\mathbf{S}_{\epsilon_y}$ and the two $\mathbf{S}_{\epsilon_b}$ consist of the adjusted estimated variances while the off-diagonal covariances are formed from the "representative" correlation coefficients, described in Appendix C, multiplied by the error standard deviations for the two altitude levels (Eq. (2.78)).

4.3.5 Null Space

Evaluation of the null space error requires an independent estimate of $\mathbf{x}^t$ and the covariance of $\mathbf{x}^t$, $\mathbf{S}_{\mathbf{x}^t}$. Ideally, they should be obtained from an instrument with superior horizontal and vertical resolution. To calculate the systematic error in a single measurement would require $\mathbf{x}^t$ be co-located spatially and temporally. For a zonal mean analysis, an ensemble of co-located $\mathbf{x}^t$ is needed, a rather demanding requirement considering the general scarcity of water vapor measurements in the upper troposphere and stratosphere.

During the time periods studied here, 1987 and 1990, only scattered *in situ* observations of water vapor were obtained in the lower stratosphere. No measurements of upper stratospheric water vapor were made during this time period. Discarding the co-location requirement and looking to remotely sensed water from satellite experiments conducted in other years improves the situation only slightly since these experiments produce data at lower resolution. In the troposphere there is a large body of *in situ* radiosonde observations that may be used to define $\mathbf{x}^t$. As shown in Sec. 5, where the radiosonde measurements are used as correlative observations, the quality of the water vapor measurements deteriorates in the upper troposphere and substantial systematic biases exist between hygrometer types. Given the lack of information on $\mathbf{x}^t$ in the stratosphere and upper troposphere, and observations of uncertain quality in the troposphere, only rough estimates of the null space error can be developed. Due to their uncertain nature, they are shown here for illustrative purposes only and will not be included in the error budget.

The zonal estimate of $\mathbf{x}^t - \overline{\mathbf{x}}$ is developed as follows. In section 5.3 it is shown that correlative comparisons between SAGE II and the most accurate (thin film capacitive) hygrometer average approximately 10% in the upper troposphere. In Chiou et al.(1993) it is shown that zonal mean comparisons of SAGE II and LIMS water vapor range from 20-25% for altitudes above the hygropause, with SAGE II higher, and may increase in the hygropause region with crossovers in zonal means occurring in some latitude bands. It is possible that much of the zonal mean difference may be due to systematic differences between the two experiments. We therefore take 10% of $\overline{\mathbf{x}}$ as a reasonable estimate of $\mathbf{x}^t - \overline{\mathbf{x}}$ for each zonal mean where $\overline{\mathbf{x}}$ is the SAGE II zonal mean.

Figure 4.13 displays the systematic null space error for the operational inversion for January 1990. The error is generally much less than 1% in the stratosphere and at most 2-3% in the troposphere. For the onion peel inversion (figure not shown) the error increases to 1% in the stratosphere but is identical in the troposphere. The systematic null space error is small because $\overline{\mathbf{x}}$ is close to $\mathbf{x}^t$ and term cancellation between the negative diagonal elements of $(\mathbf{A}-\mathbf{I})$ and positive off diagonal elements occurs.

The estimate of $\mathbf{S}_{\mathbf{x}^t}$ was developed using SAGE II observations in the stratosphere and radiosonde measurements (described in Sec. 5) in the troposphere. To facilitate the development of $\mathbf{S}_{\mathbf{x}^t}$, $\mathbf{S}_{\mathbf{x}^t}$ is expanded as follows

$$S_{\mathbf{x}^t}(i,j) = C(i,j)\sigma_{gv}(i)\sigma_{gv}(j) \tag{4.52}$$

where $\mathbf{C}$ is the correlation coefficient matrix and σ_{gv} is the standard deviation of water vapor geophysical variability.

For each radiosonde station, monthly mean water vapor mixing ratios and estimates of the associated variability σ_R were calculated on the lower eight mandatory levels (1000-200 mb), with an additional pass made through the data to screen out

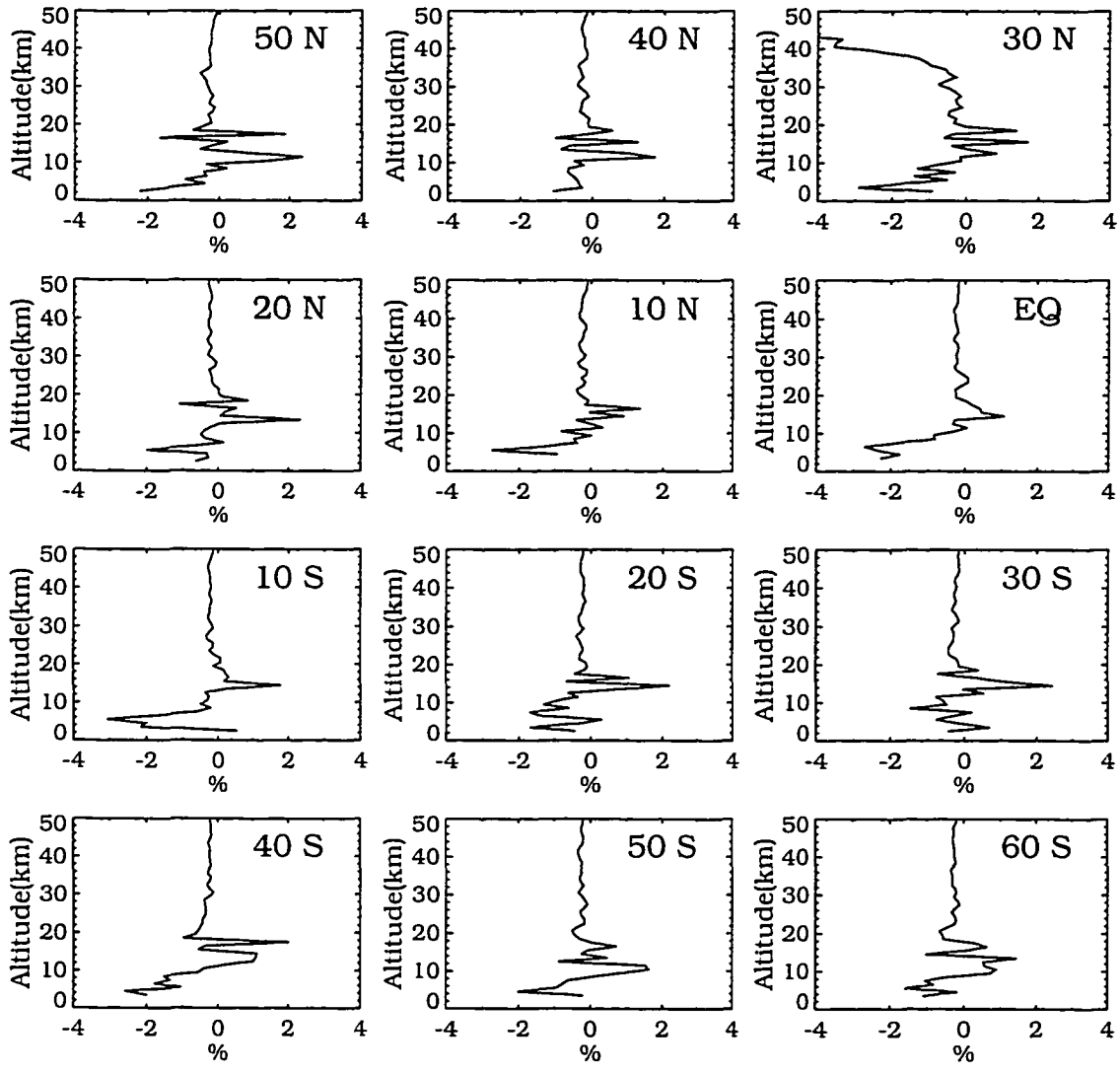


Figure 4.13: Estimated systematic null space error corresponding to the operational retrieval, January 1990.

erroneous values. SAGE II observations in the troposphere are limited to cloud free and likely drier atmospheres, a subset of the atmospheres observed by the radiosondes. To account for this effect on σ_{gv}, zonal mean values of the station ratio $\sigma_R/\overline{x}_R$ are constructed using all stations within a zone.

Then the mixing ratio variability of the drier atmospheres, σ_D, is estimated with $\sigma_D = \overline{x}_S(\sigma_R/\overline{x}_R)$, where $\overline{x}_S$ is the SAGE II zonal mean. This scaling maintains the ratio of variability to mean mixing ratio in the drier atmosphere and since $\overline{x}_S < \overline{x}_R$, it follows that $\sigma_D < \sigma_R$.

In the stratosphere, the SAGE II water vapor profiles from the operational inversion were used to calculate σ_S, the standard deviation of the inverted water vapor. When expressed relative to $\overline{x}_S$, the vertical variation of $\sigma_S/\overline{x}_S$ was similar across latitude bands, being 5-10% in the lower stratosphere and 15-20% at 45 km. The standard deviation for level i may be expressed as

$$\sigma_S = \sqrt{\sigma_{gv}^2 + \sigma_{rn}^2} \tag{4.53}$$

where σ_{rn} represents the standard deviation of the random noise. To estimate the proportion of σ_{gv} to σ_{rn} we look ahead to the results of Sec. 4.5. Comparing $\sigma_S/\overline{x}_S$ of Fig. 4.48 to Fig. 4.49, the estimated random noise uncertainty from the formal error analysis, we see that for most latitude bands $\sigma_S/\overline{x}_S$ is roughly twice the estimated noise uncertainty. This suggests σ_{gv} is three times σ_{rn}, and if it were possible to entirely remove the random noise from the measurements, leaving only σ_{gv}, σ_S would only drop $\approx$ 15%. Thus, σ_S represents a slight overestimate of σ_{gv} if the actual random error is reasonably close to the estimated. However, at 50° S and 60° S it appears that the ratio between the two terms changes, with random noise dominant, and σ_S may be a significant overestimate of σ_{gv}.

To construct σ_{gv}, we use σ_D in the troposphere up to 14.5 km and σ_S above this. If σ_D is not available in the upper troposphere σ_{gv} is defined by interpolation between the highest σ_D and lowest σ_S.

The correlation coefficient matrix in the troposphere was calculated by combining two years of January and February radiosonde data. The matrix was calculated two ways, using all the hygrometer data and then with data taken only with the thin film capacitive hygrometer (#14 in Table 5.1), believed to offer the best sensitivity and accuracy. The stratospheric correlation coefficients (above 15.5 km) were calculated directly from the SAGE II inverted water vapor data.

Figure 4.14 displays the estimated random null space error for the operational inversion. Three cases are shown in the figure corresponding to different assumptions concerning $\mathbf{C}$. The first case, given by the solid line, assumes $C_{i,i} = 1$, $C_{i,j} = 0$ for $i \neq j$, and represents a minimum information worst case. The null space error is largest in this situation because only the term $(A_{i,i} - I_{i,i})^2 C_{i,i} (\sigma_i)^2$ can contribute to $\mathbf{S}_N$. The vertical variation of the error in this case directly reflects σ_{gv}. Estimated error in the troposphere typically reaches a maximum near 5 km with values ranging from 20% to 45%, and decreases upwards to a minimum in the lower stratosphere. The errors increase upwards from there and range from 15 to 22% at 50 km.

For the second case, given by the dashed line, $\mathbf{C}$ is defined by SAGE II in the stratosphere and all hygrometer types in the troposphere, as described above. The third case, the dotted line, differs from the second only in that the tropospheric $\mathbf{C}$ was determined from the thin film capacitive hygrometers. For these two cases the negative $(A_{i,i} - I_{i,i})$ combines with adjacent positive terms (since $C_{i,j}$ is positive) to reduce $\mathbf{S}_N$ at all levels. Compared to the first case, tropospheric errors are typically reduced by half between 5 and 10 km and stratospheric errors drop by factors of two to three. For these two cases there is little difference in the troposphere because

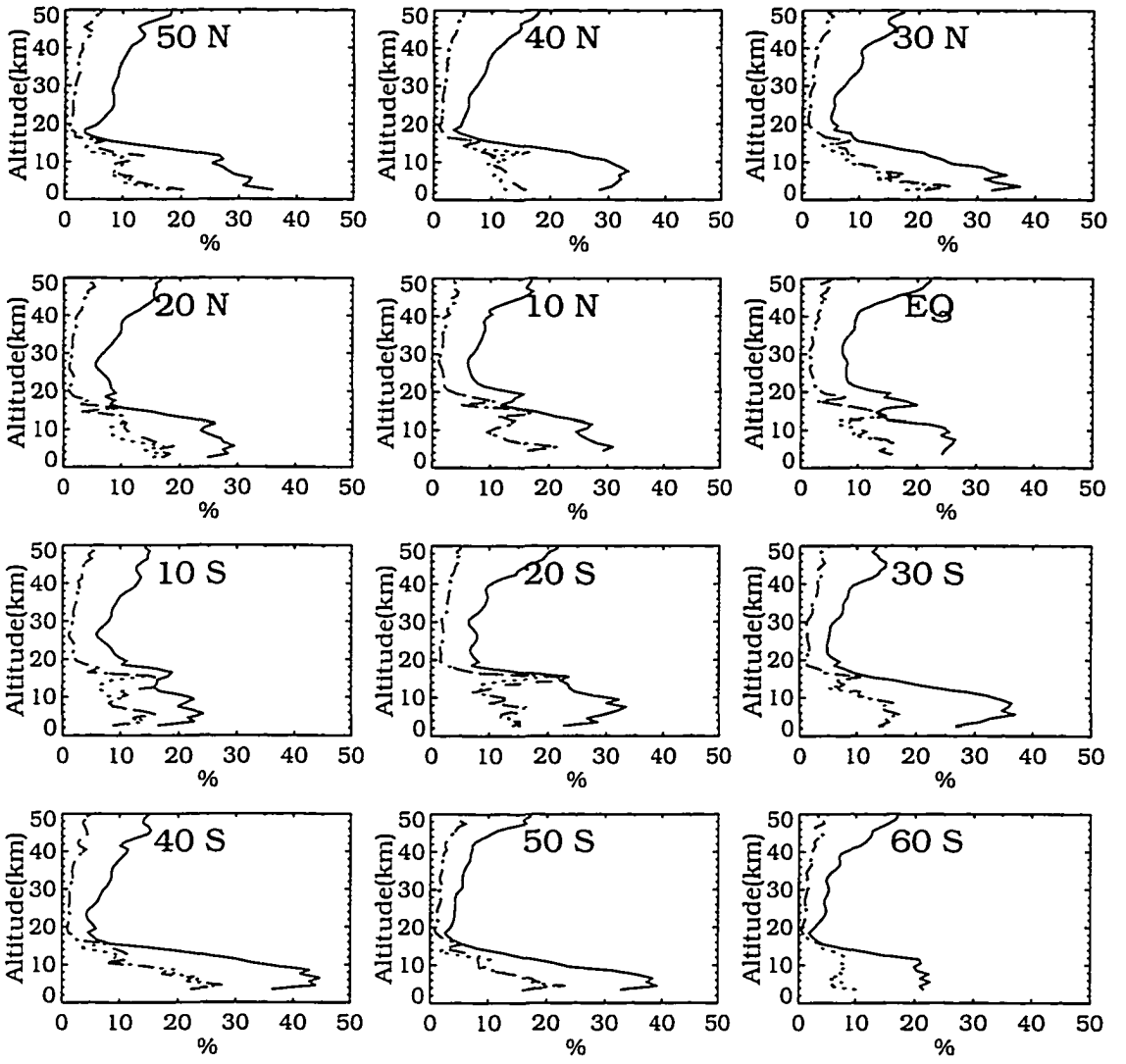


Figure 4.14: Estimated random null space error corresponding to the operational retrieval for January 1990. The correlation coefficient matrix, C, is calculated three different ways: a) C is diagonal in the stratosphere and troposphere (solid line); b) C is based on the SAGE II water vapor measurements in the stratosphere and all radiosonde hygrometers in the troposphere (dotted line); c) C is identical in the stratosphere to case b but only thin film capacitive hygrometers are used in the troposphere (dashed line).

both correlation matrices display similar high positive correlations in the vicinity of the diagonal. Away from the diagonal the correlations can be quite different but the form of $\mathbf{A}$ limits their influence on $\mathbf{S}_N$. Some similarity is to be expected since one set of correlation coefficients is a subset of the full set and in the southern hemisphere the subset may compose $\frac{1}{4}$ to $\frac{1}{2}$ of the full set for some zones.

The radiosonde hygrometer measurements are for all practical purposes point measurements and the variances and correlation coefficients likely provide an accurate picture of $\mathbf{S}_{x'}$ in the troposphere. However, for the stratospheric portion we used SAGE II water vapor measurements from the operational inversion that contain absorption and mixing ratio smoothing. Smoothing decreases σ_S and increases the off diagonal correlation coefficients. To estimate the importance of smoothing on σ_S and $\mathbf{C}$, they were recalculated with water vapor profiles from the onion peel inversion which contains only absorption smoothing. As expected, σ_S increases and $\mathbf{C}$ becomes more diagonally dominant. When these statistics are used to calculate the operational inversion null space error the stratospheric error increases. For the worst case situation, with $C_{i,i} = 1$, the error increases by 2-3% from 20 to 40 km and somewhat more than this above 40 km. In fact, this increase is roughly equivalent to the difference in $\sigma_S/\overline{x}_S$ shown in Fig. 4.48 between the operational and onion peel inversion algorithms and is due entirely to the increase in σ_S. For the second and third cases, now recalculated with a more diagonal $\mathbf{C}$, the error roughly parallels the first case shown in Fig. 4.14 and is lower by only 2-5%. Taking this process a step further, if σ_S and $\mathbf{C}$ are calculated from an inversion without absorption smoothing, one would expect to see the error increase by a similar amount and a convergence of the second and third cases to the first case. It appears that a better estimate of the stratospheric null space error is probably given by the first case in Fig. 4.14.

The best estimate of random null space error results from splicing the tropospheric error from either cases two and three (dotted line, dashed line in Fig. 4.14) to the stratospheric errors given by the first case (solid line).

4.4 Water Vapor Error Analysis

In this section, error budgets for each of the inversion models are constructed for water vapor observations in 1987 and 1990 following the formal error analysis of Sec. 2.4. Examples of $\mathbf{D_y}$ and $\mathbf{A}$, calculated with the northern middle latitude model atmosphere described in Sec. 4.3.1, are presented first. Then, systematic and random error terms for the model parameters of Table 4.1 are developed using SAGE II experimental quantities for the most part. To make the error budget realistic, $\mathbf{D_y}, \mathbf{K_b}$, and $\mathbf{A_b}$ are calculated assuming a model atmosphere defined by the inverted constituent data while the estimated measurement uncertainties for the various SAGE II channels define $\mathbf{S_{\epsilon_b}}$ and $\mathbf{S_{\epsilon_y}}$. After the error budget is presented for each inversion, the zonal variation of the error terms is examined. Since all of the inversion methods yield profiles with generally similar characteristics, discussion of the water vapor profile inversions is presented in the section following this one, with a focus on profile comparisons among the different methods.

4.4.1 Onion Peel Inversion

The contribution function $\mathbf{D_y}$, as given by Eq. (2.62), defines the sensitivity of the inverse model to changes in the slant path absorption. To demonstrate the effect of the three point slant path absorption vertical smoothing, Eqs. (4.27) and (4.28), on $\mathbf{D_y}$, $\mathbf{D_y}$ without this smoothing is examined first and is shown in Fig. 4.15 for the onion peel algorithm. The columns of $\mathbf{D_y}$ contain the response of the inverse model, a change in $\mathbf{x}^r$, to a delta function perturbation in slant path absorption. In this

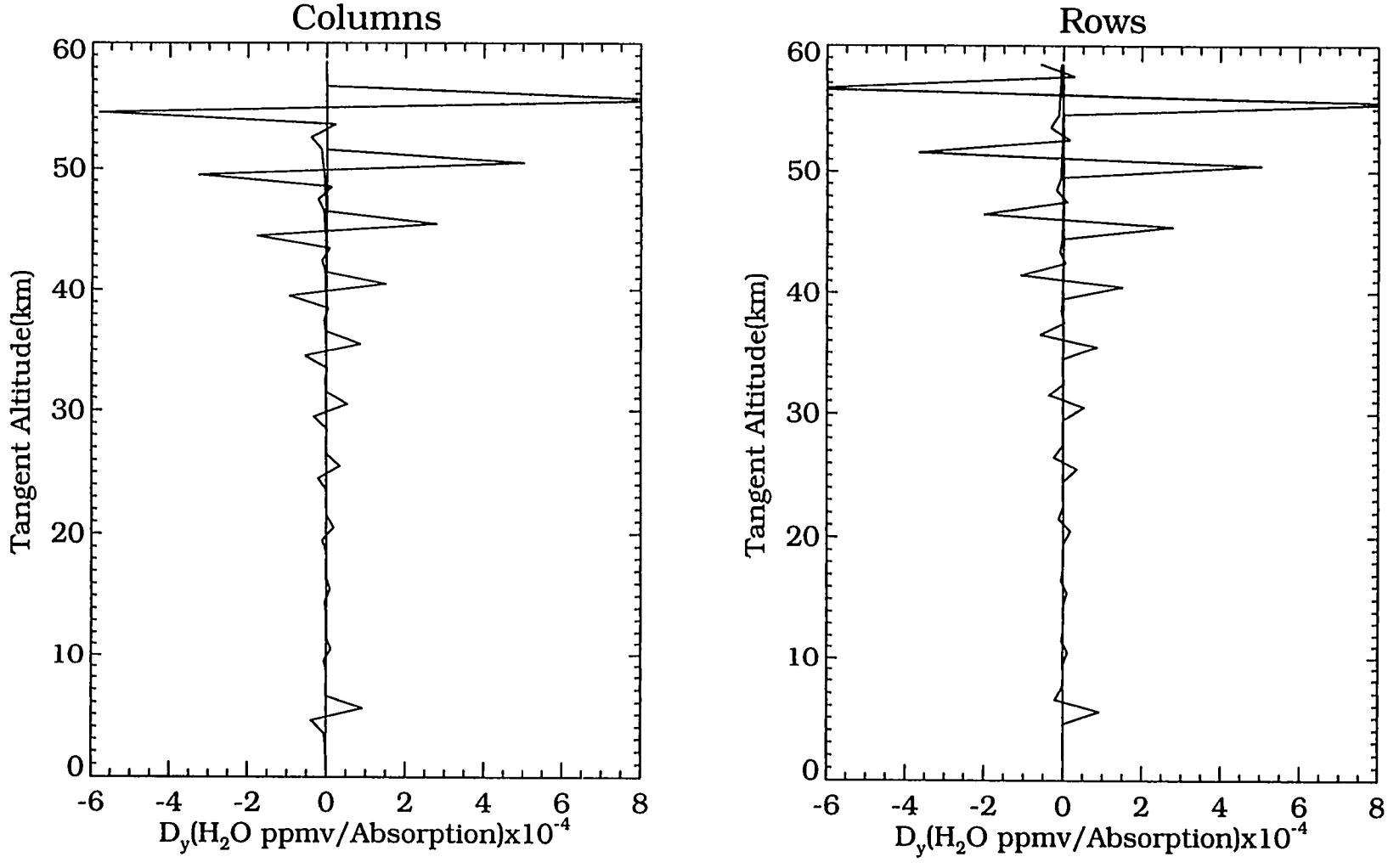


Figure 4.15: Columns and rows of the onion peel inversion contribution function matrix $\mathbf{D_y}$, every 5 km starting at 5.5 km, without the three point slant path absorption smoothing.

simple case, the column profiles peak at the perturbation layer with a corresponding delta function response and display an oscillatory response in the layers below. A positive perturbation leads to a retrieved tangent layer mixing ratio greater than the actual; absorption calculated through this layer for all lower slant paths will be too large also. The greatest impact is on the tangent layer immediately below, where the algorithm selects a smaller tangent layer mixing ratio to counterbalance the overestimate from above. Amplitude of the first negative oscillation is roughly 2/3 of the peak response and succeeding oscillations rapidly damp out. Error propagation in this case is limited primarily to one layer.

With the three point vertical smoothing $\mathbf{D_y}$ appears as shown in Fig. 4.16. The peak response values are reduced by $\approx$3 as a result of the smoothing and profile widths have almost doubled but are still relatively narrow. At the higher altitudes the profile peak has shifted to one layer higher than the perturbation layer. Error propagation no longer damps out within a single layer but now takes three to four layers.

The onion peel averaging kernel, $\mathbf{A}$, can be calculated with Eq. (2.70) now that $\mathbf{K_x}$ and $\mathbf{D_y}$ are known. The averaging kernel indicates the response of the transfer function to a change in atmospheric parameter $\mathbf{x}^t$ and is indicative of the overall experimental resolution. As before, the columns of $\mathbf{A}$ are indicative of the transfer function response to a delta function change in $\mathbf{x}$. For the case where no vertical slant path smoothing is applied, the column profiles of $\mathbf{A}$ are delta functions with peak values of unity and are not shown. Thus, the effects of a perturbation at a given level are confined to that level and the perturbation is not diminished in amplitude. The effect of vertical slant path smoothing can be seen in Fig. 4.17. Here the three point smoothing has reduced the peak value from unity to 0.45 in the stratosphere and from unity to 0.45-0.55 in the troposphere, resulting from the slightly different smoothing strengths produced by Eqs. (4.27) and (4.28). Profile

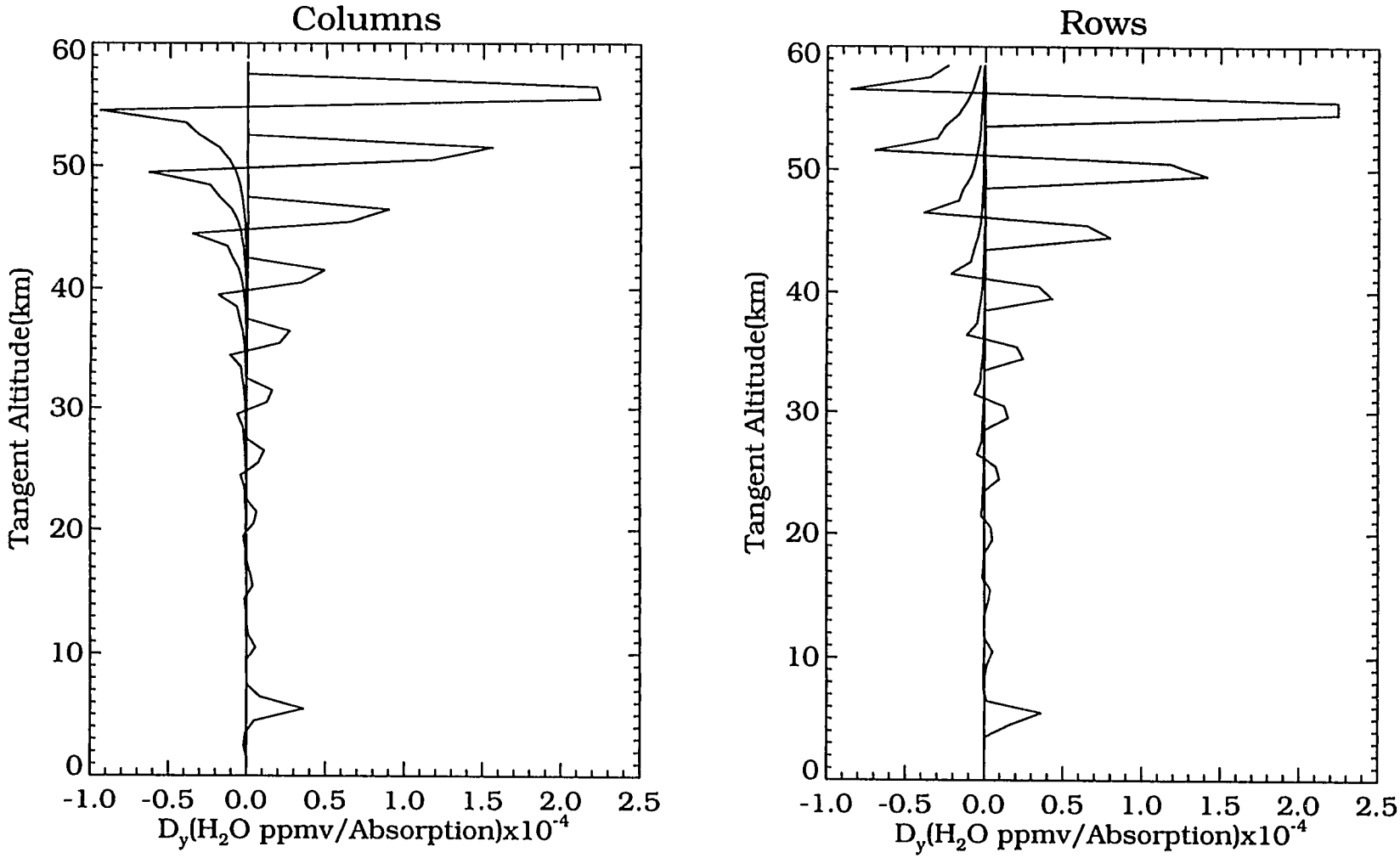


Figure 4.16: Columns and rows of the onion peel inversion contribution function matrix $\mathbf{D_y}$, every 5 km starting at 5.5 km.

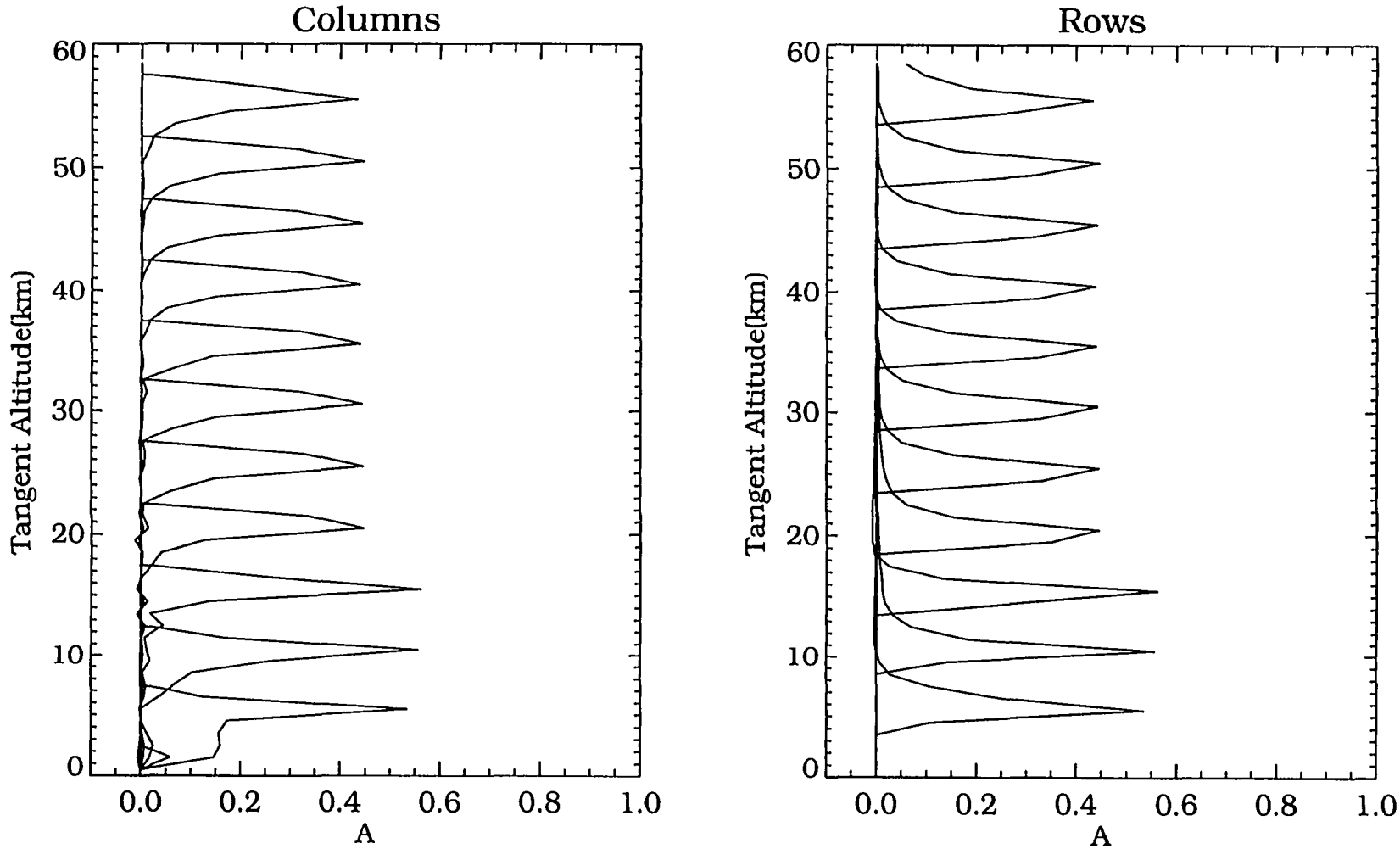


Figure 4.17: Columns and rows of the onion peel inversion averaging kernel matrix **A**, every 5 km starting at 5.5 km.

widths have broadened to give a full width-half maximum (fwhm) of roughly 2 km. Perturbations are reduced in amplitude and redistributed to adjacent levels.

Application of the formal error analysis to a large data set of individual measurements requires significant computational effort due to the number of simulated measurements containing perturbed parameters that must be inverted. A more tractable approach divides the data set into sub-groups of measurements with similar characteristics, and in this situation, monthly zonal means (10° latitude bins) capture most of the spatial and temporal variations of the error terms. Thus, monthly zonal means of the inverted constituent profiles and slant path optical depths are used to calculate $\mathbf{D_y}, \mathbf{K_b}$, and $\mathbf{A_b}$ and systematic error terms in the separation model. Monthly zonal means of the variances in $\mathbf{S}_{\epsilon_b}$ and $\mathbf{S}_{\epsilon_y}$ are based on the same group of measurements. The estimated uncertainties presented here and in the following sections are applicable to a single water vapor profile that is representative for that month and latitude zone.

Figure 4.18 displays the percent uncertainty in the retrieved water vapor mixing ratio due to systematic model parameter errors listed in Table 4.1.

The first panel displays the model error produced by the error term $[\mathcal{T}(\bar{\mathrm{x}}, \hat{\mathrm{b}}, \hat{\mathrm{c}}) - \bar{\mathrm{x}}]$ representing systematic differences between the forward and inverse model. For the SAGE II water vapor channel, the systematic bias follows from the decision to smooth the slant path absorption profile and vanishes in the unsmoothed case (figure not shown). Smoothing produces little bias in the stratosphere. However, in the troposphere where the absorption profile gradient is stronger, a negative bias slightly greater than 20% is found at 2-3 km. This decreases upwards and becomes a positive bias of 10% at 12 km.

The next three panels display errors corresponding to uncertainties in the HITRAN 82 line parameter data set. The band strength uncertainty causes a

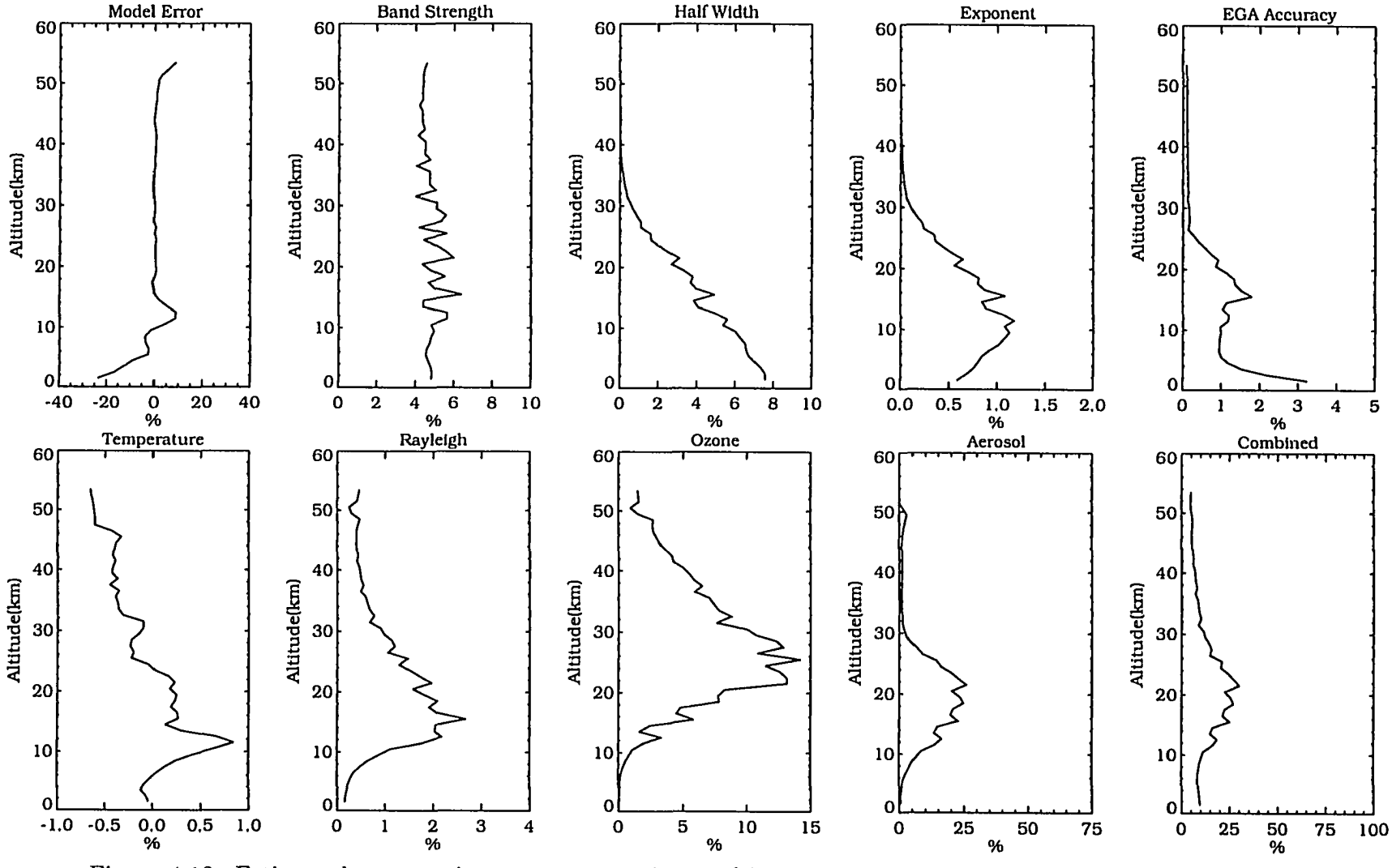


Figure 4.18: Estimated systematic water vapor onion peel inversion errors for the January 1990 40° N monthly zonal mean atmosphere.

one-to-one mixing ratio uncertainty that is nearly constant over altitude. Lorentz halfwidth uncertainties do not cause significant errors until pressure broadening becomes important in the troposphere. At the surface these errors approach 8%. Uncertainty in the exponent controlling the temperature dependence of the Lorentz halfwidth is small and contributes at most 1.2% to the error budget.

In the last top panel, uncertainties introduced by replacing line-by-line radiative transfer calculations with the EGA technique are demonstrated to be small. In the stratosphere, where the absorption curves of growth are nearly linear, the error is quite small. At the surface the retrieval error is $\approx 3\%$.

The first bottom panel displays the retrieval uncertainty due to a constant temperature uncertainty. It is less than 1% at all altitudes. Another small error, 2% or less, results from uncertainties in the Rayleigh effective channel wavelengths and is shown in the next panel.

Systematic error due to uncertainty in the ozone signal removal is shown in the next panel. This error reaches 14% in the lower stratosphere and tapers off above and below. At the peak, this error matches the uncertainty in $\Delta\delta_{2-1}^{Ozone}$ one-to-one. Systematic error in the aerosol removal term is even more important and is shown in the next panel. Here, the maximum error, found between 15 and 20 km, is almost twice as great as the uncertainty in $\Delta\delta_{2-1}^{Aerosol}$.

The last panel shows the total systematic error resulting from combining the non-deterministic terms as the square root of the sum of the squares. The deterministic model error can be removed *post retrieval* if necessary. The vertical variation of the combined error results from the larger errors; the band strength uncertainty present at all altitudes, the ozone error from 15-40 km, the aerosol error 12-28 km, and the Lorentz halfwidth error in the troposphere.

The estimated water vapor random error components corresponding to the error terms listed in Table 4.1 are shown in Fig. 4.19 for a monthly zonal mean atmosphere corresponding to January 1990 at 40° N. The estimated uncertainty is the square root of the diagonal elements of $\mathbf{S}_{\mathbf{x}^r-\mathbf{x}^t}$ expressed relative to $\mathbf{x}$ in per cent for a single water vapor profile. Random error components are calculated from quantities available directly from the SAGE II experiment, except for temperature.

The first panel of Fig. 4.19 displays the measurement uncertainty due to the random measurement error of channel 2. While the measurement error ϵ_y is approximately constant throughout the stratosphere, the water vapor slant path absorption decreases upwards and this causes the estimated water vapor uncertainty to increase upwards. Between 20 and 12 km ϵ_y increases slightly faster than the slant path absorption to cause an increase in the estimated uncertainty while the reverse occurs between 12 km and the surface. Thus, a peak uncertainty of 10% is found at 12 km that decreases immediately above and below and then increases to 20% at 46 km and 35% at 50 km.

The estimated uncertainty due to random measurement errors in the ozone channel is shown in the second panel. They are insignificant in the troposphere and at most 1% in the stratosphere. The next panel shows that tangent altitude errors are 4-6% in the stratosphere and 8-12% in the troposphere. Temperature uncertainties, shown in the next panel, are approximately 1% from the surface up to 30 km and then increase to 3% at 45 km. The aerosol uncertainty is less than 8% over all altitudes up to 50 km as shown in the next panel.

Finally, in the last panel, the random estimated uncertainties from all five error terms are combined as the square root of the sum of the squares to yield the total random estimated error. The vertical variation of the total error largely follows that of the measurement error, since it is the predominant contributor. The total random

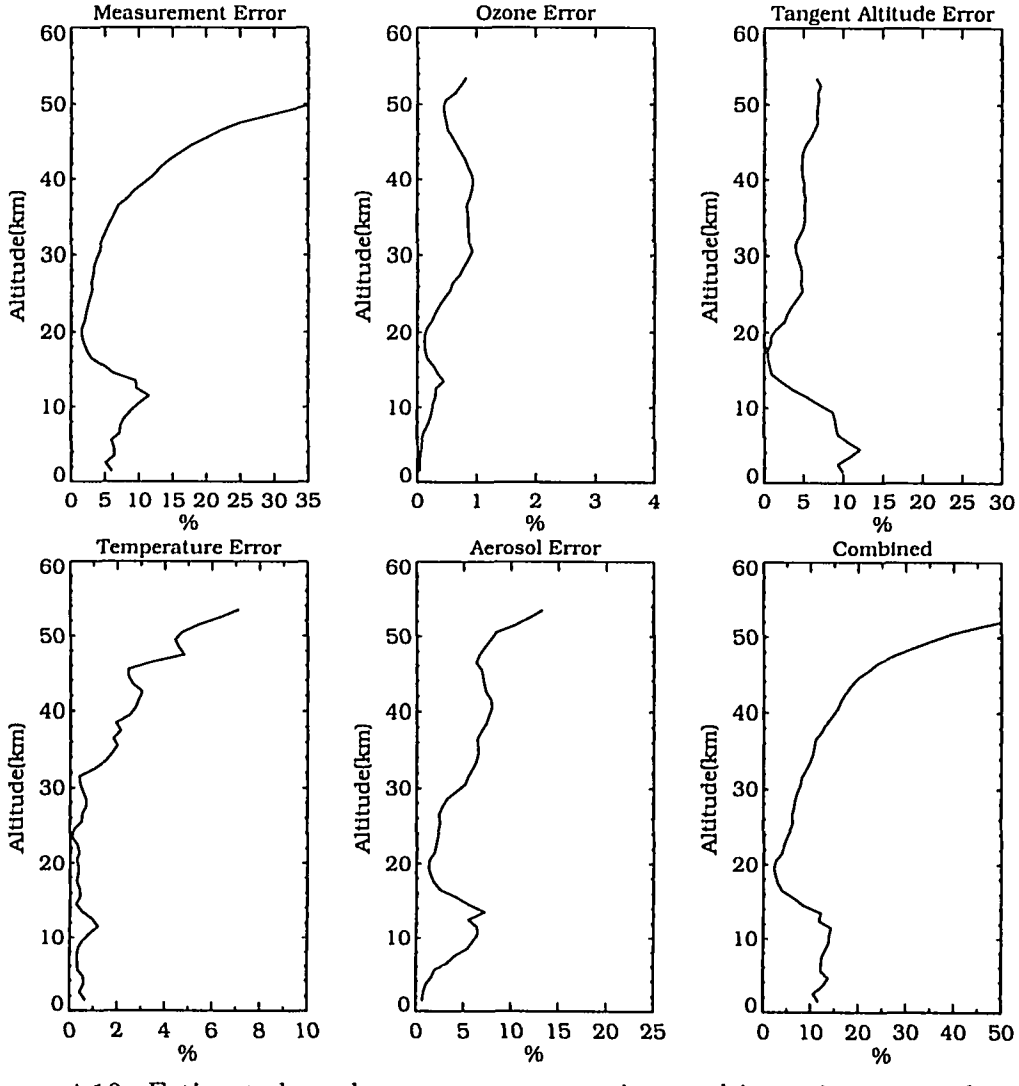


Figure 4.19: Estimated random water vapor onion peel inversion errors for the January 1990 40° N monthly zonal mean atmosphere.

error is at most 15% in the troposphere, decreases to less than 5% at 19 km and increases to 20% at 44 km.

The total systematic and random estimated uncertainties may also be combined, again using the square root of the sum of the squares, to give the total experimental water vapor uncertainty. Figure 4.20 displays the total experimental error for all latitude zones containing SAGE II measurements for January 1990. Also shown are the total systematic and random errors. The relative importance of each error is quite similar for all of the latitude zones. The total random error above 35 km and below 12 km is greater than the systematic error while the reverse is true at intermediate altitudes. The total random error increases slightly in the stratospheric equatorial region due to an increase in the measurement error. Variations of the total systematic error in the 12 to 35 km region correlates strongly with variations in the aerosol level due to the fact that aerosol modeling uncertainty is the dominant systematic error term. Thus, the maximum total systematic uncertainty increases and moves upwards in the equatorial region relative to high latitudes.

In the troposphere, the total experimental uncertainty ranges from 15-20% in most latitude zones. In the lower stratosphere, peak values of 25-35% are found that decrease upwards to minimums of 15-20% between 30 and 40 km. The total uncertainty again increases upwards and reaches 30 to 35% at 50 km.

Figures displaying the total systematic, random, and experimental estimated uncertainty for the onion peel retrieval in 1990 for other months are presented in Appendix D.

4.4.2 Operational Inversion

The contribution function $\mathbf{D_y}$, defined by Eq. (2.62), is shown in Fig. 4.21 at selected altitudes for the operational inversion. It is comparable in form to the

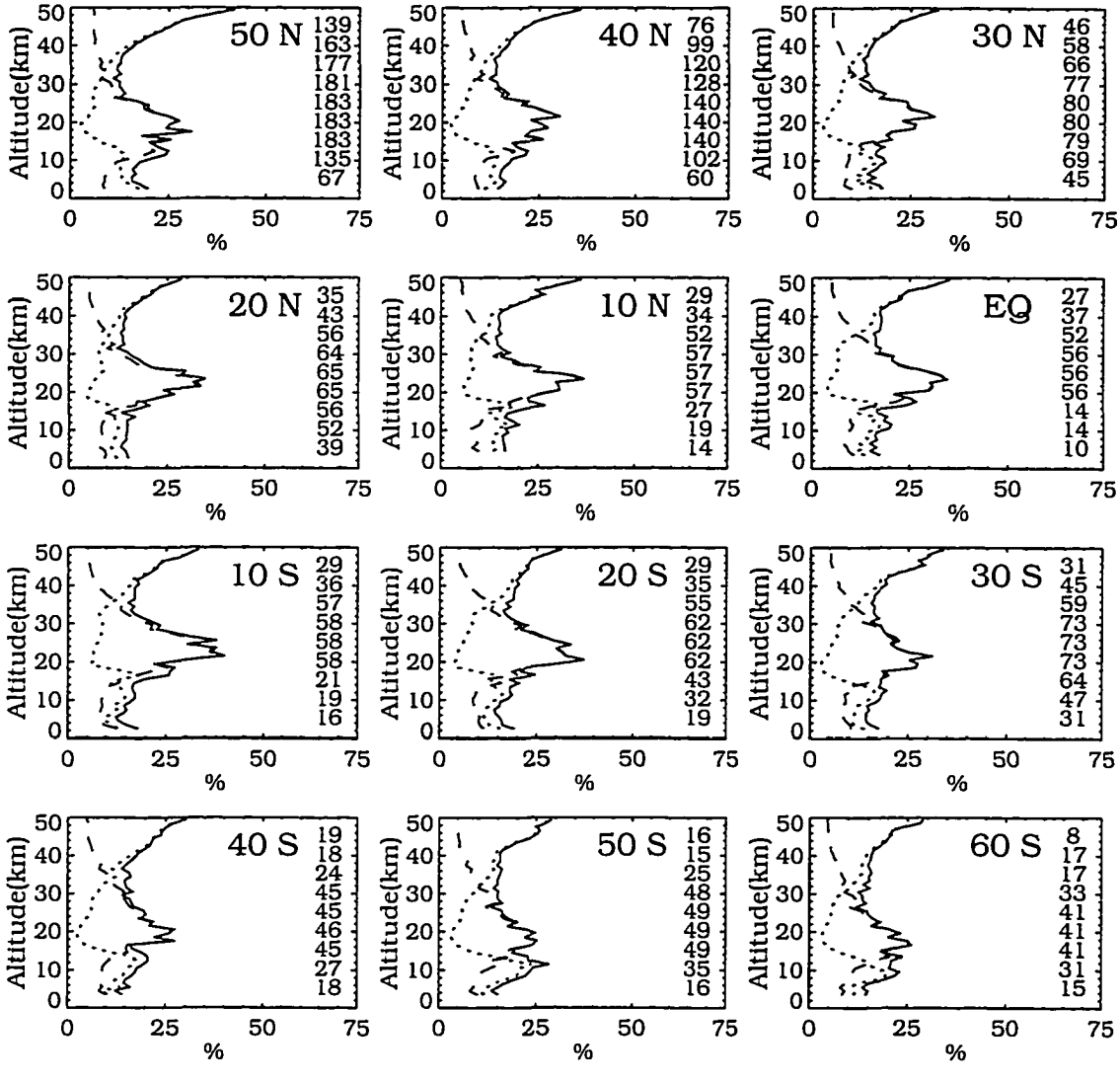


Figure 4.20: Total systematic (dashed), random (dotted), and experimental (solid) estimated uncertainty corresponding to the onion peel retrieval, January 1990. The number of points included in the zonal statistics are indicated on the right by altitude level.

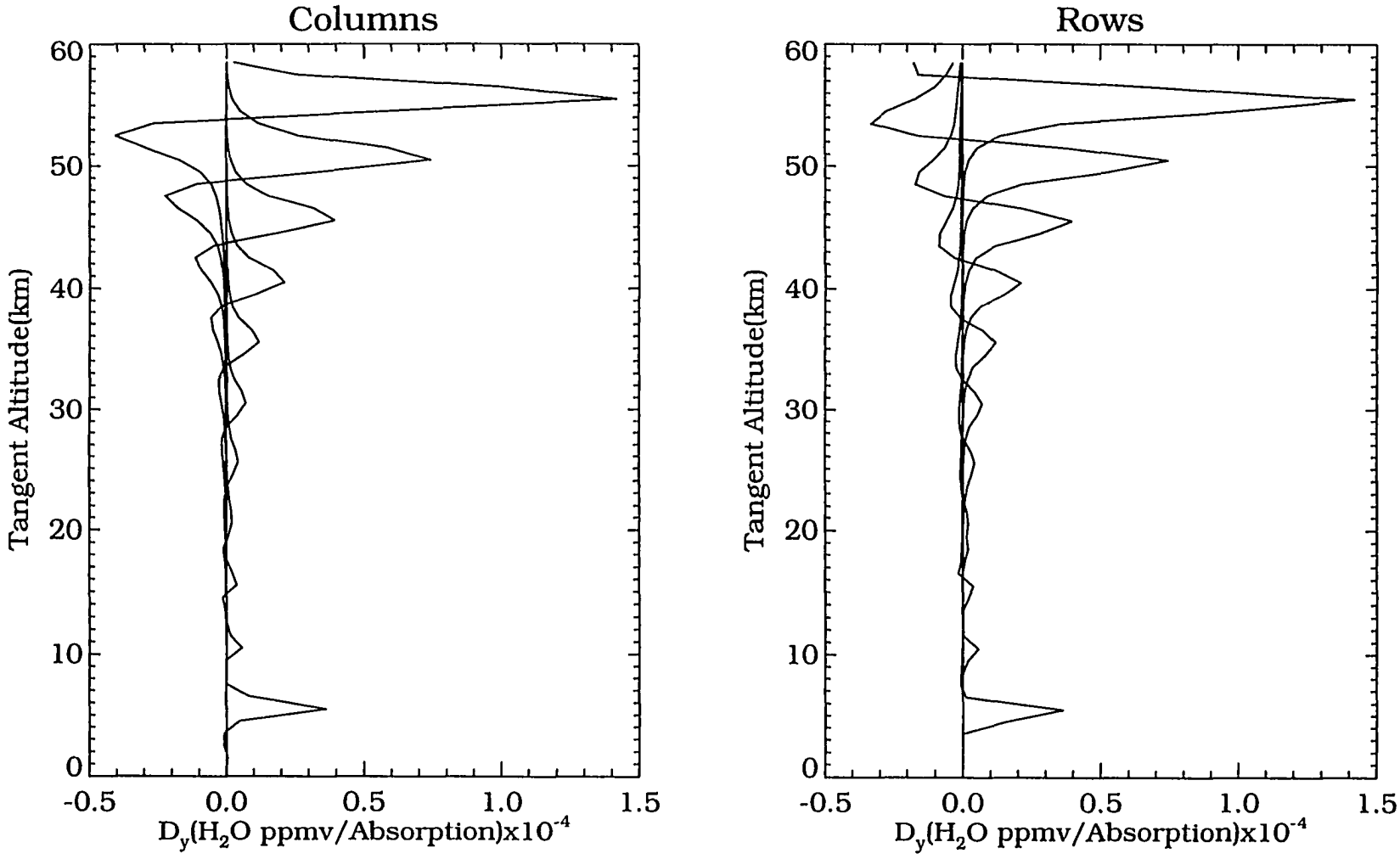


Figure 4.21: Columns and rows of the operational inversion contribution function matrix $\mathbf{D_y}$, every 5 km starting at 5.5 km.

onion peel $\mathbf{D_y}$ shown in Fig. 4.17. The three point stratospheric smoothing of U has broadened the profile functions and reduced their amplitude. The same effect appears in the averaging kernel $\mathbf{A}$ shown in Fig. 4.22. Below 18.5 km the profile functions are identical to the onion peel, while above 18.5 km the peak values are reduced to 0.2-0.25 and the function widths are 3.5 km(fwhm). If the three point slant path absorption smoothing is removed, the stratospheric functions narrow and increase in amplitude by 20% and the functions below 18.5 km become delta functions. Of the two types of stratospheric smoothing, smoothing U has the strongest effect.

Estimated systematic errors in the retrieved water vapor mixing ratio using the operational inversion are shown in Fig. 4.23 for the model parameters of Table 4.1 calculated with the monthly zonal mean atmosphere corresponding to January 1990 at 40° N. The systematic errors are nearly identical to those of the onion peel inversion shown in Fig. 4.18 except that the smoothing of U at altitudes above 18 km is clearly evident.

The estimated water vapor random error components corresponding to the error terms listed in Table 4.1 are shown in Fig. 4.24 for January 1990 at 40° N and are applicable to a single profile. Below 18 km, the estimated random errors are again identical to the onion peel while above 18 km the effects of smoothing U becomes apparent. Each of the error terms, except for temperature, show a reduction in estimated error somewhat less than $\sqrt{3}$, what one would expect from the three point smoothing of U. Random temperature error is not reduced in the stratosphere due to temperature $\mathbf{K_b}$ characteristics. Figure 4.5 shows that the standard pressure levels defining the temperature profile produce slowly varying functions in the stratosphere that are broader than the contribution function $\mathbf{D_y}$, of both inversion methods. Therefore, even though the operational contribution function is broader than the

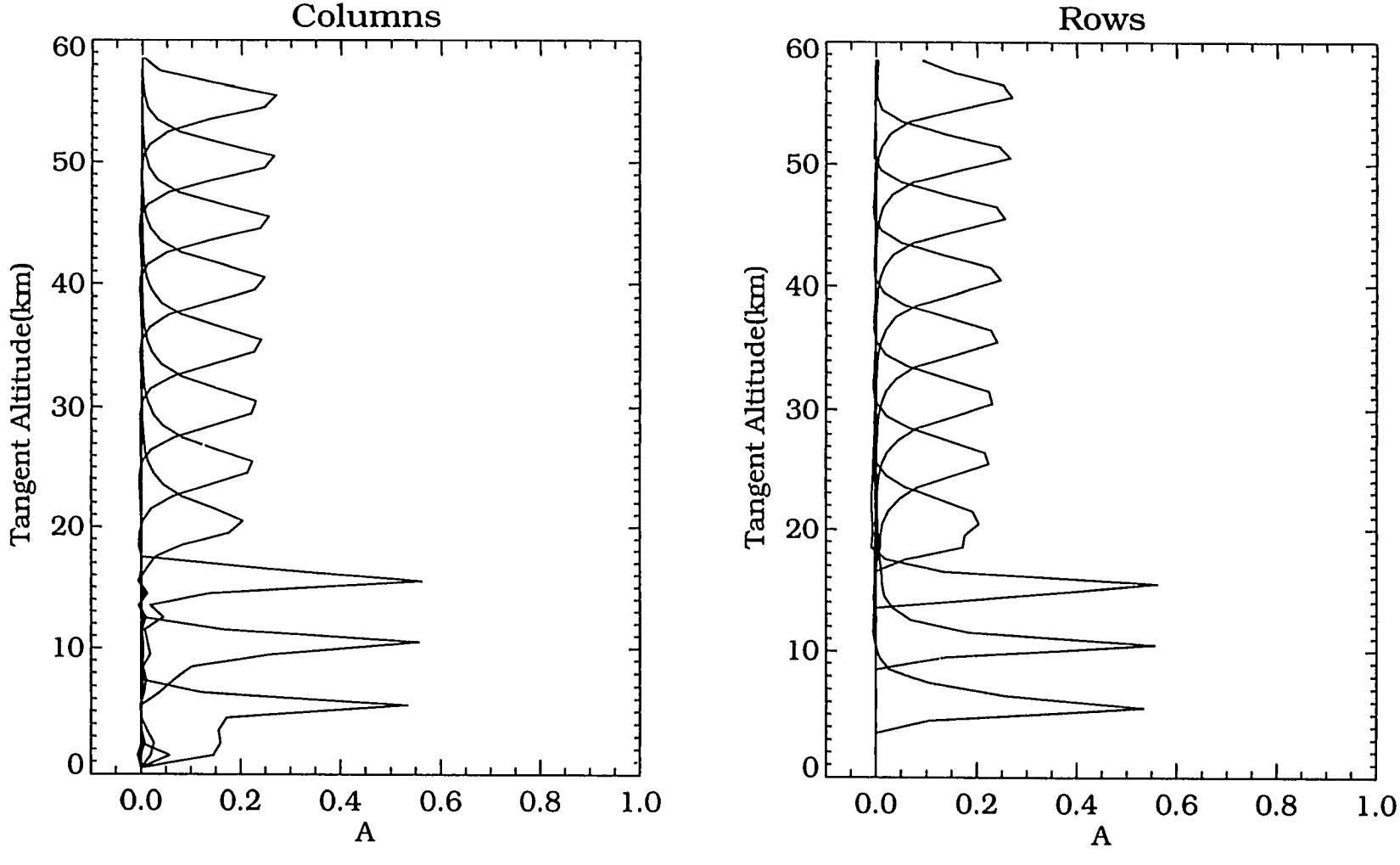


Figure 4.22: Columns and rows of the operational inversion averaging kernel matrix **A**, every 5 km starting at 5.5 km.

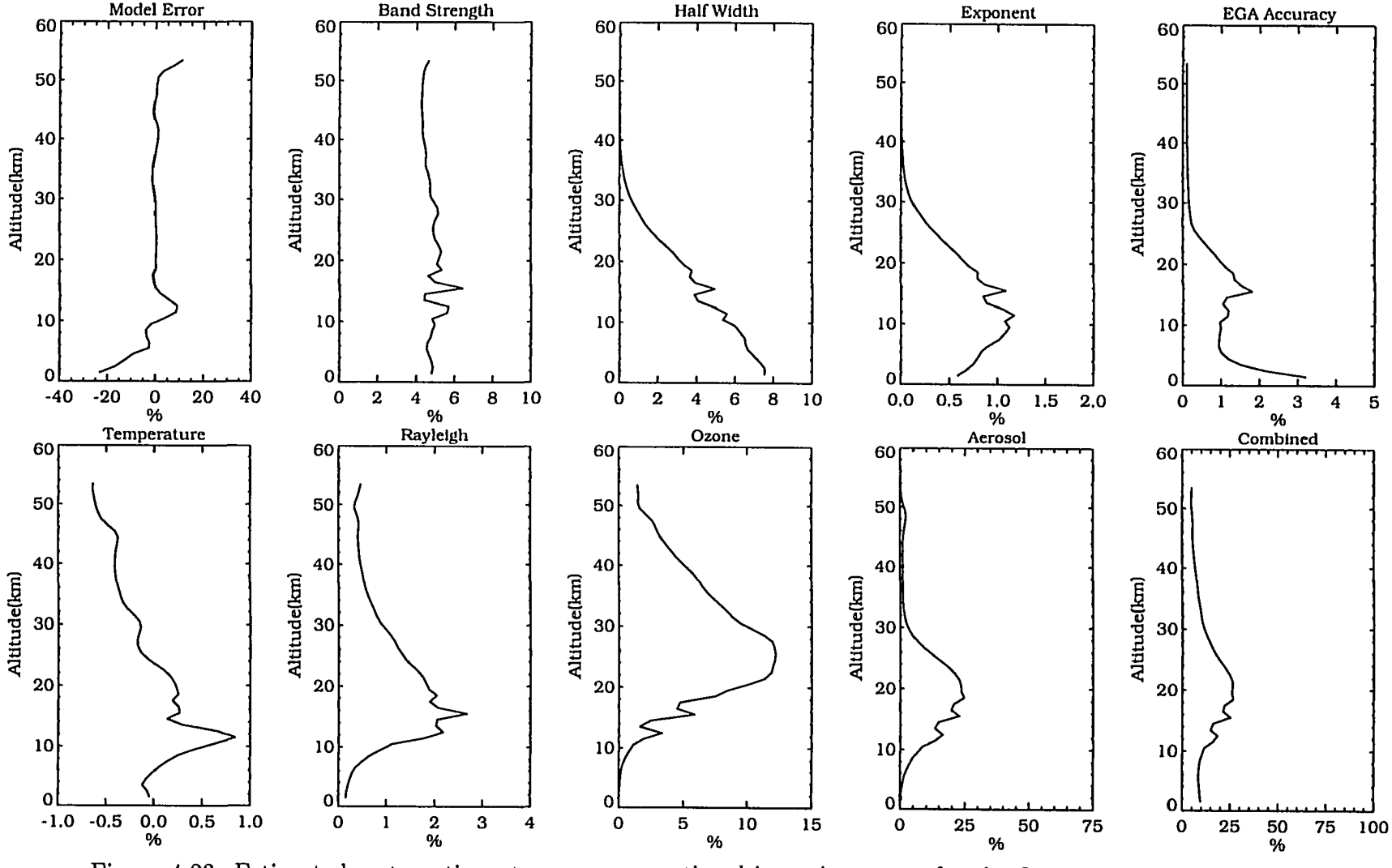


Figure 4.23: Estimated systematic water vapor operational inversion errors for the January 1990 40° N monthly zonal mean atmosphere.

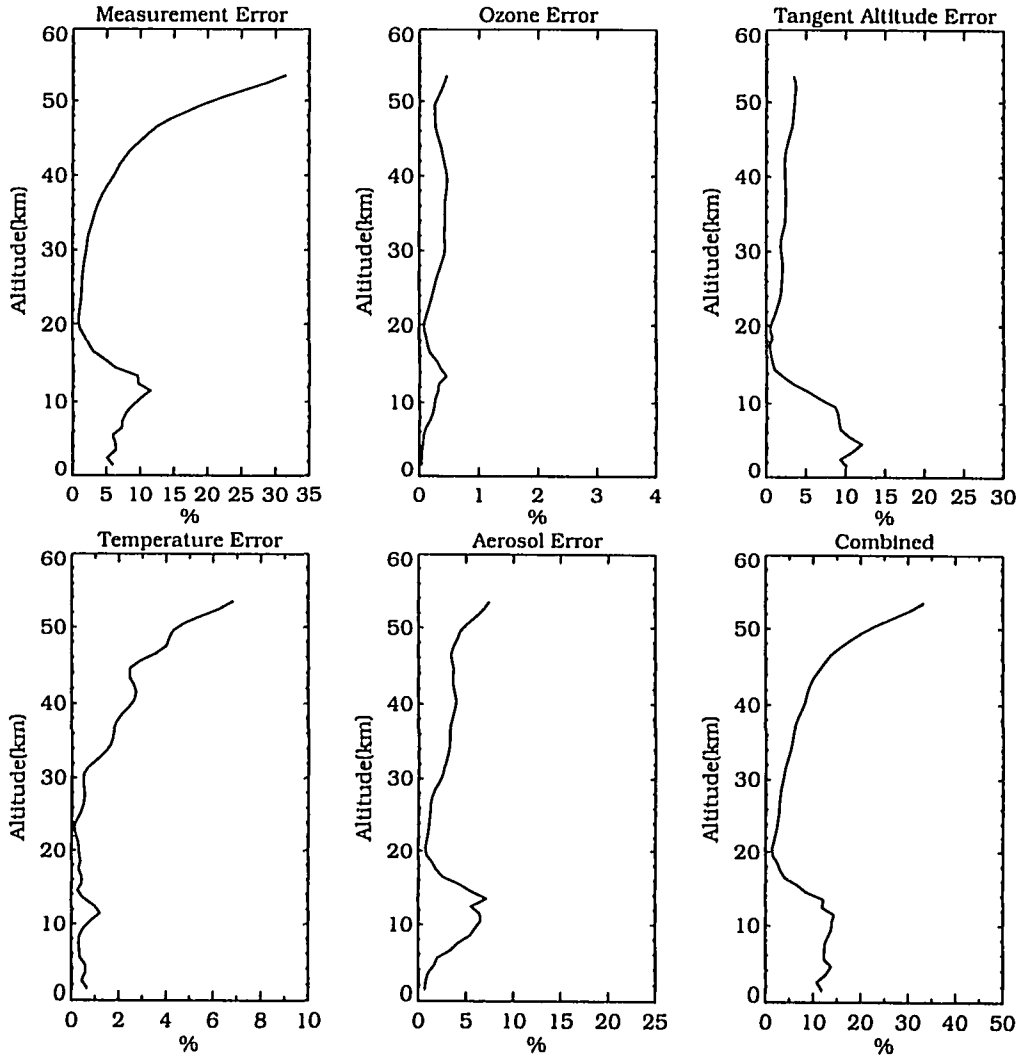


Figure 4.24: Estimated random water vapor operational inversion errors for the January 1990 40° N monthly zonal mean atmosphere.

onion peel contribution function, it is not broader than $\mathbf{K}_b$ and does not appreciably reduce the random temperature error.

The vertical variation of the total random error is similar to that observed with the onion peel method. Below 18 km the error profiles are identical, a minimum is again found near 20 km, and the three point smoothing reduces the upwards increase to 20% at 50 km.

The total experimental error for all latitude zones containing SAGE II measurements in January 1990 is shown in Fig. 4.25. The greatest change, relative to the onion peel inversion results, appears above 35 km due to the reduction of the total random error at these levels and at 50 km the total experimental error drops by 10-15%.

Figures displaying the total systematic, random, and experimental estimated uncertainty for the operational retrieval in 1990 for other months are presented in Appendix D.

Reconciliation with previous work

Application of the formal error analysis to the SAGE II water vapor observations required that each model parameter error term be properly formulated and assigned a realistic level of uncertainty. This section discusses the differences in approach, between the formal error analysis and the error analysis described in Chu et al.(1993), to formulating the error terms and assigning error levels.

Systematic error terms are treated in both analyses with sensitivity studies, the formal error analysis perturbs the forward model while in Chu et al.(1993) it is the inverse model that is perturbed. Since systematic error estimates are not produced for the inverted profiles as part of the standard data processing, we will refer to the studies and assumptions made in Chu et al.(1993) for comparison purposes.

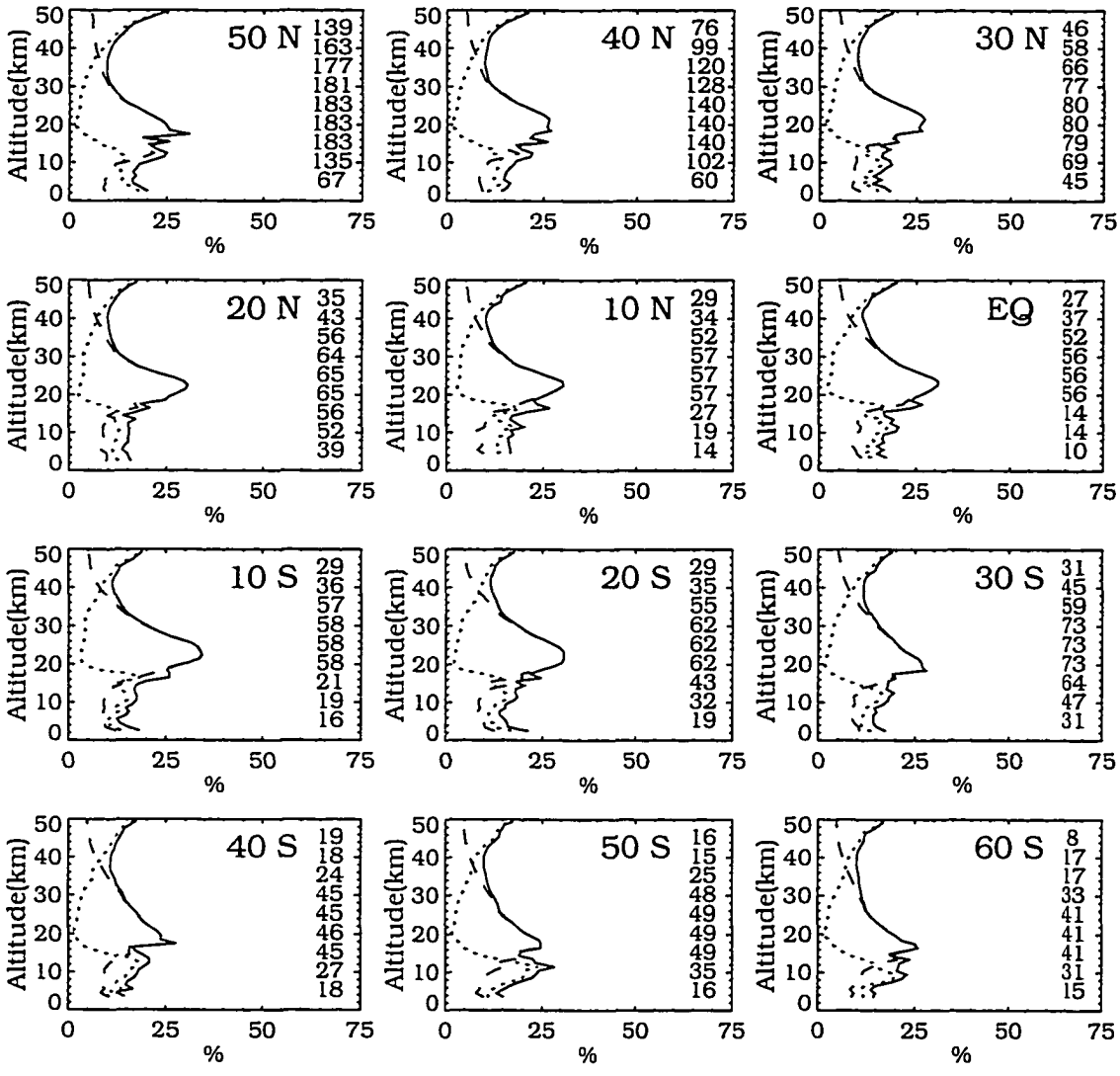


Figure 4.25: Total systematic (dashed), random (dotted), and experimental (solid) estimated uncertainty corresponding to the operational retrieval, January 1990. The number of points included in the zonal statistics are indicated on the right by altitude level.

Systematic water vapor uncertainties resulting from uncertainties in the molecular line parameters, needed to calculate the water vapor k_λ, are nearly identical. In Chu et al.(1993), the 10% constant bias due to a corresponding 10% band strength uncertainty is assumed to be correctable to zero with the help of correlative measurements in the stratosphere. Instead, we assume that the correlative measurements can reduce this uncertainty to only 5% in both the stratosphere and troposphere. The impact of 10% uncertainties in the Lorentz halfwidth and halfwidth temperature dependence exponent are identical in both methods. We further consider the importance of replacing the line-by-line radiative transfer calculations with the Emissivity Growth Approximation and find it small relative to other error terms.

Two of the three separation model error terms are unchanged. Systematic ozone and aerosol errors are formulated identically with nearly the same level of uncertainty and the resulting water vapor uncertainty is also identical. In this work, Rayleigh systematic error arises from uncertainty in the effective channel wavelengths that must be specified to calculate the Rayleigh cross sections. Water vapor uncertainty due to this error is found to be small, as is the systematic uncertainty due to temperature which is treated explicitly in the formal error analysis.

The total systematic error calculated by the two methods is very close since the terms accounting for most of the uncertainty are not that different. The total systematic error in this work is 5% higher because of the assumptions concerning the band strength. The formal error analysis does provide a more detailed look at the zonal error variation and does include the model error term, which is significant in the troposphere.

When the experiment data is operationally inverted, an estimate of the combined random error (measurement, tangent altitude, temperature, ozone, and aerosol) is made for each profile from application of the error propagation equation to an onion

peel inversion. Adjustments to the combined estimate based on the effectiveness of both forms of vertical smoothing are then made. Random temperature and tangent altitude uncertainties are formulated as random uncertainties appearing only in the Rayleigh optical depth. In the formal error analysis, random errors in temperature and tangent altitude are treated explicitly, leaving only systematic errors in the Rayleigh optical depths. Also, the effect of vertical smoothing is contained in $\mathbf{D_y}$. The formal error analysis shows that the column functions associated with the temperature $\mathbf{K_b}$ are broad in the stratosphere relative to $\mathbf{D_y}$. Thus, the random temperature errors are not reduced by smoothing as much as the other error components.

Ozone random errors are treated identically in both methods, while the aerosol random error is formulated in this work as a combination of channels 1 and 4 errors. This increases the aerosol error slightly in the stratosphere. The formal error analysis includes the correction described in Appendix B that is applied to the error terms based on SAGE II channel errors. It is not applied to the temperature or tangent altitude terms. This work also includes the "experiment" error covariances of Appendix C but the effect of this on the total random error is typically to increase it in the troposphere by 10% with some levels occasionally increased 20%.

Figure 4.26 compares the the total random estimated water vapor uncertainty between the two methods. Above 30 km both methods agree nicely while between 18 and 30 km the error correction method of Appendix B significantly reduces the estimated errors calculated in this work. From 15 to 20 km the operational analysis tends to be higher than this work. Inspection of individual error terms finds the Rayleigh term (proxy for temperature and tangent altitude errors) greater than the measurement error in the operational analysis while the reverse is true for these two terms in this work. The formal error analysis estimates higher error levels between the surface and 10 km, this is due to the tangent altitude error term.

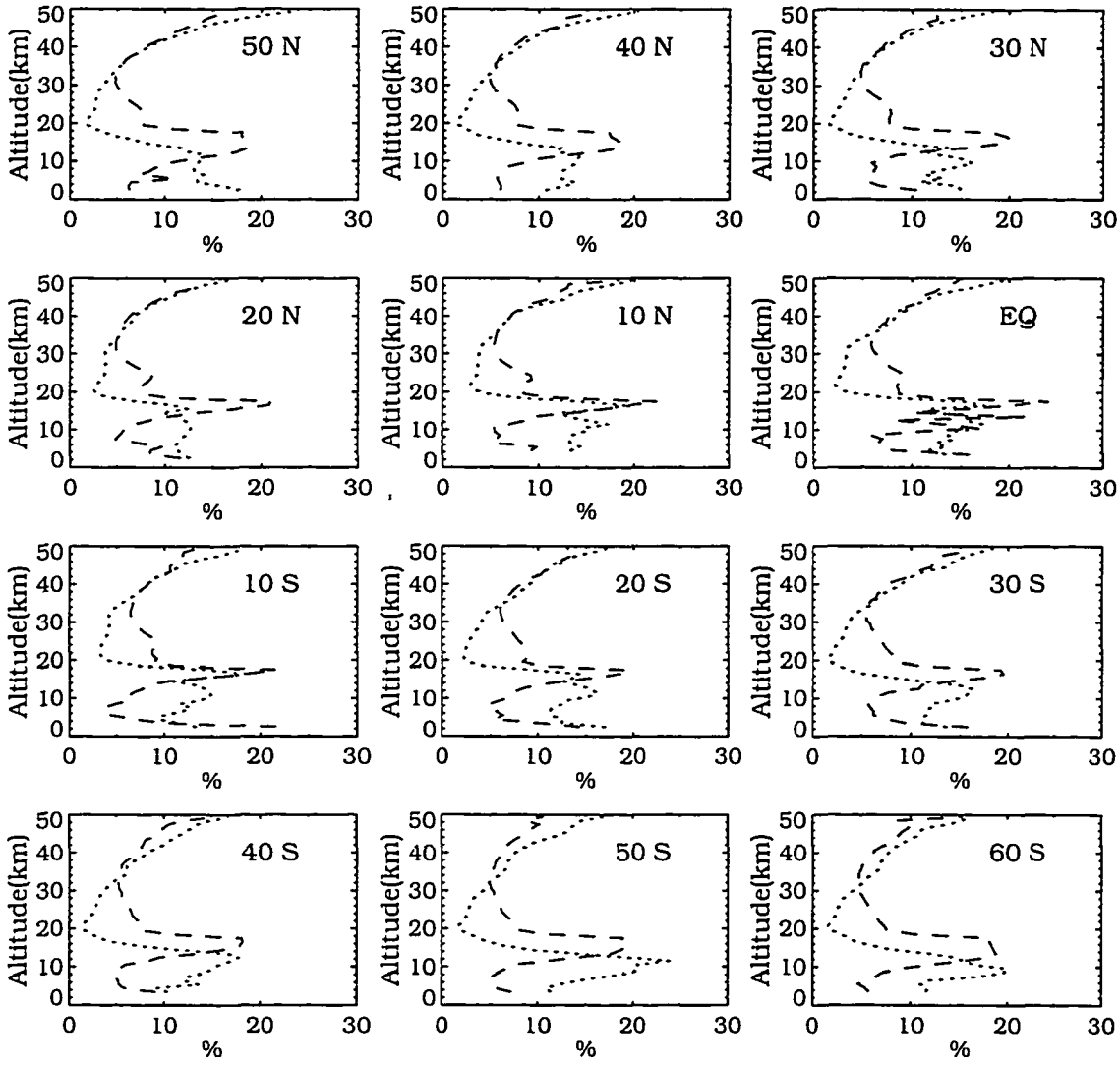


Figure 4.26: Comparison of the total random estimated water vapor uncertainty for the operational retrieval, January 1990, between the formal error analysis of this work (dotted lines) and that of Chu et al.(1993) (dashed lines).

4.4.3 Mill-Drayson(smoothed) Inversion

The contribution function for the smoothed Mill-Drayson inversion is nearly identical to the operational inversion, Fig. 4.21, and is not shown here. It follows that the averaging kernel, shown in Fig. 4.27, will also be identical.

Since the contribution function is nearly identical, the total systematic and random errors are also identical so only the total experimental error is shown in Fig. 4.28. The discussion in the previous section concerning the error characteristics of the operational inversion apply here.

4.4.4 Mill-Drayson Inversion

The contribution function for the Mill-Drayson inversion is nearly identical to the onion peel inversion, Fig. 4.16, and is not shown here. It follows that the averaging kernel, shown in Fig. 4.29, will also be identical.

Since the contribution function is nearly identical, the systematic and random errors predicted by the formal error analysis are identical to those shown in Fig. 4.18 and Fig. 4.19 for those error terms that are common to both algorithms. The discussion in the previous section concerning the error characteristics of the onion peel inversion apply here. To realize the reduction in random errors that is possible with the implicit smoothing of the Mill-Drayson algorithm we must select $\Delta_{cl} = |\epsilon_s|$ and two new terms must be included in the formal error analysis. A new systematic error arises, termed the solution error, that is caused by systematic biases in ϵ_{tc} or equivalently ϵ_{cl} since ϵ_s has no bias. Terms that reduce the random errors are referred to as implicit smoothing.

Figure 4.30 displays the uncertainty in the retrieved water vapor mixing ratio profile due to systematic model parameters. Note that the halfwidth error and exponent error are plotted together (exponent error is the smaller of the two) in the

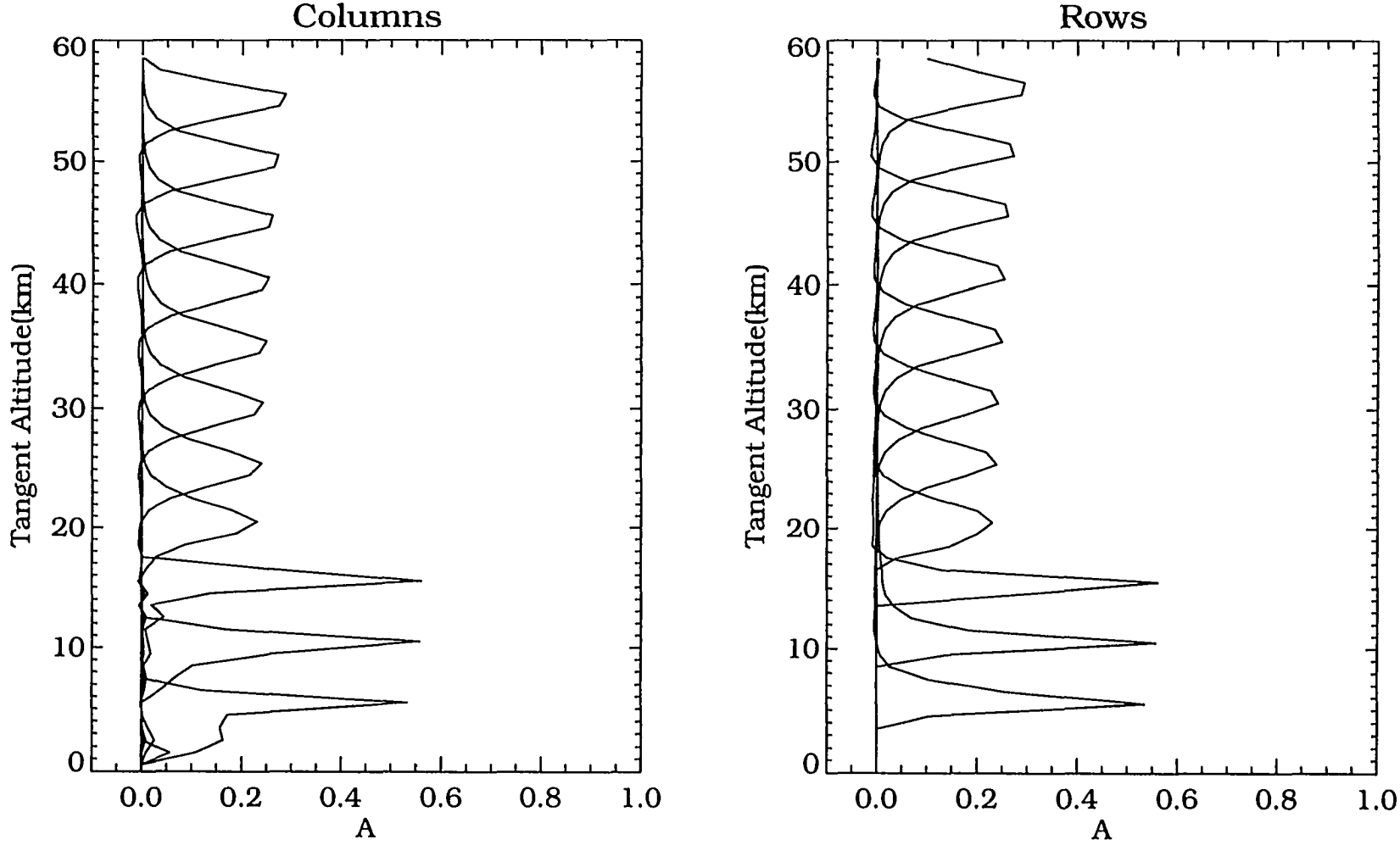


Figure 4.27: Columns and rows of the smoothed Mill-Drayson inversion averaging kernel matrix **A**, every 5 km starting at 5.5 km.

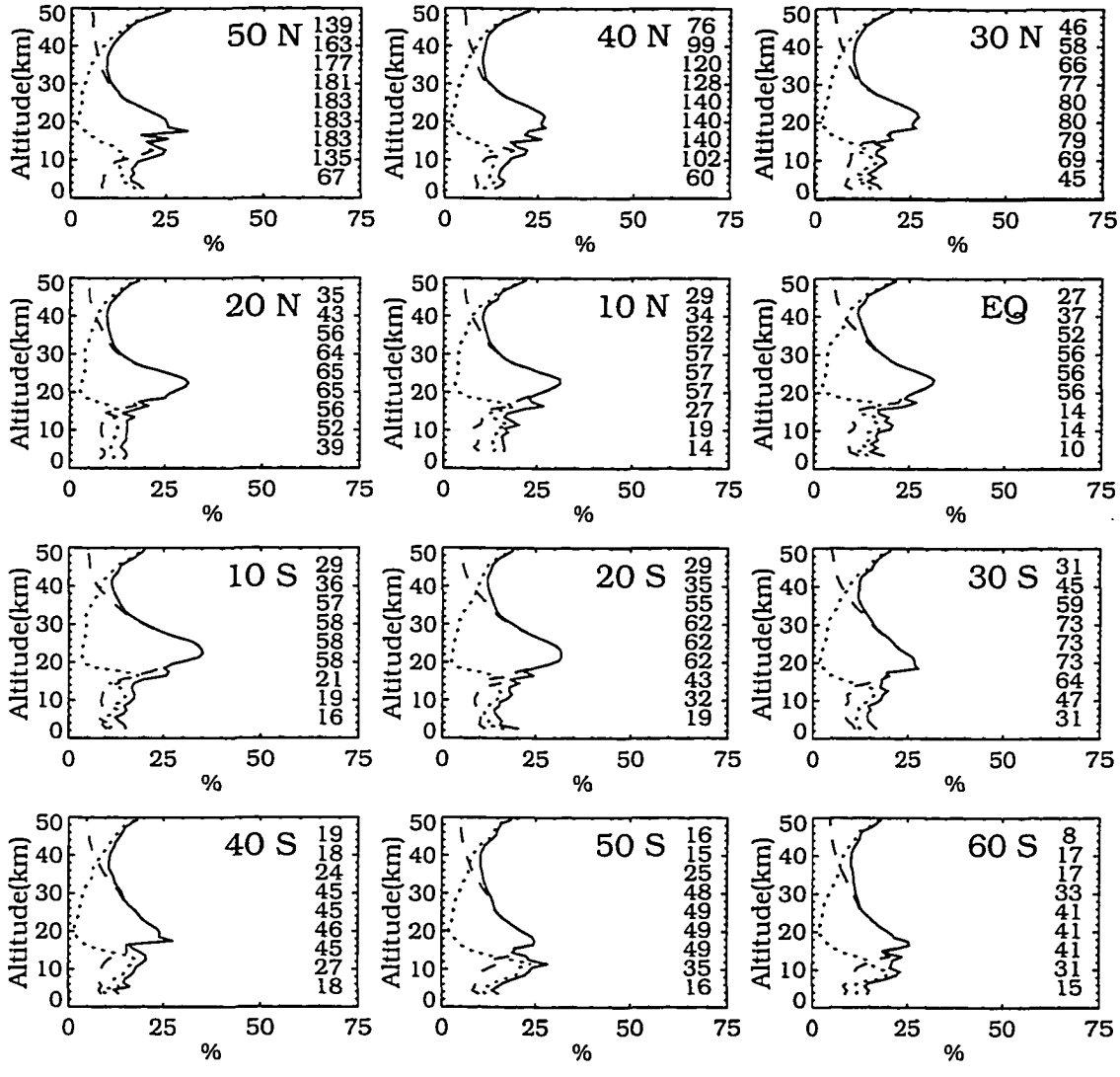


Figure 4.28: Total systematic (dashed), random (dotted), and experimental (solid) estimated uncertainty corresponding to the smoothed Mill-Drayson retrieval, January 1990. The number of points included in the zonal statistics are indicated on the right by altitude level.

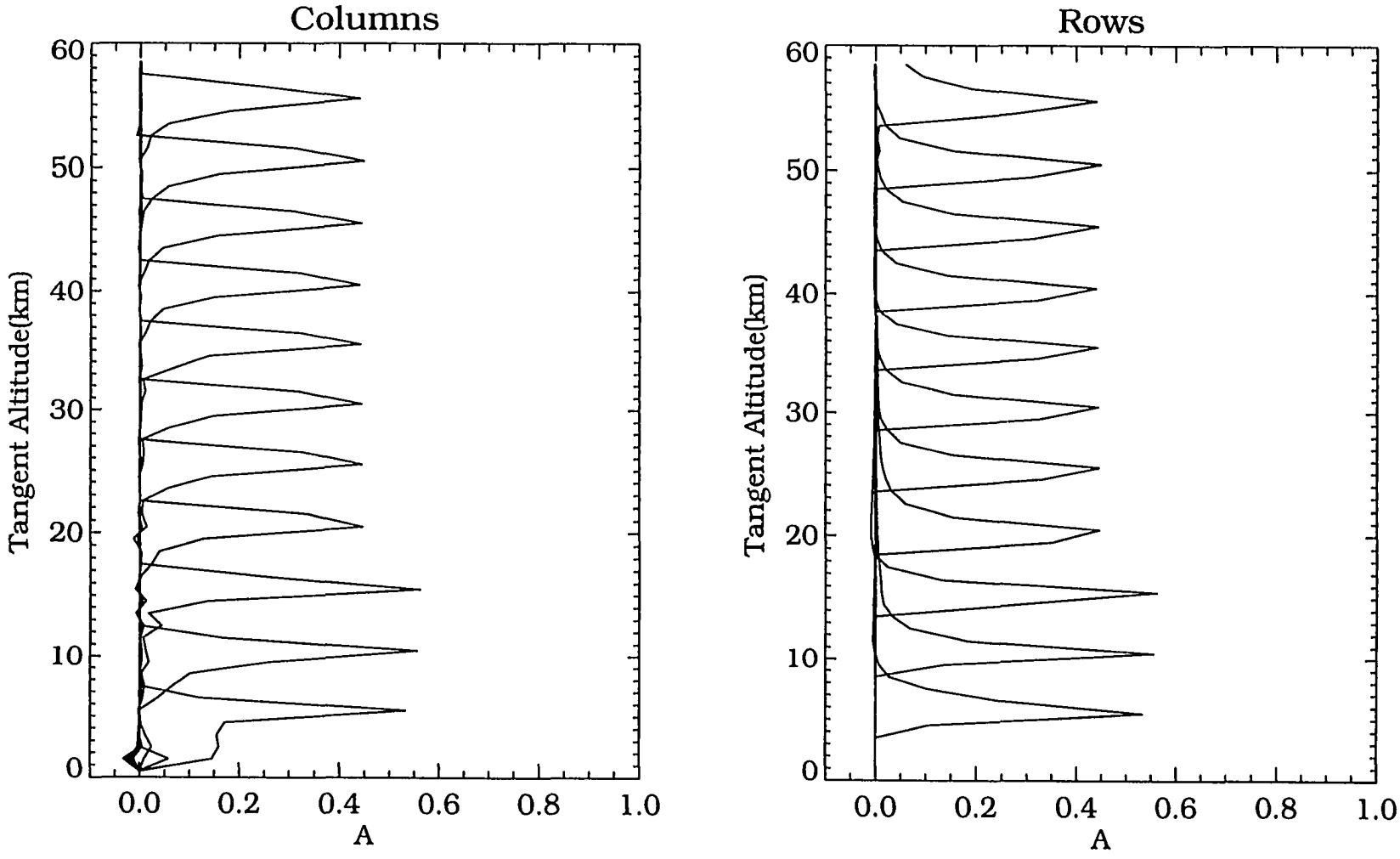


Figure 4.29: Columns and rows of the Mill-Drayson inversion averaging kernel matrix **A**, every 5 km starting at 5.5 km.

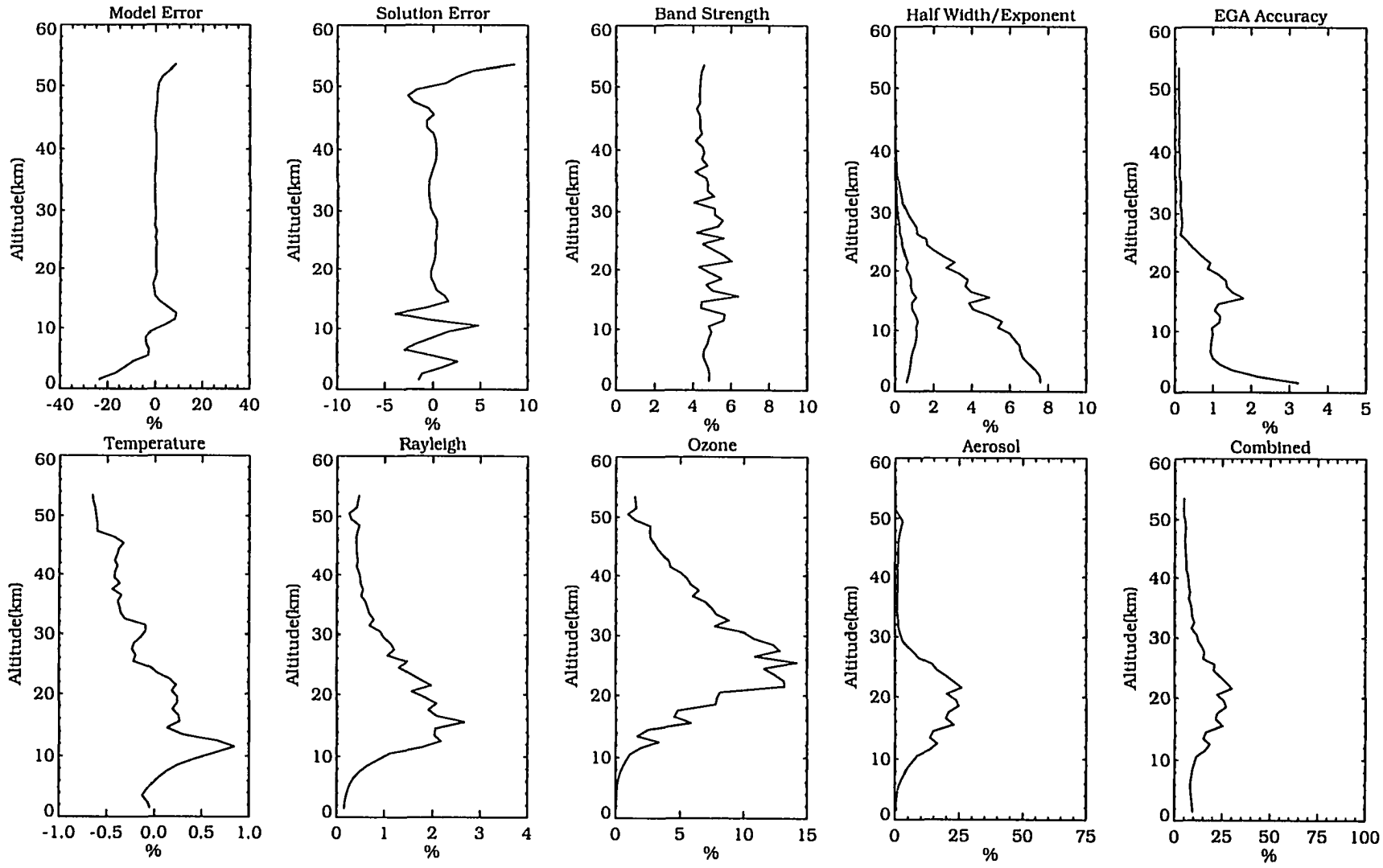


Figure 4.30: Estimated systematic water vapor Mill-Drayson inversion errors for the January 1990 40° N monthly zonal mean atmosphere.

same panel. The solution error is 4-5% at most in the troposphere and 3-4% in the upper stratosphere. The error is larger than that observed in Fig. 4.12 because the first guess profile shape differs from $\mathbf{x}^r$. The solution error in the troposphere could be reduced by decreasing $\Delta_{\mathbf{cl}}$. The effectiveness of the implicit smoothing would be reduced but this would not represent a great loss since the resulting increase in random error would still be small relative to the other error terms.

As noted in Sec. 4.3.3, additional error terms are needed to fully characterize $\mathbf{S}_{\epsilon_{\mathbf{tc}}}$ in the stratosphere. For discussion purposes, we assumed that only random measurement noise was significant in Sec. 4.3.3. Here, we also include model parameter random errors introduced by the separation model, $\mathbf{S}_{\epsilon'_{\mathbf{b}}}$, and ignoring other terms in $\mathbf{S}_{\mathbf{y}^{mi}}$ unrelated to this development leads to

$$\mathbf{S}_{\mathbf{y}^{mi}} \propto \mathbf{S}_{\epsilon_{\mathbf{mi}}} = \mathbf{S}_{\epsilon'_{\mathbf{b}}} + \mathbf{S}_{\epsilon_{\mathbf{s}}} \tag{4.54}$$

where

$$\mathbf{S}_{\epsilon'_{\mathbf{b}}} = (\frac{\partial S}{\partial \mathbf{y}^m})[\mathbf{K}_{\mathbf{b}}\mathbf{S}_{\epsilon_{\mathbf{b}}}\mathbf{K}_{\mathbf{b}}^T](\frac{\partial S}{\partial \mathbf{y}^m})^T \tag{4.55}$$

Then

$$\mathbf{S}_{\epsilon_{\mathbf{tc}}} = \mathbf{S}_{\epsilon_{\mathbf{mi}}} + \mathbf{S}_{\epsilon_{\mathbf{cl}}} + \mathbf{S}_{\epsilon_{\mathbf{mi}},\epsilon_{\mathbf{cl}}} + \mathbf{S}_{\epsilon_{\mathbf{cl}},\epsilon_{\mathbf{mi}}} \tag{4.56}$$

and analogous to Eq. (4.46), one obtains (after simplifying)

$$\mathbf{S}_{\mathbf{x}^r-\mathbf{x}^t} \propto \frac{\partial IA}{\partial \mathbf{y}^{mi}}[\mathbf{S}_{\epsilon_{\mathbf{mi}}} + \mathbf{S}_{\epsilon_{\mathbf{cl}}} + \mathbf{S}_{\epsilon_{\mathbf{mi}},\epsilon_{\mathbf{cl}}} + \mathbf{S}_{\epsilon_{\mathbf{cl}},\epsilon_{\mathbf{mi}}}]\frac{\partial IA}{\partial \mathbf{y}^{mi}}^T \tag{4.57}$$

$$\propto \mathbf{A}_{\mathbf{b}}\mathbf{S}_{\epsilon_{\mathbf{b}}}\mathbf{A}_{\mathbf{b}}^T + \mathbf{D}_{\mathbf{y}}\mathbf{S}_{\epsilon_{\mathbf{y}}}\mathbf{D}_{\mathbf{y}}^T + \frac{\partial IA}{\partial \mathbf{y}^{mi}}[\mathbf{S}_{\epsilon_{\mathbf{cl}}} + \mathbf{S}_{\epsilon_{\mathbf{mi}},\epsilon_{\mathbf{cl}}} + \mathbf{S}_{\epsilon_{\mathbf{cl}},\epsilon_{\mathbf{mi}}}]\frac{\partial IA}{\partial \mathbf{y}^{mi}}^T \tag{4.58}$$

To evaluate the additional terms appearing on the right hand side containing $\mathbf{S}_{\epsilon_{\mathbf{cl}}}, \mathbf{S}_{\epsilon_{\mathbf{mi}},\epsilon_{\mathbf{cl}}}$, without performing a detailed statistical study of the interaction of each of the model parameters and random measurement noise with $\epsilon_{\mathbf{cl}}$ for each month and zone, requires a number of simplifying assumptions. We start by recognizing that $\mathbf{S}_{\epsilon_{\mathbf{s}}}$ in the stratosphere is a diagonal matrix with zero covariances. The matrices

$\mathbf{K_b}, \mathbf{S}_{\epsilon_b}, \frac{\partial S}{\partial \mathbf{y}^m}$ are also diagonal, at least for the ozone and aerosol errors. The tangent altitude and temperature errors were calculated from a direct calculation of $\mathbf{A_b}$, solving $\mathbf{D_y} = \mathbf{A_b}\mathbf{K_b}^{-1}$ would likely yield a slightly different contribution function and this translates into a different $\frac{\partial S}{\partial \mathbf{y}^m}$. It is likely that $\frac{\partial S}{\partial \mathbf{y}^m}$ is still diagonally dominant so we will assume that $\mathbf{S}_{\epsilon_b}$ for these two parameters are also diagonal. Thus, the error patterns or correlation coefficient matrices are similar for $\mathbf{S}_{\epsilon_s}$ and $\mathbf{S}_{\epsilon'_b}$ and indistinguishable to the inversion algorithm model.

The error covariance matrices may be modeled in terms of the standard deviations and correlation coefficient matrices as follows

$$S_{\epsilon_{\mathbf{cl}}}(i,j) = \epsilon_{\mathbf{cl}}(i) C_{cl}(i,j) \epsilon_{\mathbf{cl}}(j) \tag{4.59}$$

$$S_{\epsilon_{\mathbf{mi}},\epsilon_{\mathbf{cl}}}(i,j) = \epsilon_{\mathbf{mi}}(i) C_{s,cl}(i,j) \epsilon_{\mathbf{cl}}(j) \tag{4.60}$$

for each element i, j. The matrices $\mathbf{S}_{\epsilon'_b}$ and $\mathbf{S}_{\epsilon_s}$, available from the experiment, define $\mathbf{S}_{\epsilon_{\mathbf{mi}}}$ and $\epsilon_{\mathbf{mi}}$. The correlation coefficient matrices $\mathbf{C}_{mi,cl}$ and $\mathbf{C}_{cl}$ may be calculated from the statistical database used in Sec. 4.3.3 containing only $\epsilon_{\mathbf{s}}^{ref}$ for a specific monthly mean. We will scale $\epsilon_{\mathbf{cl}}$ to other months and zones by

$$\epsilon_{\mathbf{cl}} = \epsilon_{\mathbf{cl}}^{ref} \frac{\epsilon_{\mathbf{mi}}}{\epsilon_{\mathbf{s}}^{ref}} \tag{4.61}$$

assuming that $\epsilon_{\mathbf{cl}}$ is linearly related to the total random error from all sources. Inversions of simulated data containing varying levels of random noise confirms this and also show that $\mathbf{C}_{cl}$ and $\mathbf{C}_{mi,cl}$ are consistently reproduced.

The estimated water vapor random error components corresponding to the error terms listed in Table 4.1 are shown in Fig. 4.31 for a monthly zonal mean atmosphere corresponding to January 1990 at 40° N.

The vertical variation of the implicit smoothing is shown separately in Fig. 4.31. Below 15 km, $\epsilon_{\mathbf{cl}}$ is much smaller than $\epsilon_{\mathbf{s}}$ and the diagonal covariances of $\mathbf{S}_{\epsilon_{\mathbf{mi}},\epsilon_{\mathbf{cl}}}$ tend to zero. The slant path absorption displays more curvature in the troposphere

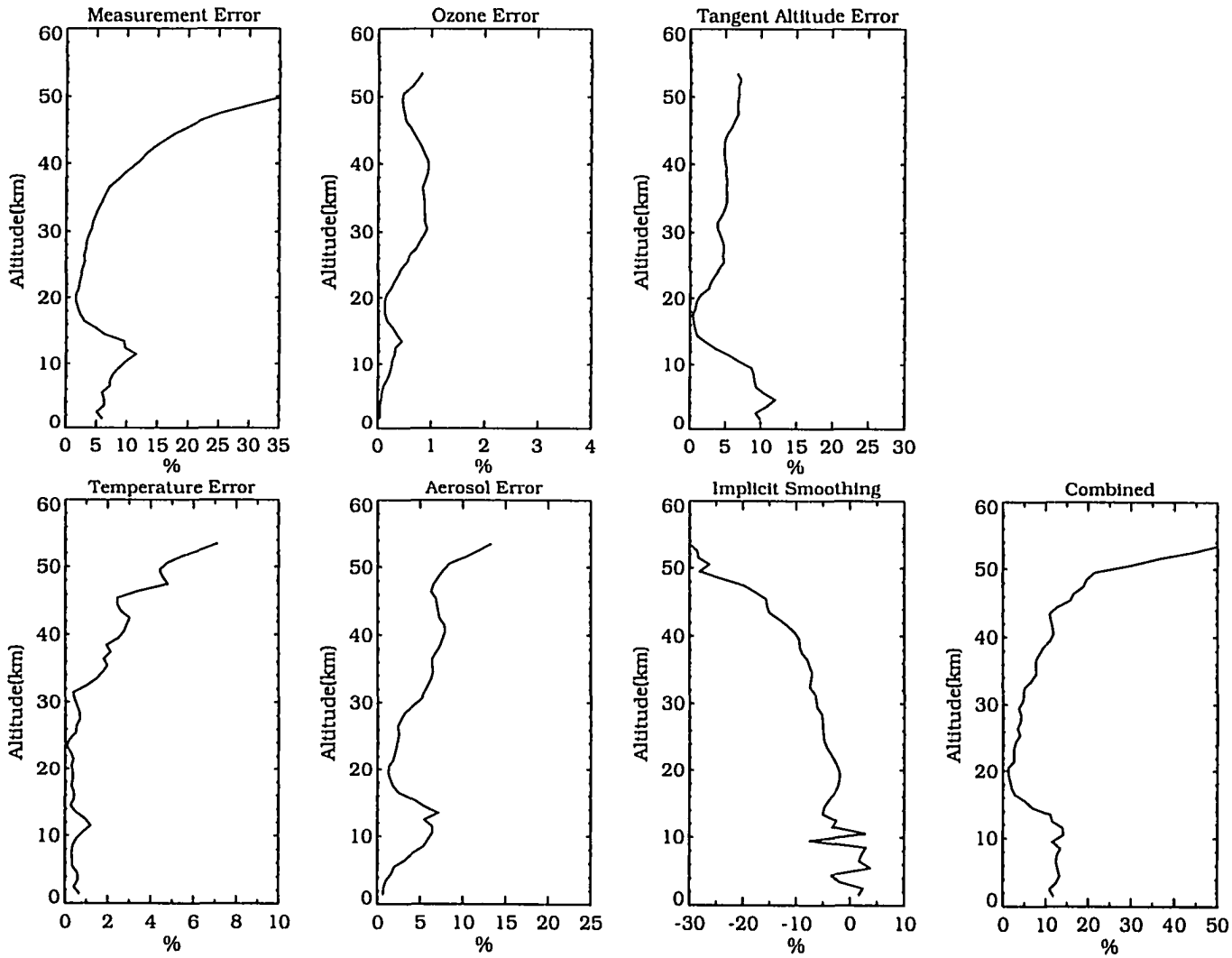


Figure 4.31: Estimated random water vapor Mill-Drayson inversion errors for the January 1990 40° N monthly zonal mean atmosphere.

than in the stratosphere and when combined with the higher signal to noise ratios makes the Mill-Drayson algorithm act like the onion peel algorithm. One might say that the random errors are no longer distributed upwards to higher levels because the rapidly increasing downwards water vapor mixing ratio causes most of the absorption to take place in the tangent layer and it is here where the full errors appear.

The implicit smoothing increases upwards throughout the stratosphere. At some altitude, depending on the month and zone, the signal to noise ratios drop to very low values and the effect of the smoothing terms diminishes. The implicit smoothing reduces the combined random error to $\approx$0.66 of the original, not quite as strong as three point vertical smoothing which would reduce it to $\approx$0.59.

The robustness of the implicit smoothing to different initial profiles and profile smoothness was tested. The implicit smoothing was not adversely affected by unrealistically low or high initial profiles or by initial profiles with strong slopes in the stratosphere. It was possible to reduce the strength of the smoothing by more than half by using a profile constructed with extreme mixing ratio variations. Since the variations were larger than what one might expect to see in a realistic profile, we will assume that the implicit smoothing shown in Fig. 4.31 is achievable for most realistic profiles.

The total experimental error is shown in Fig. 4.32. The total random error includes the effect of the implicit smoothing described by the additional terms in Eq. (4.57).

4.4.5 Sparse Grid Least Squares Inversion

The single wavelength global fit inversion reduces to a least squares solution that minimizes the slant path absorption residuals. Given the analytical least squares solution and ignoring the absorption smoothing, the inversion algorithm matrix **IA**

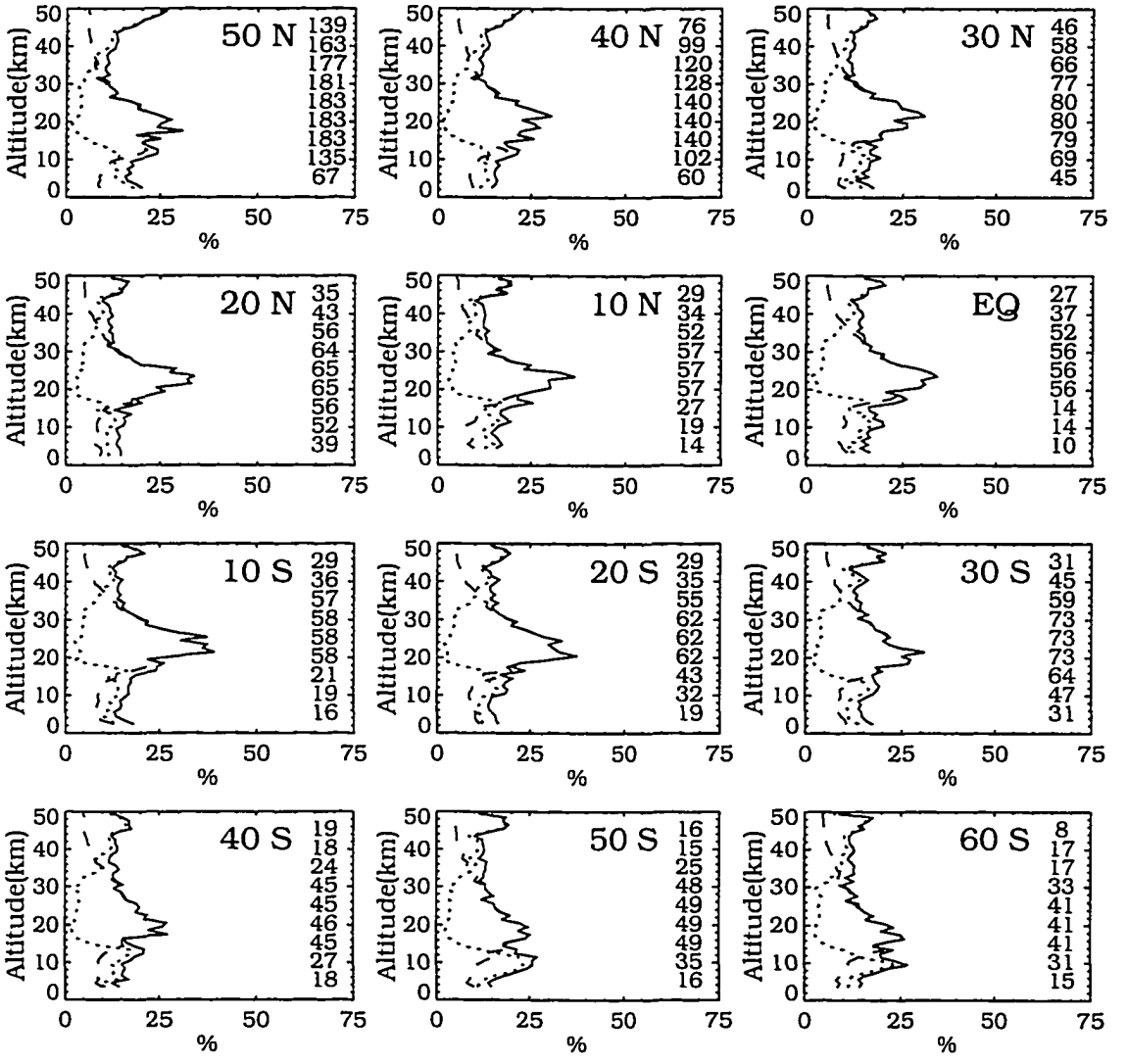


Figure 4.32: Total systematic (dashed), random (dotted), and experimental (solid) estimated uncertainty corresponding to the Mill-Drayson retrieval, January 1990. The number of points included in the zonal statistics are indicated on the right by altitude level.

may be written as

$$\mathbf{IA} = (\mathbf{K}_\mathbf{L}^T \mathbf{K}_\mathbf{L})^{-1} \mathbf{K}_\mathbf{L}^T \tag{4.62}$$

where $\mathbf{K_L}$ is another form of weighting function that may differ from $\mathbf{K_x}$. For the well posed problem where the number of unknown mixing ratios, n, matches the number of equations or measurements m, $\mathbf{IA}$ reduces to

$$\mathbf{IA} = \mathbf{K}_\mathbf{L}^{-1} = \mathbf{K}_\mathbf{x}^{-1} \tag{4.63}$$

When the least squares iterative process is allowed to proceed to the minimum, an optimal solution identical to the direct solution results. If the iterative process is terminated using a measure of the total measurement noise, the solution is no longer optimal or unique, and the analysis of Sec. 4.3.3 must be carried out to ascertain the error characteristics of the least squares method. Controlling the iterative process by specifying the desired number of significant digits in the independent parameters leads to a direct solution for a high number of digits ($>\approx 4$) while specifying fewer digits means the analysis of Sec. 4.3.3 must again be performed.

Our interest in the least squares method lies in using it to solve an over-determined systems of equations, $m > n$, created by reducing the number of levels in the stratosphere at which a solution is desired. We refer to this as the sparse-grid, least-squares (SGLS) algorithm. If $\mathbf{x}^r$ is desired, for example, at 17.5, 20.5, 23.5,..., 59.5, and 60.5 km, measurements adjacent to a given solution level will also contribute to the solution at that level and reduce the estimated errors in $\mathbf{x}^r$. Between solution levels $\mathbf{x}^r$ is filled in by interpolation and should be fairly accurate given the slow variation of the water vapor mixing ratio in the stratosphere. The amount of error reduction can be controlled by the number of solution levels to measurement levels in the stratosphere. Cases where one solution level is based on two measurements will be denoted as case 1:2. The example altitude grid given above corresponds to the 1:3 case.

The weighting function $\mathbf{K_L}$ for the 1:3 case, shown in Fig. 4.33, is now an $m \times n$ matrix. The mixing ratio altitude interpolation broadens the weighting functions relative to $\mathbf{K_x}$ in Fig. 4.4, which still relates $\mathbf{x}^t$ to $\mathbf{y}^m$. Because the system is over-determined, the row weighting functions in the stratosphere display multiple overlapping peaks. For the 1:3 case, 3 functions peak at the same altitude.

The contribution function for the 1:3 case, including slant path absorption smoothing, is an $n \times m$ matrix and is shown in Fig. 4.34 for selected columns and rows. Here, the column functions are triply overlapping. For the 41.4 km solution level, the column function resulting from a perturbation in $\mathbf{y}^m$ one level above, at 42.5 km, is plotted (dotted line) and two levels above (dashed line) while for the 29.5 km solution column functions corresponding to perturbations one level below (dotted line) and two levels below are shown (dashed line). Perturbations in $\mathbf{y}^m$ at solution levels of $\mathbf{x}^r$ produce column functions (solid lines) similar in form to those of Fig. 4.16 or Fig. 4.34 for the onion peel and operational inversions, respectively. Column functions from perturbations in $\mathbf{y}^m$ at non-solution levels exhibit oscillatory behavior above the perturbation level, broader functions (one level above 41.5 km-dotted line, two levels below 29.5 km-dashed line) and the row functions also exhibit oscillatory behavior at greater distances from the peak.

These characteristics also appear in the averaging kernels shown in Fig. 4.35. Unlike previous averaging kernels in the stratosphere, both column and row functions exhibit oscillations above and below the perturbation layer. Perturbations in $\mathbf{x}^t$ at the solution levels display a peak response at those levels and perturbations in $\mathbf{x}^t$ one level above or below a solution level will peak at that level. When peak values of $\mathbf{A}$ in the stratosphere of $\approx$0.375 are compared to $\approx$0.45 for the onion peel and $\approx$0.25 for the operational inversion, we see that this inversion will have error characteristics midway between these two and is actually closer to the onion peel inversion. This was somewhat unexpected since the same three point slant path absorption smoothing

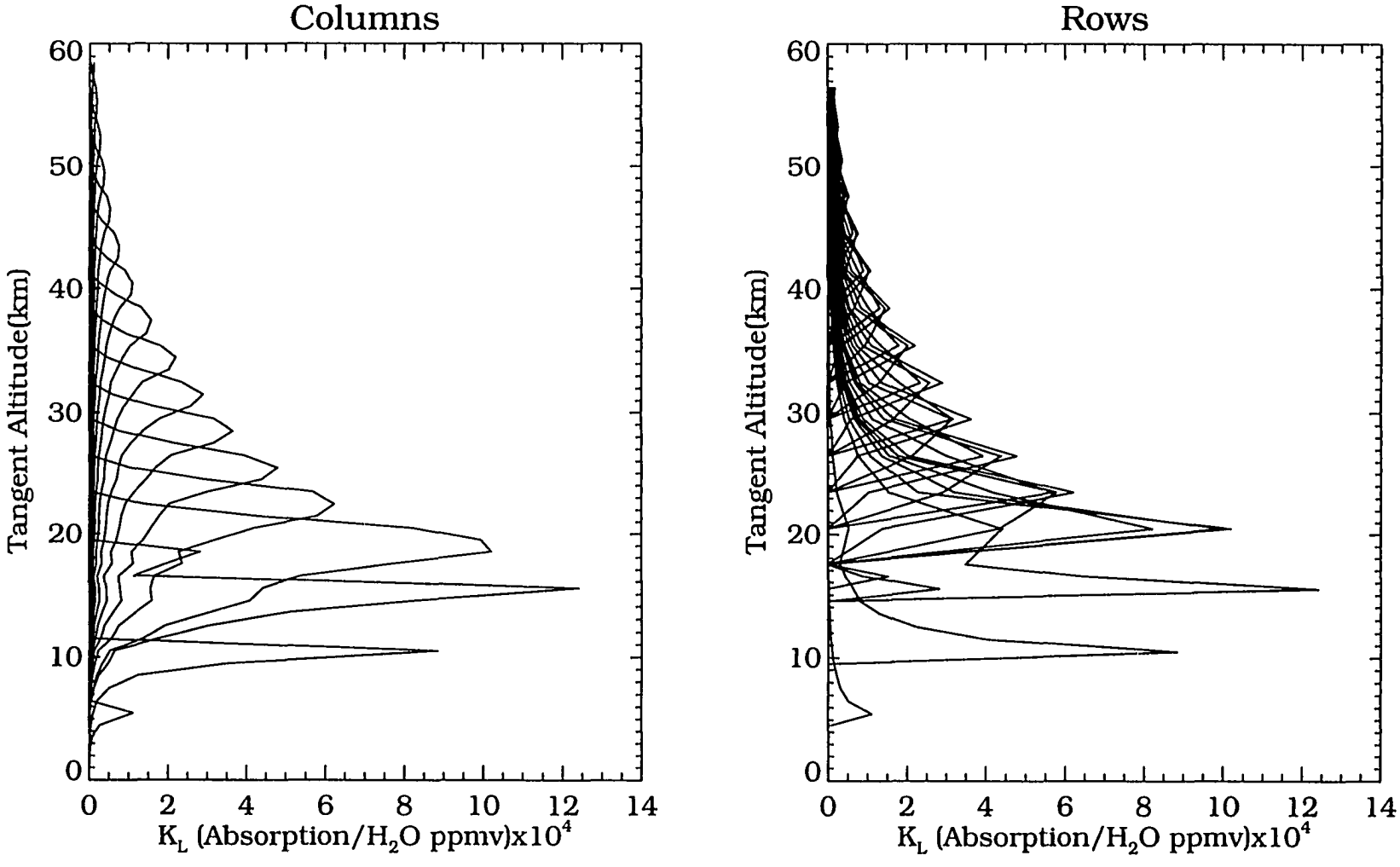


Figure 4.33: Columns and rows of the least squares weighting function matrix for the $\mathbf{K_L}$ case at 5.5, 10.5, 15.5 km, and all levels above 18.5 km. Slant path absorption smoothing is not included.

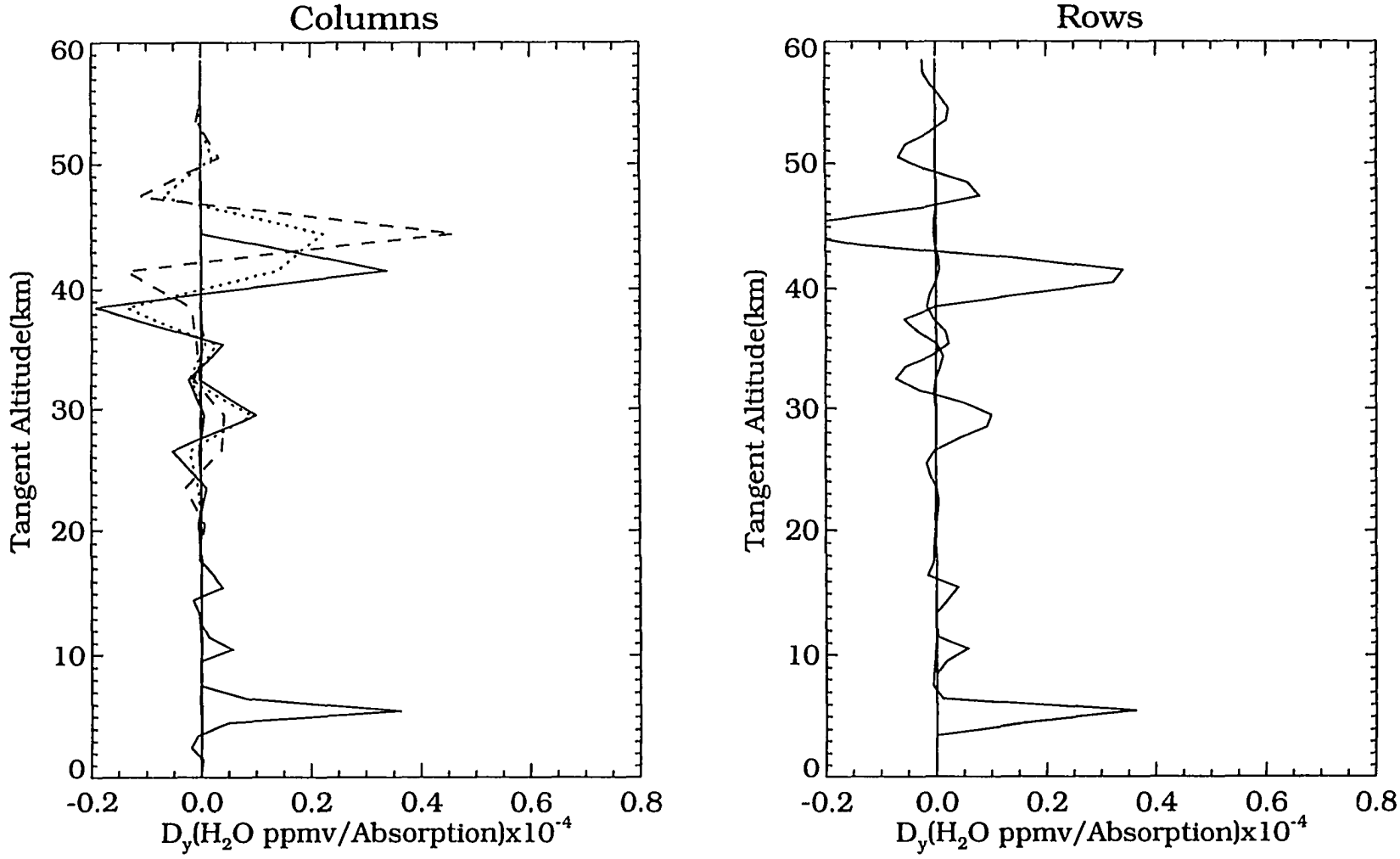


Figure 4.34: Columns and rows of the SGLS inversion contribution function matrix $\mathbf{D_y}$ for the 1:3 case at 5.5, 10.5, 15.5, 29.5, and 41.5 km. The dotted line displays the column function one level above 41.5 km and one level below 19.5 km while the dashed line corresponds to two levels above and below, respectively.

is applied in all inversion algorithms and creating an over-determined system, for the 1:3 case, should mimic the mixing ratio smoothing in the operational inversion. Without the three point absorption smoothing, peak values of **A** are ≈0.46, so the smoothing reduces **A** by a factor of 0.81. This is less than the onion peel smoothing reduction factor of 0.45. The three point absorption smoothing, a narrow function, is less effective here because the functions of $(\mathbf{K}_L^T\mathbf{K}_L)^{-1}\mathbf{K}_L^T$ are broader than those of $\mathbf{K}_x^{-1}$. Another factor that makes the peak amplitude of the functions of **A** greater than one might expect is the oscillatory nature of **A**. Negative excursions are balanced by increasing **A** at the peak.

Since the peak stratospheric values of **A** are somewhat closer to the onion peel than the operational, the estimated errors should be similar to the onion peel inversion but slightly lower, at least for random errors. For systematic errors, shown in Fig. 4.36, this inversion, the onion peel, and operational inversions yield nearly identical results.

The estimated water vapor error due to the random errors is shown in Fig. 4.37 for a monthly zonal mean atmosphere corresponding to January 1990 at 40° N. As expected, the random estimated errors are only slightly reduced from the onion peel levels.

The amount of smoothing achieved with the SGLS inversion may be varied by changing the number of solution levels. Peak stratospheric values of **A** without absorption smoothing are 0.67, 0.46, and 0.37 for the 1:2, 1:3, and 1:4 cases respectively. Including absorption smoothing these values are 0.45, 0.36 and 0.3 (estimated). Thus, the 1:2 case provides less smoothing than the onion peel inversion, the 1:3 case slightly more than the onion peel, and the 1:4 case slightly less than the operational inversion. Stability problems were experienced for the 1:4 case during the iterative process even though the onion peel profile reduced by 5% defined the initial

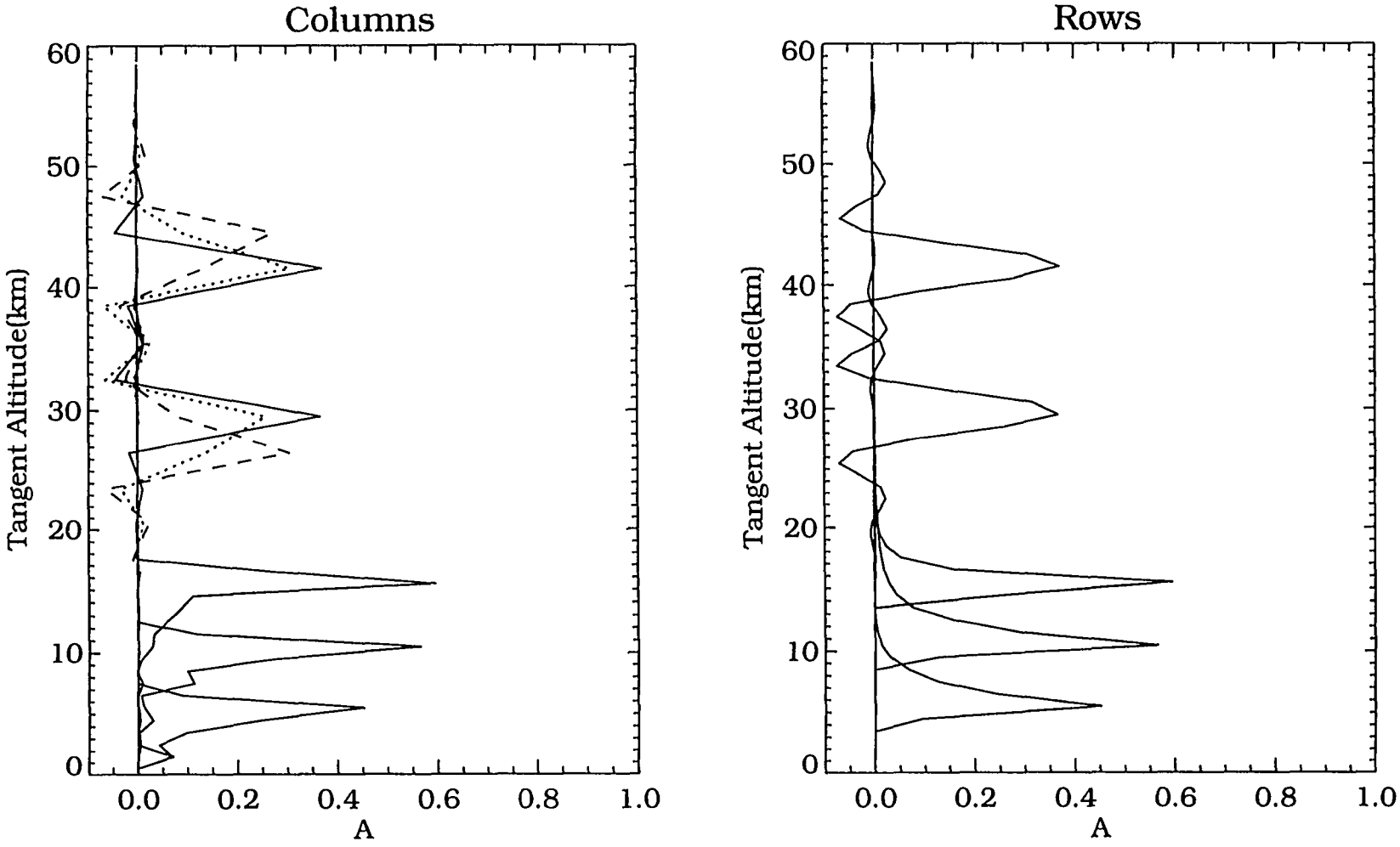


Figure 4.35: Columns and rows of the SGLS inversion averaging kernel matrix **A** for the 1:3 case at 5.5, 10.5, 15.5, 29.5, and 41.5 km. The dotted line displays the column function one level above 41.5 km and one level below 19.5 km while the dashed line corresponds to two levels above and below, respectively.

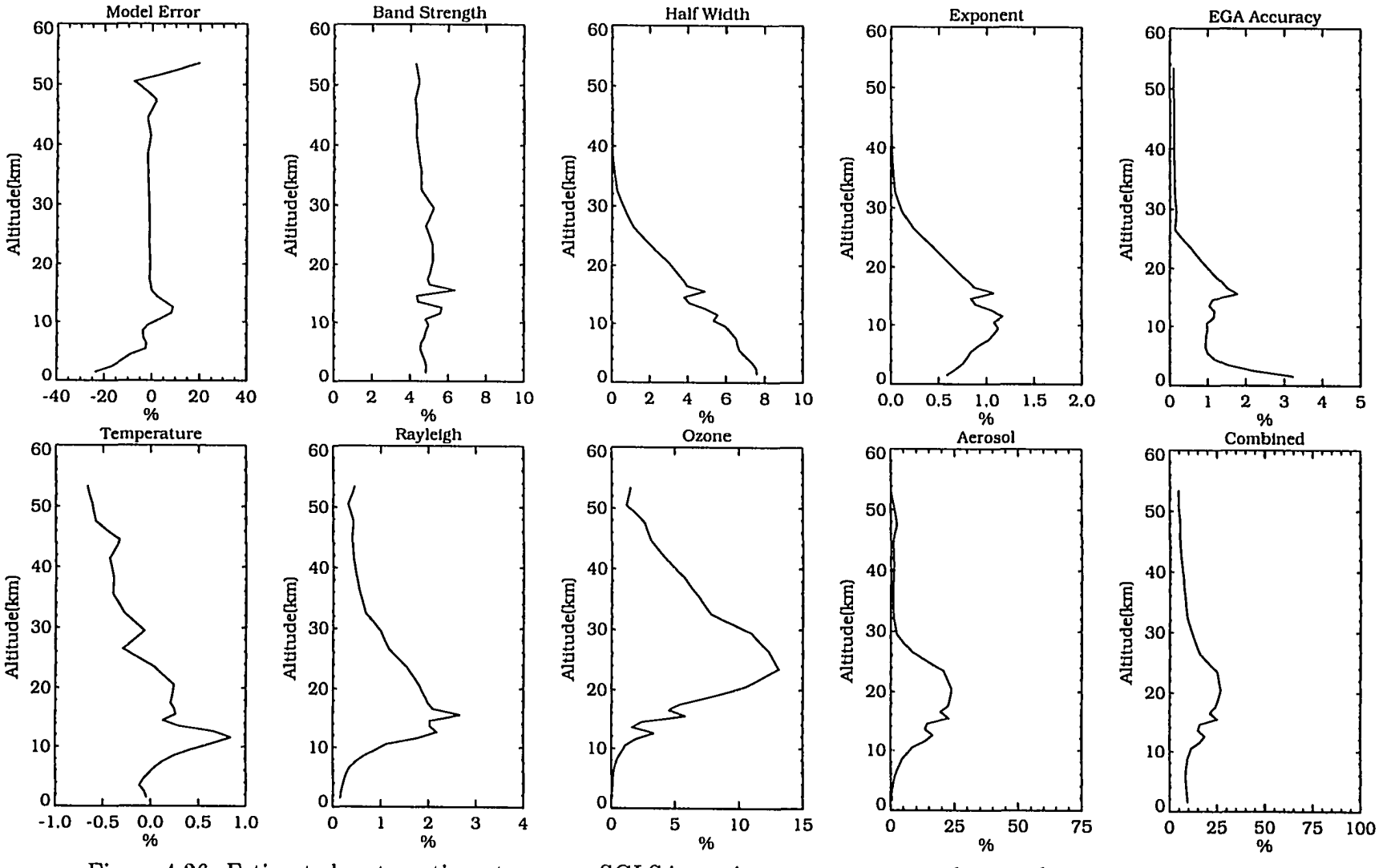


Figure 4.36: Estimated systematic water vapor SGLS inversion errors corresponding to the 1:3 case for the January 1990 40° N monthly zonal mean atmosphere.

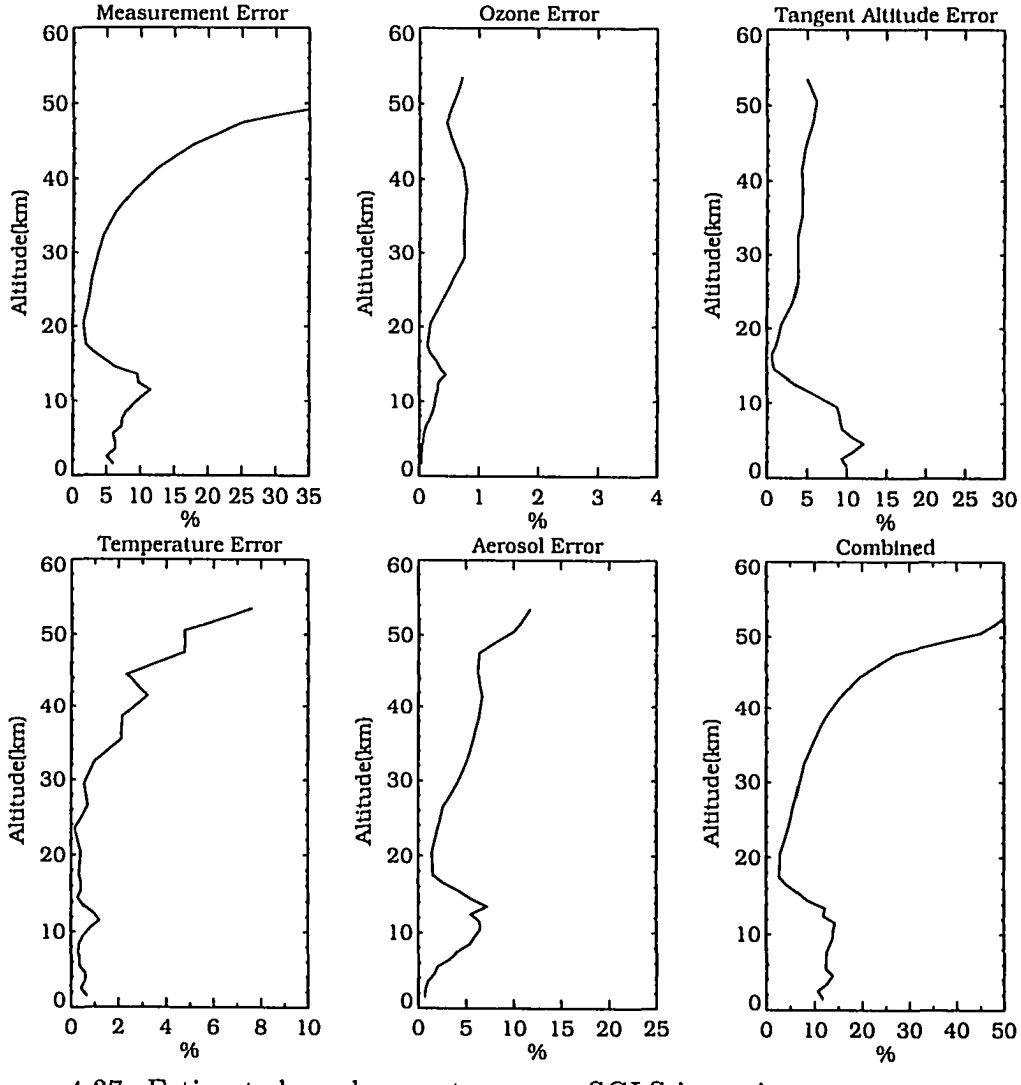


Figure 4.37: Estimated random water vapor SGLS inversion errors corresponding to the 1:3 case for the January 1990 40° N monthly zonal mean atmosphere.

profile. Although it might be possible to improve the stability by first linearizing the system and then seeking an iterative solution, or even simply applying the analytical form of the solution, this was not attempted. The non-linear least squares solution is computationally intensive and given the lack of any advantage over the previous inversion algorithms, this method will not be included in the inversion comparisons.

4.4.6 Zonal and temporal variation of error terms

Many of the error terms arise directly from uncertainties in atmospheric parameters or are expressed in some manner to atmospheric parameters; geophysical variations in the parameters, either zonally or temporally, induce variations in the error term. Other terms are related to channel measurement noise, which may change if the operating characteristics of the instrument changes. A limited view of the error terms was presented in the previous sections, individual error terms at 40° N or total errors for a single month. Here, we briefly examine the extent to which the error terms vary within a year of observations.

Some errors, significant or not, display little or no variation. Of the systematic errors, we will examine the model, ozone and aerosol errors. Systematic errors due to uncertainties in band strength, halfwidth and temperature dependence of the halfwidth are independent of time or zonal band on a percentage basis. EGA errors are constant above 12 km and vary over a 3% range below. The peak systematic temperature error doubles at the equator and shifts upward, following the tropopause, but remains small and does not contribute significantly to the total error. A similar upward shift in the peak error is seen for the systematic Rayleigh error.

Figure 4.38 displays the zonal variation of the model, ozone and aerosol errors. The model errors, resulting from smoothing the slant path absorption, tend to follow zonal variations in the water vapor mixing profile. No variation is observed in

the stratosphere. The positive bias of 10% seen at 40° N appears in other zones, tracking the zonal variation of the hygropause, and is an indicator of the change in absorption gradient between the stratosphere and troposphere. Below, the errors become increasingly negative in all zones.

The zonal variation in the systematic error due to ozone reflects the zonal distribution of the ozone slant path optical depth which in turn follows the ozone number density variation (McCormick et al., 1989). The error maximizes at the equator, slightly greater than 18%, and decreases to 12% at the higher latitudes.

The systematic error due to uncertainty in the aerosol modeling also follows the aerosol distribution as shown in Fig. 4.38. Peak errors greater than 36% at the equator drop off to 25% at the high latitudes.

Figure 4.39 displays the random error contributions from the measurement, temperature, and tangent altitude uncertainties. Random errors due to uncertainties in ozone are always less than 1% and errors due to aerosol modeling only show slight shifts in altitude with little change in peak error. Uncertainties due to measurement error display a slight zonal variation with a small increase through the equatorial middle stratosphere. NWS temperature uncertainties are constant up to 30 km with values of 3-4° C that linearly increase to 12° C at 55 km. The lack of any significant zonal variation in the temperature uncertainty is also found in the estimated random error except where the peak error tracks the tropopause/hygropause region. The estimated tangent altitude uncertainty was assumed to be constant for all altitudes, latitude zones and months. However, the water vapor random error shown in Fig. 4.39 displays zonal variation that again follows the hygropause. The water vapor uncertainty above the hygropause would be greater than that shown except for the fact that much of the error has been removed by the ozone/aerosol separation process, in effect shifting the error to the ozone and aerosol signals. Immediately

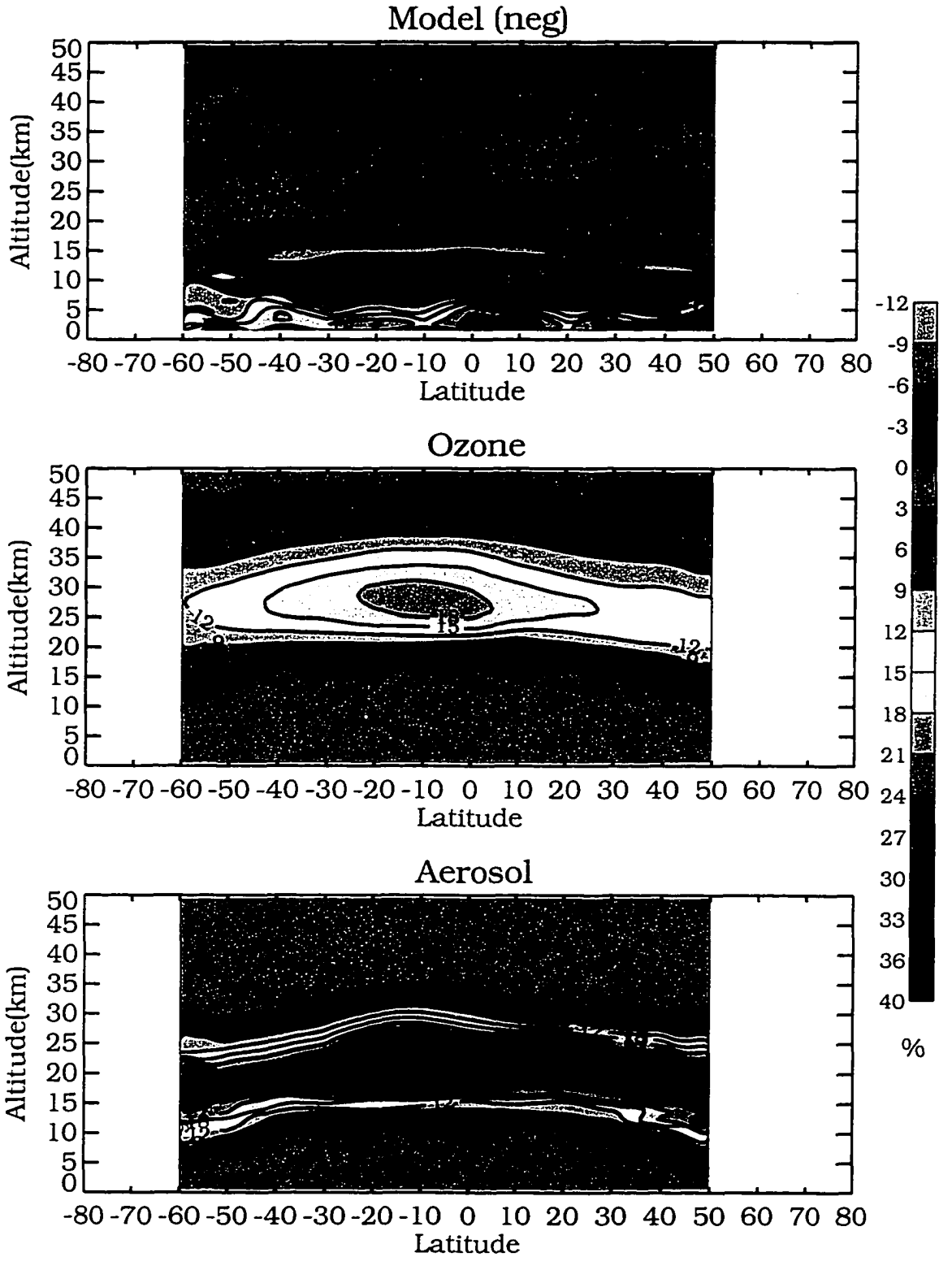


Figure 4.38: Zonal variation of estimated systematic water vapor mixing ratio errors (in percent) due to model, ozone and aerosol uncertainties for the operational inversion, January 1990.

below the hygropause the water vapor signal increases rapidly relative to ozone and aerosol, they absorb less of the tangent altitude error, and the random error increases rapidly.

Figure 4.40, displaying the zonal variation of water vapor mixing ratio errors showing the total systematic, random and experimental uncertainties for the operational inversion in January 1990, is a restatement of the information contained in Fig. 4.25. The zonal variation of the total systematic error follows primarily from the combination of the ozone and aerosol errors. The dominant component in the stratospheric portion of the total random error is the measurement error. The rapid increase in the tropospheric error again coincides with the hygropause/tropopause region. In the troposphere, the measurement, tangent altitude, and aerosol error combine to produce an error distribution with little zonal structure. Zonal variation in the total error, shown in the bottom panel, is determined by the top two panels.

Seasonal variations of ozone (McCormick et al., 1989) and temperature will cause error terms dependent on these quantities to also exhibit seasonal variations. Episodic injections of volcanic aerosols may significantly increase the total systematic estimated error while the decay of aerosols from enhanced levels produces a corresponding decrease in the estimated error. In addition, several of the errors track the zonal variation of the hygropause/tropopause, they will also track seasonal variations of this region.

Figures 4.41 and 4.42 are typical examples of the annual variations of the total systematic and random estimated errors, respectively, for 10° N, 30° N, and 50° N. The errors are constant for the most part throughout the year except for the 10° N systematic error. Here, the February 1990 eruption of Kelut dramatically increases the estimated error levels. The slow decay of the peak estimated error throughout the remainder of the year is also evident.

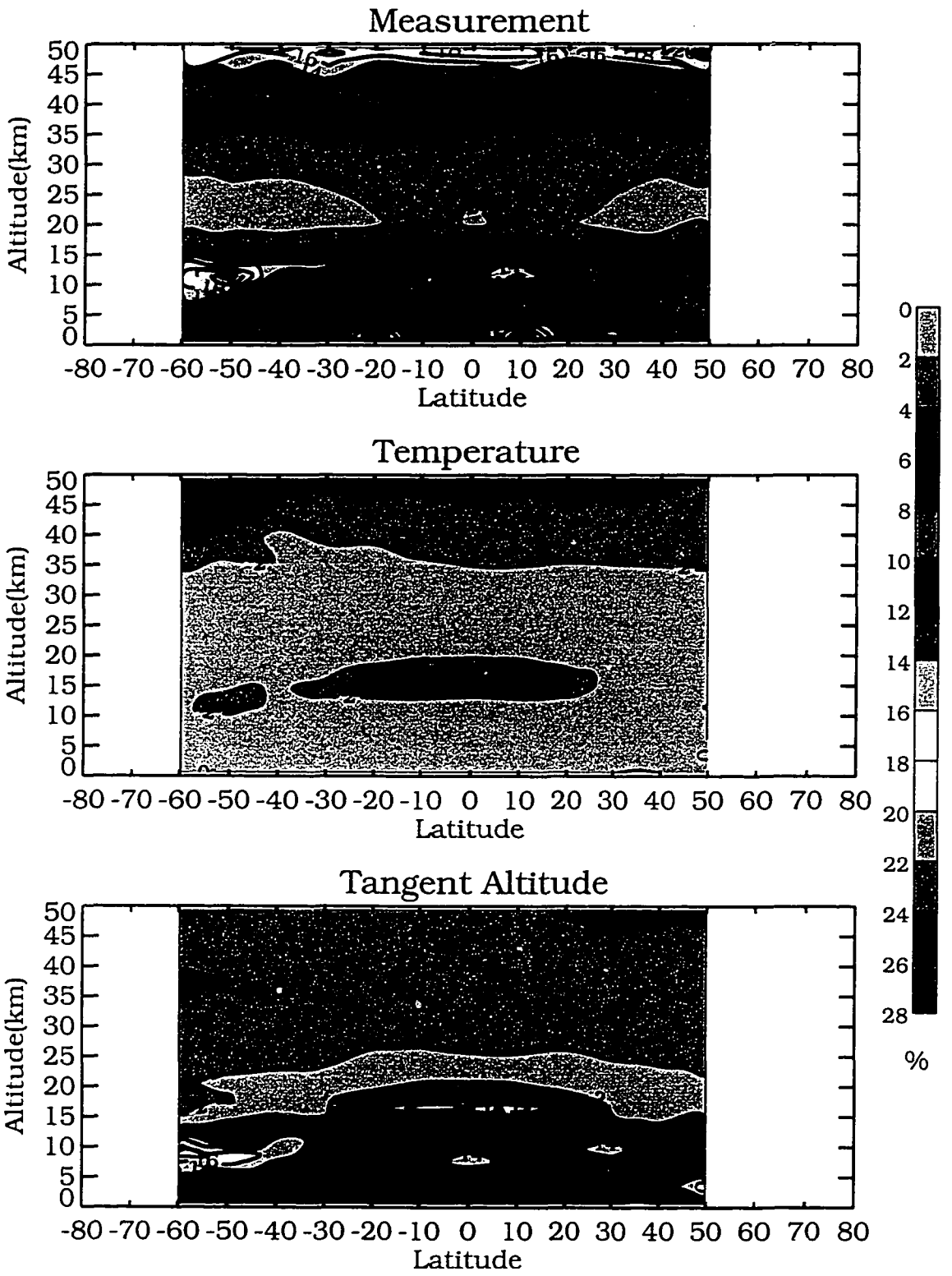


Figure 4.39: Zonal variation of estimated random water vapor mixing ratio errors (in percent) due to measurement, temperature and tangent altitude uncertainties for the operational inversion, January 1990.

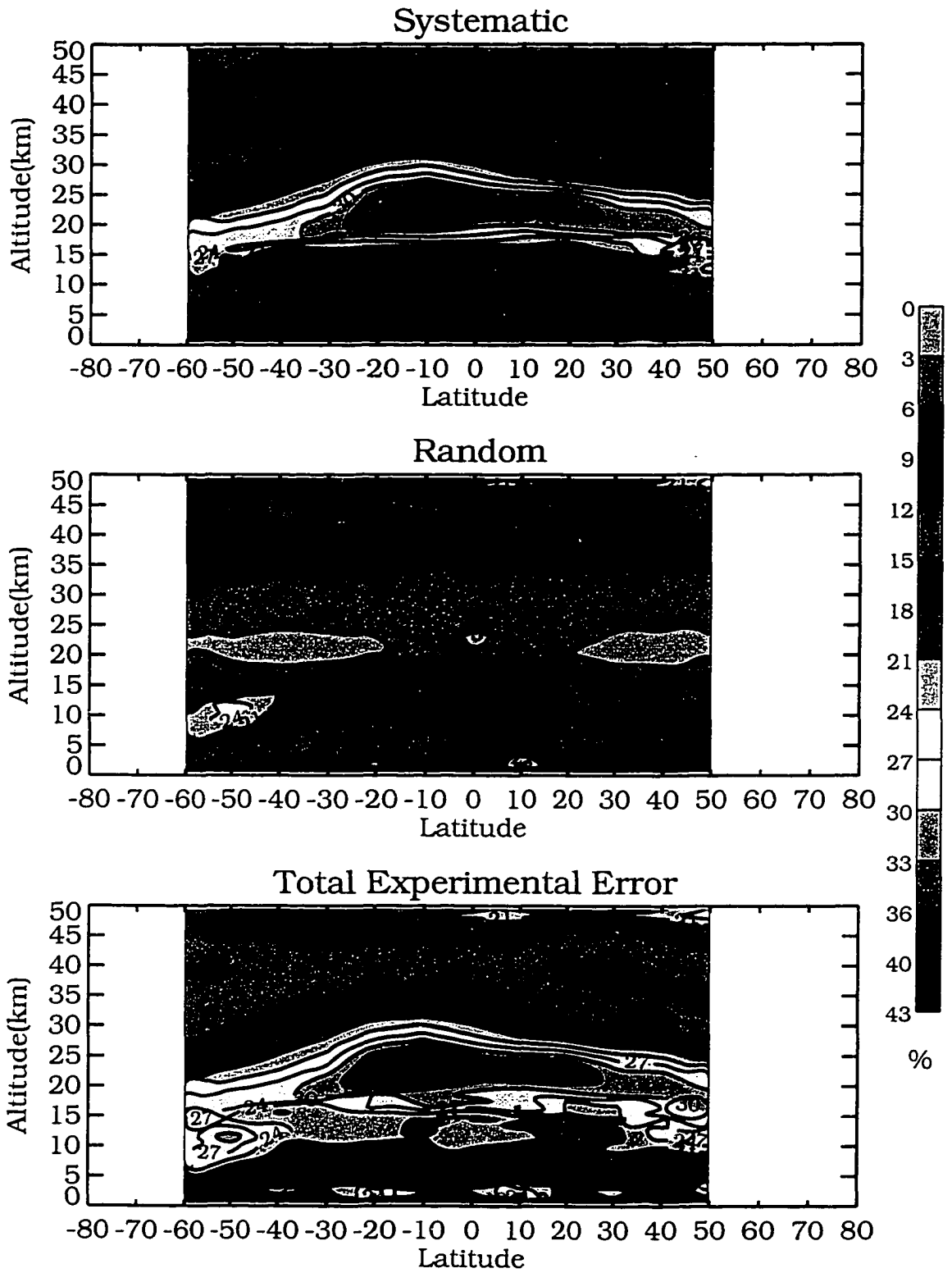


Figure 4.40: Zonal variation of estimated water vapor mixing ratio errors (in percent) showing total systematic, random and experimental uncertainties for the operational inversion, January 1990.

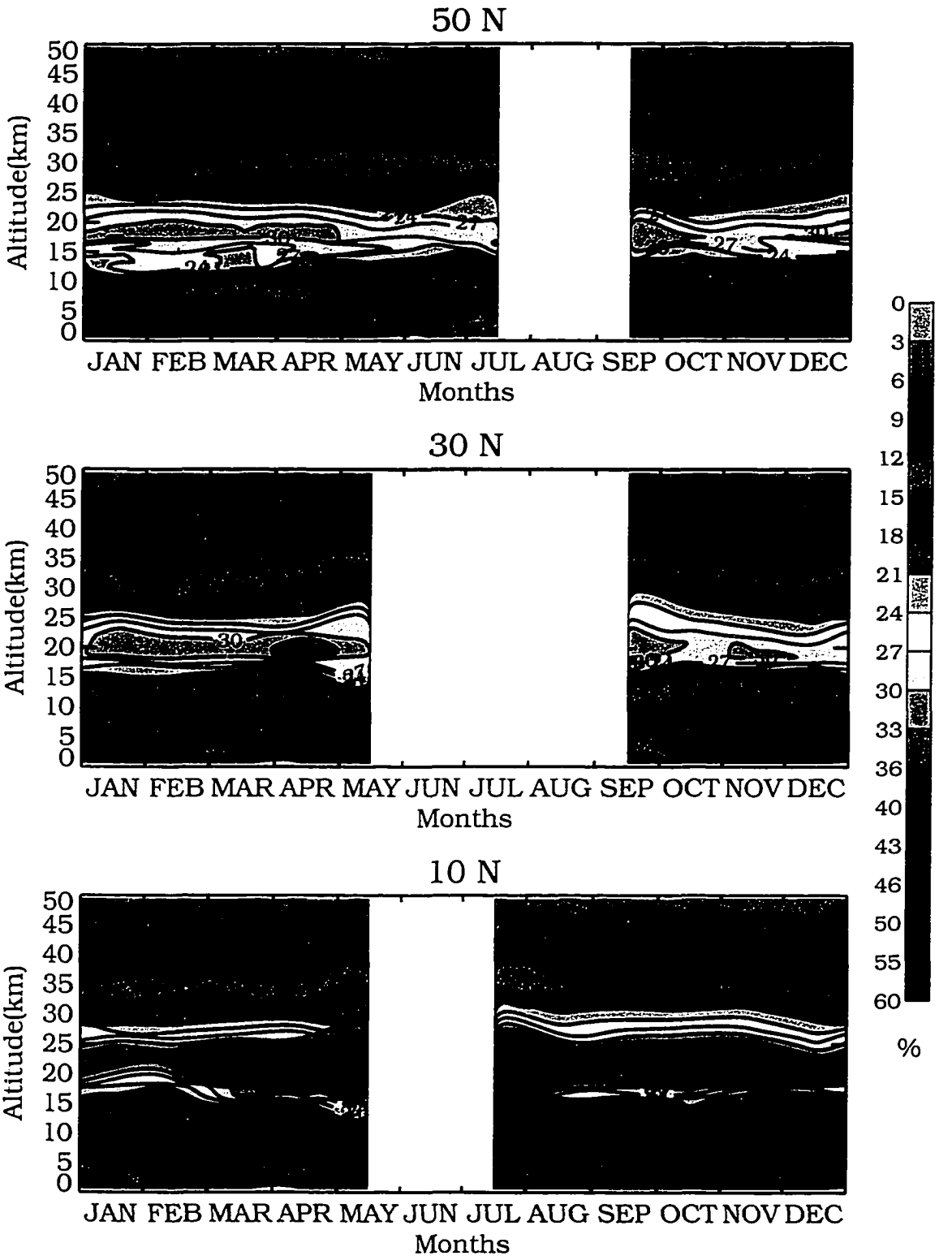


Figure 4.41: Annual variation of estimated total systematic water vapor mixing ratio errors (in percent) for the operational inversion in 1990 at 10° N, 30° N, and 50° N.

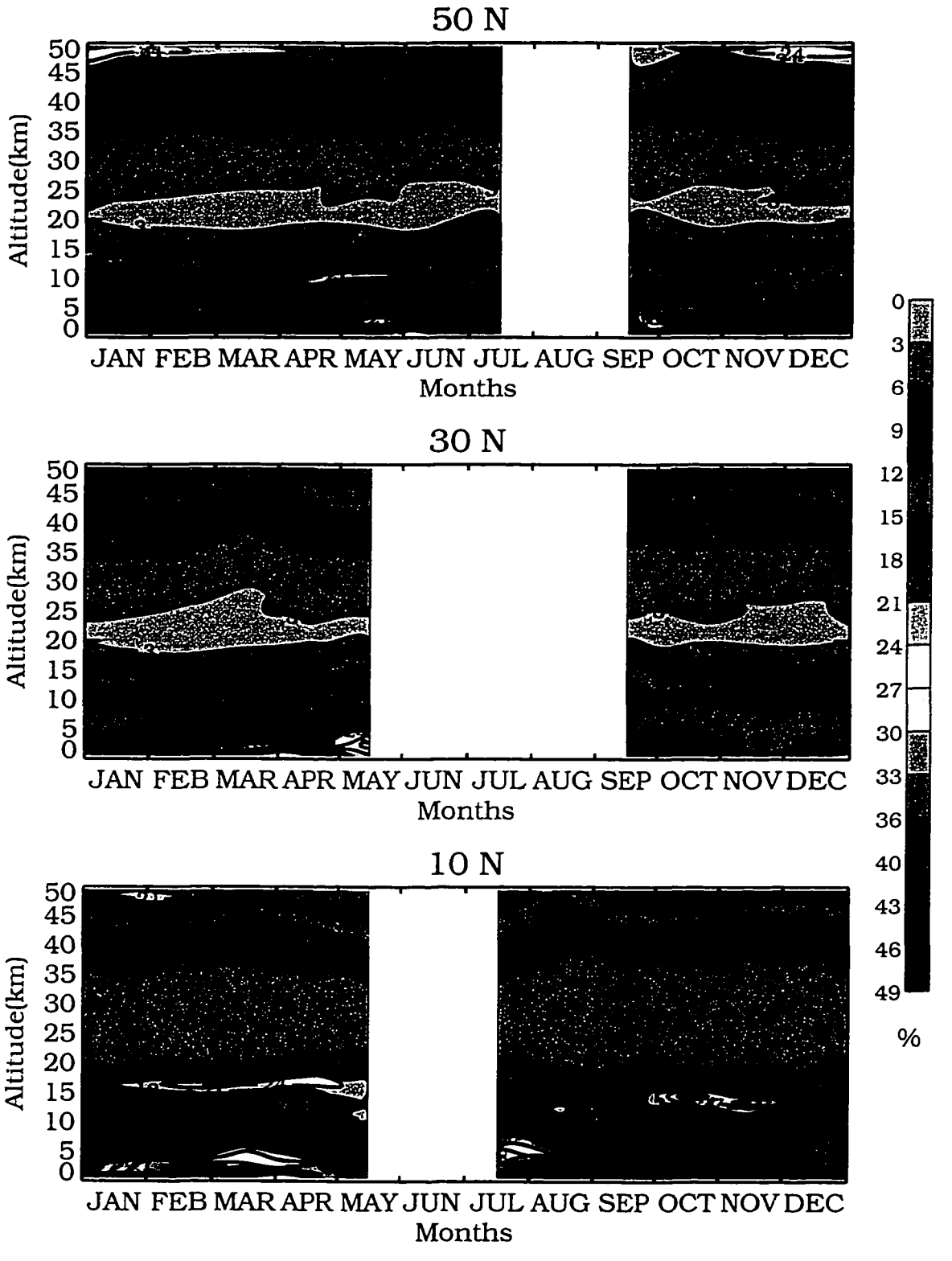


Figure 4.42: Annual variation of estimated total random water vapor mixing ratio errors (in percent) for the operational inversion in 1990 at 10° N, 30° N, and 50° N.

4.5 Inversion Results and Comparisons

Water vapor profiles inverted with the operational algorithm of Chu et al.(1993) have been examined in detail as described in Rind et al.(1992), Chu et al.(1993), Chiou et al.(1993) and Larsen et al.(1993). What we term the operational algorithm inversion algorithm in this work differs from that in Chu et al.(1993) only in the maximum number of vertical smoothing passes allowed (20 instead of 15) and the use of $\mathbf{S}_{\mathbf{y}^{mi}}$ (instead of $\mathbf{S}_{\epsilon\mathbf{y}}$) as the slant path absorption iterative convergence criterion. Since these are relatively minor changes with correspondingly small effects on the inverted water vapor, we will focus more on the profile differences between inversion types than the characteristics of the profiles and on consistency of the inverted profiles with the error analysis.

Two years (1987,1990) of observations were inverted with each algorithm and then compared on a profile to profile basis and by examining zonal means of several quantities. Due to the large number of measurement events, only a selected number of profile to profile comparisons were made. Two such examples, corresponding to equatorial and middle latitude locations, are shown in Figs. 4.43 and 4.44, respectively. Since the true water vapor mixing ratio is unknown, the onion peel, smoothed Mill-Drayson, and Mill-Drayson profiles (dotted lines) are compared to the operational profile which appears in all three panels as a solid line. The per cent difference between inverted profiles, shown on the right with a dotted line, may be viewed as a proxy for the difference between true and retrieved profiles provided one of the inversions is more heavily smoothed, such as the operational to onion peel comparison in the stratosphere. The solid lines define the random error envelope calculated using the operational error analysis of Chu et al.(1993) with the random error adjusted as described in Appendices B and C and incorporating an additional correction to account for stratospheric mixing ratio smoothing.

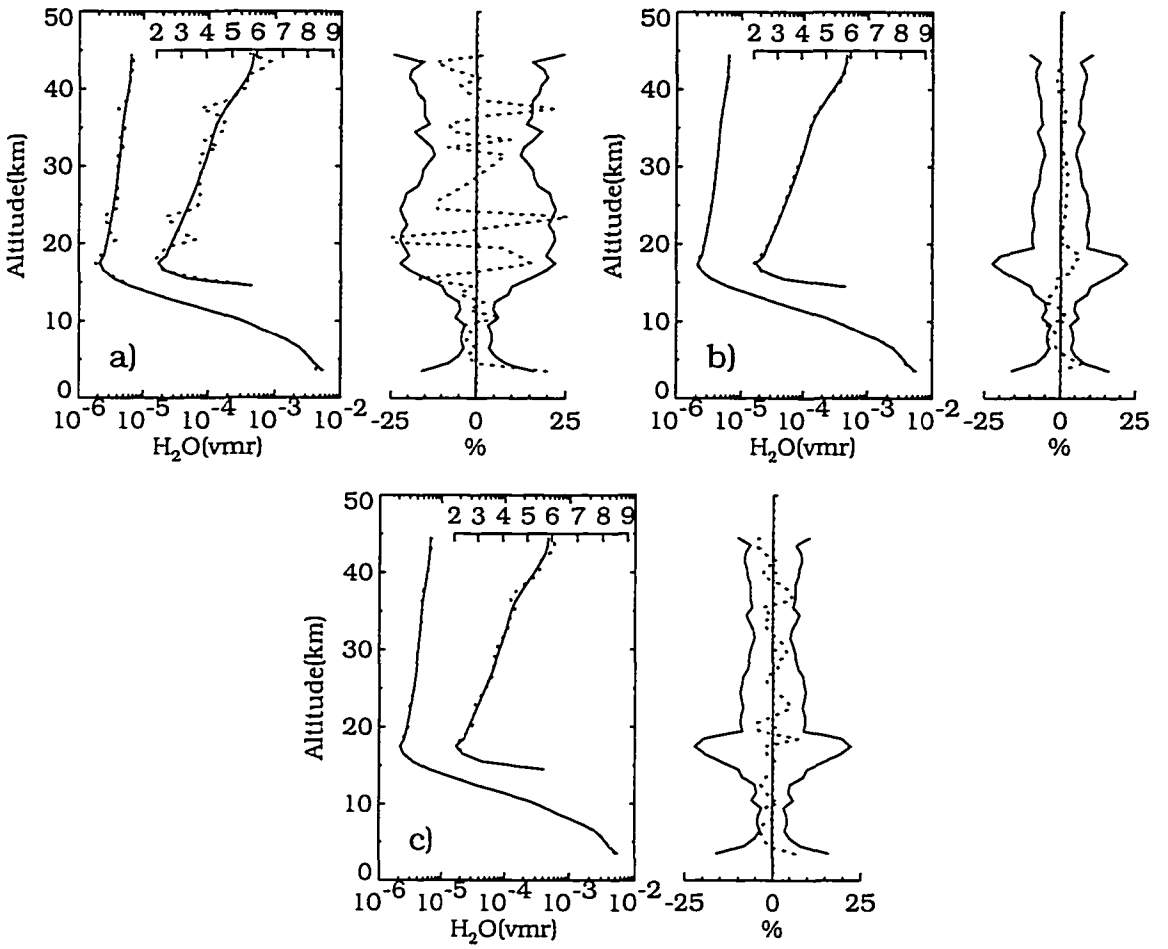


Figure 4.43: Inverted water vapor profile from the operational algorithm (solid line) are compared to a) the onion peel algorithm algorithm, b) the smoothed Mill-Drayson, and c) the Mill-Drayson algorithm (all dotted lines). Stratospheric portion of the profile is replotted on the linear inset. The percent difference between profiles is shown to the right (dotted line). The solid line defines the error envelope as described in the text. The measurement is located at 0.20° N, 144.15° W and was obtained on January 9 1990.

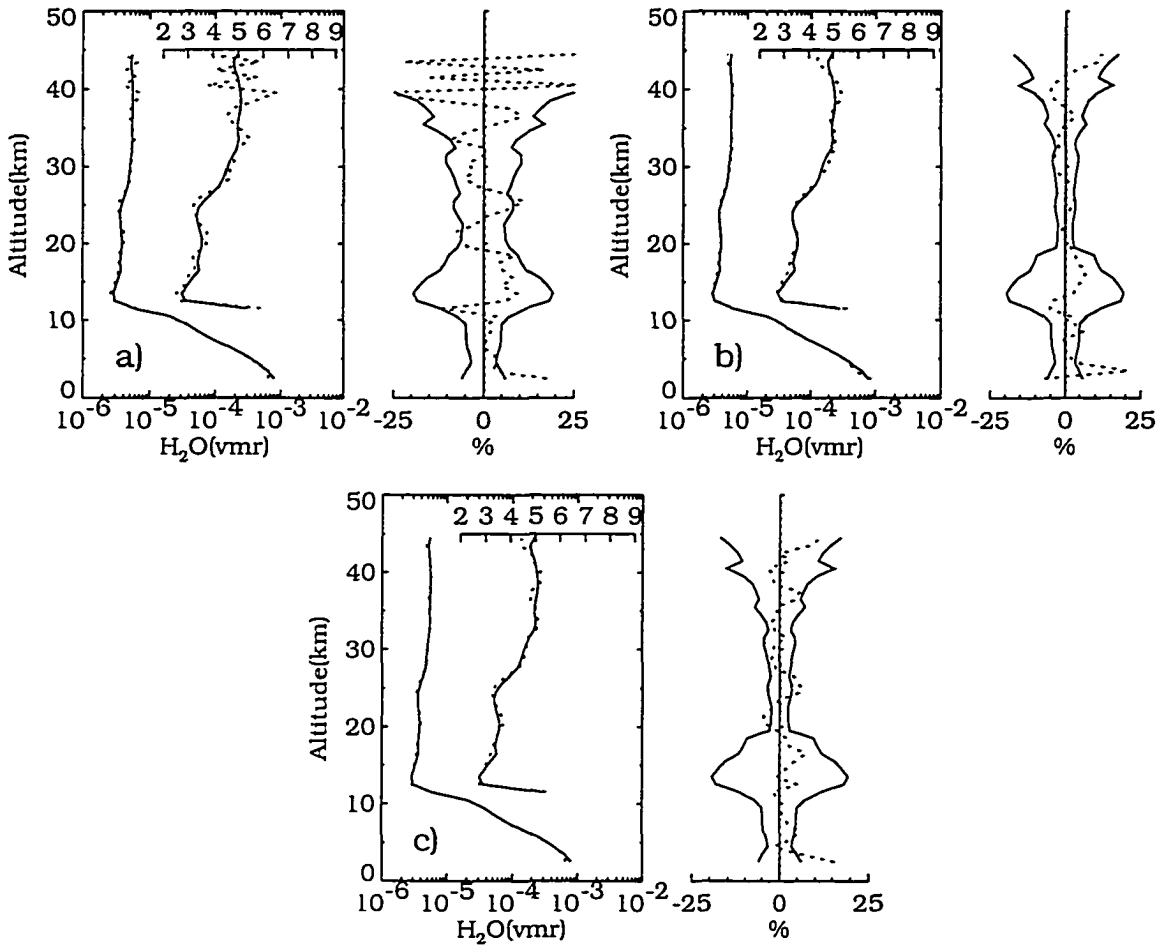


Figure 4.44: Inverted water vapor profile from the operational algorithm (solid line) are compared to a) the onion peel algorithm algorithm, b) the smoothed Mill-Drayson, and c) the Mill-Drayson algorithm (all dotted lines). Stratospheric portion of the profile is replotted on the linear inset. The percent difference between profiles is shown to the right (dotted line). The solid line defines the error envelope as described in the text. The measurement is located at 43.39° N, 16.35° E and was obtained on January 3 1990.

At most levels, the relative profile difference falls well within the estimated error envelope, indicating the inversion algorithms are performing as expected from the error analysis. The onion peel profile displays much more variability in the stratosphere than the operational as a result of the more sharply peaked contribution function. The operational and smoothed Mill-Drayson contribution function are nearly identical, as is the profile variability. The Mill-Drayson variability falls between these two cases due to the implicit smoothing. Although the profile to profile comparisons agree qualitatively with the error analysis, it is difficult to draw any conclusions about the inversion response on average. To do this we examine the monthly zonal means of the inverted profiles, associated standard deviations, and an indicator of profile curvature.

Figure 4.45 compares the monthly zonal means of the inverted water vapor profiles for the January 1990 observations. In this figure, the percent differences between the onion peel, smoothed Mill-Drayson, and Mill-Drayson are shown relative to the operational inversion. Short period variations are present in all of the comparisons and in the upper stratosphere the amplitude is nearly the same in all the zones. In the lower stratosphere the amplitude of the variations increases significantly in the equatorial region. Several of the zones show long period (top to bottom) variations of up to 5% that appear in all three comparisons suggesting that the actual variation is a feature of the operational inversion.

The amplitude of the larger short period variations is greater than what might be expected from the error analysis. The model error term causes at most a 1-2% difference between the inversion results. If many of the possible experimental systematic errors listed in Table 4.1 are actually present in the experiment, the inversions may respond differently. However, a comparison of the total systematic error between inversion algorithms finds very small differences and cannot explain those of Fig. 4.45. In Sec. 4.3.3 we stated that systematic errors in ϵ_{tc} are small and

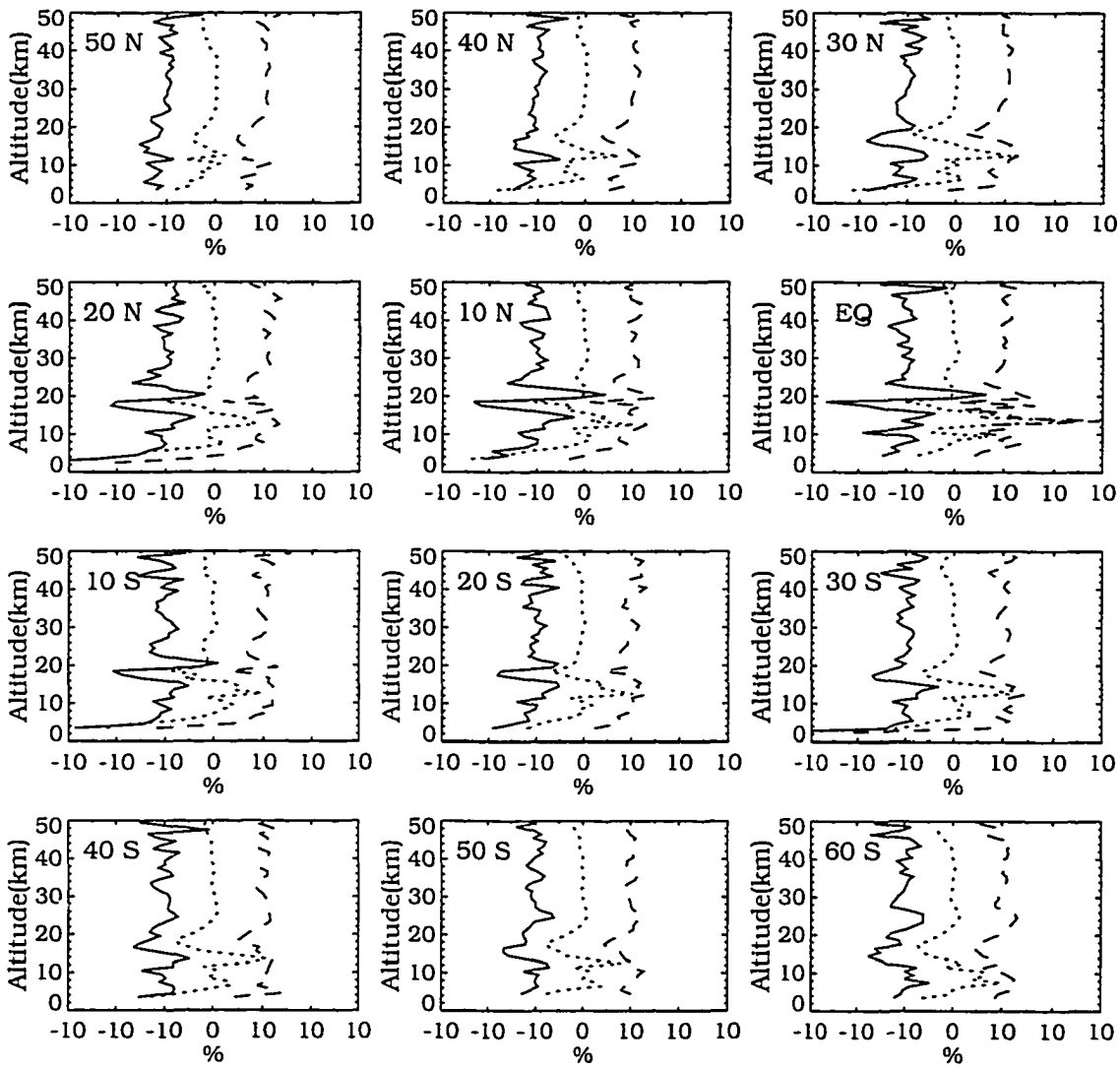


Figure 4.45: Per cent difference between the onion peel (solid line), smoothed Mill-Drayson (dotted line), and Mill-Drayson (dashed line) zonal mean water vapor relative to the operational for the January 1990 observations. Differences are offset for clarity.

could be further reduced by decreasing Δ_{cl} for all inversions except the Mill-Drayson. However, for these inversions, Δ_{cl} is given by the variances of $\mathbf{S}_{ymi}$ and some of the differences in Fig. 4.45 are caused by systematic differences in ϵ_{tc} between inversions. Re-inverting the observations with $\Delta_{cl}/5$ reduced some of the short period variations and most of the long period variations but the large equatorial variations persisted.

To determine the cause of these variations, 50 Monte Carlo simulations containing random measurement noise were inverted for each zone with Δ_{cl} given by the variances of $\mathbf{S}_{ymi}$. Figure 4.46, analogous to Fig. 4.45, shows many of the same features. In particular, the long period variations match those in the observations in several zones and the spike at 10-12 km appearing in many zones is also replicated. However, the rapid equatorial variations at 20 km are not present. Inspection of individual equatorial profiles in the observational data found some that diverged significantly from the zonal mean. It is likely that it is these profiles that are causing the large equatorial differences between inversion methods, the formal error analysis and the Monte Carlo error simulations are both based on zonal mean profiles which are usually smoother than individual profiles.

Visual inspection of individual profiles, as was done at the beginning of this section, can provide a qualitative view of profile smoothness and in turn the effectiveness of inversion smoothing. For a quantitative description of profile smoothing we developed an estimate of profile curvature γ based on the second derivative of the water vapor mixing ratio with respect to altitude, *i.e.*

$$\gamma = \left(\frac{\partial^2 \mathbf{x}^r}{\partial z^2}\right)^2 \tag{4.64}$$

Zonal mean γ are shown in Fig. 4.47 for the January 1990 observations and γ calculated from the Monte Carlo error simulations are nearly identical (not shown). As expected from the random error analysis, the operational and smoothed Mill-Drayson display identical, and least, curvature in the stratosphere while the onion

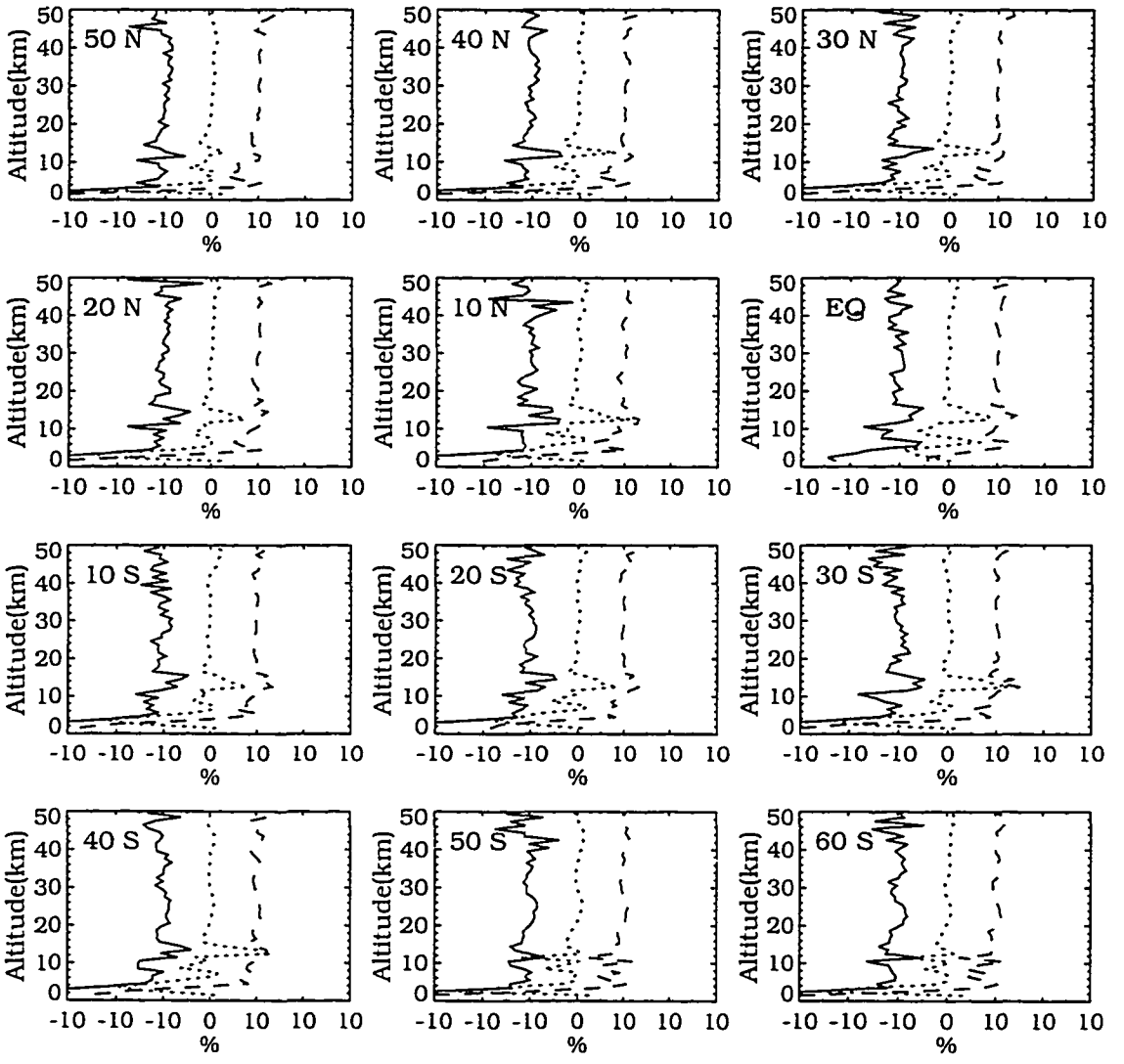


Figure 4.46: Per cent difference between the onion peel (solid line), smoothed Mill-Drayson (dotted line), and Mill-Drayson (dashed line) zonal mean water vapor relative to the operational for the January 1990 Monte Carlo random measurement noise simulations. Differences are offset for clarity.

peel displays the most. The Mill-Drayson inversions fall in the middle. Since tropospheric curvature is driven primarily by natural variability all four inversions display identical measures of curvature there.

Another measure of profile smoothness is given by the mixing ratio standard deviation. Figure 4.48 displays the zonal mean standard deviation of the January 1990 observations for each of the inversion algorithms and includes variability due to random noise and natural variability. Natural variability dominates in the troposphere and also increases the stratospheric standard deviation, reducing differences between inversions slightly. The onion peel standard deviation is always greater than the others but the Mill-Drayson standard deviations are nearly identical to the operational and smoothed Mill-Drayson. Estimated random errors from the formal error analysis that includes all error terms, shown in Fig. 4.49, agree quite well with the observational data but display a greater spread in value between inversions since natural variability is not included.

The standard deviations of the Monte Carlo random error simulations display a spread in standard deviation values between inversions similar to Fig. 4.49 except for the Mill-Drayson stratospheric deviations which fall somewhat closer to the operational and smoothed Mill-Drayson than predicted by the formal error analysis.

Overall, the characteristics of the inverted profiles, individually and in the mean, agree quite well with the formal error analysis and Monte Carlo simulations.

4.6 Section Conclusions

The error budget for the SAGE II water vapor measurements was examined in detail following Rodgers's formal error analysis. Unlike many remote sensing experiments, which dedicate one or more channels to a single atmospheric species that is the dominant contributor to the measured channel signals, many of the

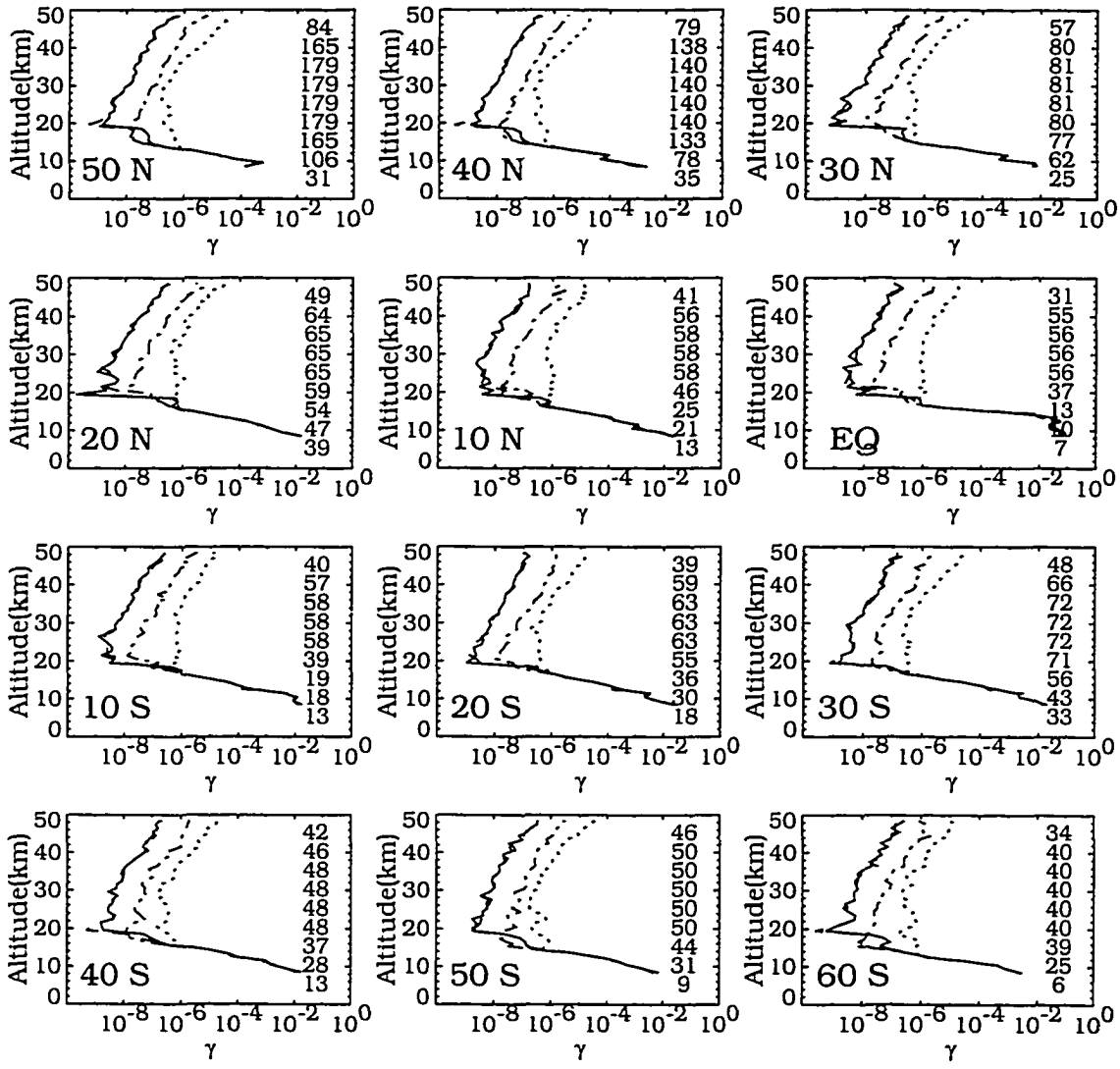


Figure 4.47: Water vapor curvature indicator γ calculated from the operational (solid line), onion peel (dotted line), Mill-Drayson (dash-dot line), and smoothed Mill-Drayson (dashed line) inversions of the January 1990 observational data. The number of points included in the zonal statistics are indicated on the right by altitude level.

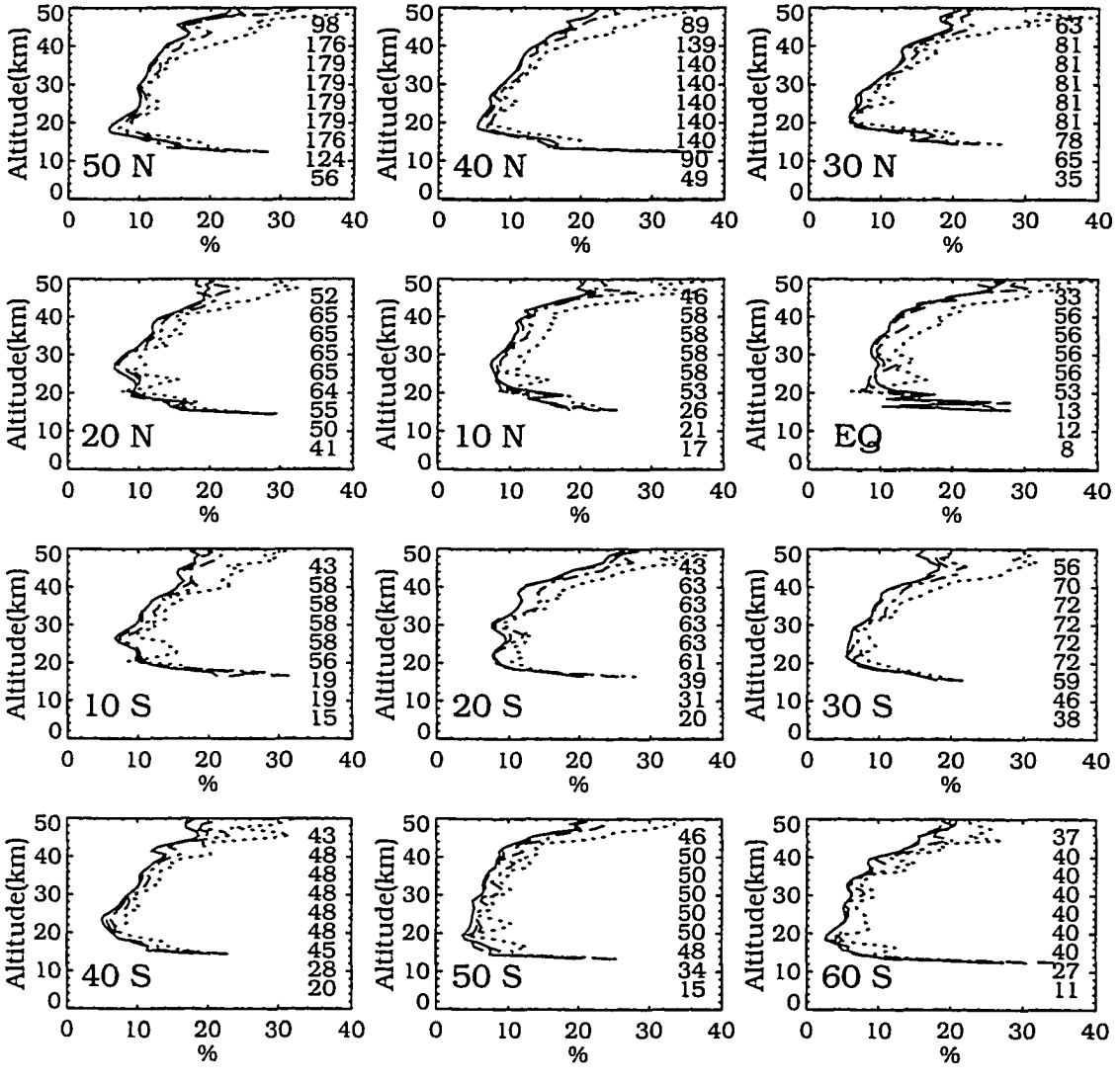


Figure 4.48: Zonal mean water vapor standard deviation, expressed as a percentage of the zonal mean, for the operational (solid line), onion peel (dotted line), Mill-Drayson (dash-dot line), and smoothed Mill-Drayson (dashed line) inversions of the January 1990 observational data. The number of points included in the zonal statistics are indicated on the right by altitude level.

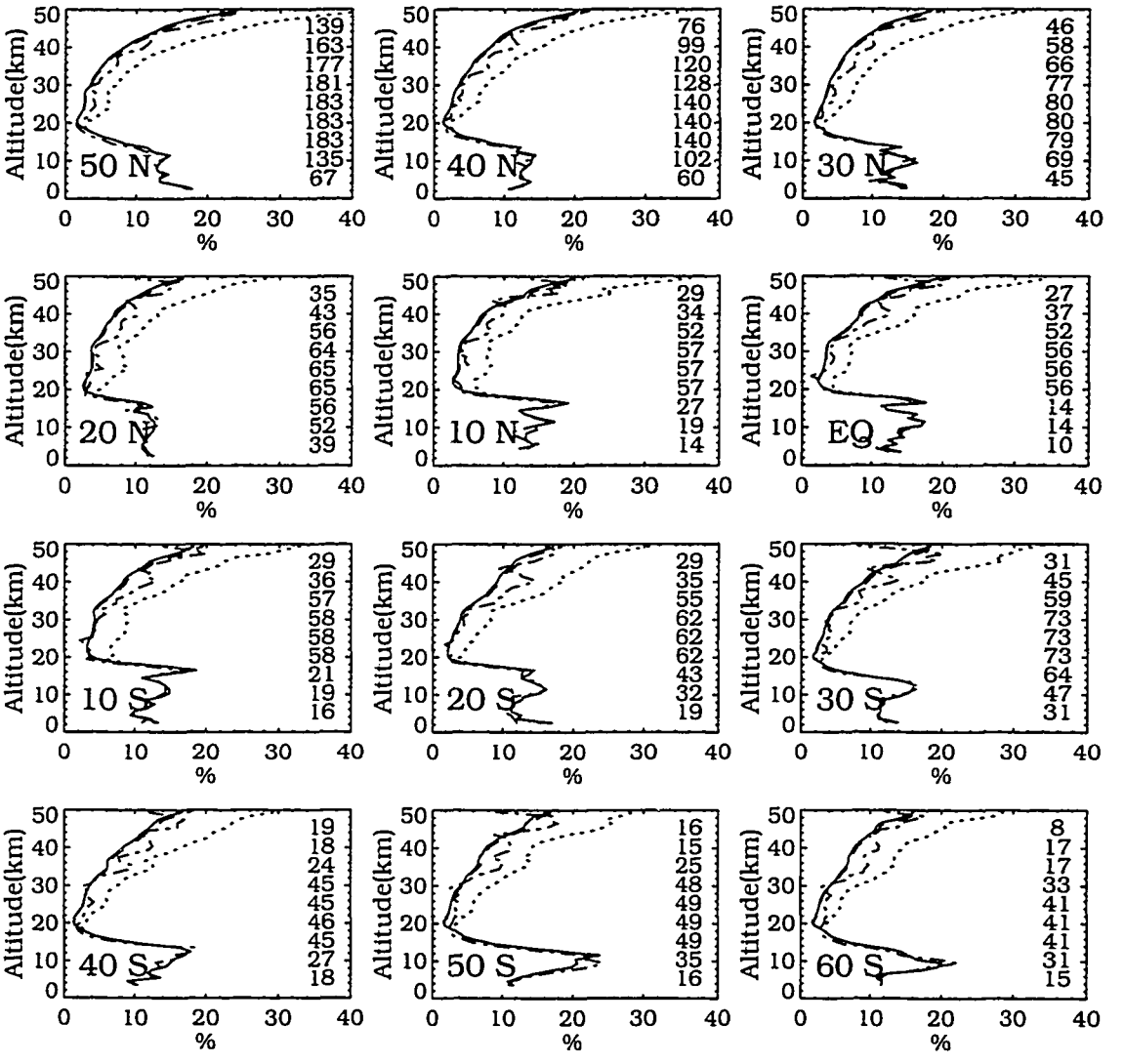


Figure 4.49: Estimated zonal mean water vapor standard deviation, expressed as a percentage of the zonal mean, for the operational (solid line), onion peel (dotted line), Mill-Drayson (dash-dot line), and smoothed Mill-Drayson (dashed line) inversions of the January 1990 formal error analysis. The number of points included in the zonal statistics are indicated on the right by altitude level.

SAGE II channels include contributions from two or more species whose importance varies with altitude and of course, all SAGE channels contain Rayleigh contributions. To account for this complex situation, the error method was refined for the water vapor channel by subdividing the inverse model into two models, a species separation model and an inversion algorithm model. This clarified the roles of certain errors and more importantly, permits random error growth tracking through the separation model. Much of the error budget development effort is concerned with error term formulations, calculating $\mathbf{K_b}$, ϵ_b, $\mathbf{S}_{\epsilon_b}$, and $\mathbf{S}_{\epsilon_y}$ for each of the latitude zones on a monthly basis with a smaller amount of time spent calculating $\mathbf{D_y}$ for each of the inversion algorithms studied and constructing error equation terms. Thus, after developing the operational inversion error budget, the process may be repeated for new or modified algorithms with little additional effort.

The important model parameters were identified and $\mathbf{K_b}$ was calculated for each. Some error terms were reformulated from those used previously in Chu et al.(1993) and several new parameters were included in the analysis. Although the estimated error levels shifted slightly in this work relative to the previous levels, the importance of each term in the error budget did not change. In this work, treating temperature as a distinct parameter instead of being folded into the Rayleigh separation error leads to a Rayleigh error term that is only systematic. The aerosol random error term was reformulated to more closely follow the empirical expression of Thomason(1991) and increases the estimated random error slightly in the stratosphere since the channel 4 random measurement noise is included.

The new uncertainties are related to the radiative transfer modeling in the inversion algorithm. The approximation of replacing the line-by-line calculations was shown to produce a small uncertainty in the inferred tropospheric water vapor. Other radiative transfer issues were also examined but not included in the error budget. For the water vapor continuum the issue is not the uncertainty in the continuum

modeling but the error caused by not including it in the EGA database. The error becomes quite significant in the lower troposphere as shown by Fig. 4.8. Also, neglecting water vapor absorption (lines and continuum) in channels 1 and 3 results in a 10% error at the surface. Because the operational EGA database, which was used to produce the archived water vapor data product, was calculated with HITRAN 82 we used the uncertainties for this version of HITRAN in the error budget. The band strength and halfwidth systematic errors would be reduced by half if HITRAN 92 were used but this would cause systematic shifts in the inferred water vapor relative to the archived data. Figure 4.8, panel (f), suggests that the archived data may be too low by up to 14% in the stratosphere and too high by up to 23% in the troposphere if HITRAN 92, continuum, and water vapor absorption in channels 1-3 represent atmospheric "truth". The stratospheric bias is caused by the difference in band strengths while the tropospheric bias is due to several causes; half width assumptions made by the radiative transfer code used to calculate the operational EGA data base, the difference in band strengths, water vapor absorption (lines and continuum) in channels 1 and 3, and most importantly the continuum in channel 2. However, it appears that the tropospheric bias may be compensated to a large extent by the systematic model error, leaving only a slight bias in the archived data.

The temperature and reference altitude model parameters presented conceptual difficulties as to the proper way to account for them in the analysis. The $\mathbf{D_y}$ and $\mathbf{K_b}$ matrices define the sensitivity of channel 2 variables. Perturbations in these two parameters will cause perturbations in the other channels and feedback into the water vapor channel through the separation model will occur. To account for this we calculated $\mathbf{A_b}$ directly instead of using $\mathbf{D_y K_b}$. Fortunately, the feedback is negative so the size of the perturbations in $\mathbf{y}^{mi}$ are significantly reduced relative to $\mathbf{y}^{m}$. As long as the perturbations appear simultaneously in all the channels and act identically, it appears that the estimates of $\Delta\delta_{2-1}^{Ozone}$ and $\Delta\delta_{2-1}^{Aerosol}$ from the separation

process also include the perturbations. For the reference altitude estimated error the end result is probably close to what could be achieved by calculating $\mathbf{K_b}$ assuming the reference atmosphere contains only water and using $\frac{\partial IA}{\partial y^{mi}}$ (identical to $\mathbf{D_y}$ for a water only atmosphere). However, for temperature, the situation is more complex because it appears in several places throughout the experiment.

The development in Sec. 4.3.3 provides a general approach to incorporating systematic and random uncertainties caused by the iterative solution method into the formal error analysis. The iterative solution is controlled by two inversion model parameters, the first guess profile and $\Delta_{\mathbf{cl}}$. For all inversion algorithms except the Mill-Drayson, it is shown that selecting $\Delta_{\mathbf{cl}} < |\epsilon_{\mathbf{s}}|$ can reduce the systematic component of error without increasing the random component. Making the same choice for the Mill-Drayson inversion algorithm increases the random component because the effectiveness of the implicit smoothing is reduced.

One of the reasons for applying the formal error analysis to the SAGE II water vapor measurements is the ability of the method to account for interlevel error correlations. When this work began, it was known that some of the channels contained correlated errors in the stratosphere that are quite apparent in the short wavelength unratioed channels. The correlations weaken as one moves toward the longer wavelength channels and are further reduced if channel ratios are inverted. Although the correlated errors in the water vapor signal are weaker than those in either channel 1 or 2, they are nonetheless still present in $\mathbf{y}^m$ and should be included in $\mathbf{S}_{\epsilon\mathbf{y}}$. Calculating $\mathbf{S}_{\epsilon\mathbf{y}}$ directly from $\mathbf{y}^m$ proved difficult due to geophysical variations in $\mathbf{y}^m$ so $\mathbf{y}^{mi}$ was examined instead. The intermediate measurement, $\mathbf{y}^{mi}$, still contains some geophysical variations due to water vapor but since $\mathbf{y}^{mi}$ is log-linear in the stratosphere it was possible to remove the variation by vertical piecewise least squares fitting of $\mathbf{y}^{mi}$. Analysis of the fitted residuals showed that the separation process removed the correlated errors, leaving the errors in $\mathbf{y}^{mi}$ independent. For

the error budget computations we assume the stratospheric errors in $\mathbf{y}^m$ are independent. In addition, the residual variances in the middle atmosphere were found to be significantly smaller than those estimated from the primary error sources for $\mathbf{y}^{mi}$. Correction factors were developed and applied to the variances of $\mathbf{S}_{\epsilon_y}$ (also to $\mathbf{S}_{\epsilon_b}$ for ozone and aerosol) which significantly reduced the estimated random error in the lower stratosphere relative to the previous results of Chu et al.(1993). The greater curvature of $\mathbf{y}^{mi}$ in the troposphere prevents vertical profile fitting but examination of $\mathbf{y}^m$ in that region suggests a slight error correlation may exist. This was incorporated in the "representative" correlation coefficient matrix shown in Fig. C.9 but the off diagonal correlations increase the estimated tropospheric random error slightly.

Estimates of the systematic and random null space error were developed from tropospheric radiosonde data and stratospheric SAGE II data for the operational inversion. Using the SAGE II monthly zonal mean of $\mathbf{x}^r$ for $\overline{\mathbf{x}}$ keeps $\overline{\mathbf{x}}$ close to the monthly zonal mean of $\mathbf{x}^t$ and the systematic error remains small. The size of the random error depends on two factors, the natural variability of $\mathbf{x}^t$ as expressed by $\mathbf{S}_{\mathbf{x}^t}$ and the form of the averaging kernels, which for the operational inversion is controlled by the absorption and mixing ratio smoothing. The tropospheric random error is large, peak values ranging from 10% to 25% are typical, and are primarily due to the natural variability of water there. Stratospheric variability is much reduced and the smoothing is stronger (peak values of $\mathbf{A}$ are half tropospheric values). The random error ranges from 3-10% between 18-21 km and increases to 15-22% at 50 km.

The systematic errors, in order of importance are: aerosol, ozone, band strength, and Lorentz halfwidth. The remaining errors contribute less than 2% individually to the total estimated systematic errors. The vertical variation and overall level of the systematic errors are nearly identical for all five inversion algorithms examined here but the small scale variations reflect the degree of smoothing. The model error, shown to be significant in the troposphere, was not included in the total error because

it is a deterministic error that can be estimated to first order for each profile and used to adjust $\mathbf{x}^r$.

Tropospheric random errors are dominated by the measurement, reference altitude, and aerosol terms and are identical for all five inversions since the averaging kernels are also identical. In the lower stratosphere the measurement, reference altitude, and aerosol errors dominate while at higher altitudes the temperature error increases but not as fast as the measurement error. The broad row and column functions composing the temperature $\mathbf{K_b}$ matrix make the temperature error insensitive to changes in the amount of smoothing between inversion algorithms. For the other error terms, error levels follow the degree of smoothing. The Mill-Drayson algorithm is an exception to this. Error levels on a term by term basis are identical to the onion peel error terms because $\mathbf{A}$ is the same, but the implicit smoothing term from the iterative solution algorithm reduces the total random error. Stratospheric total random errors from the onion peel and SGLS (case 1:3) algorithms are nearly identical and largest of the inversions studied in this work. The operational and smoothed Mill-Drayson inversion algorithms have the the lowest total error and the regular Mill-Drayson error is only slightly larger.

All of the inversion algorithms, except for the SGLS, were used to invert SAGE II data covering years 1987 and 1990. Error characteristics of the inverted profiles were compared to estimated errors from the formal error analysis and to a Monte Carlo error analysis. A comparison of monthly zonal means between inversion types found systematic differences somewhat larger than expected from the analysis in Sec. 4.3.3 or the Monte Carlo analysis. This was traced to the choice of $\Delta_{\mathbf{cl}} = |\epsilon_s|$ which caused systematic differences in ϵ_{tc} and to some equatorial profiles that differed radically from the corresponding zonal mean. An estimate of profile smoothness applied to the inverted profile data confirmed the relative degree of stratospheric smoothing expected from the averaging kernel for each inversion algorithm and from the Monte

Carlo analysis. A comparison of profile standard deviations found similar results. The inversion algorithms studied here, when applied to actual data, produce mixing ratio profiles with error characteristics that closely follow those expected from the formal error analysis.

Two methods were used to obtain the iterative solution to $|\overline{A}^m - \overline{A}^c(U^n)| < \Delta_{cl}$, Chahine's predictor and the secant method. The secant method converges faster, typically within three or four iterations, but occasionally predicts negative mixing ratios that must be replaced with more realistic values. Chahine's predictor sometimes failed to converge in the troposphere within the allotted 20 iterative steps when the initial profile turned out to be a poor starting point. When the operational, onion peel, or smoothed Mill-Drayson used $\Delta_{cl} << \epsilon_s$ to decrease systematic errors in ϵ_{tc}, the secant method provided a solution in fewer iterative steps. For the Mill-Drayson inversion both methods were equally effective.

Although we have not done so, the inversion algorithm could be further modularized into a slant path absorption smoothing model, an iterative solution model, and a mixing ratio smoothing model. This viewpoint does provide additional insight into the differences in error characteristics –primarily random– between inversion algorithms and allows one to predict the error characteristics of other algorithms formed from these modules with a reasonable degree of success. For instance, the absorption and mixing ratio smoothing in the operational algorithm could be replaced by a single equivalent smoothing of either absorption or mixing ratio.

All of the inversions examined in this study are suitable candidates for inverting the SAGE II water vapor data. Based on the Monte Carlo error studies and from application to two years of observational data it appears that the inversion algorithms are extremely stable, at least for the level of measurement noise found in the SAGE II data. One exception to this was the SGLS 1:4 case which exhibited

convergence problems that were likely related to the computational formulation and which could probably be fixed with some additional work.

Each of the inversions contain profile smoothing in some form that are experimenter controllable to varying degrees. The slant path absorption smoothing, common to all algorithms, is easily controlled by the experimenter as is the mixing ratio smoothing in the operational and smoothed Mill-Drayson algorithms. The implicit smoothing of the Mill-Drayson algorithm may be controlled by Δ_{cl}. Selecting $\Delta_{cl} << \epsilon_s$ makes the algorithm act more like the onion peel algorithm. In this situation, systematic errors in ϵ_{tc} decrease but the total random error increases. The smoothing of the SGLS algorithm is probably the most difficult to adjust if smoothing intermediate to the integral cases (1:2,1:3) is desired because of the additional programming complexity.

In the past, the amount of smoothing was driven by the desire to reduce the random errors below some acceptable level without considering the null space or model error, both of which increase with increasing smoothing. If the gradient of $\mathbf{x}^r$ is minimal, such as for water vapor in the stratosphere, the model error is insensitive to the level of smoothing but for strong gradients the model error can become significant and sensitive to smoothing, as is the case for water vapor in the troposphere. The tradeoff between the total random error and null space depends on the relative size (diagonal elements) of $\mathbf{S}_N$ to $\mathbf{S}_B + \mathbf{S}_E$ which for the operational algorithm is given by Figs. 4.14 and 4.25, respectively. When $\mathbf{S}_N \approx \mathbf{S}_B + \mathbf{S}_E$, $\mathbf{S}_N + \mathbf{S}_B + \mathbf{S}_E$ is nearly independent of the amount of smoothing and the figures show that this holds true in the troposphere and upper stratosphere. In the lower stratosphere, however, $\mathbf{S}_N$ is greater than $\mathbf{S}_B + \mathbf{S}_E$ by factors of 2 to 3 and $\mathbf{S}_N + \mathbf{S}_B + \mathbf{S}_E$ could be reduced by decreasing the smoothing strength.

We have shown that the total systematic and random errors are not materially affected by the choice of the Chahine or secant predictor to perform the iterative solution. We have also shown that the systematic error for all algorithms is generally the same and that the random error level can be controlled by the amount of smoothing. One might ask, if the smoothing is adjusted in each of the algorithms to produce identical random error levels, what are the relative merits of each algorithm? To achieve a given error level, the averaging kernels will have approximately the same peak values for all the algorithms except for the Mill-Drayson algorithm in the stratosphere, where implicit smoothing would allow for reduced explicit smoothing. Thus, the stratospheric model and null space errors would be smaller for the Mill-Drayson algorithm relative to the other algorithms but would be identical in the troposphere. The remaining algorithm differences are not error related but concern execution time.

The onion peel algorithm required the shortest execution time. The operational, Mill-Drayson smoothed, and Mill-Drayson algorithms took more time than the onion peel because of the multiple vertical passes through the atmosphere needed to satisfy the convergence criteria. The SGLS, because of the non-linear least squares iterative solution process, required much more time than any of the others. Execution times based on inverting 100 profiles were determined for the overall inversion (includes setup overhead, file manipulations and inversion of other species) and for the inversion algorithm by itself. The overall execution time was 1.75 times the onion peel for the operational and smoothed Mill-Drayson, 1.85 times for the Mill-Drayson, and 11.5 times for the SGLS 1:3 case. Corresponding ratios were 7,7.6, and 80 for execution times consumed within the algorithms. No attempt was made to optimize the algorithms except for the SGLS where slant path absorptions were calculated only for tangent altitudes affected by the mixing ratio being perturbed. The convergence error limit, Δ_{cl}, was set to, $|\epsilon_s|$, for all but the SGLS which iterated to three

significant digits and Chahine's predictor was used for the operational and Mill-Drayson smoothed algorithms. Convergence to three significant digits, with further adjustments to appear in the fourth digit, implies the mixing ratio is known to 0.1% at all altitude levels. Relaxing the convergence criteria to two significant digits, or 1%, results in an iterative solution error comparable to the other algorithms. Execution times are still 5.8 and 37 times greater than the onion peel for the overall and inversion algorithm, respectively.

The implicit smoothing of the Mill-Drayson algorithm was perhaps the greatest surprise from the formal error analysis. For it to be effective requires a good profile shape match between the true and initial mixing ratio profile and a constant or upwards increasing random error. As an extreme example, assume that the profile shapes are identical and that the relative random error smoothly increases upwards. Then, if the estimated random mixing ratio error is 5% at 20 km and 10% at 30 km, the actual mixing ratio standard deviation will be 5% at 30 km and no iterative adjustment will ever be made at 30 km because the convergence criteria is always satisfied. The reduction in random error from the estimated 10% to the actual 5% is described by what we term implicit smoothing. In general, the random error at any level is given by the error at which the last iterative adjustment was made. If the estimated random error suddenly drops to 2% at 30 km, many of the predicted mixing ratios from the last iterative solution will fall outside the 2% error limit and the solution would require additional iterations. At levels where iterative adjustments are made, the estimated and actual random errors are equal and revert to the onion peel error. Mismatched profile gradients between the true and initial can also trigger iterative adjustments. Significant gradient mismatch is more likely to occur in the troposphere where the gradients are large for water vapor. This and decreasing random error from 12 to 20 km forces iterative adjustments on each level in the troposphere.

When the Mill-Drayson inversion algorithm is applied to observational data with varying profile shapes and changing measurement errors, the adjustment levels will also change from profile to profile. As the number of iterative adjustments increases on each level in a group of profiles, the actual random error approaches the onion peel error. Unlike the explicit absorption or mixing ratio vertical smoothing, which forces random error reduction, there is no guarantee that the level of implicit smoothing present in a group of profiles will be found in a single profile. We also note that increasing the absorption vertical smoothing increases the likelihood of adjustment due to gradient mismatch and likely decreases the implicit smoothing strength. We have not investigated the tradeoff between explicit and implicit smoothing in this situation.

The smoothed Mill-Drayson and SGLS inversion algorithms initially appeared to be reasonable variations of existing algorithms that might offer some advantages. However, the explicit mixing ratio smoothing dominates the implicit smoothing in the smoothed Mill-Drayson to yield inversion errors identical in size and character to the operational algorithm. Although this variation of the SGLS algorithm performed as expected, it offers no advantages compared to the other algorithms and required an inordinate amount of execution time.

SECTION 5

EXPERIMENT VALIDATION

The final step in any measurement experiment is reconciliation of the measured data set with other independent measurements. The error analysis, even though it may have been carried out meticulously, will not identify situations where knowledge of the underlying physical processes are incomplete. Not only will the measurements be incorrect in some unknown manner but this uncertainty will not be reflected in the error estimate. The ideal comparison data set would consist of data collocated in space and time obtained with an instrument providing similar horizontal and vertical resolution but based on an entirely different observational technique. Only occasionally does this situation occur and invariably some compromise is necessary for any comparison to take place.

This section is concerned with validating the inversion results of Sec. 4 through comparison to other independent measurements of water vapor in the troposphere. The majority of water vapor observations in the troposphere have been obtained with the global radiosonde network. The large number of reports provided by the network each day across the globe presents numerous opportunities for validation comparisons. The high data density and extensive historical database make the radiosonde hygrometers a logical choice for validation studies. Radiosonde hygrometer errors and operational limitations are evaluated in Sec. 5.1. This information will be used to construct a high quality radiosonde data set for use in Sec. 5.3. Next, climatological comparisons are made in Sec. 5.2 and then profile comparisons are performed following relatively strict temporal and spatial criterion in Sec. 5.3. A technique to increase the number and quality of correlative profile pairs by

incorporating information from air mass isentropic trajectories is developed in Sec. 5.4 and applied to correlative radiosonde observations in the remaining sections.

5.1 Upper Air Radiosonde Observations

The measurement of tropospheric humidity has been carried out routinely for many years by the upper air radiosonde network. Several different types of simple, inexpensive hygrometers with accuracy sufficient for synoptic meteorology are now flown on the radiosonde balloon packages. Measurements can be made in clear and cloudy skies. Observations are made at 0, 6, 12, and 18z, with the majority taken at 0z and 12z for most stations. Although overall coordination of the upper air network is provided by the World Meteorological Organization, individual countries select the type of sensor to fly and how to analyze and report the observations. The sensors are updated periodically as improved instrumentation becomes available and individual stations oftentimes alternate sensor types. Thus, the radiosonde relative humidity observations compose a data set that is not entirely homogeneous. This not only complicates the analysis for this work but reduces the usefulness of the data set for climate studies (Elliot and Gaffen, 1991).

Monthly upper air data tapes covering 1987 obtained from the National Climate Data Center (NCDC) in Asheville, N.C., were scanned for active stations. Three categories (mandatory, significant and tropopause) in each upper air land station report were decoded and retained if temperature, dewpoint depression and geopotential height were available together on two or more of the first eight mandatory pressure levels (1000, 850, 700, 500, 400, 300, 250 and 200 mb). Data on significant levels above 50 mb and tropopause data missing either temperature or dewpoint depression were deleted. Altogether, 1155 stations showed activity ranging from 1 to almost 1300 reports in 1987, only 928 stations had 10 or more reports.

The station locations are shown in Fig. 5.1. Station density is greatest over the northern hemispheric land masses and minimal over the oceans. Table 5.1, which follows the nomenclature of NMC Office Note 29 (NMC Format for Observational Data, revised 4/30/87), lists the many different radiosondes in use in 1987. We experienced some difficulty identifying the hygrometer associated with each radiosonde package in 1987. The last column in Table 5.1 lists our identifications based on sometimes conflicting information found in the published literature and WMO reports augmented with information gathered as part of the project to track historical changes in radiosonde instruments and practices (Gaffen, 1993). Some confusion arises from the fact that while a radiosonde is manufactured in a particular country, it may be deployed anywhere globally, and the country of manufacture may be using an entirely different radiosonde. For example, the Mesural radiosonde manufactured in France is identified as #5 when deployed in France and #15 for other locations. In 1987 however, several French stations started flying Vaisala radiosondes (Gaffen, 1993) and reports from these stations were given the #14 identifier. Figure 5.2, which displays frequency of hygrometer use, broken down by latitude and by sensor type, reflects the high northern hemispheric station density and the tendency of some instrument types to dominate particular latitude bands. The greatest number of reports come from the USSR sensors(20, 21), followed by the Chinese(19) and then the US(1). The Vaisala(14, 4) is also used extensively worldwide.

Radiosonde hygrometer accuracy and precision are typically addressed through laboratory testing and field trials. Laboratory testing under controlled conditions provides additional information on hysteresis effects and sensor response times but field trials where two or more radiosondes (with the same or different hygrometers) are flown on the same balloon give results that are more representative of operational conditions. A comprehensive review of hygrometer measurement quality is beyond the scope of this study; here we present the measurement uncertainty assigned to

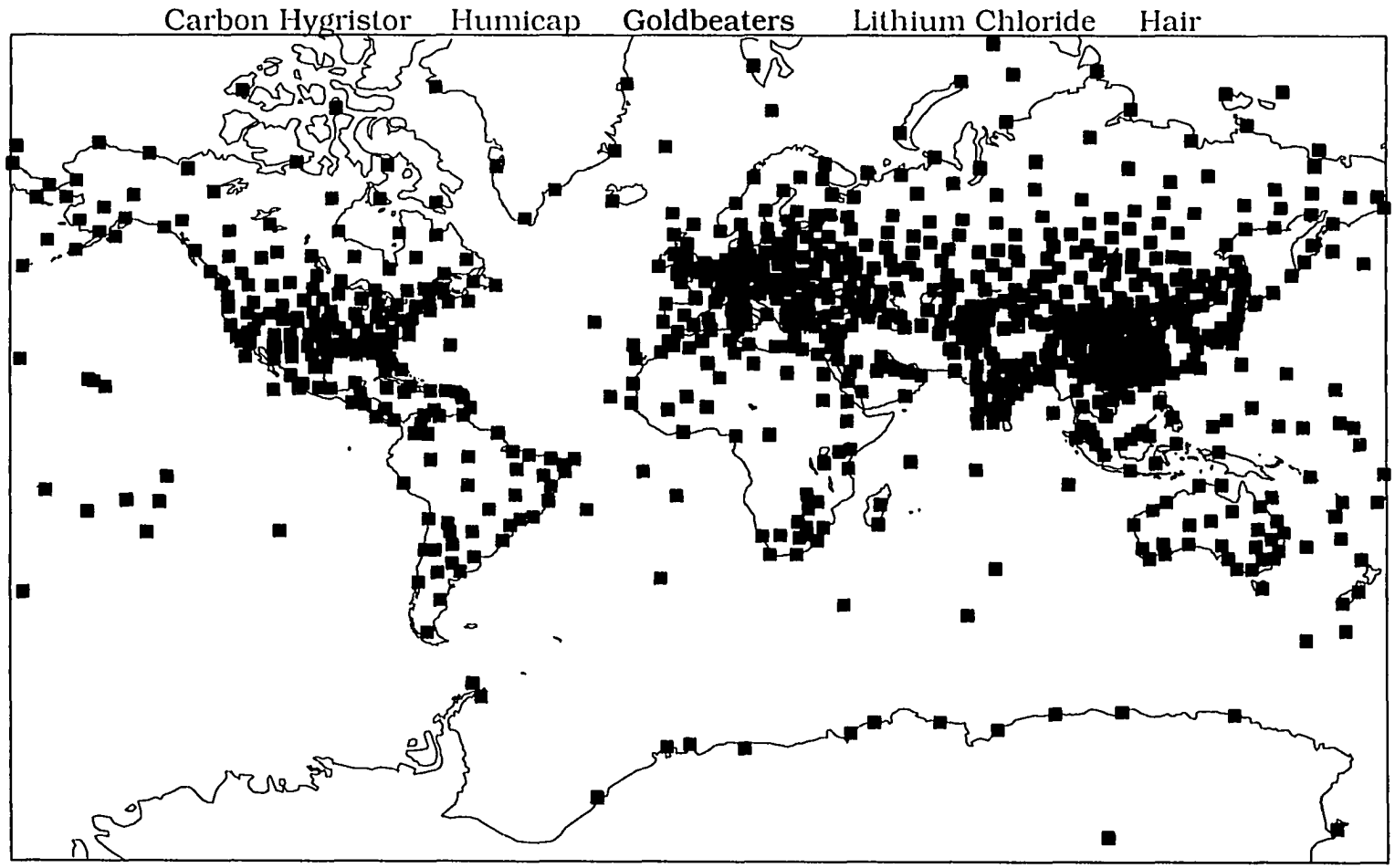


Figure 5.1: Locations of active radiosonde stations in 1987 by hygrometer type. Stations with fewer than 10 reports in 1987 are not shown.

Table 5.1: Hygrometer Types

NMC Office Note 29 Identifier	Radiosonde	Hygrometer
1	US-NOAA VIZ	Carbon Hygristor
3	US-VIZ AN/AMT-4B	Carbon Hygristor
4	Finland-Vaisala	Thin Film Capacitive
5	France-Mesural	Goldbeaters Skin
7	W.Germany-Graw/Sprenger	Hair
9	Japan-Meisei or Oki	Carbon Strip
11	UK-Met. Office	Goldbeaters Skin
12	USSR-Meteorite A-22 III and IV	Goldbeaters Skin
13	US-NOAA VIZ 403 Mhz	Carbon Hygristor
14	Vaisala	Thin Film Capacitive
15	Mesural	Goldbeaters Skin
16	Australia-Phillips	Carbon Hygristor
18	Canada-Sangamo	Carbon Hygristor
19	China-Shanghai	Goldbeaters Skin
20	USSR-RKZ-2 and -5	Goldbeaters Skin
21	USSR-Unknown	Goldbeaters Skin
22	India-Met.Service	Lithium Chloride
24	N.Korea-Jinyang(VIZ)	Carbon Hygristor
25	Switzerland-Met. Lab.	Goldbeaters Skin
26	Czechoslovakia-Vinohrady	Goldbeaters Skin
27	US-NOAA VIZ 1680 Mhz	Carbon Hygristor

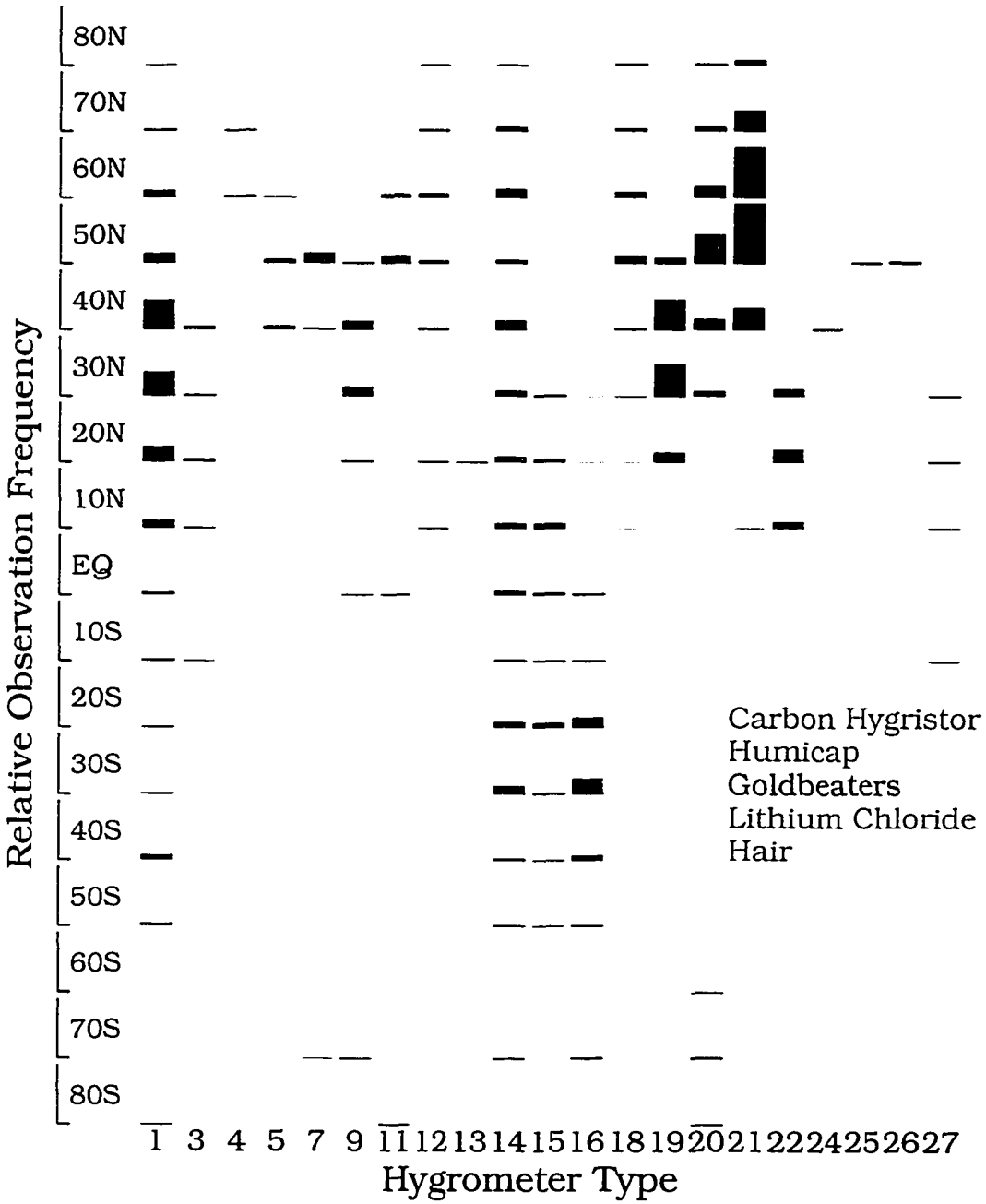


Figure 5.2: Frequency of radiosonde observations resolved by hygrometer type and latitude band for 1987.

each hygrometer type for this work and in a following section discuss some of the systematic biases uncovered by radiosonde intercomparisons between hygrometers.

The VIZ radiosonde, employing the carbon hygristor, has a stated (private communication, M.Friedman, VIZ, 1990) relative humidity accuracy of ± 7% from 0-20%, ±5% from 20-90% and ± 7% from 90-100%. The precision is stated to be ± 3% (rms). Although the U.S. National Weather Service reporting practice (Radiosonde Observations, 1981) does not report humidity for ambient temperatures below -40° C and records dewpoint depression as 30° C for relative humidities below 20%, other users of this hygrometer(for example, Canadian stations) oftentimes report humidity at colder and dryer conditions as will be shown later. In a review of the VIZ radiosonde, Pratt(1985) notes that for temperatures above -30° C the response time of the hygrometer element convolved with typical balloon ascent rates gives an approximate vertical resolution of 0.15 km, this increases at colder temperatures due to the slower sensor response time.

The accuracy of the Vaisala Humicap, a thin film capacitive sensor, is stated by the manufacturer (private communication, K. Goss, Vaisala, 1990) to be 2-3% over 0-100% relative humidity at ambient temperatures down to -90° C with a precision of 2%. Numerous twin flights made at Vaisala in November 1982 (Antikainen and Hyvonen, 1983) support the reported level of precision. Because of this sonde's wide operating range, many users report relative humidity under all conditions. The response time of the Humicap is thought to be as good as or better than the carbon hygristor (Schmidlin, 1988).

For the remaining three hygrometers we develop estimated accuracies from a variety of contemporary and historical sources. The SONDEX 1981 intercomparison (Richner and Philips, 1982) lists an accuracy of ±10% over a measurement range of 10-100% relative humidity for goldbeaters skin. Muller (1965), in his review of

organic hygrometers, notes that the increased lag time at low temperatures makes this sensor unsuitable for radiosonde use at temperatures below -20° C. Wexler (1970) shows an uncertainty of ±3% that increases below 0° C to ± 10% at -40° C covering a range of 2-100% for rolled and treated hair. The rolled and treated hair is thought to respond adequately down to 20% relative humidity in stratospheric conditions and becomes nonresponsive at temperatures below -55° C (Muller,1965). The response time of rolled and treated hair is considerably faster than goldbeaters skin (Muller,1965). The lithium chloride hygrometer, developed as a replacement for the untreated hair hygrometer and used by the U.S. National Weather Service from 1943 to 1965, has a faster response time than the hair hygrometer but not as fast as the carbon hygristor. Its accuracy is approximately ±5% with a cutoff below 40% relative humidity at and below -40° C (Mathews, 1965a,b). The accuracies and operational limits cited here should be viewed solely as guides; slight changes in instrument configurations, manufacturing, operational procedures, and reporting practices can have a significant impact on the final data product.

The hygrometers in Table 5.1 are thought to respond to relative humidity with respect to water at all temperatures and standard practice is to analyze data from them assuming this type of response. Recent work partially explains this behavior as due to enhanced binding energy caused by a thin film of water over a polarized surface that reduces the vapor pressure (Anderson, 1995). We converted the dewpoint temperature depressions on the upper air tapes to mixing ratios assuming saturation over water rather than convert the SAGE II mixing ratios to relative humidity or dewpoint depression since it is not clear whether this conversion should be done over water or ice. We also assume, in our discussion of the results, that the radiosonde data is taken in ascent. Unfortunately, this information is not retained in the NCDC data.

Time series scatter plots of dewpoint depression, relative humidity, volume mixing ratio, and temperature on the first six mandatory levels clearly show the analysis and reporting procedures in use at each station when sufficient data exists to reveal a pattern. The predominant reporting procedures, which follow from the known and suspected operational limits of specific hygrometers, are summarized in Table 5.2. In many instances, relative humidity is reported as often as temperature, indicat-

Table 5.2: Summary of Reporting Procedures

Case	Description
A	No Deletions or Alterations
B	Data Insufficient to Establish Reporting Procedure
C	When R.H.[1] <20%, D.P.=30° C
D	When D.P.[2] >30° C, D.P.=30° C
E	When R.H.<10%, R.H.=10%
F	When R.H.<20%, R.H.=20%
G	Data Deleted When T<-40° C
H	Data Deleted When R.H.<20%
I	Data Deleted When R.H.<10%

1-Relative Humidity
2-Dewpoint Depression

ing that the particular station believes the hygrometer is producing acceptable data under all atmospheric conditions. In other instances we found stations that did not report relative humidity when the temperature dropped below -40° C or when the relative humidity fell below 10% or 20%. Of far greater concern to this work are those stations that insert plausible but artificial data when the hygrometer is thought to be operating in a questionable region. Relative humidities were found set to 10% or 20%, dewpoint depressions to 30° C and in a great many cases the dewpoint depression was set to 30° C only when the relative humidity fell below 20%. Combinations of the cases itemized in Table 5.2 were also found occasionally, such as G and H, or G and C. As one would expect from Table 5.2, most of the data deletions and alterations occur in the cold and dry conditions of the upper

troposphere. Inability to provide observations under these conditions introduces a moist bias of unknown size in radiosonde derived climatologies. Potentially, the bias may be significant, if one views the study of Elliot and Gaffen (1991) as indicative of the upper tropospheric situation. Taking Canadian hygrometer data at four stations and reanalyzing it with the more restrictive "US Rules", they found significant change (5% to 45%) in specific humidity at 500 mb for the 1989 cold season.

Table 5.3 summarizes the reporting procedures by hygrometer type for 1987. Note that data from the same type of hygrometer may be subjected to a number of different analysis procedures. Even data for the Vaisala, which is thought to maintain sensitivity in the upper troposphere, is interpreted several different ways. In the work that follows, we have removed artificial data whenever possible. In situations where stations alternated sonde types, the most restrictive screening standard was applied. However, reports from stations that observe infrequently may still contain some artificial data. The station time series in some cases also display a number of subtle biases such as a suspicious tendency for some quantities (dewpoint depression, relative humidity) to cluster around a certain level. No attempt was made to categorize or screen data from stations with these types of biases.

As a further check on data quality, we employed a method similar to that used by Oort(1983) in his development of global water vapor distributions from radiosonde observations. For each station, monthly means and standard deviations of temperature, geopotential height, and water vapor mixing ratio were calculated on the eight mandatory levels. The station means and standard deviations are refined by making a second pass through the data while excluding values that lie more than four standard deviations from the mean. The same exclusion test applied to individual reports will eliminate grossly incorrect values and has been incorporated in the analyses that follow.

Table 5.3: Frequency of Reporting Procedure by Hygrometer Type

NMC Office Note 29 Identifier	Cases A	B	C	D	E	F	G	H	I	Additional Screening of A and B
1	40	31	121	4	1	2	70	2	2	R.H.<20%
3	7	3	8	1	0	0	4	0	0	R.H.<20%
4	3	0	0	0	0	0	0	0	0	T<-65° C
5	0	0	0	0	0	1	8	0	0	
7	15	0	0	0	0	0	0	0	0	T<-40° C
9	13	0	0	0	0	0	9	1	0	T<-40° C
11	2	1	0	0	0	0	12	0	0	T<-40° C
12	38	1	0	0	0	0	1	0	0	T<-40° C
13	1	0	0	0	0	0	0	0	0	R.H.<20%
14	94	25	4	11	12	0	4	0	0	T<-65° C
15	10	6	1	0	8	0	3	0	1	T<-40° C
16	3	4	0	34	0	0	0	0	1	T<-40° C
18	35	3	0	0	0	0	3	0	0	T<-60° C
19	117	7	1	0	0	0	0	0	0	T<-40° C
20	61	0	1	0	1	0	132	0	0	T<-50° C
21	75	5	1	0	0	0	119	0	0	T<-50° C
22	30	3	0	0	0	0	0	1	0	R.H.<10%
24	1	0	0	0	1	0	0	0	0	R.H.<20%
25	1	0	0	0	0	0	0	0	0	
26	2	0	0	0	0	0	0	0	0	T<-40° C
27	1	0	6	0	0	0	0	0	0	R.H.<20%

Numbers do not add to 1155 because of multiple hygrometer use and combined analysis procedures in use at each station.

Subsequent use of the screened radiosonde data uncovered some remaining problems. Based on a first pass inspection of the profile pair comparisons, discussed in a following section, it was obvious that questionable data had bypassed the screening procedures. More commonly, at the higher altitudes the radiosonde profiles sometimes reverse and show increasing mixing ratios with altitude. This vertical variation is most likely caused by diminishing instrument sensitivity. The last column in Table 5.3 lists additional quality checks for removing suspect data applied to case A and B stations. These were developed from the profile pair comparisons and from the predominant screening case shown in Table 5.3 with some modifications. Although this improved the automatic screening results, it was decided that for the profile pair comparisons each radiosonde profile would be examined visually with suspect points flagged subsequently. We limited this final subjective screening to removing data above the inflection point marking the reversal from decreasing to increasing mixing ratios.

5.2 Climatological Comparison

One of the first steps we took in the validation process was a comparison of the SAGE II upper tropospheric climatology to the Global Atmospheric Circulation Statistics (GACS) developed by Oort(1983), which, in the upper troposphere, is based entirely on radiosonde observations. GACS zonal means were constructed using 10 years of observations, from 1963 to 1973 in both hemispheres, by first calculating monthly station averages over this time period, developing global gridded fields from the irregularly spaced stations and then averaging zonally along the grid. This procedure differs from the manner in which the SAGE II zonal means are developed, where all values in the latitude band for the month are averaged together, but both techniques should produce essentially the same distribution.

Because of the observational biases inherent with both techniques, we expected SAGE II to be systematically drier than the radiosonde climatology. This is indeed the case as can be seen in Fig. 5.3 where January 1987 SAGE II land/ocean monthly zonal mean mixing ratios are shown along with GACS and 1987 radiosonde land zonal means . Although the SAGE II data match the general shape of the GACS distribution, with maxima at all altitude levels located between 0°10° S, the SAGE II mixing ratios are roughly half of the GACS climatology. The close agreement between GACS and the radiosonde zonal means shows that the GACS climatology adequately portrays the January 1987 distribution. Remaining differences are likely due to interannual variations, instrumentation changes over the intervening 15 years, changes in station and observational density or possibly differences in screening and analysis procedures. When comparing SAGE II January 1987 to other years, 1986, 1988, 1989, and 1990, we again find that 1987 is fairly typical. Inspection of the seasonal variations at several latitude bands finds SAGE II mimicking the radiosonde seasonal variations with the systematic bias appearing as a consistent feature between the two data sets. Interpretation of these results is difficult and far from conclusive. The SAGE II climatology should be drier than the radiosondes because it is cloud free. The radiosondes, because they cannot consistently provide observations in the dry upper troposphere, produce a moist climatology relative to the true climatology. Thus, the observed difference could be geophysically based but slightly exaggerated. On the other hand, systematic instrument biases could still exist in either data set.

A number of factors specific to SAGE II that have the potential to introduce systematic errors were considered in an effort to explain these systematic differences. Some of these have since been eliminated as a result of ozone and aerosol correlative and validation studies. For instance, a SAGE II altitude registration error of approximately 1-2 km would explain much of the bias, but an error of this

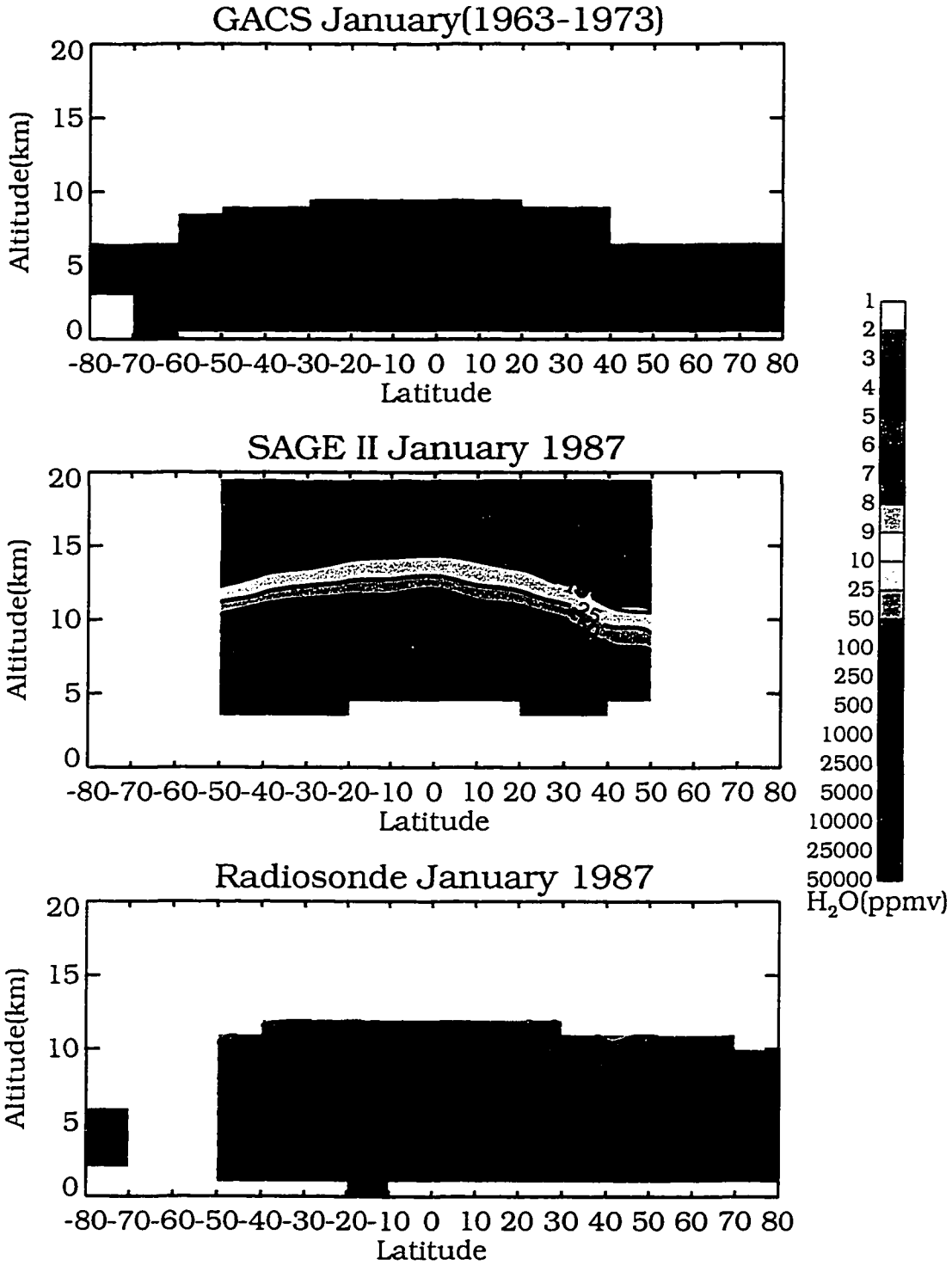


Figure 5.3: Monthly zonal mean water vapor mixing ratios(ppmv) for a) SAGE II (January 1987), b) Global Atmospheric Circulation Statistics (January 1963-1973) and c) radiosondes (January 1987). The radiosonde zonal mean contains data above 300 mb passing the screening criterion while 300 mb is the cutoff for the GACS zonal mean.

size is not supported by the ozone or aerosol correlative measurements and is much greater than the estimated altitude uncertainty of 0.2 km (Cunnold et al., 1989). In a similar manner, the correlative frost point hygrometer measurements show that certain uncertainties specific to the water vapor channel, such as line strength errors or channel bandwidth uncertainties, are probably not major contributors to the systematic differences. As shown in the previous section, there are altitude dependent sensitivities caused by uncertainties in air broadened halfwidths and temperature correction of the halfwidths that may come into play in the upper troposphere but again these are too small.

In the following section a more detailed comparison with the radiosondes is conducted in order to identify the cause of this large climatological difference. By developing correlative pairs of SAGE II and radiosonde measurements, observing the same air mass, we remove most of the observational sampling biases. Any remaining differences should be the result of systematic instrument biases.

5.3 Point Profile Matching

Useful correlative criterion must take into account the time and space variability of the gas. In the stratosphere, maximum spatial separations of 1000 km and time periods up to 8 hrs are acceptable criterion for ozone (Cunnold et al., 1989) and water vapor (Russell et al., 1984). However, convective activity in the troposphere in conjunction with strong vertical gradients of water vapor makes water vapor highly variable on small scales. Spatial correlations of radiosonde water vapor mixing ratios at DeBilt in the Netherlands (Van Maanen,1981) on the 500 mb level were found to have exponential decay with length scales of 480 km for winter and 330 km for summer. Kitchen(1989) found that a significant portion of the water vapor variance occurs on time scales less than 4 hrs in the troposphere. Taking

into consideration these scale sizes we limited the spatial separation to 250 km and temporal separation to ± 6 hours. This represents a compromise to some extent between the need to retain a significant number of profile pairs and the desire to minimize spatial and temporal variations. Even with this relatively tight criterion, we have not eliminated convective scale effects and are likely susceptible to mesoscale (200-300 km) variations. This is more of a concern for the lower troposphere than the upper troposphere. Vertical profile overlap was also required of each pair and multiple matches of radiosonde profiles to a single SAGE II profile were allowed. Of the approximately 10,200 SAGE II profiles and 559,000 radiosonde profiles, 814 pairs were found satisfying all screening criterion in 1987. The 814 correlative pairs thus consist for the most part of measurements taken in clear sky conditions over land due to the sampling biases for SAGE II and the radiosondes, respectively.

5.3.1 Correlative Pair Comparisons

All 814 correlative profile pairs were examined for general trends and features. To illustrate the general characteristics of the comparisons we selected a limited set of comparisons for presentation. The comparisons, chosen to push the operational limits of both instruments, focused on extremely dry conditions, those close to the GACS climatology, or those where SAGE II penetrates to the lower troposphere.

Figure 5.4 displays six comparisons between SAGE II and the VIZ carbon hygristor. In each comparison, SAGE II (solid line) and the radiosonde observations (plusses) are plotted as volume mixing ratio versus altitude. The error bars indicate the SAGE II estimated uncertainty (±1 standard deviation) due to random errors and estimated radiosonde accuracy (as discussed in a previous section). The monthly station averages are the open circles and the dashed line is the GACS climatology interpolated to the time and location of the measurement. The solid circles on the

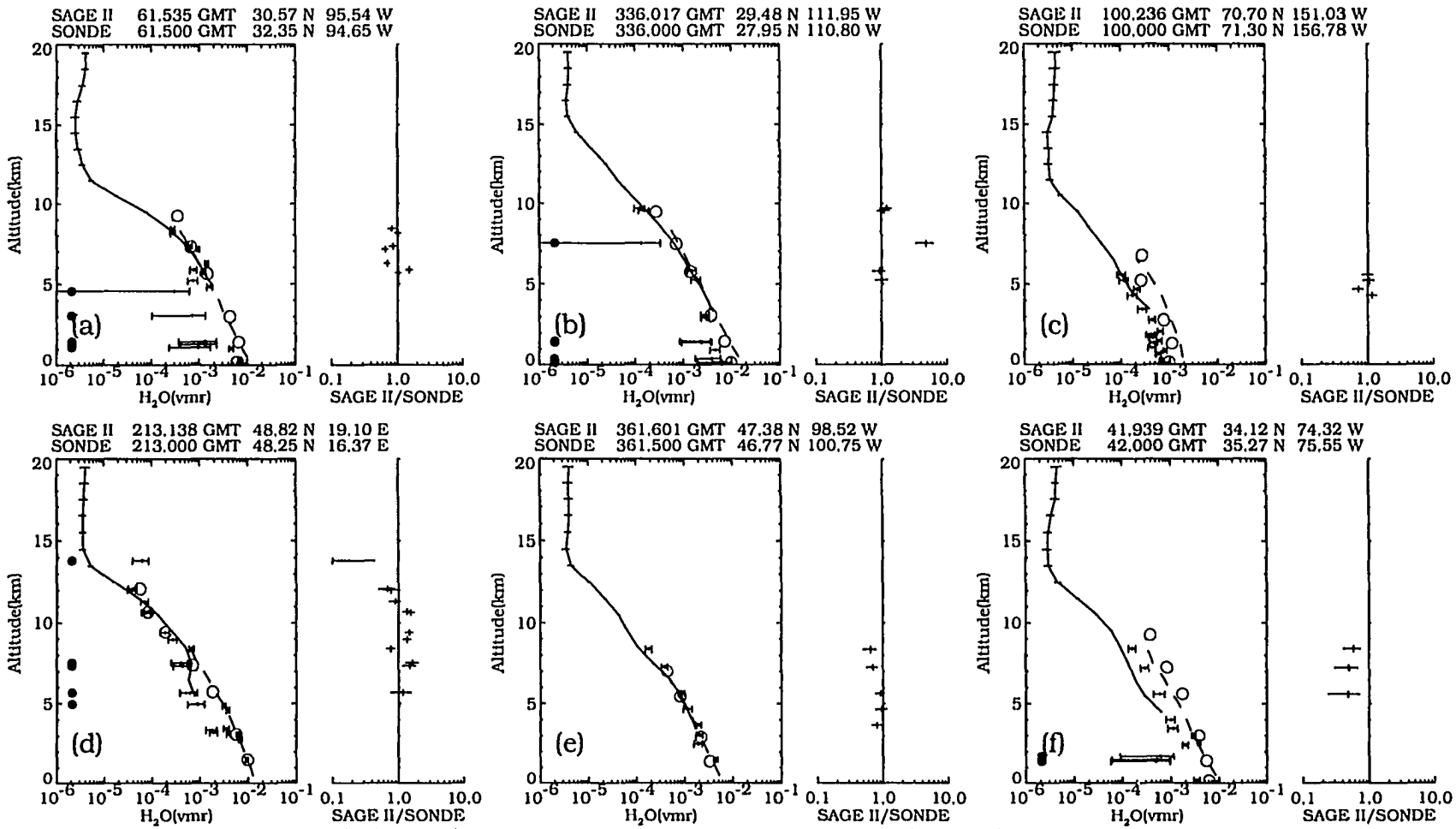


Figure 5.4: SAGE II (solid line) and the VIZ radiosonde observations (plusses) are plotted as volume mixing ratio versus altitude. The error bars indicate the SAGE II estimated uncertainty (random error, $\pm 1\sigma$) and estimated radiosonde accuracy. Open circles indicate monthly station averages and the dashed line is the GACS climatology interpolated to the measurement time and location. The solid circles on the left mark radiosonde observations that failed the screening criterion. On the right is shown the ratio of SAGE II to the radiosonde.

left mark radiosonde observations that failed either the automatic or visual screening quality checks. On the right is shown the ratio of the SAGE II measurement to the radiosonde. The spatial separation between the SAGE II subtangent point and radiosonde is 215.1, 204.1, 218.4, 210.7, 181.8, 170.1. km and the measurement time difference is 0.84, 0.40, 5.67, 3.32, 2.42, 1.47 hrs for correlative pairs a-f, respectively.

In the first comparison, Fig. 5.4a, good agreement is found over the 5.5-8.5 km altitude range with SAGE II generally within 35% of the radiosondes. A number of the radiosonde observations below 5 km failed the relative humidity criterion when they dropped below 20%, above this the relative humidity increased to the 30-80% level. Both sets of observations along with the monthly station averages are similar to GACS in the overlap region. Figure 5.4b, shows excellent agreement with the SAGE II profile overlapping the radiosonde observations, station averages, and climatology. The radiosonde relative humidity falls between 20-30% with a few values dropping below the 20% cutoff. This comparison validates the ability of SAGE II to measure water vapor in the lower troposphere, down to 3.5 km at average water vapor levels. Figure 5.4c displays a high latitude springtime comparison where both sets of measurements are considerably drier than the monthly station averages or even the GACS climatology. In this case SAGE II again measures water vapor in the lower troposphere but in drier than average conditions. The next comparison, Fig. 5.4d, was chosen because SAGE II does a reasonable job following the vertical variation appearing in the radiosonde observations between 5.5-9.5 km. In the strictest sense we should not be using the lower four radiosonde observations that failed the screening criterion but in this case they fall only slightly below the 20% relative humidity level. Figure 5.4e shows an excellent comparison that is similar to Fig. 5.4b. Finally, in Fig. 5.4f we show a comparison where SAGE II is much drier than the radiosondes, the station averages and the GACS climatology. This is a common situation that occurs to varying degrees with most radiosonde types.

Table 5.4: Multiple Radiosonde Matches
November 20 1987

Figure	GMT	Latitude	Longitude	Station	Type	Separation Km	Separation Time(hrs)
		SAGE II					
	324.66	50.2 N	4.59 E				
		Radiosonde					
11a	324.50	52.10 N	5.18 E	06260	14	213.7	-3.77
11b	324.46	50.80 N	4.35 E	06447	1	67.4	-4.77
11c	324.46	50.03 N	5.40 E	06476	14	61.2	-4.77
11d	324.46	48.77 N	2.02 E	07145	5	245.2	-4.77
11e	324.46	48.68 N	6.22 E	07180	5	207.1	-4.77
11f	324.46	51.40 N	6.97 E	10410	7	212.9	-4.77
11g	324.58	49.70 N	7.33 E	10618	7	204.0	-1.77
11h	324.75	52.10 N	5.18 E	06260	14	213.7	2.23

Correlative profile comparisons for other hygrometer types may be found in Larsen et al.(1993) that also show very good agreement in some cases, poor agreement in many others, and a tendency for SAGE II to report lower levels of water vapor than the radiosondes.

The high station density in the northern hemisphere provided a number of cases where a single SAGE II water vapor profile was paired with a series of radiosonde profiles. In Fig. 5.5, eight radiosonde soundings satisfying the correlative criterion in western Europe, show the extent that spatial and temporal variations, along with whatever instrumental differences that may exist, can influence the correlative comparisons. The observational details, listed in Table 5.4, show that four different hygrometer types were used and that one station repeated a sounding six hours later. The large spread in the radiosonde values bracket the SAGE II measurements with differences at times greater than the estimated uncertainties of both instruments. The shift to drier conditions at the station with two soundings (Fig. 5.5a,11h) is indicative of the changes occurring over periods of six hours and in this case may

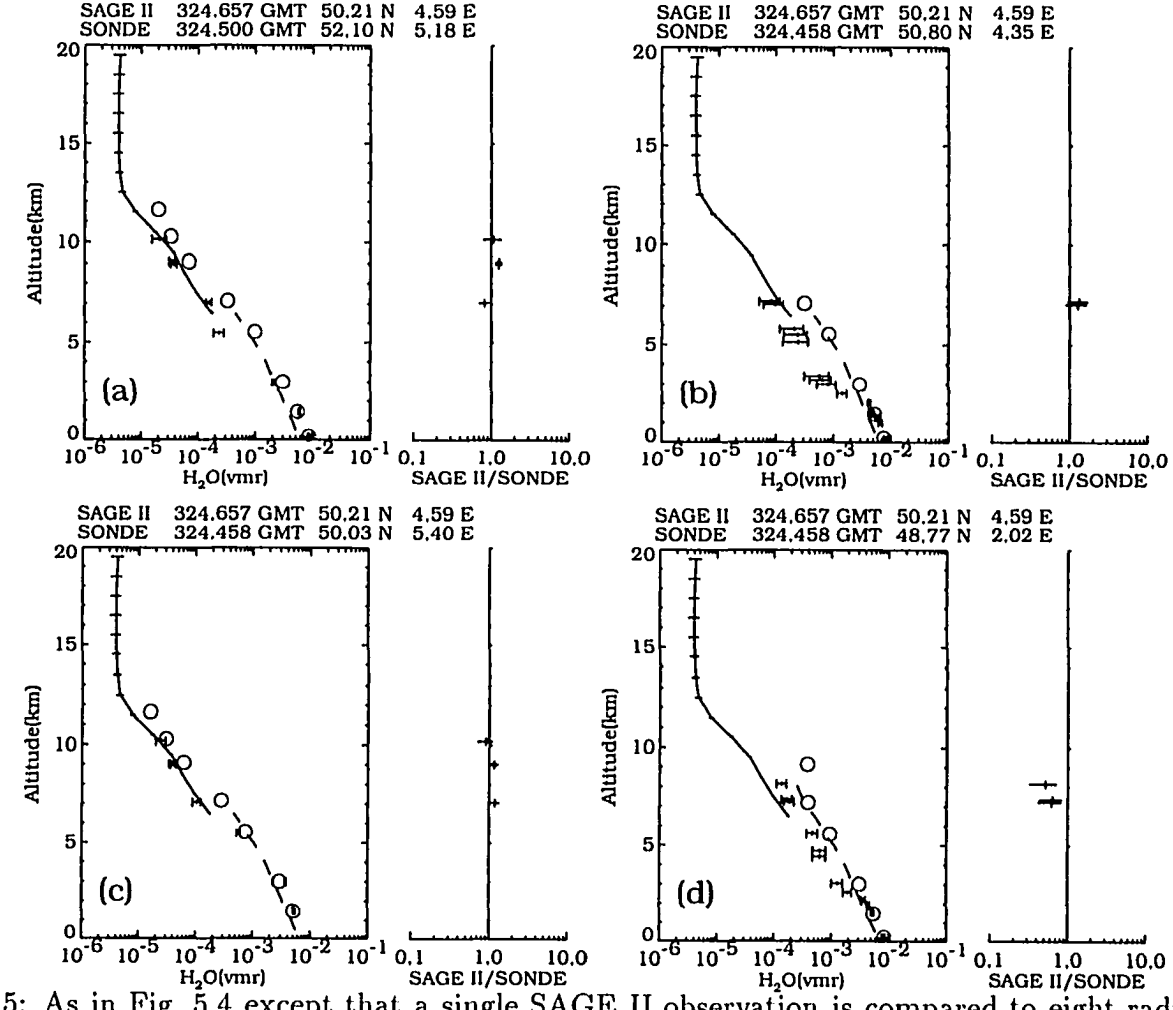


Figure 5.5: As in Fig. 5.4 except that a single SAGE II observation is compared to eight radiosonde soundings satisfying the correlative criterion on November 20 1987.

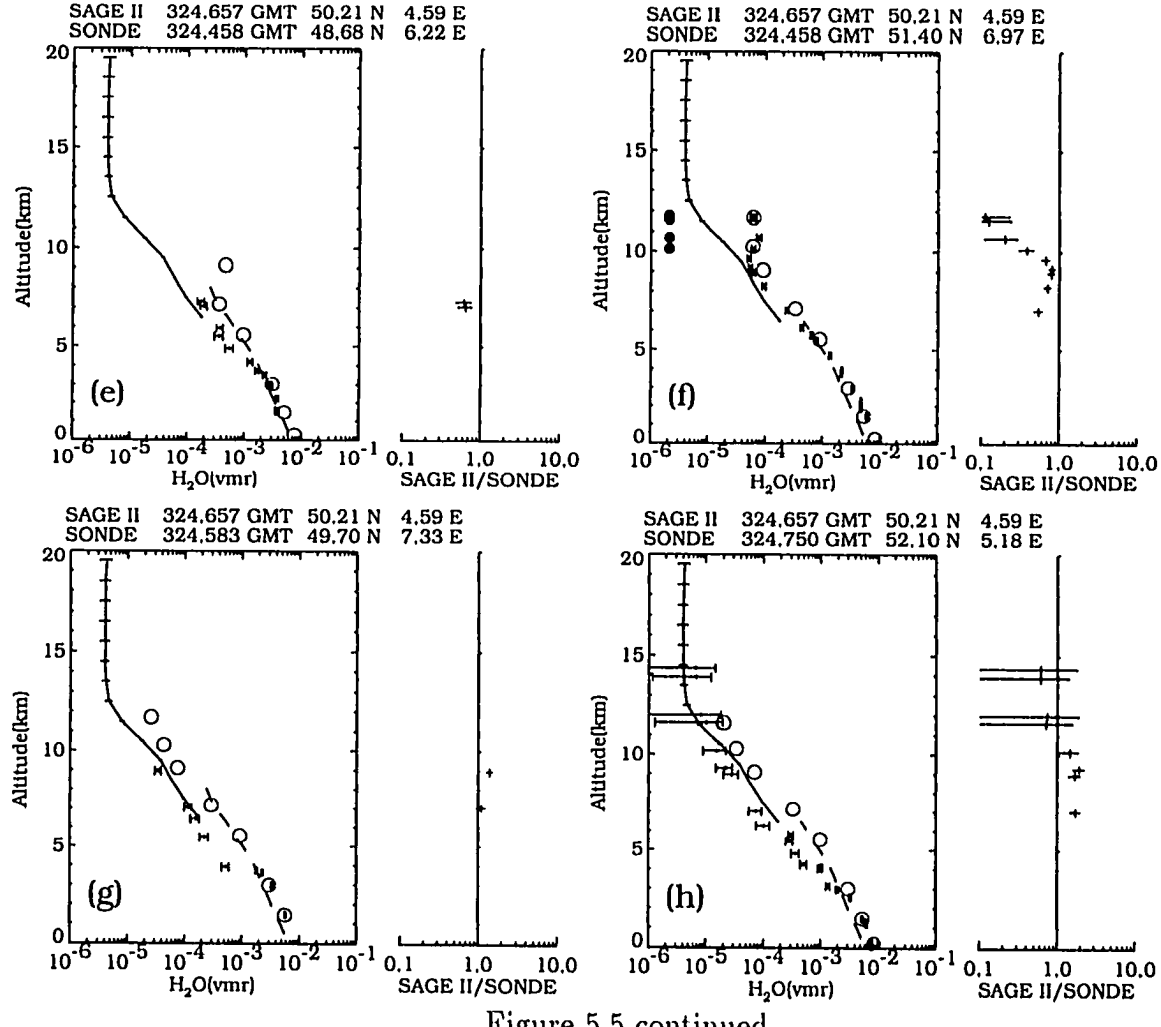


Figure 5.5 continued.

have occurred much more rapidly than this since SAGE II agrees quite well with the sounding taken almost five hours earlier but is wetter than the later sounding even though the time separation is smaller (a little more than two hours).

The water vapor distribution at 6.5 km displayed in Fig. 5.6 surrounding the SAGE II location incorporates additional radiosondes at distances greater than 250 km. The SAGE II line of sight, shown here as a black line extending out to 125 km on either side of the subtangent point, approximates the SAGE II horizontal resolution. The software package used to contour the non-uniformly spaced radiosonde data tends to create minima/maxima at places in the field where there are no stations, this is caused by the sparseness of the radiosonde network and by rapid mixing ratio changes between stations. In spite of these artificial features, the resulting water vapor field clearly shows the presence of strong horizontal gradients at several locations in the field. Actual gradients might be found to be much greater than those in Fig. 5.6, if it was possible to fully resolve the field. These single level gradients are slightly stronger than those reported for integrated column water vapor across frontal systems (McMurdie and Katsaros, 1991). The four stations (b,c,d,f) bracketing the SAGE II line of sight reported values of 138, 193, 280, and 340 ppmv, respectively and the projected line of sight intersects the contoured field at levels from 175 to 275 ppmv. Approximately half of the horizontal line of sight overlays the 175 ppmv contour and a simple average of the contoured field along the line of sight yields 196 ppmv, which is close to the reported SAGE II value of 178 ppmv. The close agreement must be partially fortuitous and is likely based on error cancellation associated with uncertainties due to measurement errors, the contouring algorithm and measurement locations for the radiosondes (balloon drift) and SAGE II (ephemeris information). In addition to the strong spatial variations shown in Fig. 5.6 the temporal evolution of the water vapor field was also quite significant on this day. At 0z the same four stations (b,c,d,f) reported values of 181,

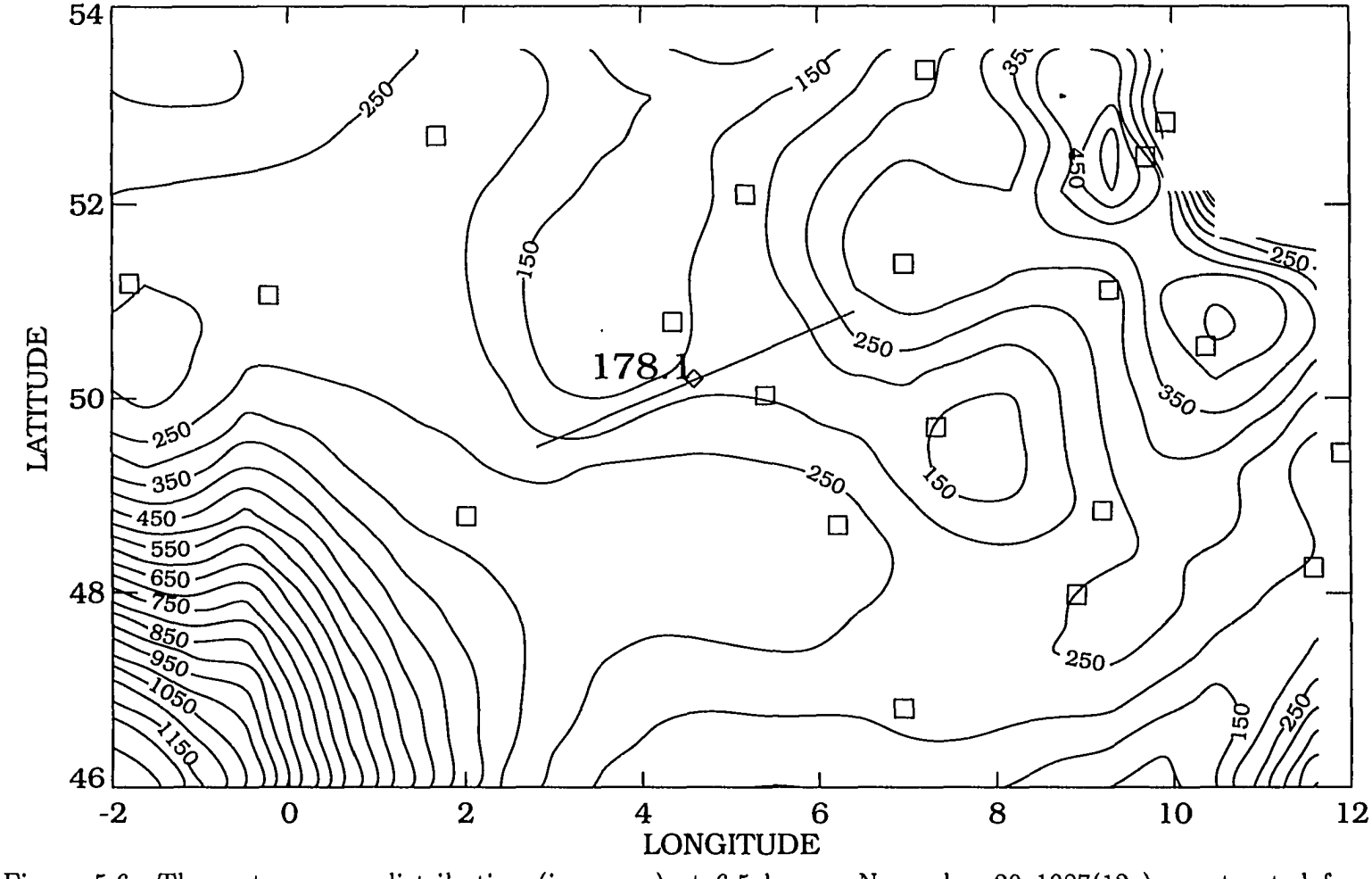


Figure 5.6: The water vapor distribution (in ppmv) at 6.5 km on November 20 1987(12z) constructed from the applicable radiosondes listed in Table 5.4 with additional radiosondes(squares) beyond the 250 km distance criterion used to fill in the field. SAGE II horizontal resolution is superimposed (solid line, $\approx$ 250 km) on the map. The SAGE II value is shown next to the subtangent point (diamond).

297, 464, and 197 ppmv, respectively, and values greater than 500 ppmv were found south of 48° while values greater than 850 ppmv occurred east of 8.5° with several stations reporting 1300 ppmv or above. At 0z on November 21 the location of high mixing ratios (greater than 500 ppmv) was found west of 2° while stations b,c,d,f reported values of 222, 147, 585, and 428 ppmv, respectively.

While we can find many good comparisons between SAGE II and the radiosonde hygrometers, there are also a large number of comparisons with differences greater than one would expect based on the estimated observational uncertainties. Figure 5.6 shows that strong gradients on scales smaller than the SAGE II horizontal resolution occur and this may explain some of the less satisfactory profile pair comparisons. It also follows that the time and space correlative criterion that we have used here may not be sufficiently strict for upper tropospheric water vapor and further emphasizes the difficulties one may encounter when comparing remote and *in situ* observations. To achieve a better understanding of the correlative profile pair differences in a general sense they are examined statistically in the following section.

5.3.2 Correlative Pair Climatology

To see if the set of correlative pairs might yield more consistent climatologies we calculated annual zonal means using only paired data at each altitude. The result is shown in Fig. 5.7 for 10° latitude bands for which sufficient values remained. For reference, we include the land based annual averaged GACS climatology (dashed line) and the SAGE II annual average (solid line) calculated with all 1987 data over land and ocean. The SAGE II annual average is further separated into that solely over land (dotted line). Radiosonde data were interpolated to SAGE II measurement altitudes. In most cases the correlative radiosonde hygrometers (squares) produce a drier climatology than GACS. This is to be expected since these observations are

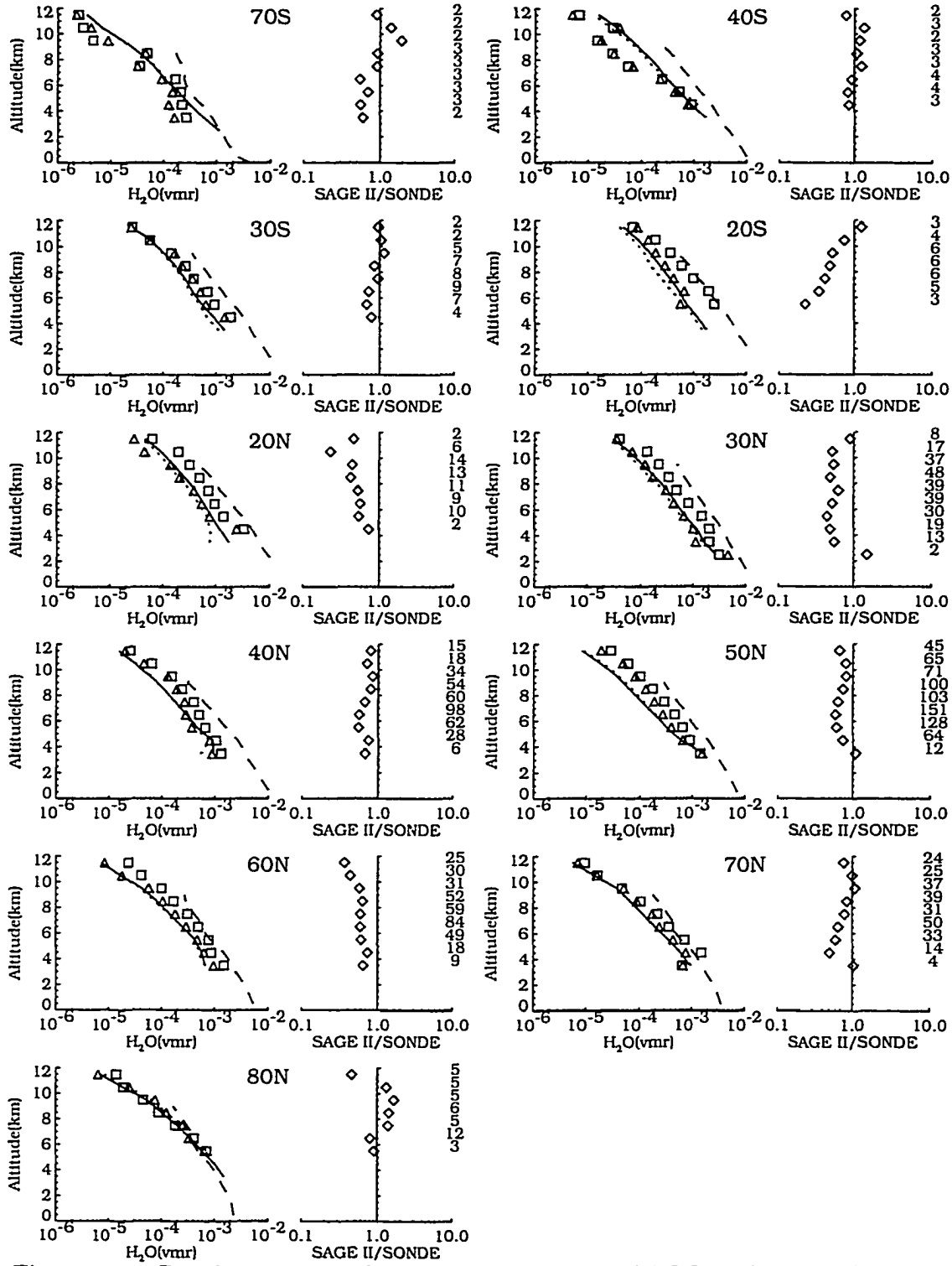


Figure 5.7: Zonal mean correlative pairs for 1987. SAGE II (triangles) and radiosondes (squares) are both drier than the yearly averaged Oort climatology (dashed line). Annual zonal means derived from the full SAGE II data set (land/ocean-solid line,land-dotted line) are also shown. On the right is shown the ratio of SAGE II to radiosonde and the number of paired values at each altitude.

paired to the cloud free (and drier) SAGE II observations. The correlative SAGE II pairs (triangles) generally exhibit greater zonal means than the annual land/ocean average except at 20° N and 30° N. However, even here they are still greater than the land average. The profile pair climatologies show somewhat better agreement than that shown in Fig. 5.3, but the shift to drier levels in the radiosonde data is negated in many cases by a similar shift in the SAGE II data. In the highly populated northern hemispheric latitude bands the differences still approach 50% at some altitudes. For the latitude bands shown, the differences between the SAGE II ocean/land averages and land averages are too small to explain the climatological differences shown in Fig. 5.3. We attempted to reduce the remaining profile pair differences by tightening the correlative criterion, reducing both the time and distance separation. Only slight improvements were achieved with drastic reductions (by 1/2) in these criterion. Further investigation of this lack of sensitivity was done by making scatter plots of mixing ratio differences and percent mixing ratio differences versus distance and time separation (not shown here). They revealed only a slight positive correlation with distance. That is, decreasing differences with decreasing separation distance were observed.

Figure 5.8 resolves the correlative values by hygrometer type. The radiosonde profile data were interpolated to the SAGE II measurement altitudes and all paired values up to 20 km were included in the SAGE II versus radiosonde mixing ratio scatter plots. The first panel is a composite of all the hygrometers and several of the hygrometers had too few points to plot. The group of hygrometers with the majority of the paired values, 1, 14, 18, 19, 20,and 21, display two types of clusters. The first cluster, from hygrometers 1, 18, 19, 20, 21, are consistently wetter than SAGE II while the second, from the Vaisala (14), tends to be drier than SAGE II. The range of mixing ratios in Fig. 5.8 result from the low altitude cloud cutoff of SAGE II (high mixing ratios) and the high altitude cutoff due to radiosonde operational limitations

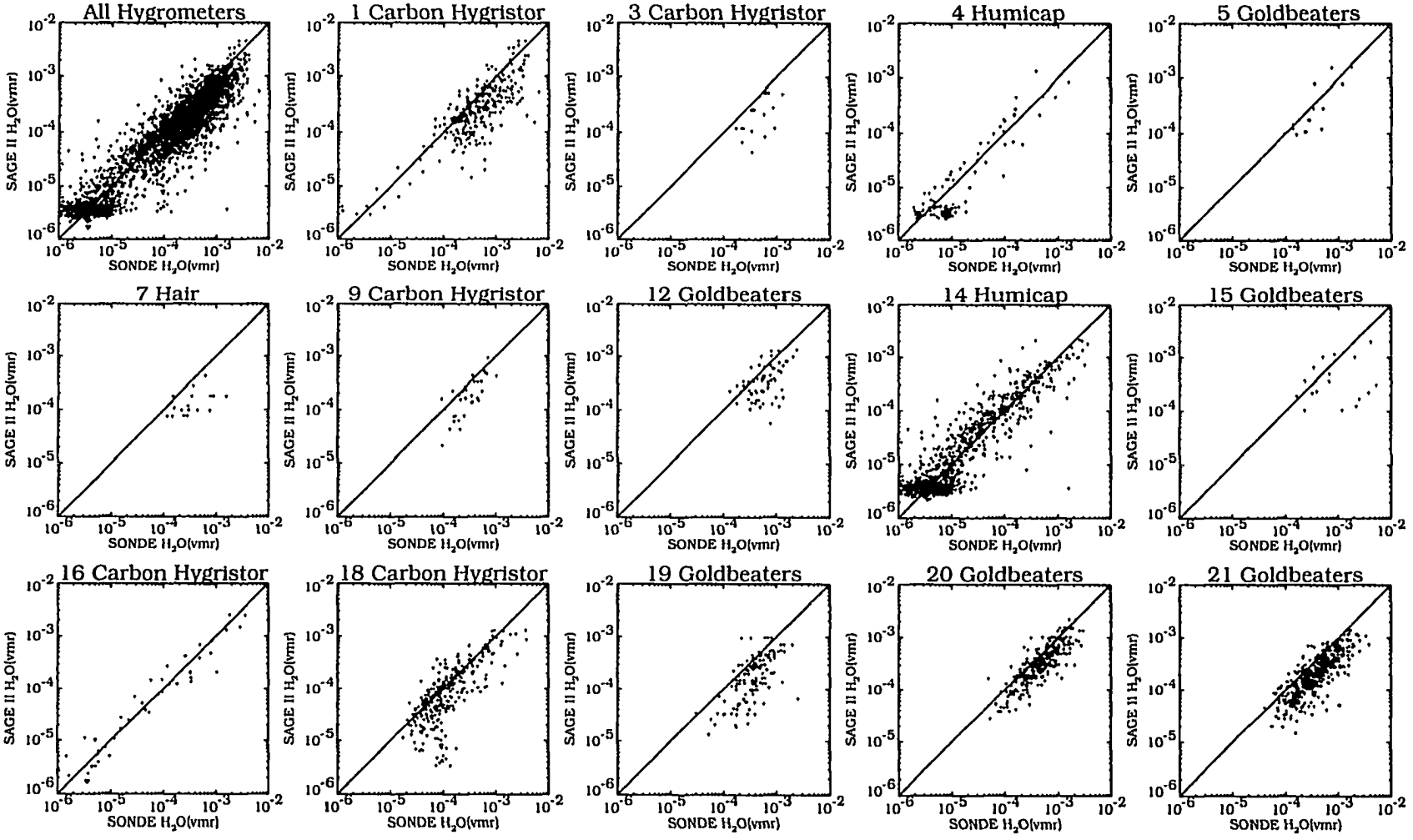


Figure 5.8: SAGE II versus radiosonde scatter plots of all paired mixing ratio values for 1987 resolved by instrument type. Radiosonde hygrometer numbering follows the NMC Office Note 29 as listed in Table 5.1 except for the first figure which is a composite of all the hygrometers.

(low mixing ratios). The use of Canadian reporting rules (hygrometer 18) with the carbon hygristor extends the measurements to lower mixing ratios, as can be seen by comparison to hygrometer 1. Likewise, the Vaisala (14), with its extended operational range relative to other hygrometers, also retains low mixing ratios. Of the carbon hygristors shown in the figure (1, 3, 16, 18, 24) all except 16 indicate higher moisture levels than SAGE II on average. Measurements from radiosondes using goldbeaters skin (11, 12, 19, 20, 21) are all clustered to higher values relative to SAGE II. The single radiosonde using the hair hygrometer (7) also follows this trend. Both thin film capacitive hygrometers (4, 14) have points clustered around dry conditions. For 14, there appears to be a shift in the average from dry values at low mixing ratios to an average similar to SAGE II when mixing ratios are $\approx 7{\cdot}10^{-4}$. The composite figure, formed from all hygrometer data, shows both types of clusters. At low mixing ratios the cluster average tends to be drier than SAGE II (due to Vaisala data) while at higher mixing ratios the cluster average is wetter (all others). This shift in cluster averages is also reflected in Fig. 5.7 at some latitude bands, such as 40° and 50° N. It does not occur in all latitude bands, since the contribution of each hygrometer type varies as shown in Fig. 5.2. Each figure displays a great deal of scatter which we believe to be caused by strong variability on small scales over short time periods as previously shown in Fig. 5.6 for the November 20 case. Even though we are comparing SAGE II observations that represent average conditions along the line of sight to *in situ* radiosondes that can respond to the small scale features in the highly variable water vapor field, one would expect the plotted points to fall equally on both sides of the diagonal line in Fig. 5.8, in the absence of any instrumentation bias. We note that the gold beater hygrometers, which have the longest response time, show the tightest grouping of the clustered points and that the thin film capacitive hygrometers, with the fastest response time, have the most scatter. One may interpret this as the thin film capacitive hygrometers responding

better to localized conditions while the gold beater hygrometers tend to report higher levels of water on ascent in the drier upper altitudes because of the longer response time. One can also interpret the carbon hygristor as intermediate to these two hygrometers since it is believed to have a response time ranging from somewhat longer to comparable to the thin film capacitive hygrometer.

In Fig. 5.9 we show global annual averages of the correlative data shown in Figs. 5.7 and 5.8 separated by hygrometer type. The first figure again is a composite of all the data followed by the goldbeaters skin, carbon hygristor, and thin film capacitive. The hair and lithium chloride hygrometers are not shown due to a lack of data. The composite figure is essentially what one would get by averaging the latitude bands in Fig. 5.7 together. Both the carbon hygristor and goldbeaters skin hygrometers display considerably greater mixing ratios than SAGE II and are driving the composite profile to greater mixing ratios due to the large number of points contributed by each. The thin film capacitive hygrometers, on the other hand, show a markedly different response. Between 8.5 and 11.5 km they tend to report equivalent or lower mixing ratios than SAGE II while between 4.5 and 6.5 km they are higher but only by 20-30%. This is encouraging since the thin film capacitive hygrometers are thought to be the best hygrometer currently in use. The operational limits listed in Table 5.3 for this hygrometer may be overly optimistic as evidenced by the average response above 12.5 km. If we view the SAGE II observations as a transfer standard, it follows that the radiosonde data set in the upper troposphere is not entirely homogeneous because the hygrometers on which this data set is based appear to have significantly different operating responses. It is curious that the response characteristics of the carbon hygristor is similar to that of goldbeaters skin when it should probably be more like the thin film capacitive hygrometer based on its response time, operational limitations and balloon intercomparison trials (Schmidlin, 1988).

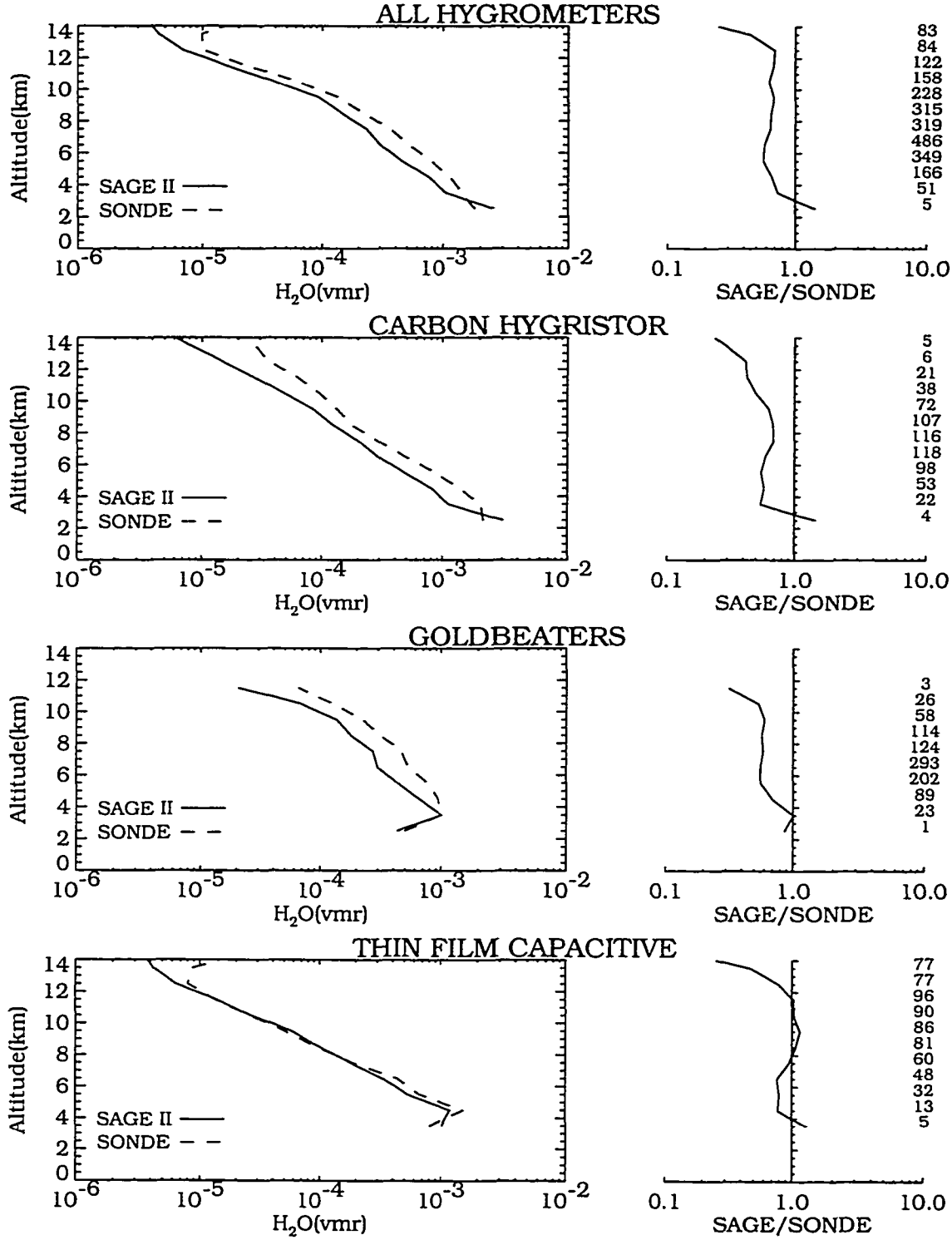


Figure 5.9: Global mean correlative pairs for 1987 resolved by hygrometer type. SAGE II profiles are the solid lines and the dashed lines are the radiosondes. The first figure incorporates all paired data. On the right is shown the ratio of SAGE II average to the radiosonde average along with the number of paired values at each altitude.

The scatter of the profile pair mixing ratios shown in Fig. 5.8 is quite high because pairs at all latitudes for all of 1987 are included. When points are scattered over an order of magnitude, one can expect some bias in the arithmetic mean. In addition, outlier points on the high side will have an inordinate affect on the arithmetic mean. If the SAGE and radiosonde mixing ratios contain similar variability, the bias will be present in both means and approximately cancel in the SAGE/radiosonde ratio. However, the faster response time of the thin film capacitive hygrometer makes it respond more accurately to small scale water vapor features and the data will display more variability than the other hygrometers. Taking the mean of the SAGE radiosonde ratio, defined for each data pair, improves the situation only slightly. The ratio displays slightly less variability than the mixing ratios individually but the ratio of SAGE to the thin film capacitive still displays more variability than the other hygrometers.

One could perform the comparison using physical units that are restricted to narrower ranges such as relative humidity or dewpoint temperatures. The hygrometers respond to relative humidity; this is converted to dewpoint for the upper air data base and then to volume mixing ratio in this work. The primary measurement of SAGE II is water vapor number density which is easily converted to mixing ratio, but the best approach to converting to relative humidity is uncertain since this could be done relative to ice or water. An alternative approach that narrows the mixing ratio range and mimics the conversion to relative humidity is based on calculating the mean of log[volume mixing ratio] followed by reconversion to mixing ratio. This also reduces the effect of the high outlier points and produces a more stable mean for comparison purposes.

Figure 5.10 shows the global annual mean obtained with this approach. The mean mixing ratios show a significant decrease in all four cases. The ratio of SAGE to radiosonde above 9 km shifts closer to unity but is unchanged below this for the

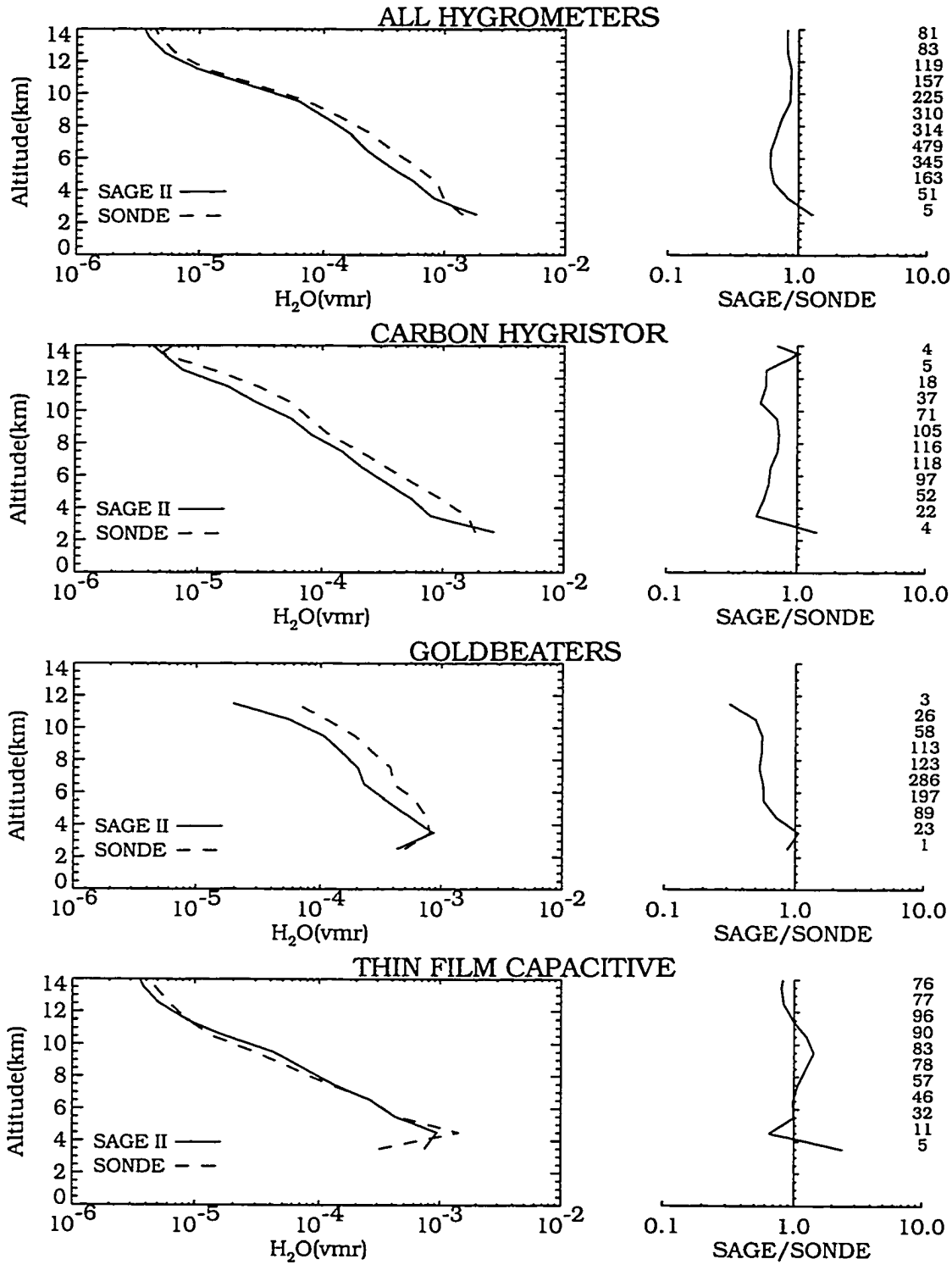


Figure 5.10: Similar to Fig. 5.9 except that global means are calculated from log[volume mixing ratio].

carbon hygristor and goldbeaters hygrometers. The thin film capacitive ratio also displays the shift towards unity above 10 km and an increase in the ratio below. The SAGE II radiosonde comparison is still better for the thin film capacitive and qualitative differences in the comparisons between hygrometer types are similar.

5.4 Isentropic Trajectories

Previous use of air mass trajectory analyses has been confined to characterizing the movement of air parcels and constituents in only a most general manner. For example, to better understand atmospheric constituent measurements at Mauna Loa Observatory the long range transport was examined with cluster analysis of 10 day isobaric back trajectories on 700 and 500 mb pressure levels (Harris and Kahl, 1990). It was found that many of the dominant transport features repeat from year to year and that transport from or through anthropogenic source regions is unlikely to occur. Middle atmosphere trajectory studies have tracked particle movement in the polar and tropical regions through a simulated stratospheric sudden warming (Hsu, 1980) and for the winter of 1978/79 using satellite observations of temperature (Lyjak and Smith, 1987). The evolution of the Mount Pinatubo volcanic plume and associated mixing examined with isentropic trajectories found good agreement with observations (Fairlie, 1993). In this work we will attempt to use the trajectory calculations in a more demanding manner by incorporating them into the matching scheme as a way to increase the number of correlative points.

Whether or not trajectory analyses can improve or enhance measurement validation depends on how well calculated trajectories represent the true air parcel movement. Given a three-dimensional wind field it is a straightforward process to calculate Lagrangian trajectories. With a high order finite difference scheme the time integration can be done quite accurately both forward and backward. The

greatest uncertainties in the calculations arise from the limited spatial and temporal resolution of the wind fields, from uncertainties in each of the wind components, and from simplifying model approximations.

Current global wind data sets are a blend of radiosonde and satellite observations assimilated in operational weather models and output on model grids at fixed time intervals. Observational space and time sampling limitations coupled with model discretization act to filter small scale features from the operational product. This is more of a concern for the troposphere than the stratosphere (Austin and Tuck, 1985) where circulation features tend to be of large scale. While trajectory sensitivity to grid coarseness has been studied by degrading high resolution model output wind fields for trajectories in the boundary layer (Rolph and Draxler, 1990), troposphere (Merrill et al., 1986), and the stratosphere (Knudsen and Carver, 1994); accuracy results have not been reported for the range of potential temperature surfaces considered here or the relatively short 24 hr travel times used in this study. However, Knudsen and Carver(1994) present relative spherical distance errors (absolute spherical distance relative to trajectory length) for adiabatic 475 K trajectories as a function of travel time for the high latitude northern hemisphere winter/spring of 1991/92 for changes in several meteorological parameters. The 24 hour trajectory experienced less than 0.5% relative error when: the time step was changed from 15 to 5 minutes; horizontal resolution was changed from 1.5°to 0.75°; or diabatic effects were included. Changing the temporal resolution from 6 hrs to 1 hr produced a 1.5% error and decreasing the number of model levels by half for the uppermost levels caused a 3.5% error. In the same paper, two comparisons of model trajectories to actual balloon trajectories (451-467 K) of 1.83 and 2.14 days duration found relative errors of 8% and 12%, respectively.

Horizontal winds calculated with the geostrophic approximation have an accuracy of approximately 10% (Austin and Tuck, 1985). Vertical wind components

are typically much smaller and difficult to measure. Isobaric trajectories in regions of strong wind shear may diverge significantly from actual when vertical winds are poorly known, although this may be less of a concern for short travel times. To circumvent the lack of vertical wind knowledge one can choose a coordinate system in which the vertical wind component is naturally small. Trajectories calculated on isentropic surfaces (Danielson, 1961) prove to be a good choice for adiabatic conditions. Thus, our model assumes diabatic processes such as radiative heating/cooling may be ignored and the air parcel does not experience vertical mixing or condensation. Radiative processes have longer time scales than transport processes and become important only over periods of time much greater than those considered here. Vertical mixing and condensation processes produce flow across isentropic surfaces and are more likely to occur at lower tropospheric levels. Determining when these small scale processes occur from the lower resolution global data sets at hand is difficult to impossible. We note that condensation may not be much of a concern because the SAGE II measurement locations are the trajectory starting points and are dryer than the atmosphere as a whole. A dry air parcel would have to experience significant ambient changes in pressure and temperature for condensation to occur.

The isentropic Lagrangian trajectory package selected for use here may be classified as kinematic (Merrill et al., 1986) in that only the wind fields are used in the calculations. The trajectory package is based on a fourth order Runge-Kutta time integration scheme. It uses the European Center for Medium Range Forecasting (ECMWF) analysis gridded fields (2.5° latitude, 9.7° longitude) at 0z and 12z of winds, temperature and geopotential height on 15 pressure levels (1000-10 mb).

Selecting the meteorological data set also fixes the temporal and spatial resolution so the only parameter under our control that affects accuracy is the time step. A series of trajectories were calculated with decreasing time steps and the change in distance between corresponding times was examined. At the end of a 24 hr

integration the shift in the trajectory location was less than 10 km between the 1/8 hr and 1/4 hr time step calculation. A time step of 1/2 hr was chosen to reduce computation time and storage requirements. The shift in location was generally less than 20 km when compared to the 1/8 hr time step case at 24 hrs. As the air parcel travel time increased, uncertainty in the location increased because of the temporal and spatial resolution of the meteorological data and the correlative criterion effectively broadens.

Potential temperature, conserved for adiabatic and inviscid conditions, is used as the vertical coordinate in the trajectory package. Potential temperature is defined by

$$\Theta = T(\frac{1000}{P})^{.286} \tag{5.1}$$

Vertical motion of the air parcel across Θ levels will not occur provided diabatic and mixing processes are small. Thus, air parcel movement will be limited to the horizontal on constant potential temperature surfaces. Figure 5.11 displays the variation of the monthly zonal mean of theta versus altitude for January 1989. Potential temperature increases upwards due to the inverse pressure relationship and is further modified by the temperature distribution. Stratospheric potential temperature surfaces are closely aligned to altitude while the tropospheric surfaces display considerable vertical altitude shifts over latitude. Generally air parcel movement will be more zonal than meridional, due to the relative wind strengths, and latitude changes will be less than 10°for 24 hr trajectories in the equatorial and middle latitude regions. At the higher latitudes the latitude range may be greater than 10°. Altitude shifts of the potential temperature surfaces over 10° can reach 2 km and may be even greater when longitudinal variations are considered.

To maintain trajectory accuracy, at least initially, potential temperatures are calculated from the SAGE II NWS supplied temperature and pressure profiles. Initial

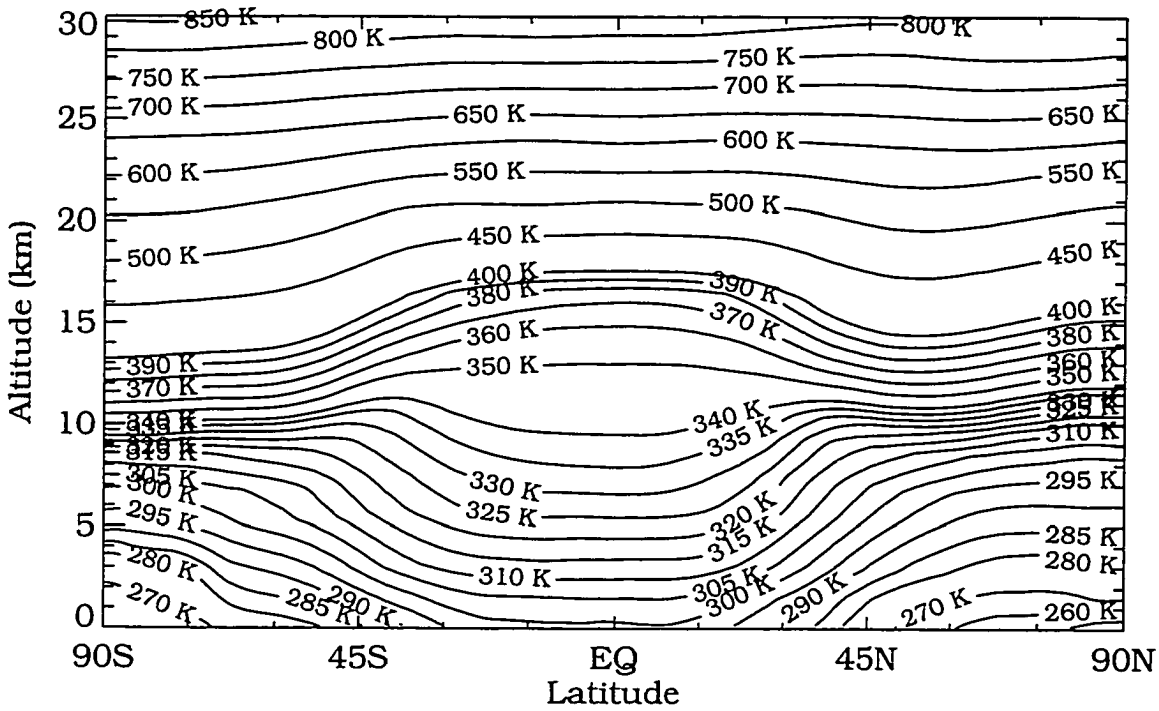


Figure 5.11: Monthly zonal mean potential temperature for January 1989 calculated with the ECMWF gridded data.

air mass latitudes and longitudes are calculated for each of the trajectory Θ surfaces by interpolating over the function of subtangent location versus Θ. Forward and backward trajectories are calculated for 24 hr travel times with 1/2 hr time steps. Air mass pressure, temperature, and altitude are stored for each time step.

5.5 Trajectory Profile Matching

5.5.1 The Isentropic Trajectory Correlative Technique

The correlation criterion used to determine whether points along the trajectory are in close proximity, in space and time to a radiosonde observation, is depicted in Fig. 5.12 for a generic trajectory. Solid circles indicate the calculated air mass locations at each time step (shown here for a 1 hr time step) over 48 hrs. The SAGE II measurement time, t_0, is the trajectory starting point where $t_0 - t_{-24}$ corresponds to the backward portion and $t_0 - t_{+24}$ corresponds to the forward portion. The distance between the radiosonde and each air mass location can be expressed as a function of time. The correlative criterion is satisfied when the separation distance is less than 250 km, indicated by the envelope in Fig. 5.12, and $|t_S - t_R| < 6$ hrs. The correlative criterion used in Sec. 5.3 would appear as a circle of radius 250 km centered at t_0.

5.5.2 Climatological Comparisons

The potential temperature levels for the air mass trajectory calculations were chosen taking into consideration the radiosonde ceiling for useful data, the low frequency of SAGE II observations near the surface and the high latitude cutoff for the SAGE measurements. From Fig. 5.11 we see that selecting levels of 270, 290, 300, 310, 320, 330, 340, 345, 355, 370, 400, 430, 460, 500, and 550 K provides reasonable vertical coverage in the troposphere.

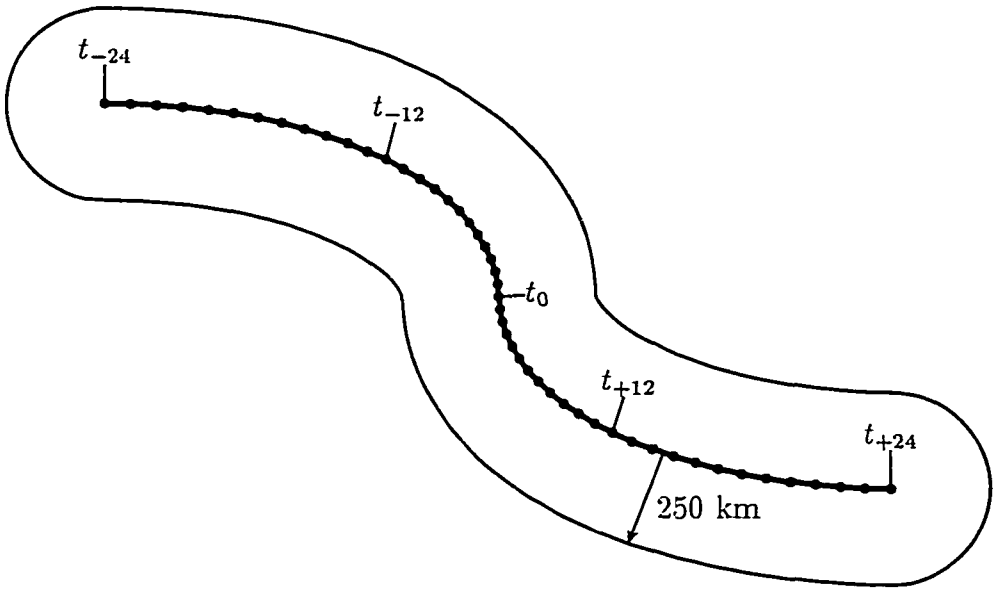


Figure 5.12: Trajectory correlation geometry.

The starting location for each trajectory is the SAGE II subtangent location, which is a function of altitude. To preserve the trajectory accuracy the subtangent locations on each surface were calculated as follows. A potential temperature profile was calculated from the collocated National Weather Service temperature and pressure profiles used in the inversion. The altitude of each potential temperature surface was determined by interpolation and then used in the interpolation of subtangent latitude and longitude.

Isentropic trajectories were calculated for the 1987 SAGE II measurement locations and then the trajectory and radiosonde data sets were examined for matches using the technique of the previous section. Scatter plots of SAGE versus radiosonde mixing ratios for the first six months of 1987 are shown in Fig. 5.13 for the case where no screening is applied to the radiosonde data. The scatter plots display patterns that for the most part are consistent within hygrometer types. Both humicap hy-

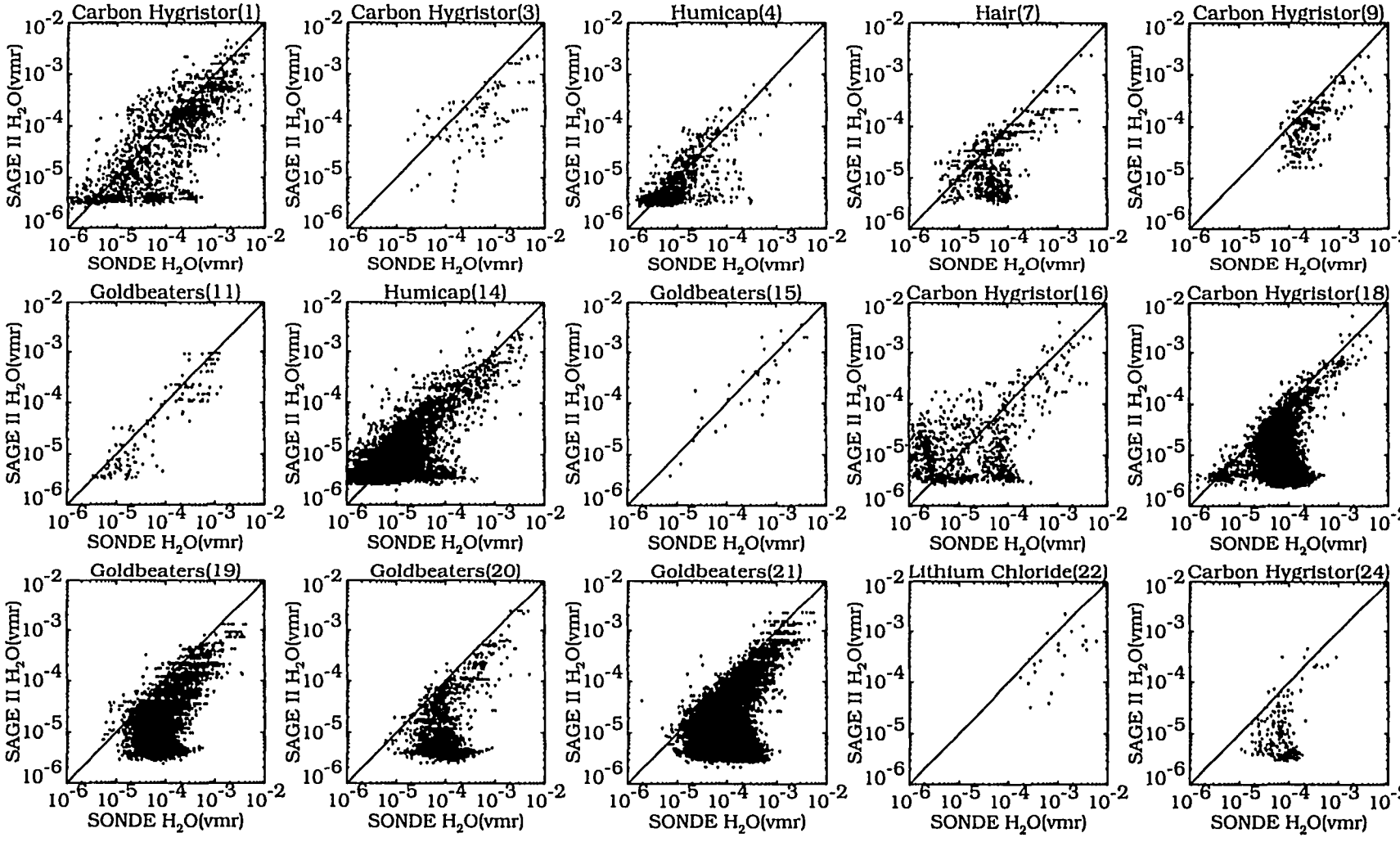


Figure 5.13: Scatter plots of SAGE II trajectory versus radiosonde paired mixing ratio values on all theta surfaces for the first six months of 1987 resolved by instrument type. No screening applied to radiosonde data. Radiosonde hygrometer numbering follows the NMC Office Note 29 as listed in Table 5.1.

grometers run slightly wetter for SAGE mixing ratios greater than $1 \cdot 10^{-4}$ (low altitudes) but are dryer from $1 \cdot 10^{-4}$ to $2 \cdot 10^{-5}$. At SAGE mixing ratios of $8 \cdot 10^{-5}$ the humicap shows a broad range of values that is interpreted as a loss of hygrometer sensitivity. The goldbeater hygrometers display a very consistent pattern, on average wetter than SAGE at all altitude levels with the bias becoming extreme at the highest altitudes. The hair hygrometers display a similar pattern while there are not enough points for the lithium chloride hygrometer to discern a pattern. The carbon hygristor pattern is the least consistent. Hygrometer (1) tends to follow the humicap pattern, hygrometers (9,18,24) display goldbeater patterns and hygrometer (16) displays a rather unique pattern.

Since a greater number of correlative match points are available, we decided to refine the hygrometer screening process summarized in Table 5.3 by tightening some of the screening tests. In addition to the screening of Table 5.3, the screening was modified in the following manner. All goldbeater data was dropped when $T < -40°$ C. This primarily affected hygrometers (20, 21). All carbon hygristor data (except 18) with R.H.$< 20\%$ was also deleted. This represented a change from temperature screening to relative humidity for hygrometers (9, 16). For all hygrometers, data were dropped for pressures less than 200 mb, except for the humicap (4, 14), which was dropped at pressures less than 160 mb. It was also discovered that some artificial dew point depression data had been inserted and this was removed. In addition to the temperature screen, data from hygrometer (18) was deleted if R.H.$< 10\%$, temperatures were below 220 K and when the dew point temperature was below 210 K.

In addition to the automatic screening, visual screening, as described in Sec. 5.1, was also performed. During the matching process the temperature difference between the trajectory air mass and radiosonde was checked. The match was rejected if the temperature difference was greater than 7° C. The match was also rejected

if the altitude difference between the air mass and radiosonde exceeded 4 km. The scatter plots between SAGE and radiosonde incorporating all screening is shown in Fig. 5.14. The patterns are similar to those of Fig. 5.8 and most of the comments concerning that figure apply. It appears that the screening has effectively removed the obviously questionable data except for hygrometer (21).

The global means for each hygrometer type derived from the data shown in Fig. 5.14 are shown in Fig. 5.15. Due to sampling biases on the isentropic grid the global mean profiles no longer display a typical water vapor profile shape. The sampling bias was confirmed by calculating equivalent global means with the Oort data set at the match pair times and locations. When compared to Fig. 5.10, we note that the agreement between SAGE II and the thin film capacitive hygrometer improves from 5 to 12 km and worsens above 12 km. The bias widens by approximately 10% for the goldbeaters hygrometer and also widens for the carbon hygristor. The number of matched profile pairs with this approach increases by rough factors of two, three, and ten for the goldbeaters, carbon hygristor, and thin film capacitive, respectively. The expansion of profile pairs follows from the increased spatial area over which matches are allowed, which is in turn related to trajectory length. The distance along the trajectory from the SAGE II subtangent location to the match point is shown in Fig. 5.16. No matches were found for the 270 K surface and only a few remained after screening on the 500 K and 550 K surfaces. The greatest number of matches were found on surfaces between 300 K and 355 K where some air masses traveled 4000 km before encountering a nearby radiosonde observation.

One way to judge the quality of the air mass trajectory is by comparing an air mass and radiosonde property at a match point. Given the strong water vapor gradient in the troposphere we chose to examine the altitude difference. The distribution of altitude differences for each surface is shown in Fig. 5.17. The distributions are somewhat asymmetrical with a tendency for the air mass altitude to occasionally

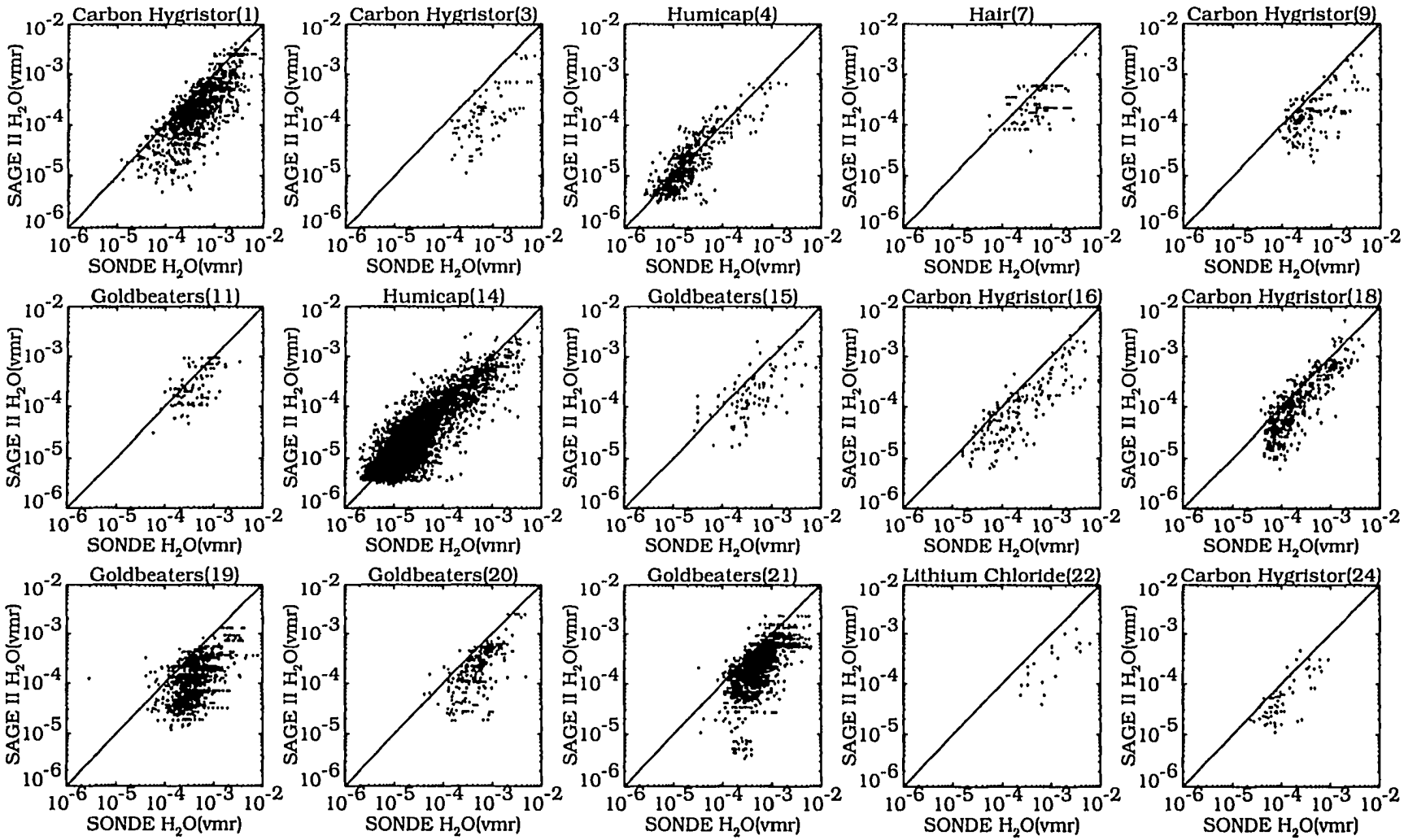


Figure 5.14: Scatter plots of SAGE II trajectory versus radiosonde paired mixing ratio values on all theta surfaces for 1987 resolved by instrument type. Radiosonde hygrometer numbering follows the NMC Office Note 29 as listed in Table 5.1.

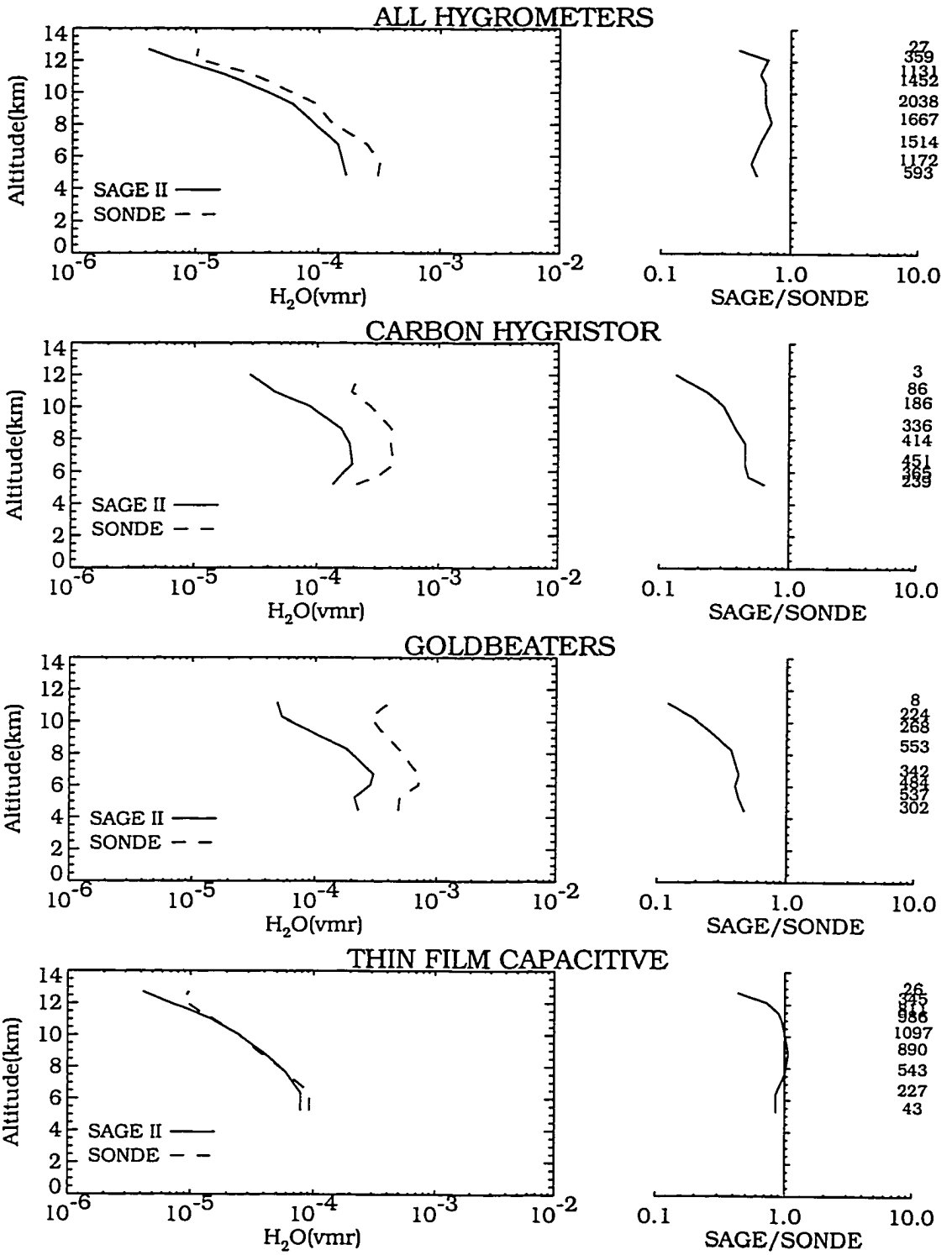


Figure 5.15: Similar to Fig. 5.10 except for trajectory to radiosonde pair matching.

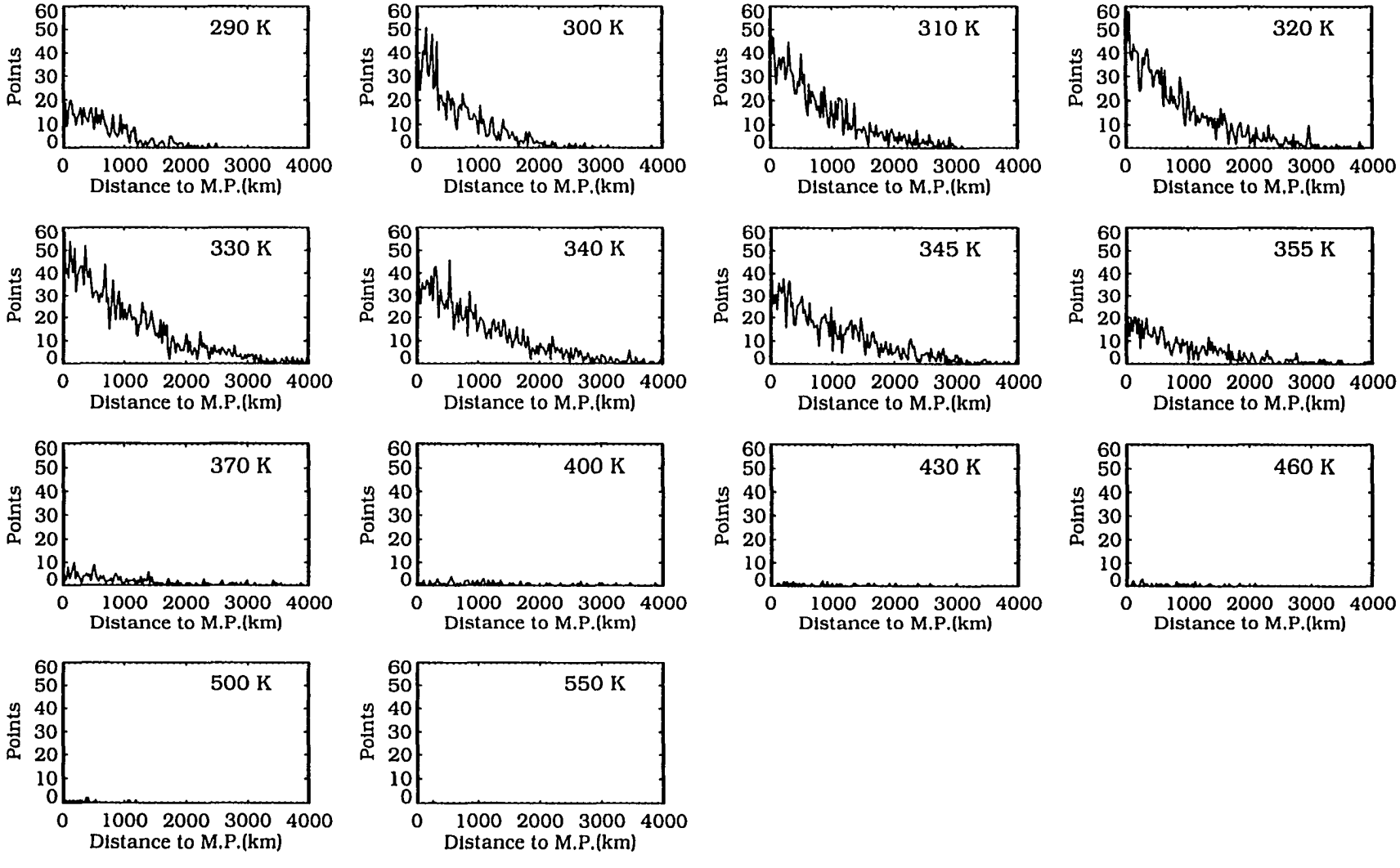


Figure 5.16: Frequency distribution of distance along trajectory from SAGE II subtangent to match point.

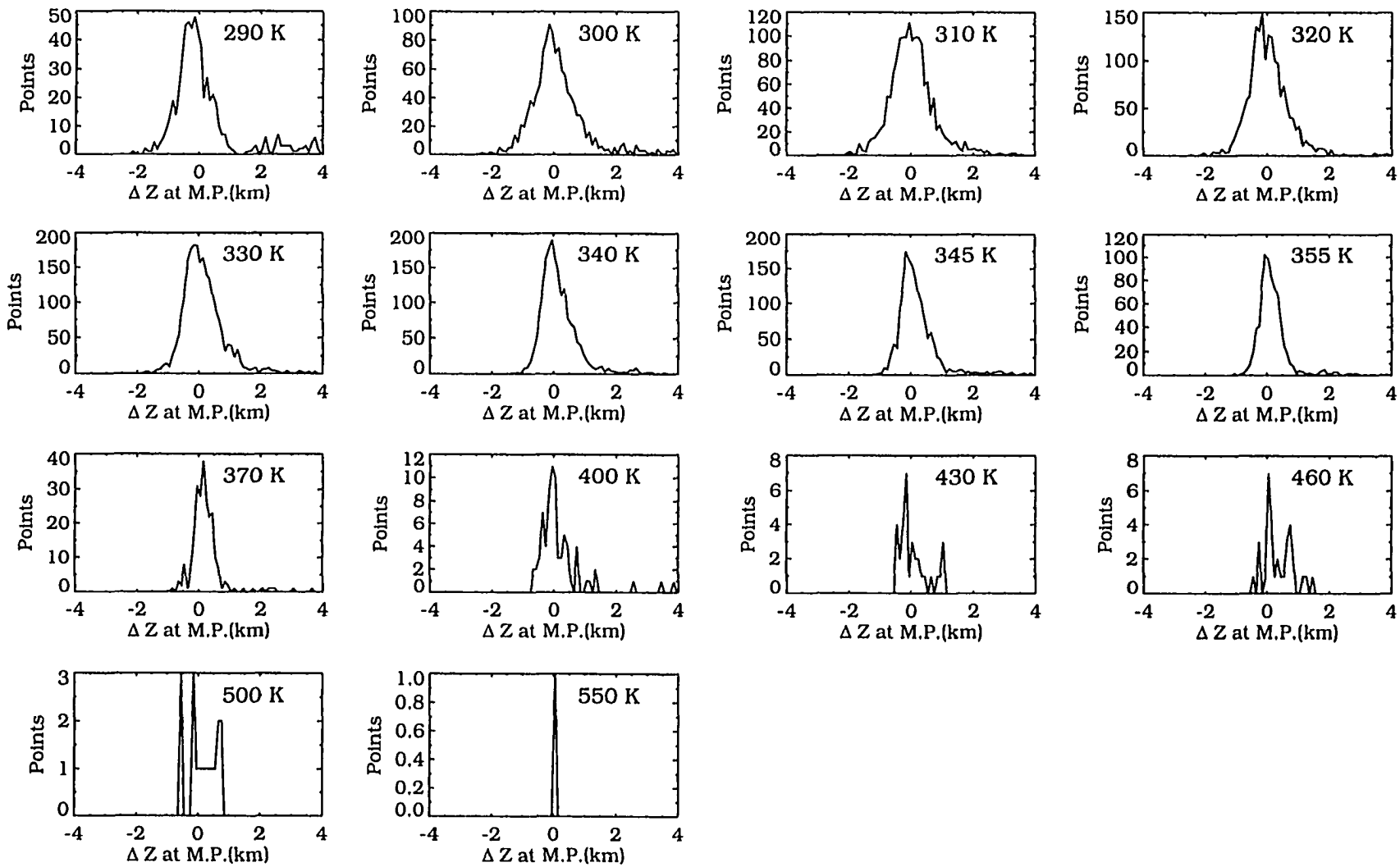


Figure 5.17: Frequency distribution of altitude separation (trajectory - radiosonde) at match point.

exceed the radiosonde by more than 2 km. The distribution spread increases at lower altitudes, probably due to enhanced dynamic activity and the inability of the ECMWF data set to reproduce small scale features resulting from that activity. To determine a reasonable maximum separation we examined three situations. The first was the spread in the stratosphere which is about 1 km at most. For the second, the comparison leading up to Fig. 5.10 was repeated using potential temperature instead of altitude for the vertical grid with the same correlative criterion. The altitude spread on the lower surfaces was about 1 km. Lastly, we examined the mean altitude shift on each surface as a function of altitude difference. For a 1 km difference the mean altitude shift on all surfaces generally remained much less than 0.1 km.

When matches with altitude differences greater than 1 km were eliminated, the SAGE II-radiosonde mixing ratio scatter plots can be represented as shown in Fig. 5.18 for the fully screened case. The patterns are very similar to those of Fig. 5.14. The resulting global mean mixing ratio, shown in Fig. 5.19, is also nearly identical to Fig. 5.10. Thus, it appears that the global means are not particularly sensitive to trajectory degradation as measured by the altitude difference.

Another approach to judging trajectory degradation involves examining the change in global means as the maximum allowable air mass distance from the subtangent to the match point is varied. One might expect questionable matches to occur more often as this distance increases. The maximum cutoff distance was selected at 2000, 1500, 1000, 500, and 250 km. The global means varied only slightly from those shown in Fig. 5.15.

A common approach to screening radiosonde data is to accept all data except for that above 300 mb. The scatter plots appear as shown in Fig. 5.20 with this approach. This removes much of the questionable data, primarily in the upper

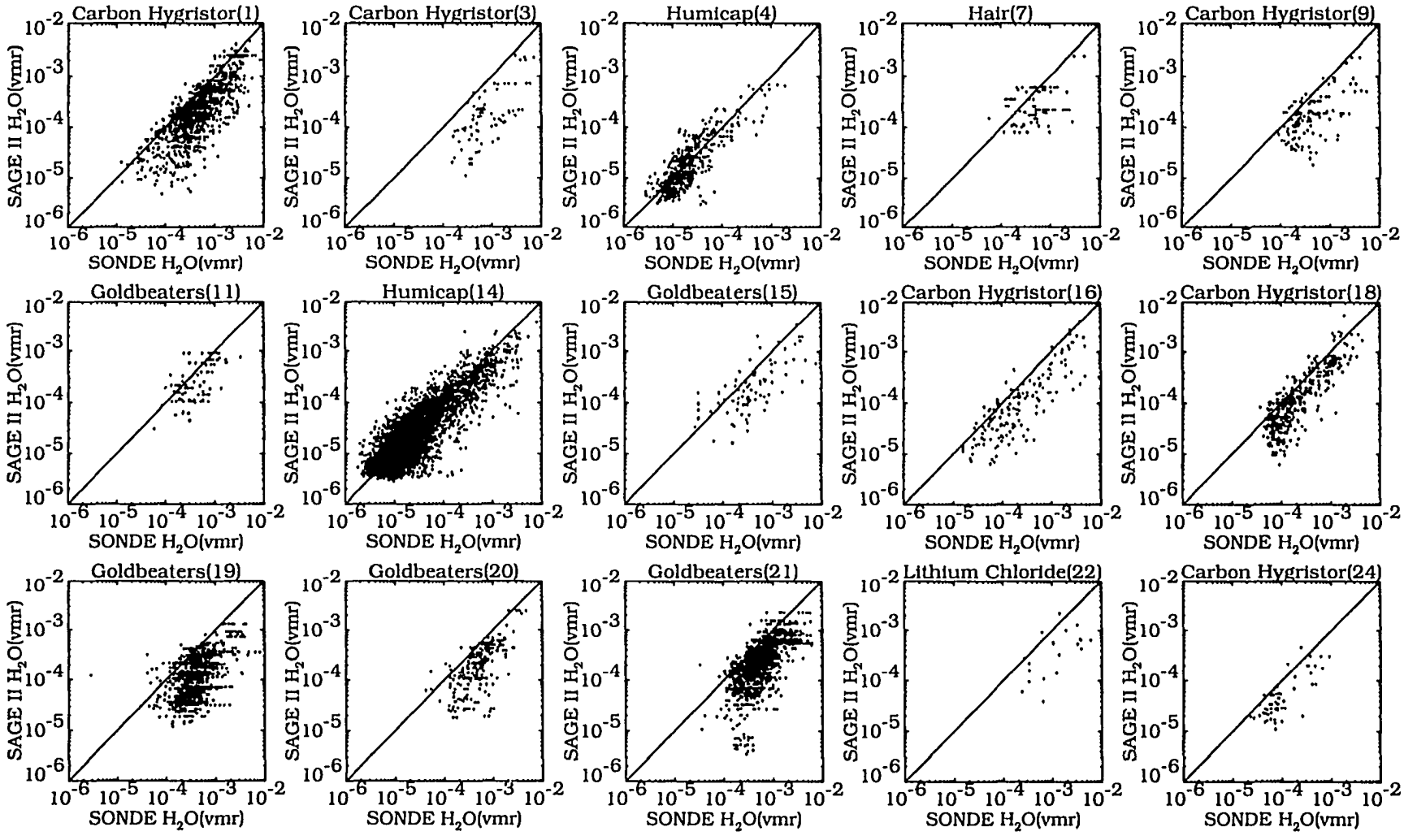


Figure 5.18: Scatter plots of SAGE II trajectory versus radiosonde paired mixing ratio values on all theta surfaces for 1987 resolved by instrument type. Data pairs with $|\Delta Z| > 1$ km eliminated. Radiosonde hygrometer numbering follows the NMC Office Note 29 as listed in Table 5.1.

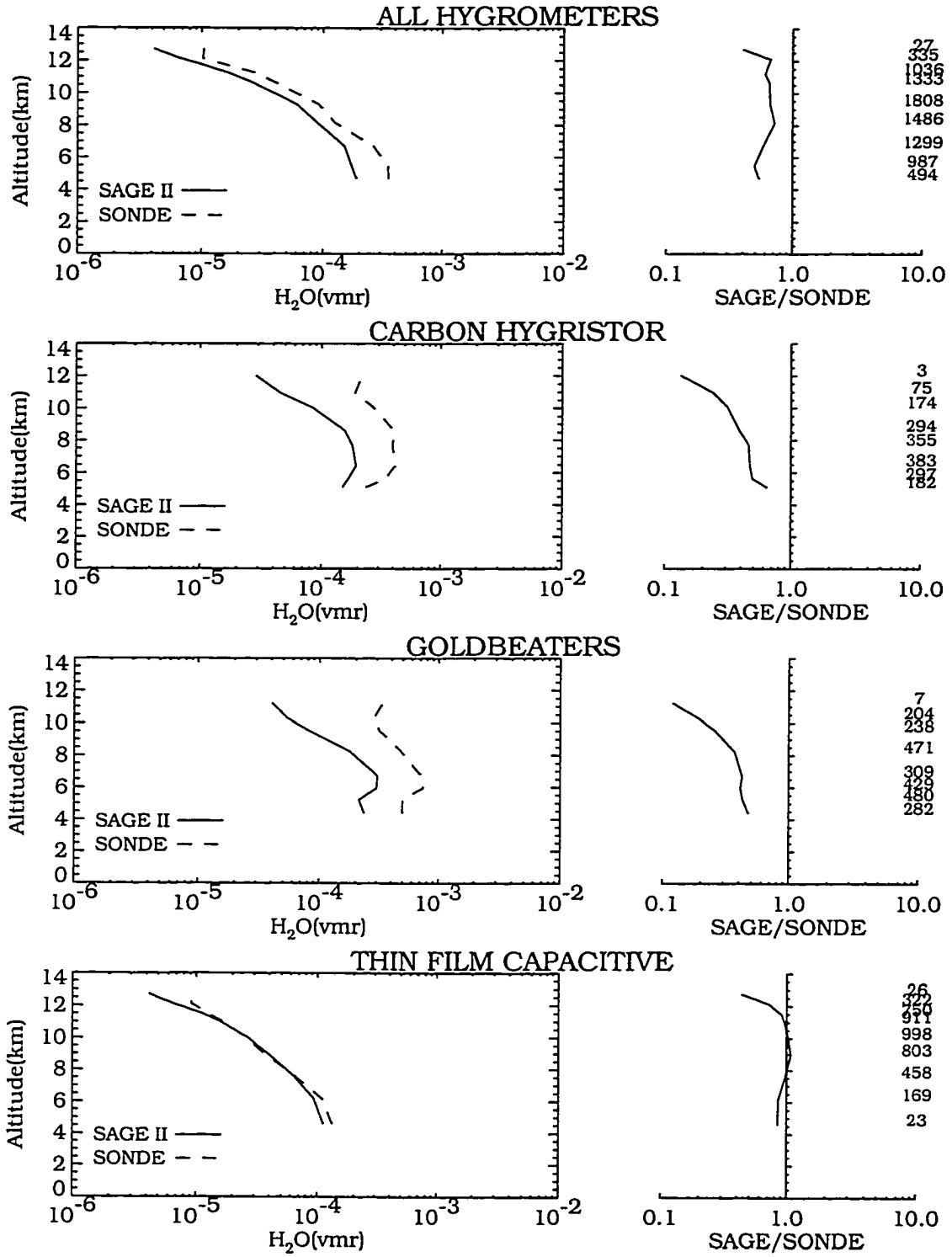


Figure 5.19: Similar to Fig. 5.15 except that data pairs with $|\Delta Z| >$ 1 km eliminated.

troposphere, for most of the hygrometers except for (18) and (21) which still show evidence of a wet bias at the higher altitudes. If the 300 mb cutoff is combined with the full screening, the scatter plots appear as shown in Fig. 5.21. The global mean mixing ratio is shown in Fig. 5.22. The wet bias of the goldbeaters and carbon hygristor hygrometers increase slightly in the 10 km region when compared to the results shown in Fig. 5.15. The humicap now shows a consistent wet bias over all altitudes of approximately 20%.

Incorporating isentropic trajectories into the validation process significantly increases the number of match pairs, as can be seen by comparing the number of data pairs in in Fig. 5.10 to Fig. 5.15. The spread in the SAGE II versus radiosonde scatter diagrams are approximately the same, suggesting the data pairs are appropriately matched, and the overall conclusions regarding the SAGE versus radiosonde measurement biases also do not change. While we have confirmed the results of Sec. 5.3 with a larger data set, extensive computations related to the isentropic trajectories and an ancillary data set (wind fields) were required. An alternative approach to increasing the number of data pairs is simply to expand the match radius. The point profile matching of Sec. 5.3 was repeated on the potential temperature surfaces used here with temporal correlative criteria of 6 hrs and maximum spatial separation of 250, 500, and 1000 km. The SAGE versus radiosonde scatter plots at 500 km display patterns nearly identical to those using the 250 km criteria, however the 1000 km pattern shows signs of blurring and probably represents a maximum for the spatial criteria. The global mean profiles do not change significantly for the three cases. The number of data pairs produced by the isentropic trajectory correlation technique corresponds to the number produced by a point profile matching radius of approximately 750 km. The original 250 km criteria was selected based on our knowledge of water vapor correlation distances in the troposphere. Since the scatter plots do not change appreciably between the 250 and 1000 km cases the degree of

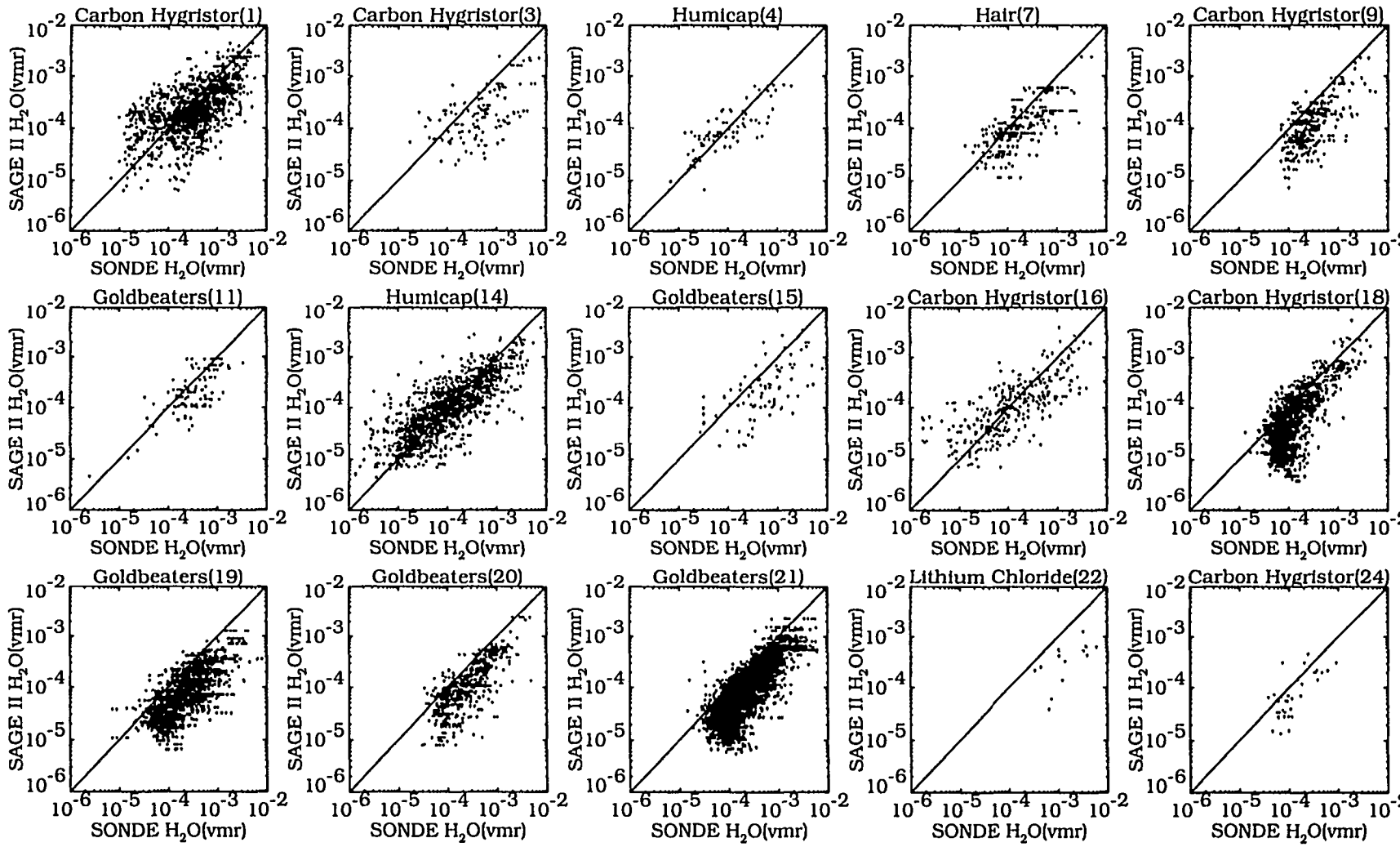


Figure 5.20: Scatter plots of SAGE II trajectory versus radiosonde paired mixing ratio values on all theta surfaces for 1987 resolved by instrument type. No screening applied to radiosonde except for 300 mb cutoff. Radiosonde hygrometer numbering follows the NMC Office Note 29 as listed in Table 5.1.

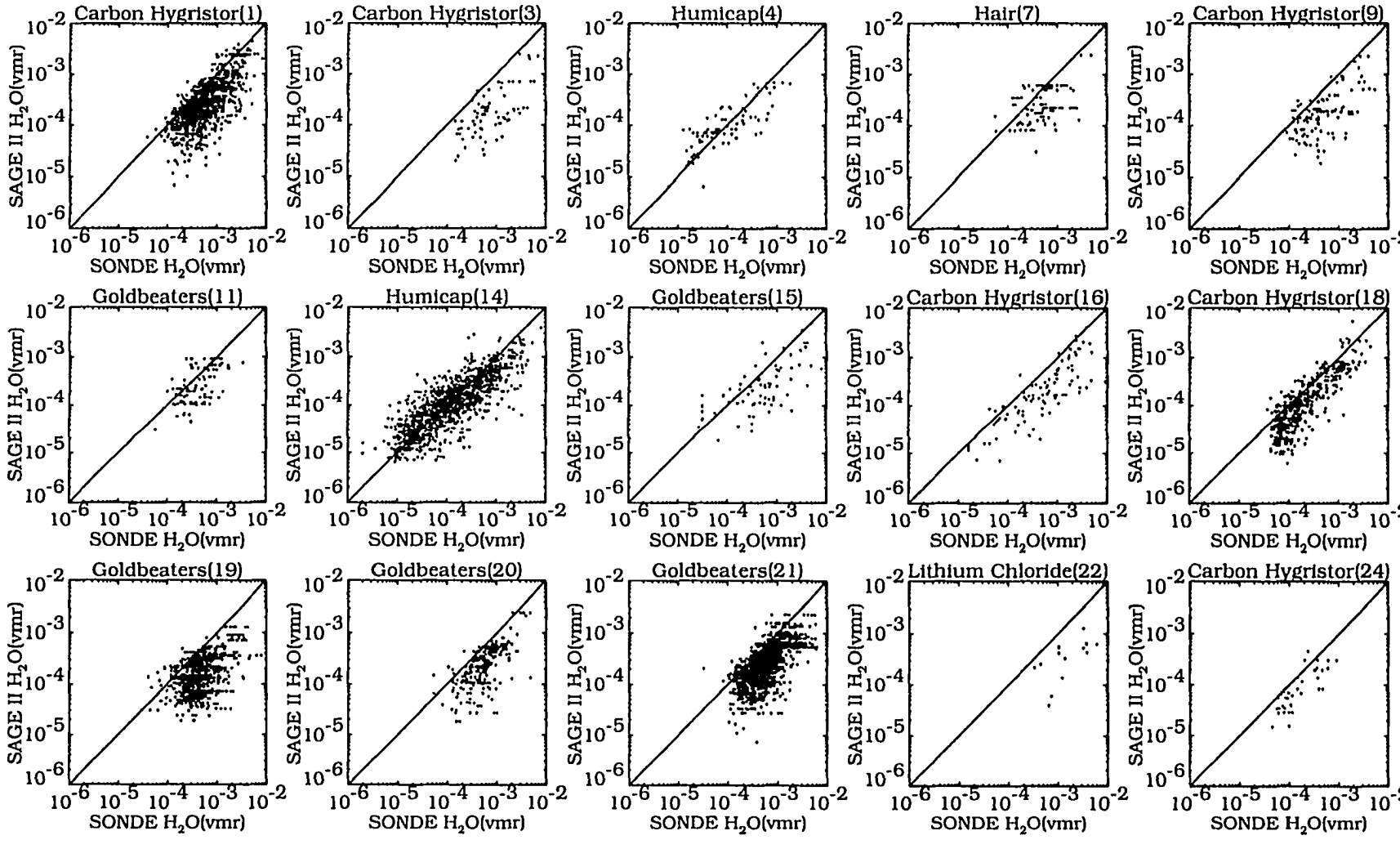


Figure 5.21: Scatter plots of SAGE II trajectory versus radiosonde paired mixing ratio values on all theta surfaces for 1987 resolved by instrument type. Full screening and 300 mb cutoff applied. Radiosonde hygrometer numbering follows the NMC Office Note 29 as listed in Table 5.1.

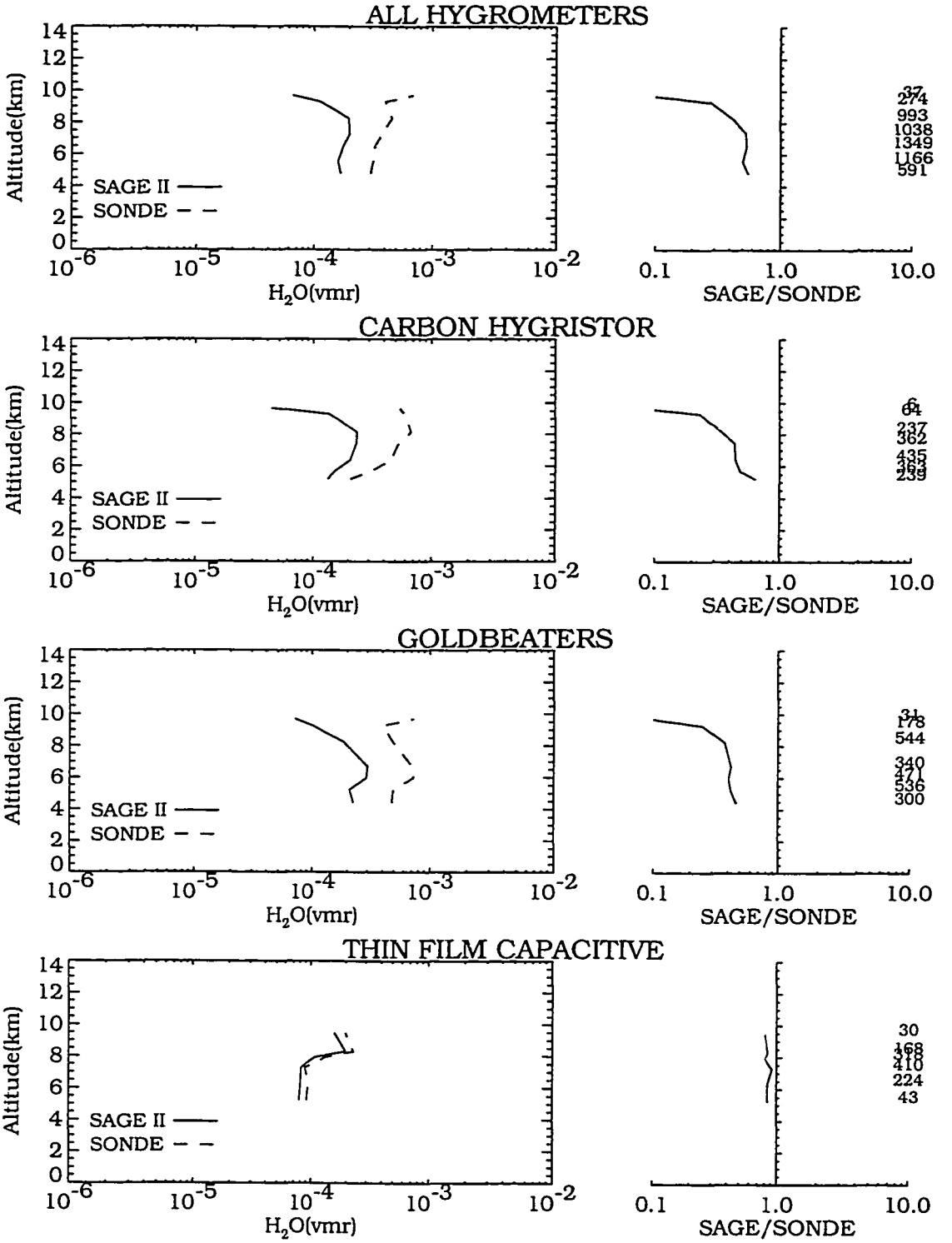


Figure 5.22: Similar to Fig. 5.15 except that full screening and 300 mb cutoff applied.

correlation between the data pairs is probably similar and probably not very high. It appears that the small scale processes actively reduce the degree of correlation on scales less than 250 km.

5.6 Section Conclusions

Results of a study comparing the SAGE II upper tropospheric water vapor to radiosonde observations from both a climatological and correlative profile pair viewpoint have been presented. SAGE II provides clear sky land/ocean observations while the radiosondes provide mostly land observations under clear/cloudy conditions. The observational biases of both instruments are described in detail and the effects of the biases on the correlative comparisons are assessed. A critical review of the radiosonde measurement capability was performed to establish the useful operating range and accuracy of each hygrometer type, whenever possible. Except for the thin film capacitive type, all hygrometers have poor response characteristics in the cold dry upper troposphere, thus low relative humidities will be under-sampled and climatologies derived from radiosonde observations will be biased to values higher than the true atmospheric state. In addition, reporting practices at individual radiosonde stations can range from reporting all data, whether good or bad, to inserting artificial values when the hygrometer data are suspect. We reverse-engineered the reporting practices of many of the stations and combined this with the hygrometer operating ranges to develop an automated screening criterion that was applied to the 1987 radiosonde data. This isolated most, but not all, of the questionable values so we instituted visual screening of the radiosonde profiles for the smaller set of correlative comparisons. While it is possible, perhaps likely, that subtle errors remain in the screened radiosonde data, we believe the screening improves the overall data

quality through removal of completely spurious data and places the radiosonde data on a firmer footing in the upper troposphere.

A straight comparison of monthly zonal means between SAGE II (1987), GACS (1963-1973), and radiosondes (1987) for January showed the SAGE II mixing ratios to be approximately half the level of either radiosonde climatology. We found that this holds true for the remainder of 1987 and other years. Although we do not have a good estimate of the possible size of the clear sky to clear/cloudy sky water vapor differences, we believe this climatological difference was suspiciously large and might be the result of additional observational or instrumental biases. Unable to identify a single error source or combination of error sources that would satisfactorily explain a significant fraction of the difference we then decided to use the radiosonde data as correlative observations.

From the 1987 SAGE II and radiosonde data we found slightly more than 800 coincidences satisfying our correlative criterion of spatial and temporal separations less than 250 km and 6 hours, respectively. This effectively limits the correlative profiles to clear sky, land observations. Profile pair comparisons in some cases showed excellent agreement but in many cases the comparisons showed significant disagreements. Examination of the water vapor distribution at constant altitude developed from the radiosondes showed strong variations over horizontal distances smaller than the SAGE II resolution.

Annual zonal means based on the correlative profile pairs still displayed large differences in most latitude bands. Recalculating the SAGE II zonal means over land produced small changes relative to the land/ocean zonal means. Calculating a global annual mean, grouped by hygrometer type, revealed the source of most of the remaining bias to lie with the hygrometers. In a climatological or averaged sense, SAGE II agrees best with the thin film capacitive hygrometer, currently thought

to be the most accurate and responsive hygrometer in operational use. Of course the fact that two different instruments report the same value does not guarantee an accurate observation was made, but it is encouraging and improves confidence in both.

Isentropic trajectory calculations were used to expand the number of data pairs and confirmed the results from the profile pair comparisons. This work represents the first application of trajectory calculations in the validation of large satellite data sets. Although the approach works, equivalent results for water vapor may be achieved by simply expanding the spatial matching distance. It may be possible to further increase the number of data pairs by increasing the trajectory travel time without decreasing the quality of the matches but we have not attempted that. Incorporating trajectories into the validation process may be more effective for species whose mixing ratios are more conserved than water vapor.

As pointed out in Elliot and Gaffen (1991) the historical record of water vapor in the U.S. is not homogeneous. This work provides another picture of radiosonde non-homogeneity, in this case the global data set for a single year. If the thin film capacitive hygrometer continues to supplant other hygrometers, as it seems to be doing in the Western hemisphere, a rather strong trend to a dryer upper troposphere will appear in the radiosonde archives. This work also suggests a similar but smaller shift will occur in the middle troposphere. It also follows that radiosonde based global maps of upper tropospheric water may be skewed because of geographic grouping by country or continent.

SECTION 6

RESULTS AND CONCLUSIONS

Increasing recognition of the importance of atmospheric water vapor to global warming issues has led in recent years to the development of new remote sensing experiments and reanalysis of existing measurements with new techniques and algorithms. An improved understanding of the SAGE II water vapor measurements is realized in this study through the progressive steps of algorithm development, error analysis and measurement validation.

The formal error analysis of Rodgers was applied to the SAGE II water vapor channel. The inverse model was split into two sub-models, mirroring the two-step process consisting of interfering species signal removal and inversion of slant path absorptances to water vapor mixing ratio profile. One advantage to splitting the inverse model is that it allows comparison of estimated slant path absorption random errors to actual errors. Error terms were reformulated to fit the formal error analysis framework and several new sources of error were considered.

Radiative transfer errors were examined in detail. It was shown that substituting the EGA approximation for line by line calculations resulted in a small error. Large errors were shown to result from ignoring the water vapor continuum and water absorption in other channels. The archived data was processed with water vapor cross-sections for only channel 2 that were based on the HITRAN 82 spectroscopic line parameters and did not include continuum absorption. It was shown that the archived data may be too low in the stratosphere by 14% and too high in the troposphere by 23% if our current knowledge of water vapor absorption better represents the true atmospheric water vapor absorption. That is, the absorption cross-sections are calculated for channels 1, 2, and 3 using HITRAN 92 and the continuum is

included in all three channels. It was pointed out that the tropospheric bias may be balanced by the systematic model error.

Considerable effort was expended to limit and remove the effects of geophysical variations in the water slant path absorption. A least squares segmented fit was determined to be the best approach to characterizing the random noise at this stage of the analysis. It was shown from analysis of the fitting residuals that the estimated random errors in the lower stratosphere were much lower than estimated. Correction factors were developed and included in the error budget. Interlevel error correlations were not found in the stratosphere and only weak hints of correlations were found in the troposphere.

Estimates of systematic and random null space errors were developed from tropospheric radiosonde data and SAGE II stratospheric data for the operational inversion. The systematic error was small because the SAGE II monthly zonal mean water vapor was used for the mixing ratio reference state. The tropospheric random error was significant, ranging from 10% to 25%, and results from the natural variability of water in that region. Stratospheric random errors are 3-10% at 18-21 km and 15-22% at 50 km.

The formal error analysis was used to study the error characteristics of four other inversion algorithms in addition to the operational. Two of these were developed specifically for this work. The first two are the commonly used onion peel and Mill-Drayson algorithms. For the third algorithm the Mill-Drayson algorithm was modified to include stratospheric mixing ratio smoothing (similar to the operational) and for the fourth, a sparse vertical grid non-linear least squares method was developed. Error characteristics of the smoothed Mill-Drayson were shown to be nearly identical to those of the operational method. The sparse grid contribution function was found to be insensitive to grid point density and computationally intensive.

The Mill-Drayson contribution function and averaging kernels are identical to the onion peel, yet Monte Carlo simulations with realistic measurement noise levels showed the Mill-Drayson profiles displayed much smaller variances than the onion peel. This variance reduction is termed implicit smoothing and differs in nature from the explicit slant path absorption and mixing ratio smoothing. To explain implicit smoothing the iterative solution method was examined in detail for each of the algorithms. The convergence criteria can be set artificially small, to reduce systematic errors due to the solution method, without influencing the level of random errors for all algorithms except the Mill-Drayson. It was shown that doing this in the Mill-Drayson inversion increased the random errors to onion peel levels. Implicit smoothing can be significant when the iterative process proceeds from low noise to high noise levels. In cases where this progression is reversed, implicit smoothing does not occur and the random error levels correspond to onion peel levels.

SAGE II data from 1987 and 1990 were inverted with the onion peel, Mill-Drayson, and smoothed Mill-Drayson algorithms. Changes in profile standard deviations and profile smoothness from algorithm to algorithm followed that expected from the formal error analysis and Monte Carlo analysis.

Climatological and correlative comparisons of the SAGE II upper tropospheric water vapor to radiosonde hygrometer data for 1987 were performed. The climatological comparisons show the SAGE II measurements are approximately half of the radiosonde measurements. Since this difference appeared to be greater than what might be expected from clear sky versus clear/cloudy sky sampling differences, the radiosonde data set was used in correlative comparisons. First however, the observational capabilities of each hygrometer was reviewed and station reporting procedures examined in order to develop data screening methods for the hygrometers. From a single year of observations, 1987, 800 profile pairs satisfying the correlative criterion

were found. Individual profile pair comparisons displayed a range of agreement from very good to very poor.

Zonal and global means calculated from the coincident pairs continued to show a dry bias in the SAGE II observations relative to radiosonde. The systematic bias was shown to be a wet bias present in all hygrometer types except for the thin film capacitive. Agreement between SAGE II and the thin film capacitive was found to be quite good and the cause of the moist bias appears to be the slow response times of the other hygrometers.

Of 45,000 to 50,000 radiosonde reports and 900 SAGE II measurements typically obtained each month only 65 coincidences were found each month on average. Zonal and temporal relationships in the comparisons cannot be established with a data set this small. Also, we suspected that the high degree of scatter in the SAGE versus radiosonde scatter plots and poor profile comparisons might be caused by comparing observations from air masses with different time histories and ambient conditions. To increase the number of coincidences and possibly confine both observations to the same air mass, isentropic trajectories were incorporated in the validation procedure. Trajectories were calculated forward and backward over 24 hrs, starting at the SAGE subtangent locations on potential temperature surfaces spanning the upper troposphere and lower stratosphere. The number of coincident data pairs increased dramatically due to the increased horizontal coverage but scatter in the paired data did not decrease. An alternative approach, expanding the spatial correlative criteria in the profile pairs to 750 km also increased the number of data pairs by about the same amount. The quality of the data pairing, judged from the scatter diagram, was similar to the trajectory pairs. Including trajectories in the matching process does not seem to offer much of an improvement over the standard approach to correlative studies. The slight improvement provided by this approach must be weighed against the burden of acquiring global wind fields, developing software and

performing the trajectory calculations. However, correlative data developed from both the isentropic trajectory matching technique and from expanding the spatial criteria to 750 km confirms the results of the much smaller data set based on the 250 km criteria.

REFERENCES

Adler-Golden, S., Lee, J., and Goldstein, N., 1992, "Diode Laser Measurements of Temperature Dependent Line Parameters for Water Vapor Near 820 nm," *J. Quant. Spectrosc. Radiat. Transfer*, Vol. 48, pp. 527-535.

Allario, F., Hoell, J. M., Katzberg, S. J., and Larsen, J. C., 1980, "Measurements of Stratospheric Trace Constituents by Laser Heterodyne Spectroscopy," *Applied Physics*, Vol. 23, pp. 47-56.

Allario, F., Katzberg, S. J., and Larsen, J. C., 1980, "Sensitivity Studies and Laboratory Measurements for the Laser Heterodyne Spectrometer Experiment," Heterodyne Systems and Technology Conference, March 25-27, Williamsburg, VA, NASA CP-2138, pp. 221-240.

Anderson, P. S., 1995, "Mechanism for the Behavior of Hydroactive Materials Used in Humidity Sensors," *J. Atmos. Oceanic Technol.* Vol. 12, pp. 662-667.

Anderson, S. M., Maeder, J., and Mauersberger, K., 1991, "Effect of Isotopic Substitution on the Visible Absorption Spectrum of Ozone," *J. Chem. Phys.*, Vol. 94, pp. 6351-6357.

Angell, J. K., Elliot, W. P., and Smith, M. E., 1984, "Tropospheric Humidity Variations at Brownsville, Texas and Great Falls, Montana, 1958-80," *J. Climate and Appl. Meteorol.*, Vol. 23, pp. 1286-1295.

Antikainen, V. and Hyvonen, V. 1983, "The Accuracy of Vaisala RS 80 Radiosonde, Proceedings of Fifth Symposium Meteorological Observations and Instrumentation," pp. 134-140, April 11-15, Toronto, Canada.

Armstrong, B. H., 1967, "Spectrum Line Profiles, The Voigt Function," *J. Quant. Spectrosc. Radiat. Transfer*, Vol. 7, pp. 61-88.

Austin, J. and Tuck, A. F., 1985, "The Calculation of Stratospheric Air Parcel Trajectories using Satellite Data," *Q. J. R. Meteorol. Soc.*, Vol. 111, pp. 279-307.

Barnett, T. P., Schlesinger, M. E., and Jiang, X., 1991, Workshop Proceedings , *Greenhouse Gas Induced Climate Change: A Critical Appraisal of Simulations and Observations*, M. Schlesinger, Ed., Elsevier, Amherst, Mass., pp. 537-558.

Benedict, W. S., and Kaplan, L. D., 1959, "Calculation of Line Widths in H_2-N_2 Collisions," *J. Chem. Phys.*, Vol. 30, pp. 388-399.

Bohren, C. F., and Huffman, D. R., 1983, "Absorption and Scattering of Light by Small Particles," John Wiley and Sons, New York, New York.

Brown, K. M., and Dennis, J. E., 1972, "Derivative Free Analogues of the Levenberg-Marquardt and Gauss Algorithms for Nonlinear Least Squares Approximations," *Numerische Mathematik*, Vol. 18, pp. 289-297.

Carlotti, M., 1988, "Global-fit Approach to the Analysis of Limb-Scanning Atmospheric Measurements," *Applied Optics*, Vol. 27, pp. 3250-3254.

Chahine, M. T., 1972, "A General Relaxation Method for Inverse Solution of the Full Radiative Transfer Equation," *J. Atmos. Sci.*, Vol. 29, pp. 741-747.

Chevillard, J.-P., Mandin, J.-Y., Flaud, J.-M., and Camy-Peyret, C., 1989, "$H_2^{16}O$: Line Positions and Intensities between 9500 and 11500 cm^{-1}. The Interacting Vibrational States (041), (220), (121), (022), (300), (201), (102), and (003)," *Can.J.Phys.*, Vol. 67, pp. 1065-1084.

Chiou, E. W., McCormick, M. P., McMaster, L. R., Chu, W. P., Larsen, J. C., Rind, R., and Oltmans, S., 1993, " Intercomparison of Stratospheric Water Vapor Observed by Satellite Experiments: Stratospheric Aerosol and Gas Experiment II Versus Limb Infrared Monitor of the Stratosphere and Atmospheric Trace Gas Molecule Spectroscopy," *J. Geophys. Res.*, Vol. 98, pp. 4875-4887.

Chiou, E. W., Remsberg, E. E., Rodgers, C. D., Munro, R., Bevillacqua, R. M., McCormick, M. P., and Russell, J. M., 1996, "Proposed Reference Model for Middle Atmosphere Water Vapor," *Advances in Space Research*, Vol. 18, pp 59-89.

Chu, W. P., and McCormick, M. P., 1979, "Inversion of Stratospheric Aerosol and Gaseous Constituents from Spacecraft Solar Extinction Data in the .38-1.0 μm Wavelength Region," *Applied Optics*, Vol. 18, pp. 1404-1413.

Chu, W. P., 1983, "Calculations of Atmospheric Refraction for Spacecraft Remote-Sensing Applications," *Applied Optics*, Vol. 21, pp. 721-725.

Chu, W. P., McCormick, M. P., Lenoble, J., Brogniez, C., and Pruvost, P., 1989, "SAGE II Inversion Algorithm," *J. Geophys. Res.*, Vol. 94, pp. 8339-8351.

Chu, W. P., Chiou, E. W., Larsen, J. C., Thomason, L. W., Rind, D., Buglia, J. J., Oltmans, S., McCormick, M. P., and McMaster, L. R., 1993, "Algorithms and Sensitivity Analyses for Stratospheric Aerosol and Gas Experiment II Water Vapor Retrieval," *J. Geophys. Res.*, Vol. 98, pp. 4857-4866.

Clough, S. A., and Kneizys, F. X., 1979, "Convolution Algorithm for the Lorentz Function," *Applied Optics*, Vol. 18, pp. 2329-2333.

Clough, S. A., Kneizys, F. X., Rothman, L. S., and Gallery, W. O., 1981, "Atmospheric Spectral Transmittance and Radiance: FASCOD1B," Society of Photo-Optical Instrumentation Engineers, SPIE Vol. 277, Atmospheric Transmission, pp. 152-166.

Clough, S. A., Kneizys, F. X., Anderson, G. P., Shettle, E. P., Chetwynd, J. H., Abreu, L. W., 1989a, " FASCOD3: Spectral Simulation," IRS'88: Current Problems in Atmospheric Radiation, Proceedings of the 1988 International Radiation Symposium, Editors J.Lenoble and J-F. Geleyn, Deepak Publishing, Hampton, Va.

Clough, S. A., Worsham, R. D., Smith, W. L., Revercomb, H. E., Knuteson, R. O., Woolf, H. W., Anderson, G. P., Hoke, M. L., Kneizys, F. X., 1989b, "Validation of Fascode Calculations with HIS Spectral Radiance Measurements," IRS'88: Current Problems in Atmospheric Radiation, Proceedings of the 1988 International Radiation Symposium, Editors J.Lenoble and J-F. Geleyn, Deepak Publishing, Hampton, Va.

Clough, S. A.,Kneizys, F. X., and Davies, R. W., 1989c, "Line Shape and the Water Vapor Continuum," *Atmospheric Research*, Vol. 23, pp. 229-241.

Connor, B. J. and Rodgers, C. D., 1989, "A Comparison of Retrieval Methods: Optimal Estimation, Onion-Peeling, and a Combination of the Two," Proceedings on Advances in Remote Sensing Retrieval Methods, Williamsburg, Va., December 15-18, 1987, published as RSRM 87 Advances in Remote sensing Retrieval Methods, A.Deepak, H.E. Fleming, and J.S. Theon editors, A. Deepak Publishing, Hampton, Va.

Cunnold, D. M., Chu, W. P., Barnes, R. A., McCormick, M. P., and Veiga, R. E., 1989, "Validation of SAGE II Ozone Measurements," *J. Geophys. Res.*, Vol. 94, pp. 8447-8460.

Danielson, E. F., 1961, "Trajectories: Isobaric, Isentropic and Actual," *J. of Meteorol.*, Vol. 18, pp. 479-486.

Dave, J. V., 1968, Subroutines for Computing the Parameters of Electromagnetic Radiation Scattered by a Sphere, IBM Scientific Center, Palo Alto, California, Report No. 320-3237, May 1968.

Davies, R. W. and Oli, B. A., 1978, "Theoretical Calculations of H_2O Linewidths and Pressure Shifts: Comparison of the Anderson Theory with Quantum

Many-Body Theory for N_2 and Air Broadened Lines," *J. Quant. Spectrosc. Radiat. Transfer*, Vol. 20, pp.95-111.

Delbouille, L., and Roland, G., 1963, "Photometric Atlas of the Solar Spectrum from λ 7498 to λ 12016," *Mémoires de la Société Royale Des Sciences de Liège*, Special Volume 4.

Del Genio, A. D., Lacis, A. A., and Ruedy, R.A., 1991, "Simulations of the Effect of a Warmer Climate on Atmospheric Humidity," *Nature*, Vol. 351, pp. 382-385.

Drayson, S. R., 1976, "Rapid Computation of the Voigt Profile," *J. Quant. Spectrosc. Radiat. Transfer*, Vol. 16, pp. 611-614.

Edlén, B., 1966, "The Refractive Index of Air," *Metrologia*, Vol 12, pp. 71-80.

Elliot, W. P., Smith, M. E., and Angell, J. K., 1991, "On Monitoring Tropospheric Water Vapor Changes using Radiosonde Data," Workshop Proceedings , *Greenhouse Gas Induced Climate Change: A Critical Appraisal of Simulations and Observations*, M. Schlesinger, Ed., Elsevier, Amherst, Mass., pp. 311-328.

Elliot, W. P., and Gaffen, D. J. 1991, "On the Utility of Radiosonde Humidity Archives for Climate Studies," *Bull. Am. Meteorol. Soc.*, Vol. 72, pp. 1507-1520.

Fairlie, T. D. A., "Three Dimensional Transport Simulations of the Dispersal of Volcanic Aerosol from Mount Pinatubo," Masters Thesis, Old Dominion University, August, 1993.

Gaffen D., J., Barnett, T. P., and Elliot, W. P., 1991, "Space and Time Scales of Global Tropospheric Moisture," *J. of Climate*, Vol. 4, pp. 898-1008.

Gaffen, D., 1993, "Historical Changes in Radiosonde Instruments and Practices," Instruments and Observing Methods Rep. No. 50, WMO/TD-No.541, World Meteorological Organization, 123 pp.

Gamache, R. R., and Rothman, L. S., 1988, "Temperature Dependence of N_2- Broadened Halfwidths of Water Vapor: The Pure Rotation and ν_2 Bands," *J. Mol. Spect.*, Vol. 128, pp. 360-369.

Geller, M. A., Wu, M. F., and Gelman, M. E., 1983, "Troposphere-Stratosphere (Surface-55km) Monthly Winter General Circulation Statistics for the Northern Hemisphere- Four Year Averages," *J. Atmos. Sci.*, Vol. 40, pp. 1334-1352.

Gille, J. C., and Bailey, P. L., 1989, "Vertical Resolution and Error Components of Ozone Retrievals from Measurements of Infrared Limb Emission," Proceedings

on Advances in Remote Sensing Retrieval Methods, Williamsburg, Va., December 15-18, 1987, published as RSRM 87 Advances in Remote sensing Retrieval Methods, A.Deepak, H.E. Fleming, and J.S. Theon editors, A. Deepak Publishing, Hampton, Va.

Giver, L. P., Gentry, B., Schwemmer, G., and Wilkerson, T. D., 1982, "Water Absorption Lines, 931-961 nm: Selected Intensities, N_2-Collision Broadening Coefficients, Self-Broadening Coefficients, and Pressure Shifts in Air," *J. Quant. Spectrosc. Radiat. Transfer*, Vol. 27, pp. 423-436.

Goody, R. M., 1964, Atmospheric Radiation, I Theoretical Basis, Oxford at the Clarendon Press.

Gordley, L. L. and Russell, J. M., 1981, "Rapid Inversion of Limb Radiance data using an Emissivity Growth Approximation," *Applied Optics*, Vol. 20, pp. 807-813.

Harris, J. M., and Kahl J. K., 1990, "A Descriptive Atmospheric Transport Climatology for the Mauna Loa Observatory Using Clustered Trajectories," *J. Geophys. Res.*, Vol. 95, pp. 13,651-13,667.

Hense, A., Krahe, P., and Flohn, H., 1988, "Recent Fluctuations of Tropospheric Temperature and Water Vapour Content in the Tropics," *Meteorol. Atmos. Phys.*, Vol. 38, pp. 215-227.

Herman, B. M., Caudill, T. R., Flittner, D. E., Thome, K. J., and Ben-David, A., 1995, " Comparison of the Gauss-Seidel Spherical Polarized Radiative Transfer Code With Other Radiative Transfer Codes," *Applied Optics*, Vol. 34, pp. 4563-4572.

Hooper, A. H., 1975, "Upper-Air Sounding Studies, Volume I: Studies on Radiosonde Performance," WMO/No. 394, Technical Note No. 140.

Hsu, C. P., 1980, "Air Parcel Motions During a Numerically Simulated Sudden Stratospheric Warming," *J. Atmos. Sci.*, Vol. 37, pp. 2768-2792.

Hui, A. K., Armstrong,B.H. and Wray,A.A., 1978, "Rapid Computation of the Voigt and Complex Error Functions," *J. Quant. Spectrosc. Radiat. Transfer*, Vol. 19, pp. 509-516.

Humlicek, J., 1979, "An Efficient Method for Evaluation of the Complex Probability Function: The Voigt Function and its Derivatives," *J. Quant. Spectrosc. Radiat. Transfer*, Vol. 21, pp. 309-313.

Junge, C. E., Chagnon, C. W., and Manson, J. E., 1961, "Stratospheric Aerosols," *J. Meteorology*, Vol. 18, pp. 81-108.

Kasten, F., 1968, "Falling Speed of Aerosol Particles," *J. Appl. Meteorol.*, Vol. 7, pp. 944-947.

Kitchen, M., 1989, "Representativeness Errors for Radiosonde Observations," *Q. J. R. Meteorol. Soc.*, Vol. 115, pp. 673-700.

Klim, A., 1981, "A Comparison of Methods for the Calculation of Voigt Profiles," *J. Quant. Spectrosc. Radiat. Transfer*, Vol. 26, pp. 537-545.

Knudsen, B. M., and Carver, G. D., 1994, "Accuracy of the Isentropic Trajectories calculated for the EASOE Campaign," *Geophys. Res. Lett.*, Vol. 21, pp. 1199-1202.

Kondrat'ev,K.Ya, 1973, "Radiation Characteristics of the Atmosphere and the Earth's Surface," NASA TTF-678, Amerind Publishing, New Delhi, India.

Larsen, J. C., Chiou, E. W., Chu, W. P., McCormick, M. P., McMaster, L. R., Oltmans, S., and Rind, D., 1993, "A Comparison of the SAGE II Tropospheric Water Vapor to Radiosonde Measurements," *J. Geophys. Res.*, Vol. 98, pp. 4897-4917.

Lashof, D. A., 1989, "The Dynamic Greenhouse: Feedback Processes That May Influence Future Concentrations of Atmospheric Trace Gases and Climate Change," *Climate Change*, Vol. 14, pp. 213-242.

Lenoble, J., "Radiative Transfer in Scattering and Absorbing Atmospheres: Standard Computational Procedures," Chapter 3, pp. 247-255, Deepak Publishing, Hampton, Va., 1985.

Linz, P., 1985, "Analytical and Numerical Methods for Volterra Equations," SIAM, Philadelphia.

Lyjak, L. V., and Smith, A. K., 1987, "Lagrangian Mean Circulation in the Stratosphere," *J. Atmos. Sci.*, Vol. 44, pp. 2252-2266.

Mauldin, L. E., McCormick, M. P., McMaster, L. R., and Vaughan, W. R., 1985, "The Stratospheric Aerosol and Gas Experiment II (SAGE II) Design and In-Orbit Performance," *Proceedings, SPIE 2nd Intl. Tech. Sympos. on Optical and Electro-Optical Appl. Sci. and Eng.*, Cannes, France, Paper No. 589018.

Mathews, D. A., 1965a, "Review of the Lithium Chloride Radiosonde Hygrometer," *Humidity and Moisture*, Vol. 1, pp. 219-227, A. Wexler editor, Reinhold, New York.

Mathews, D. A., 1965b, "Some Research on the Lithium Chloride Radiosonde Hygrometer and a guide for making it," *Humidity and Moisture*, Vol. 1, pp. 228-247, A. Wexler editor, Reinhold, New York.

McCormick, M. P., Swissler, T. J., Chu, W. P., and Fuller, W. H., 1978, "Post-Volcanic Stratospheric Aerosol Decay as Measured by Lidar," *J. Atmos. Sci.*, Vol. 35, pp. 1296-1303.

McCormick, M. P., Hamill, P., Pepin, T. J., Chu, W. P., Swissler, T. J., and McMaster, L. R., 1979, "Satellite Studies of the Stratospheric Aerosol," *Bull. Am. Meteorol. Soc.*, Vol. 60, pp. 1038-1046.

McCormick, M. P., Zawodney, J. M., Veiga, R. E., Larsen, J. C., and Wang, P. H., 1989, "An Overview of SAGE I and II Ozone Measurements," *Planet.Space Sci.*, Vol. 37, pp. 1567-1586.

McKee, T. B., Whitman, R. I., and Lambiotte, J. J., 1969, "A Technique to Infer Atmospheric Water Vapor Mixing Ratio from Measured Horizon Radiance Profiles," NASA TN D-5252.

McKeen, S. A., Liu, S. C., and Kiang, C. S., 1984, "On the Chemistry of Stratospheric SO_2 from Volcanic Eruptions," *J. Geophys. Res.*, Vol. 89, pp. 4873-4881.

McMaster, L. R., 1986, "Stratospheric Aerosol and Gas Experiment (SAGE II)," *Proceedings, Sixth Conference on Atmospheric Radiation*, Williamsburg, Va., May 13-16, pp. j46-j48.

McMurdie, L. A. and Katsaros, K. B., 1991, "Satellite-Derived Integrated Water Vapor Distribution in Oceanic Midlatitude Storms: Variations with Region and Season," *Monthly Weather Review*, Vol. 119, pp. 589-605.

Merrill, J. T., Bleck, R, and Boudra, D., 1986, "Techniques of Lagrangian Trajectory Analysis in Isentropic Coordinates," *Monthly Weather Review*, Vol. 114, pp. 571-581.

Mill, J. D., and Drayson, S. R., 1978, "A Nonlinear Technique for Inverting Limb Absorption Profiles," Developments in Atmospheric Science, 9, Remote Sensing of the Atmosphere: Inversion Methods and applications, A.L.Fymat and V.E.Zuev eds., Elsevier, pp. 123-135.

Muller, H. G., 1965, "Humidity Sensors from Natural Materials, Humidity and Moisture," Vol. 1, pp. 574-577, A. Wexler editor, Reinhold, New York.

Oltmans, S. J., and Hofmann, D. J., 1995, "Increase in Lower Stratosphere Water Vapour at a Mid-latitude Northern Hemisphere Site from 1981 to 1994," *Nature*, Vol. 374, pp. 146-149.

Oort, A. H., 1983, "Global Atmospheric Circulation Statistics, 1958-1973," NOAA Professional Paper 14, 180 pp., National Oceanic and Atmospheric Administration, Rockville, Md.

Peixoto, J. P., and Oort, A. H., 1983, "The Atmospheric Branch of the Hydrological Cycle and Climate, Variations in the Global Water Budget, D.Reidel, Hingham, Ma., 5-65.

Pierluissi, J. H., Vanderwood, P. C., and Gomez, R. B., 1977, "Fast Calculational Algorithm for the Voigt Profile," *J. Quant. Spectrosc. Radiat. Transfer*, Vol. 18, pp. 555-558.

Pinnick, R. G., Rosen, J. M., and Hofmann, D. J., 1973, "Measured Light- Scattering Properties of Individual Aerosol Particles Compared to Mie Scattering Theory," *Applied Optics*, Vol. 12, pp. 37-41.

Pinnick, R. G., Rosen, J. M., and Hofmann, D. J., 1976, "Stratospheric Aerosol Measurements III: Optical Model Calculations," *J. Atmos. Sci.*, Vol. 33, pp. 304-314.

Pratt, R. W., 1985, "Review of Radiosonde Humidity and Temperature Errors," *J. Atmos. Oceanic Techn.*, Vol. 2, pp. 404-407.

Pruvost, P., Ovarlez, J., Lenoble, J., and Chu, W. P., 1993, "Comparison of Stratospheric Aerosol and Gas Experiment II and Balloon-Borne Stratospheric Water Vapor Measurements," *J. Geophys. Res.*, Vol. 98, pp. 4889-4896.

"Radiosonde Observations," Federal Meteorological Handbook No.3, June 1, 1981.

Richner, H. and Philips, P. D., 1982, "The Radiosonde Intercomparison SONDEX Spring 1981, Payerne," *PAGEOPH*, Vol. 120, pp. 852-1198.

Rind, D., Chiou, E. W., Chu, W. P., Oltmans, S., Lerner, J., Larsen, J. C., McCormick, M. P., and McMaster, L. R., 1993, "SAGE II Water Vapor Observations," *J. Geophys. Res.*, Vol. 98, pp. 4835-4856.

Rinsland, C. P., Goldman, A., Murcray, F. J., Murcray, D. G., Smith, M. A. H., Seals, R. K., Larsen, J. C., and Rinsland, P. L., 1983, "Stratospheric Tempera-

ture Profile from Balloon-borne Measurements of the 10.4μm band of CO_2," *J. Quant. Spectrosc. Radiat. Transfer*, Vol. 30, pp. 327-334.

Rinsland, C. P., Gunson, M. R., Zander, R., and Lopez-Puertas, M., 1992, "Middle and Upper Atmosphere Pressure-Temperature Profiles and the Abundance of CO_2 and CO in the Upper Atmosphere From ATMOS/Spacelab 3 Observations," *J. Geophys. Res.*, Vol. 97, pp. 20479-20495.

Rodgers, C. D., 1990, "Characterization and Error Analysis of Profiles Retrieved from Remote Sounding Measurements," *J. Geophys. Res.*, Vol. 95, pp. 5587-5595.

Rolph, G. D., and Draxler, R. R., 1990, "Sensitivity of Three- Dimensional Trajectories to the Spatial and Temporal Densities of the Wind Field," *J. Appl. Meteorol.*, Vol. 29, pp. 1043-1054.

Ross, R. J., and Elliot, W.P., 1996, "Tropospheric Water Vapor Climatology and Trends Over North America: 1973-93," *J. of Climate*, Vol. 9, pp. 3561-3574.

Rothman, L. S., Gamache, R. R., Tipping, R. H., Rinsland, C. P., Smith, M. A. H., Benner, D. C., Devi, V. M., Flaud, J.-M., Camy-Peyret, C., Perrin, A., Goldman, A., Massie, S. T., Brown., L. R., and Toth, R. A., 1992, "The HITRAN Molecular Database: editions of 1991 and 1992," *J. Quant. Spectrosc. Radiat. Transfer*, Vol. 48, pp. 469-507.

Rothman, L. S., Gamache, R. R., Goldman, A., Brown., L. R., Toth, R. A., Pickett, H. M., Poynter, R. L., Flaud, J.-M., Camy-Peyret, C., Barbe, A., Husson, N., Rinsland, C. P., and Smith, M. A. H., 1987, "The HITRAN Database: 1986 edition," *Applied Optics*, Vol. 26, pp. 4058-4097.

Rothman, L. S., Gamache, R. R., Barbe, A., Goldman, A., Gillis, J. R., Brown., L. R., Toth, R. A., Flaud, J.-M., and Camy-Peyret, C., 1983a, "AFGL Atmospheric Absorption Line Parameters Compilation: 1982 edition," *Applied Optics*, Vol. 22, pp. 2247-2256.

Rothman, L. S., Goldman, A., Gillis, J. R., Gamache, R. R., Pickett, H. M., Poynter, R. L., Husson, N., and Chedin, A., 1983b, "AFGL Trace Gas Compilation: 1982 edition," *Applied Optics*, Vol. 22, pp. 1616-1627.

Rothman, L. S., 1981, "AFGL Atmospheric Absorption Line Parameters Compilation: 1980 edition," *Applied Optics*, Vol. 20, pp. 791-795.

Russell, J. M. and Gordley, L. L., 1979, "The Simultaneous Inference of Stratospheric NO_2-H_2O and HNO_3-CF_2CL_2 Using Limb Sounding Radiometry," *J.*

Atmos. Sci., Vol. 36, pp. 2259-2266.

Russell, J. M., Gille, J. C., Remsburg, E. E., Gordley, L. L., Bailey, P. L., Fischer, H., Girard, A., Drayson, S. R., Evans, W. F. J., and Harries, J. E., 1984, "Validation of Water Vapor Results Measured by the Limb Infrared Monitor of the Stratosphere Experiment on NIMBUS 7," *J. Geophys. Res.*, Vol. 89, pp. 5115-5124.

Russell, P. B., Swissler, T. J., McCormick, M. P., Chu, W. P., Livingston, J. M., and Pepin, T. J., 1981, "Satellite and Correlative Measurements of the Stratospheric Aerosol I: An Optical Model for Data Conversions," *J. Atmos. Sci.*, Vol. 38, pp. 1279-1294.

Schmidlin, F. J., 1988, "WMO International Radiosonde Intercomparison Phase II, 1985," WMO/TD-No. 312, Instruments and Observing Methods Report No. 29.

Schreier, F., 1992, "The Voigt and Complex Error Function: A Comparison of Computational Methods," *J. Quant. Spectrosc. Radiat. Transfer*, Vol. 48, pp. 743-762.

Shettle, E. P., and Fenn, R. W., 1979, "Models for the Aerosols of the Lower Atmosphere and the Effects of Humidity Variations on Their Optical Properties," AFGL-TR-79-0214, Environmental Research Papers, No. 676, Air Force Geophysics Laboratory, Hanscom AFB, Ma., Sept. 20, 1979.

Smithsonian Meteorological Tables, 1951, Sixth Revised Edition, Smithsonian Institution, Washington, D.C.

Swensson, J. W., Benedict, W. S., Delbouille, L., and Roland, G., 1970, "The Solar Spectrum from λ 7498 to λ 12016," *Institute d'Astrophysique de l'Université de Liège*, Special Volume 5.

Thomason, L. W., 1991, "A Diagnostic Stratospheric Aerosol Size Distribution Inferred from SAGE II Measurements," *J. Geophys. Res.*, Vol. 96, pp. 22,501-22,508.

Tiwari, S. N., 1978, "Models For Infrared Atmospheric Radiation," *Advances in Geophysics*, Vol. 20, Academic Press, New York.

Toon, O. B., and Pollack, J. B., 1976, "A Global Average Model of Atmospheric Aerosols for Radiative Transfer Calculations," *J. Appl. Meteorol.*, Vol. 15, pp. 225-246.

Turco, R. P., Whitten, R. C., and Toon, O. B., 1982, "Stratospheric Aerosols: observations and theory," *Rev. Geophys. Space Res.*, Vol. 20, pp. 233-279.

Twitty, J. T., Rarig, P. L., and Thompson, R. E., 1980, "A Comparison of Fast Codes for the Evaluation of the Voigt Profile Function," *J. Quant. Spectrosc. Radiat. Transfer*, Vol. 24, pp. 529-532.

Twomey, S., 1977, "Introduction to the Mathematics of Inversion in Remote Sensing and Indirect Measurements," Developments in Geomathematics 3, Elsevier Scientific Publishing Co., New York, NY.

"U.S. Standard Atmosphere, 1976," 1976, NOAA-S/T 76-1562, Superintendent of Documents, US Government Printing Office, Washington, D.C., 20402.

Van De Hulst,H.C., 1957, Light Scattering by Small Particles, New York, Wiley.

Van Maanen, J., 1981, "Objective Analysis of Humidity by the Optimum Interpolation Method," *Tellus*, Vol. 33, pp. 113-122.

Varanasi, P., 1994, "Thermal Infrared Absorption of Atmospheric Radiation by Water Vapor and its Effect on Climate," Modern Developments in Energy, Combustion and Spectroscopy, Eds. F.A. Williams, A.K. Oppenheim, D.B. Olfe, and M. Lapp, Pergamon Press, New York, NY., Chapt. 11, pp. 125-146.

Wexler, A., 1970, "Measurement of Humidity in the Free Atmosphere Near the Surface of the Earth, Meteorological Observations and Instrumentation," AMS Monographs, Vol. 11, pp. 262-282.

Wiscombe, W. J., 1979, Mie Scattering Calculations, Improvements in Techniques and Fast, Vector-Speed Computer Codes, NCAR Technical Note.

WMO, 1988, "Report of the International Ozone Trends Panel 1988," World Meteorological Organization, Global Ozone Research and Monitoring Project, Report No. 18, Vol. 1.

Yue, G. K., McCormick, M. P., and Chiou, E. W., 1991, "Stratospheric Optical Depth Observed by the Stratospheric Aerosol and Gas Experiment II: Decay of the El Chichon and Ruiz Volcanic Eruptions," *J. Geophys. Res.*, Vol. 96, pp. 5209-5219.

APPENDICES

APPENDIX A

FORMAL ERROR ANALYSIS EQUATION SUMMARY

The upper portion of Table A.1 summarizes the linearized equations for the forward model, inverse model, and transfer function governing $\mathbf{y}^m$ and $\mathbf{x}^r$. Immediately beneath each are shown the covariances $\mathbf{S}_{\mathbf{y}^m}$ or $\mathbf{S}_{\mathbf{x}^r}$, aligned by terms. In the lower portion the error equation is restated, split into systematic and random components to emphasize the dual nature of the model parameters and null space errors, followed by the error covariances $\mathbf{S}_{\mathbf{x}^r-\mathbf{x}^t}$ arising from the random components.

The covariance equations $\mathbf{S}_{\mathbf{y}^m}$ or $\mathbf{S}_{\mathbf{x}^r}$ estimate the covariances in a single profile of $\mathbf{y}^m$ or $\mathbf{x}^r$ provided the reference states are stationary. While appropriate for error studies, it must be realized these covariances are not consistent with covariances calculated directly from a set of $\mathbf{y}^m$ or $\mathbf{x}^r$ in a real experiment where the reference states may no longer be stationary due to geophysical variability. The correct form of $\mathbf{S}_{\mathbf{y}^m}$ in this situation requires an additional term $\mathbf{S}_{\mathbf{F}}$ that accounts for the covariance of $F(\overline{\mathbf{x}}, \hat{\mathbf{b}})$ caused by geophysical variations of $\mathbf{b}$.

Table A.1: Equation Summary

Model Equations

Forward Model

$$\begin{aligned} \mathbf{y}^m &= F(\mathbf{x}^t, \mathbf{b}^t) + \epsilon_\mathbf{y} = F(\overline{\mathbf{x}}, \hat{\mathbf{b}}) + \mathbf{K}_\mathbf{x}(\mathbf{x}^t - \overline{\mathbf{x}}) + \mathbf{K}_\mathbf{b}(\mathbf{b}^t - \hat{\mathbf{b}}) + \epsilon_\mathbf{y} \\ \mathbf{S}_{\mathbf{y}^m} &= \mathbf{K}_\mathbf{x}\mathbf{S}_{\mathbf{x}^t}\mathbf{K}_\mathbf{x}^T + \mathbf{K}_\mathbf{b}\mathbf{S}_{\epsilon_\mathbf{b}}\mathbf{K}_\mathbf{b}^T + \mathbf{S}_{\epsilon_\mathbf{y}} \end{aligned}$$

Inverse Model

$$\begin{aligned} \mathbf{y}^{mi} &= S(\mathbf{y}^m, \hat{\mathbf{b}}, \hat{\mathbf{c}}) = S(\overline{\mathbf{y}}^m, \hat{\mathbf{b}}, \hat{\mathbf{c}}) + \tfrac{\partial S}{\partial \mathbf{y}^m}(\mathbf{y}^m - \overline{\mathbf{y}}^m) \\ \mathbf{S}_{\mathbf{y}^{mi}} &= (\tfrac{\partial S}{\partial \mathbf{y}^m})\mathbf{S}_{\mathbf{y}^m}(\tfrac{\partial S}{\partial \mathbf{y}^m})^T = (\tfrac{\partial S}{\partial \mathbf{y}^m})[\mathbf{K}_\mathbf{x}\mathbf{S}_{\mathbf{x}^t}\mathbf{K}_\mathbf{x}^T + \mathbf{K}_\mathbf{b}\mathbf{S}_{\epsilon_\mathbf{b}}\mathbf{K}_\mathbf{b}^T + \mathbf{S}_{\epsilon_\mathbf{y}}](\tfrac{\partial S}{\partial \mathbf{y}^m})^T \end{aligned}$$

$$\begin{aligned} \mathbf{x}^r &= IA(\mathbf{y}^{mi}, \hat{\mathbf{b}}, \hat{\mathbf{c}}) = IA(\overline{\mathbf{y}}^{mi}, \hat{\mathbf{b}}, \hat{\mathbf{c}}) + \tfrac{\partial IA}{\partial \mathbf{y}^{mi}}(\mathbf{y}^{mi} - \overline{\mathbf{y}}^{mi}) \\ \mathbf{S}_{\mathbf{x}^r} &= (\tfrac{\partial IA}{\partial \mathbf{y}^{mi}})\mathbf{S}_{\mathbf{y}^{mi}}(\tfrac{\partial IA}{\partial \mathbf{y}^{mi}})^T = (\tfrac{\partial IA}{\partial \mathbf{y}^{mi}})(\tfrac{\partial S}{\partial \mathbf{y}^m})[\mathbf{K}_\mathbf{x}\mathbf{S}_{\mathbf{x}^t}\mathbf{K}_\mathbf{x}^T + \mathbf{K}_\mathbf{b}\mathbf{S}_{\epsilon_\mathbf{b}}\mathbf{K}_\mathbf{b}^T + \mathbf{S}_{\epsilon_\mathbf{y}}](\tfrac{\partial S}{\partial \mathbf{y}^m})^T(\tfrac{\partial IA}{\partial \mathbf{y}^{mi}})^T \end{aligned}$$

$$\begin{aligned} \mathbf{x}^r &= I(\mathbf{y}^m, \hat{\mathbf{b}}, \hat{\mathbf{c}}) = I(\overline{\mathbf{y}}^m, \hat{\mathbf{b}}, \hat{\mathbf{c}}) + \mathbf{D}_\mathbf{y}(\mathbf{y}^m - \overline{\mathbf{y}}^m) = I(\overline{\mathbf{y}}^m, \hat{\mathbf{b}}, \hat{\mathbf{c}}) + \mathbf{D}_\mathbf{y}\mathbf{K}_\mathbf{x}(\mathbf{x}^t - \overline{\mathbf{x}}) + \mathbf{D}_\mathbf{y}\mathbf{K}_\mathbf{b}(\mathbf{b}^t - \hat{\mathbf{b}}) + \mathbf{D}_\mathbf{y}\epsilon_\mathbf{y} \\ \mathbf{S}_{\mathbf{x}^r} &= \mathbf{D}_\mathbf{y}\mathbf{S}_{\mathbf{y}^m}\mathbf{D}_\mathbf{y}^T = \mathbf{A}_\mathbf{x}\mathbf{S}_{\mathbf{x}^t}\mathbf{A}_\mathbf{x}^T + \mathbf{A}_\mathbf{b}\mathbf{S}_{\epsilon_\mathbf{b}}\mathbf{A}_\mathbf{b}^T + \mathbf{D}_\mathbf{y}\mathbf{S}_{\epsilon_\mathbf{y}}\mathbf{D}_\mathbf{y}^T \end{aligned}$$

Transfer Function

$$\begin{aligned} \mathbf{x}^r &= \mathcal{T}(\mathbf{x}^t, \mathbf{b}^t, \hat{\mathbf{c}}) + \mathbf{D}_\mathbf{y}\epsilon_\mathbf{y} = \mathcal{T}(\overline{\mathbf{x}}, \hat{\mathbf{b}}, \hat{\mathbf{c}}) + \mathbf{A}_\mathbf{x}(\mathbf{x}^t - \overline{\mathbf{x}}) + \mathbf{A}_\mathbf{b}(\mathbf{b}^t - \hat{\mathbf{b}}) + \mathbf{D}_\mathbf{y}\epsilon_\mathbf{y} \\ \mathbf{S}_{\mathbf{x}^r} &= \mathbf{A}_\mathbf{x}\mathbf{S}_{\mathbf{x}^t}\mathbf{A}_\mathbf{x}^T + \mathbf{A}_\mathbf{b}\mathbf{S}_{\epsilon_\mathbf{b}}\mathbf{A}_\mathbf{b}^T + \mathbf{D}_\mathbf{y}\mathbf{S}_{\epsilon_\mathbf{y}}\mathbf{D}_\mathbf{y}^T \end{aligned}$$

Error Equation

$$\begin{aligned} \mathbf{x}^r - \mathbf{x}^t &= [\mathcal{T}(\overline{\mathbf{x}}, \hat{\mathbf{b}}, \hat{\mathbf{c}}) - \overline{\mathbf{x}}] + (\mathbf{A} - \mathbf{I})(\mathbf{x}^t - \overline{\mathbf{x}}) + \mathbf{D}_\mathbf{y}\mathbf{K}_\mathbf{b}(\mathbf{b}^t - \hat{\mathbf{b}}) + \mathbf{D}_\mathbf{y}\epsilon_\mathbf{y} \\ \text{Systematic} \quad & [\mathcal{T}(\overline{\mathbf{x}}, \hat{\mathbf{b}}, \hat{\mathbf{c}}) - \overline{\mathbf{x}}] + (\mathbf{A} - \mathbf{I})(\mathbf{x}^t - \overline{\mathbf{x}}) + \mathbf{D}_\mathbf{y}\mathbf{K}_\mathbf{b}(\mathbf{b}^t - \hat{\mathbf{b}}) \\ \text{Random} \quad & (\mathbf{A} - \mathbf{I})(\mathbf{x}^t - \overline{\mathbf{x}}) + \mathbf{D}_\mathbf{y}\mathbf{K}_\mathbf{b}(\mathbf{b}^t - \hat{\mathbf{b}}) + \mathbf{D}_\mathbf{y}\epsilon_\mathbf{y} \\ \mathbf{S}_{\mathbf{x}^r - \mathbf{x}^t} &= (\mathbf{A} - \mathbf{I})\mathbf{S}_{\mathbf{x}^t}(\mathbf{A} - \mathbf{I})^T + \mathbf{A}_\mathbf{b}\mathbf{S}_{\epsilon_\mathbf{b}}\mathbf{A}_\mathbf{b}^T + \mathbf{D}_\mathbf{y}\mathbf{S}_{\epsilon_\mathbf{y}}\mathbf{D}_\mathbf{y}^T \\ \mathbf{S}_{\mathbf{x}^r - \mathbf{x}^t} &= \mathbf{S}_N + \mathbf{S}_B + \mathbf{S}_E \end{aligned}$$

APPENDIX B

ERROR VARIANCES

The success of the formal error analysis in estimating inversion uncertainty depends greatly on accurate knowledge of the statistical characteristics of the model parameters and measurement error as described by the error covariance matrices $\mathbf{S}_{\epsilon_\mathbf{b}}$ and $\mathbf{S}_{\epsilon_\mathbf{y}}$. The error covariance matrices can be specified completely in terms of error variances and a correlation coefficient matrix. This Appendix describes the steps taken to improve variance estimates while Appendix C contains the development of the correlation coefficient matrix. Appendices B and C use the error terms as formulated in Chu et al.(1993) and differ from the formulation given in Sec. 4 primarily in the treatment of the temperature uncertainty. In Sec. 4 temperature uncertainties are treated explicitly while in Chu et al.(1993) they appear only in the Rayleigh error.

An estimate of $\epsilon_\mathbf{y}$ is supplied for each $\mathbf{y}^m(z_{th})$ resulting from the 1 km transmission averaging procedure. Calculation of $\mathbf{S}_{\epsilon_\mathbf{y}}$, needed for the error equation, from a group of transmission or absorption profiles is complicated by geophysical variability in $\mathbf{y}^m$ due to geophysical variability in $\mathbf{x}^t$ (water vapor) and $\mathbf{b}$ (primarily Rayleigh, ozone, and aerosol). Even assuming zero covariances for $\mathbf{S}_{\epsilon_\mathbf{y}}$ with variances provided by the estimated $\epsilon_\mathbf{y}$, one must still develop estimates of $\mathbf{S}_{\epsilon_\mathbf{b}}$ for the error equation which for some $\mathbf{b}$ may be difficult.

Another approach is to start with the intermediate measurement $\mathbf{y}^{mi}$, the signal inverted by IA, and calculate $\mathbf{S}_\mathbf{y}^{mi}$. Since the species separation process has removed much of the geophysical variability one is left with

$$\mathbf{S}_{\mathbf{y}^{mi}} \approx \left(\frac{\partial S}{\partial \mathbf{y}^m}\right)[\mathbf{K}_\mathbf{b}\mathbf{S}_{\epsilon_\mathbf{b}}\mathbf{K}_\mathbf{b}^T + \mathbf{S}_{\epsilon_\mathbf{y}}]\left(\frac{\partial S}{\partial \mathbf{y}^m}\right)^T \tag{B.1}$$

provided $\mathbf{S}_{\mathbf{x}^t}$, the covariance of $\mathbf{x}^t$, is small. At this stage in the inversion process $\mathbf{S}_{\mathbf{y}^{mi}}$ is a combination of covariances arising from model parameter uncertainties (found in the separation model S) with random measurement error and can even be used in the error equation without having to separate the two terms or even specify individual $\mathbf{S}_{\epsilon_b}$. A direct calculation of $\mathbf{S}_{\mathbf{y}^{mi}}$ from the data rather than from estimates of $\mathbf{S}_{\epsilon_b}$ and $\mathbf{S}_{\epsilon_y}$ is almost always preferable. Estimates may not accurately reflect the true uncertainty, being either too small or too large, affecting the error covariances at this stage and the final error estimates in $\mathbf{x}^r$. Some error cancellation may occur leading to a smaller $\mathbf{S}_{\mathbf{y}^{mi}}$ than one would estimate. A good estimate of the variance of $\mathbf{y}^{mi}$ is also needed to control the iteration process in IA.

To develop the best estimates for $\mathbf{S}_{\epsilon_y}$ and $\mathbf{S}_{\epsilon_b}$ (found in S), the following approach was taken. Since most of the terms in the separation model are based on channel 1 or channel 3 data, reasonable estimates of $\mathbf{S}_{\epsilon_b}$ are available which, when combined with $\mathbf{S}_{\epsilon_y}$, provide an estimate of $\mathbf{S}_{\mathbf{y}^{mi}}$. The estimated $\mathbf{S}_{\mathbf{y}^{mi}}$ are compared to the measured $\mathbf{S}_{\mathbf{y}^{mi}}$, leading to correction factors for $\mathbf{S}_{\epsilon_b}$ and $\mathbf{S}_{\epsilon_y}$.

In Appendix C it is shown that $\mathbf{S}_{\mathbf{x}^t}$ dominates the experimental error and a number of approaches developed with the goal of reducing the influence of $\mathbf{S}_{\mathbf{x}^t}$ are presented there. A segmented profile fitting technique on $\mathbf{y}^{mi}$ gave the best results which lead to a nearly diagonal "representative" correlation coefficient matrix. Segmented fitting worked well in most of the stratosphere, except for the equatorial middle stratosphere, but in the troposphere the strong variability of water vapor produced slant path absorption profiles with significant curvature that the fitting process was unable to follow.

Figure B.1 displays the ratio of the zonal mean estimated $\mathbf{S}_{\mathbf{y}^{mi}}$ (actually the square root of the mean diagonal elements σ^2) to the measured $\mathbf{S}_{\mathbf{y}^{mi}}$ for the 50° latitude bands. In the upper part of the stratosphere, the estimated and measured

errors agree closely, with the estimated slightly less than the measured. Below this, the ratio increases with decreasing altitude, indicating a significant over estimation of the error. Note the break in the curves at 30.5 km between the two fitting regions and that even at these latitudes there was one month when the fitting process broke down(50° S, ratio $\approx$ 0.1).

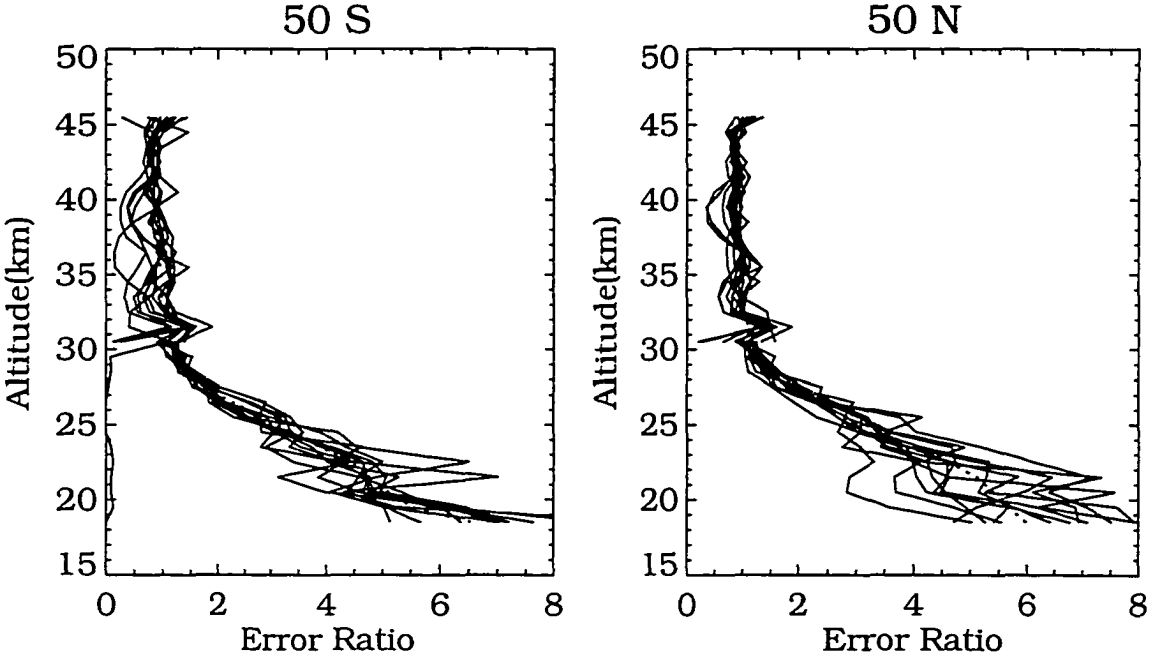


Figure B.1: Ratio of average estimated $\sqrt{\overline{\sigma^2}}$ to those from the segmented profile fitting. Each solid line corresponds to a monthly zonal mean in 1987 at 50° S and 50° N. The dotted line in each panel is an average of the monthly profiles.

The following steps were taken to correct the error covariances σ^2 of $\mathbf{S}_{\mathbf{y}^{mi}}$. In the stratosphere estimated σ^2 within a particular zone were corrected using the zonal mean error ratio curve, such as (Fig. B.1). When the error ratio was undefined or poorly defined, a default ratio (dotted line in Fig. B.1) was applied. No adjustments were made to the estimated variance in the troposphere since the fitting process provided no additional information there. For consistency, the correction ratio must

also be applied to the variances of the model parameters or measurement error if these are calculated separately.

Figure B.2 displays the change in the random measurement error when this correction ratio is applied. At the high latitudes the measurement error becomes nearly constant over altitude in the stratosphere. The reduction in error is not nearly as significant in the equatorial middle stratosphere.

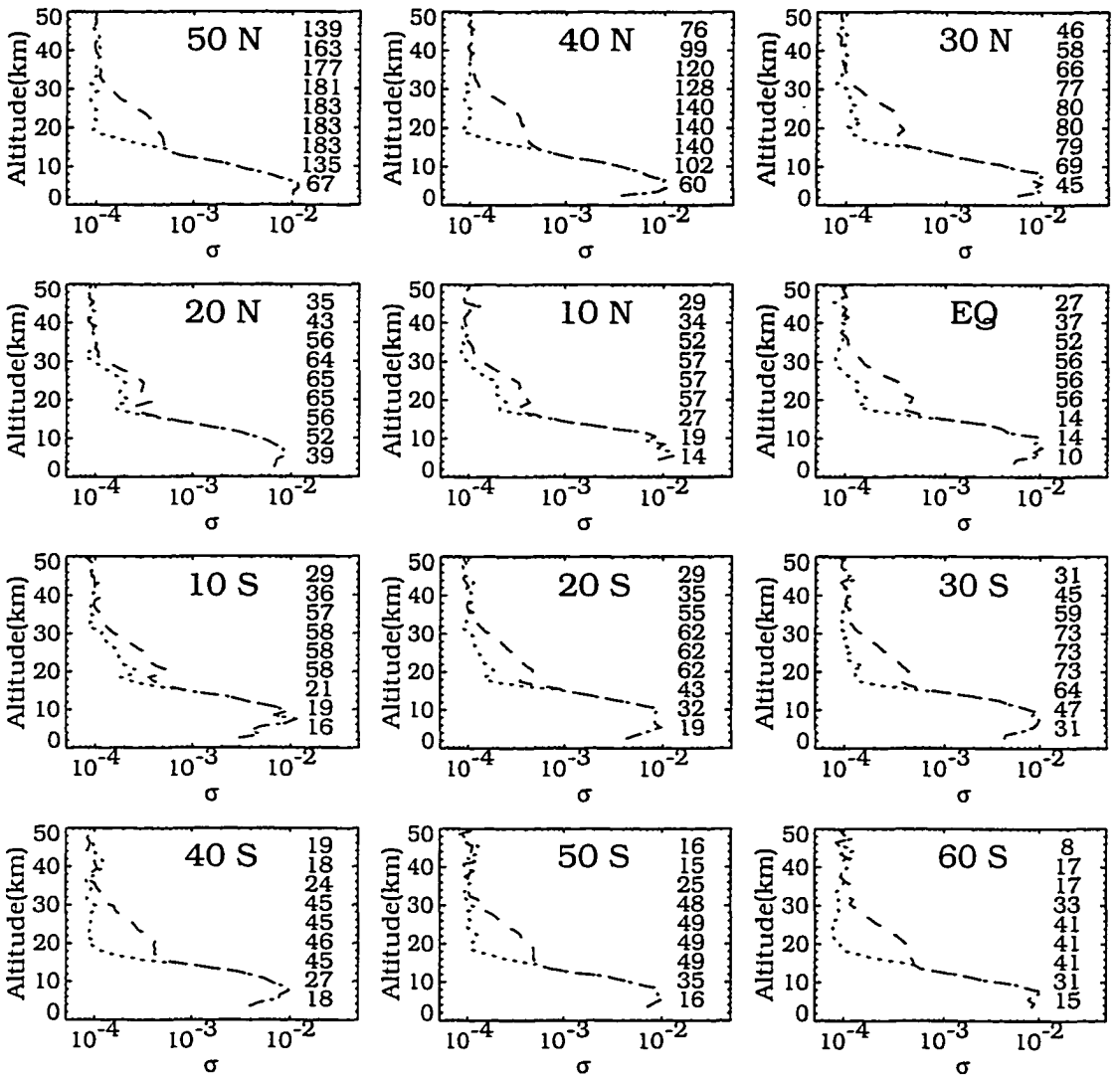


Figure B.2: Estimated standard deviation of the random measurement before (dashed line) and after (dotted line) correction ratio applied for January 1990. The number of points included in the zonal statistics are indicated on the right by altitude level.

APPENDIX C

ABSORPTION CORRELATION COEFFICIENTS

This Appendix describes the error covariance calculations of the slant path absorption random errors and factors leading to the construction of a "representative" error covariance matrix. While much can be learned about an experiments' error characteristics through laboratory testing, it is not always feasible to configure and exercise an instrument for all conditions experienced in an actual deployment and there is no guarantee that the operational characteristics will not change between the laboratory and deployment. Determining the experiment error characteristics, in the form of error covariances $\mathbf{S}_{\epsilon_\mathbf{y}}$, from the measurement data itself would be a straightforward task except for the existence of natural variability contained within the atmospheric and model parameters. Modifying $\mathbf{S}_{\mathbf{y}^m}$ from Table A.1 to include a term for geophysical variability, the covariances of the measurements $\mathbf{y}^m$ are

$$\mathbf{S}_{\mathbf{y}^m} = \mathbf{S}_F + \mathbf{K}_\mathbf{x}\mathbf{S}_{\mathbf{x}^t}\mathbf{K}_\mathbf{x}^T + \mathbf{K}_\mathbf{b}\mathbf{S}_{\epsilon_\mathbf{b}}\mathbf{K}_\mathbf{b}^T + \mathbf{S}_{\epsilon_\mathbf{y}} \tag{C.1}$$

The first term on the right accounts for the covariances of $F(\overline{\mathbf{x}}, \hat{\mathbf{b}})$ caused by the natural variability of $\hat{\mathbf{b}}$. The second term is geophysical covariance resulting from the natural variability of water vapor, $\mathbf{S}_{\mathbf{x}^t}$, transformed through the forward model. $\mathbf{S}_{\mathbf{x}^t}$ rarely, if ever, vanishes. The model parameter covariances $\mathbf{S}_{\epsilon_\mathbf{b}}$ in the third term are error covariances due to uncertainties in $\mathbf{b}$. The geophysical covariances in the Rayleigh, ozone and aerosol model parameters are the major contributors to $\mathbf{S}_F$ and also dominate the other model parameter covariances. In general, the first three terms are much greater than the last term. Some advantage accrues from using $\mathbf{y}^{mi}$ to determine $\mathbf{S}_{\epsilon_\mathbf{y}}$. From Eq. (4.13) we have

$$\mathbf{S}_{\mathbf{y}^{mi}} = \mathbf{S}_S + (\frac{\partial S}{\partial \mathbf{y}^m})[\mathbf{K}_\mathbf{x}\mathbf{S}_{\mathbf{x}^t}\mathbf{K}_\mathbf{x}^T + \mathbf{K}_\mathbf{b}\mathbf{S}_{\epsilon_\mathbf{b}}\mathbf{K}_\mathbf{b}^T + \mathbf{S}_{\epsilon_\mathbf{y}}](\frac{\partial S}{\partial \mathbf{y}^m})^T \tag{C.2}$$

for the covariances of $\mathbf{y}^{mi}$ where $\mathbf{S}_S$ accounts for the covariance of $S(\overline{\mathbf{y}}^m, \hat{\mathbf{b}}, \hat{\mathbf{c}})$ due to geophysical variations remaining after the species separation process. Since the Rayleigh, ozone, and aerosol contributions have been removed at this stage, so have most of their geophysical variations and $\mathbf{S}_S$ should be small. Of course, any systematic errors existing during the species separation process in either of these will reintroduce a geophysical random error component. If $\mathbf{S}_{\mathbf{x}^t}$ can be removed or sufficiently reduced, $\mathbf{S}_{\mathbf{y}^{mi}}$ will be a measure of the combined random error contained in $\mathbf{y}^{mi}$ due to random measurement and model parameter error. Model parameters found in IA may still contribute random error which is not included in $\mathbf{S}_{\mathbf{y}^{mi}}$. Note that there is no need to partition the covariances between $\mathbf{S}_{\epsilon_\mathbf{b}}$ and $\mathbf{S}_{\epsilon_\mathbf{y}}$ since the error equation can accommodate the combination.

This appendix first shows that geophysical covariances $\mathbf{S}_{\mathbf{x}^t}$ dominate the experimental errors and presents several approaches that were studied in an attempt to reduce $\mathbf{S}_{\mathbf{x}^t}$. A segmented profile fitting technique was found to offer the best estimate of the random errors in $\mathbf{y}^{mi}$.

Computationally, one takes the set of slant path absorptions given by Eq. (4.20), $\overline{A}^{\mathrm{H_2O}}, (\mathbf{y}^{mi})$ and applies Eq. (2.77) for the error covariance matrix $\mathbf{S}_{\mathbf{y}^{mi}}$. The error covariance calculated in this manner includes the random measurement error $\epsilon_\mathbf{y}$ found in $T_{2/1}(z_{th}), (\mathbf{y}^m)$ and all random errors introduced by the species separation as given by Eq. (4.7). Geophysical variations of the water vapor absorption are also included through the term $\mathbf{S}_{\mathbf{x}^t}$ and may be a significant source of covariance that cannot be removed entirely. Since it is likely that the errors will be functions of time and measurement location, the measurement data set was partitioned into zonal monthly segments. The error covariances range over 4 orders of magnitude, complicating efforts to plot and identify correlation patterns, so they are shown here as correlation coefficients (Eq. (2.78)).

The zonal slant path water vapor absorption correlation coefficients after removal of Rayleigh, ozone, and aerosol components were calculated for 1987 and 1990. March and September, examples selected to span most of both hemispheres, are shown in Figs. C.1-C.2 and Figs. C.3-C.4 for 1987 and 1990, respectively. The number of paired values along the diagonal are shown above the x-axis every 10 km, starting at 0.5 km, to provide a rough indication of the number of data pairs involved in each correlation. Coefficients are not calculated for less than 15 pairs and the number of pairs decreases away from the diagonal. Also shown in each figure in the lower right are the correlation coefficients for all zones combined. The color scale is similar in all figures but may not be shown if all latitude bins contain data.

Examination of the figures finds little commonality in the correlation patterns whether comparing zones for a single month, a single zone through a year (or both years) or even a single month zone year-to-year. The altitude range over which high correlations are found is surprisingly large. For a solar occultation experiment with a narrow field of view one would expect to find correlated levels limited to a neighborhood of about 2-3 km. Natural geophysical variations are very likely causing the random error correlations to appear greater than they actually are, extend farther off the diagonal than they should, and introduces seasonal and annual variations. Evidence for this may be found in the general features of these figures.

A recurring pattern in many of the zonal plots, such as March 1987 for 60° S, finds the tropospheric absorption displaying high correlation levels (yellow and red) up to 6-7 km off the diagonal that narrows to 1-2 km around the tropopause and then increases again in the stratosphere. Oftentimes there is a relatively strong region of negative correlation (blue) between the absorption in the lower stratosphere and upper troposphere, but in most cases the correlation between stratospheric and tropospheric absorption values are low (green). This pattern appears in almost all of the monthly global correlations and typically is more pronounced than in individual

zones. The correlative pattern shows the stratospheric and tropospheric absorption values are independent of one another, indicating the dynamical and chemical processes controlling the water vapor distribution are different and that measurement error correlations are insignificant. Within the the stratosphere or troposphere the absorption values tend to move in concert, thus the high levels of correlation. The region of narrow correlations, which can be viewed as a transitional region between the stratospheric and tropospheric regimes, may provide the best indicator of absorption random error. In this region the absorption profiles have low variability as indicated by the relative standard deviation (standard deviation/mean).

Three approaches were taken to mitigate the geophysical component of the variances, that is, reduce $\mathbf{S}_{\mathbf{x}'}$. The first attempted to account for large feature profile variations by performing the covariance calculations in a coordinate system that might better follow natural profile variations than the altitude coordinate grid. Success in this approach would produce higher correlation values near the diagonal. For a number of trial months transformed to pressure coordinates, it was found that the correlation coefficient pattern actually diffused in the lower stratosphere, with minimal change elsewhere, and did tighten slightly above 25 km in potential temperature coordinates.

The second approach involved restricting the covariance calculations to times and locations where the measurement locations do not change much, and one may sample the same atmosphere repetitively to develop the necessary statistics in a relatively short time period or where the atmospheric state and the water vapor distribution are unchanging. Inspection of Fig. 3.1 suggests that periods of sweep reversal might meet these criterion, such as summer and late fall in the Northern Hemisphere (May-Sept) or in the Southern Hemisphere (Nov-Mar) at high latitudes. Correlations for two short time periods of 12 days or less, with latitude spreads of 6° to 12°, in both hemispheres were calculated using only sunset or sunrise data. Again, the

correlation patterns were relatively similar compared to the nearest corresponding monthly zonal values, indicating that latitudinal and longitudinal variations still existed in the data sample. The equatorial region, at least in the troposphere, is another possible location to investigate since the seasonal and latitudinal variations are minimal Oort(1983). However, since the SAGE II measurement locations move through the tropics quickly and altitude coverage in the troposphere is limited by clouds, it is difficult to accumulate sufficient data in a month and a longer time period is necessary. Annual zonal correlative coefficients are shown in Figs. C.5 and C.6 for 1987 and 1990, respectively, where it can be seen that the pattern in this region has narrowed such that high levels of correlation are found within 2-4 km of the diagonal. While an improvement, it is still likely an overestimate of the correlation coefficients in the troposphere.

The third approach attempts to remove geophysical variations through segmented least squares fits of the absorption profiles with correlation coefficients calculated from the residuals. Inspection of the absorption profile found the logarithm of the absorption approximately linear over altitudes between 25-45 km globally. At middle and high latitudes linearity extended down to lower altitudes (minimum of 15 km). In the equatorial region, the profiles still exhibited a great deal of curvature between 18 and 25 km, as did the tropospheric portion at all latitudes. Two calculational schemes were set up to exploit these profile characteristics. The first used two profile segments, 4-11 km in the troposphere and 25-45 km in the stratosphere while the second further sub-divided the stratosphere into two segments, 18-29 km and 30-45 km. For each scheme, second, third, and fourth order polynomials were fit to the measured absorption data in each segment and correlation coefficients were calculated with the residuals. Higher order polynomials were not tested because of the limited number of data points in the segments, fitting noise would simply remove the information sought in this analysis.

Figures C.7 and C.8 illustrate the results of this approach for March 1987 from the two segment and three segment fitting schemes. At middle and higher latitudes, where there are sufficient data points, a consistent pattern appears in all latitude zones in the stratosphere. In both schemes, the off-diagonal correlations are reduced to almost zero. Some artifacts resulting from the fitting process are present in both figures. Correlations using data points close to the segment endpoints tend to be highly negative (blue) and appear at 26-28 km in the two segment scheme and 18-19 km, 29-30 km in the three segment scheme. The small region of high correlation (yellow) at 28 km in the two segment scheme is not found in the three segment scheme and is also an artifact. In addition to these small scale features there are also some larger scale features that should not be interpreted as real. Regions of slightly negative correlations (light blue) that sometimes parallel the diagonal to produce banded structure shift position relative to the diagonal depending on the polynomial order used in the fit. When the correlation patterns are averaged relative to the diagonal to produce a mean correlation profile for each zone, profiles from all zones were very similar but changed shape with the order of the fit.

In spite of these limitations of the segmented fitting process, it appears that a strong case can be made that the random errors in the stratospheric slant path absorption are uncorrelated or independent in all latitudes. Inspection of correlation coefficients for other months and years finds similar results, implying the coefficients are time independent. As a final check of this conclusion, synthetic absorptances with correlated errors were analyzed with the segmented fit procedures. Both fitting schemes found the correlated errors and even reproduced the shape of the large scale artifacts. Unfortunately, the segmented fitting procedure did not work very well in the troposphere, due to the high curvature there, as can be seen in Figs. C.7 and C.8.

A conservative "representative" correlation coefficient matrix can now be constructed for the SAGE II slant path water vapor absorption applicable to all times and locations. Based on the results of the segmented fitting analysis, the off-diagonal elements are set to zero between 18 and 45 km. Some of the zonal plots shown in Figs. C.1-C.4 suggest that this is a reasonable assumption to make down to 15 km. In the troposphere, the annual equatorial zonal coefficients provide some guidance as probable upper limits but are not definitive. Locations outside of the equatorial region will be treated in a similar manner since no information is available there. It is assumed that the correlation coefficients are 0.6 adjacent to the diagonal and 0.3 when 2 km off the diagonal and independent of time. It is likely that as the instrument scans the smaller solar image through the troposphere and refraction affects the tangent altitude determination to a greater extent, the variances and covariances will increase in the troposphere. The correlation matrix corresponding to these assumptions is shown in Fig. C.9.

Although the correlation coefficients are assumed to be independent of time and location, the variances are not and it follows that the covariances will also vary. Covariances for particular times and locations can be constructed by multiplying this "representative" correlation matrix by the appropriate variances.

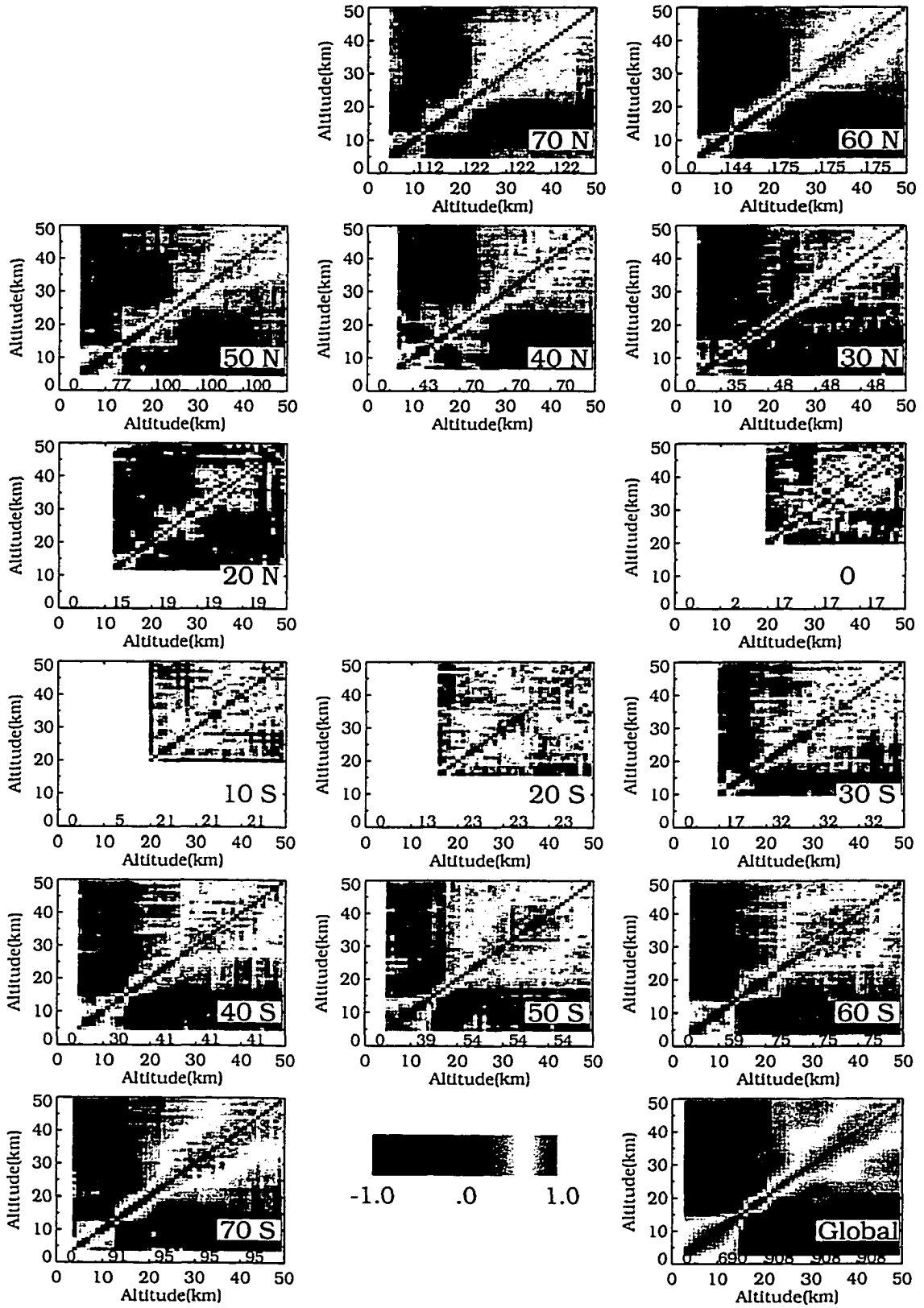


Figure C.1: Zonal water vapor absorption correlation coefficients for March 1987. See text for description.

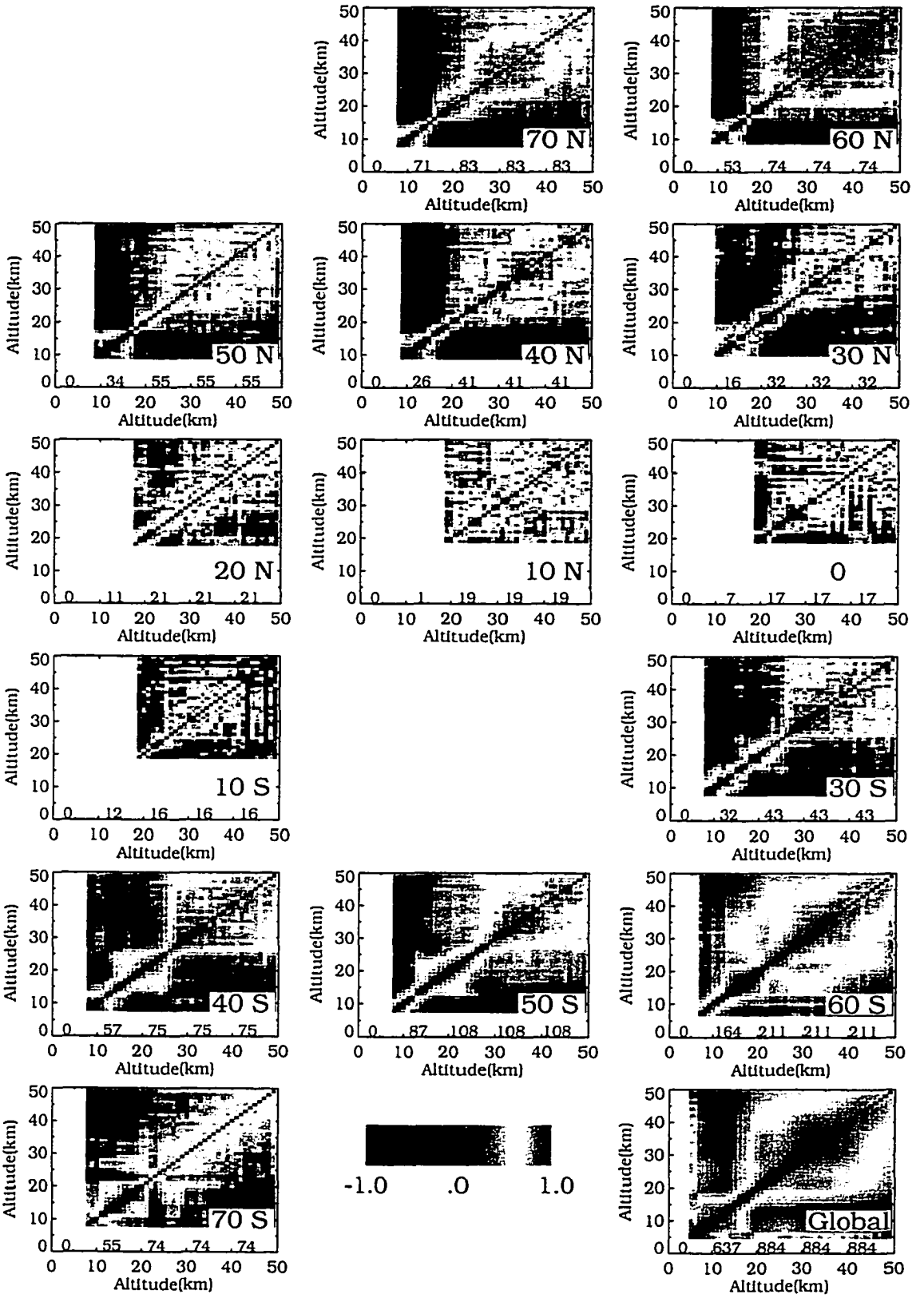


Figure C.2: As in Fig. C.1 except for September 1987.

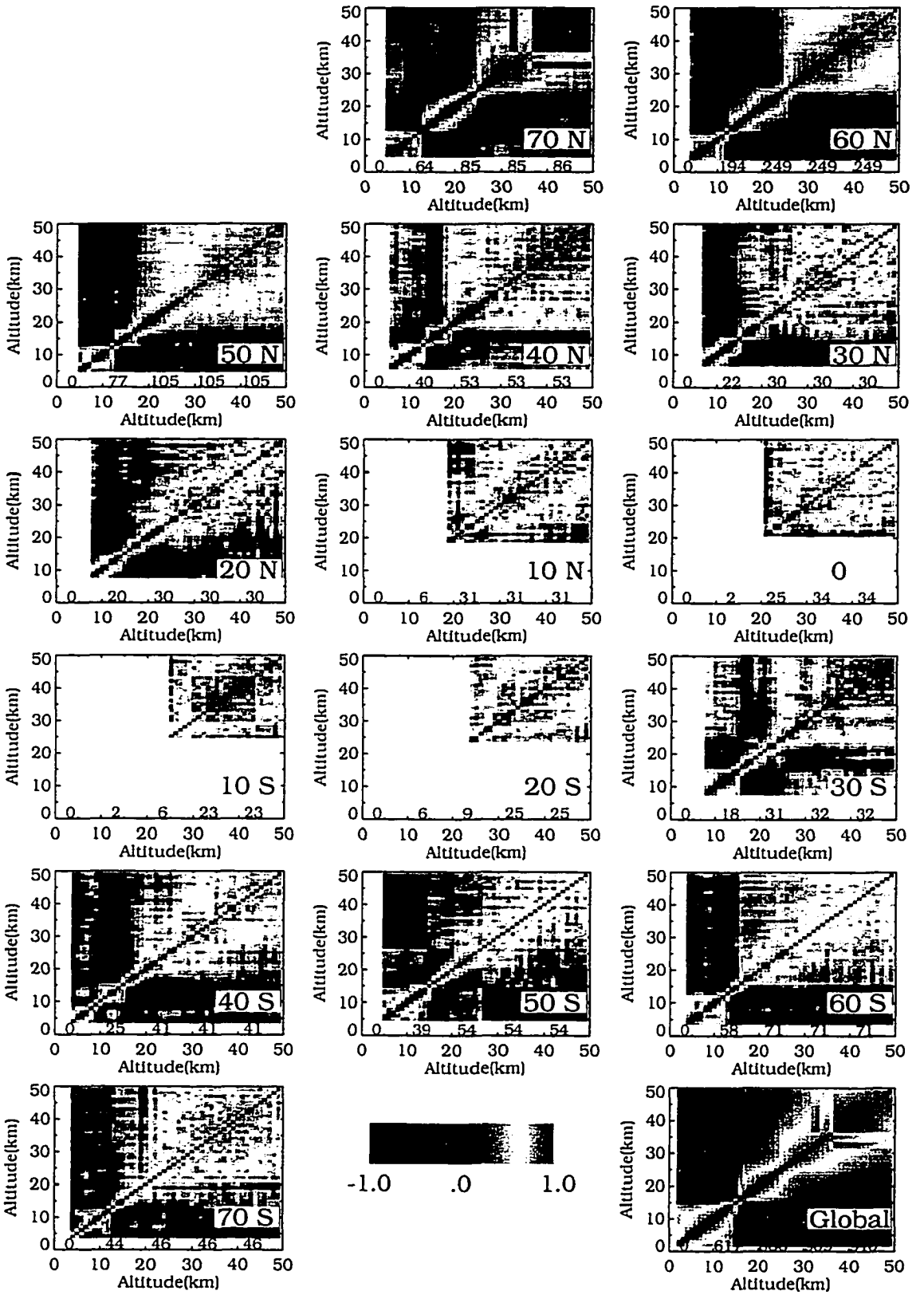


Figure C.3: As in Fig. C.1 except for March 1990.

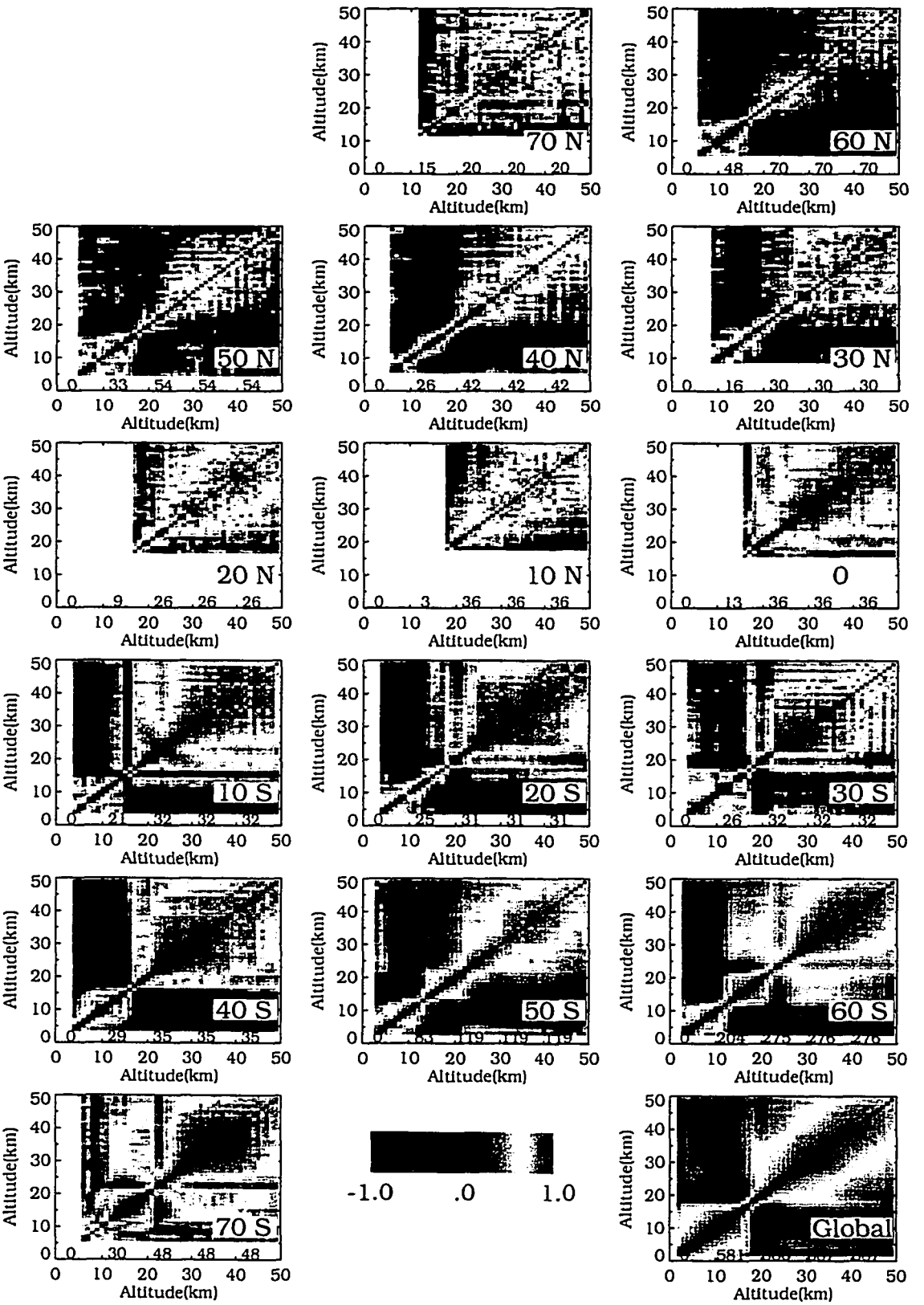


Figure C.4: As in Fig. C.1 except for September 1990.

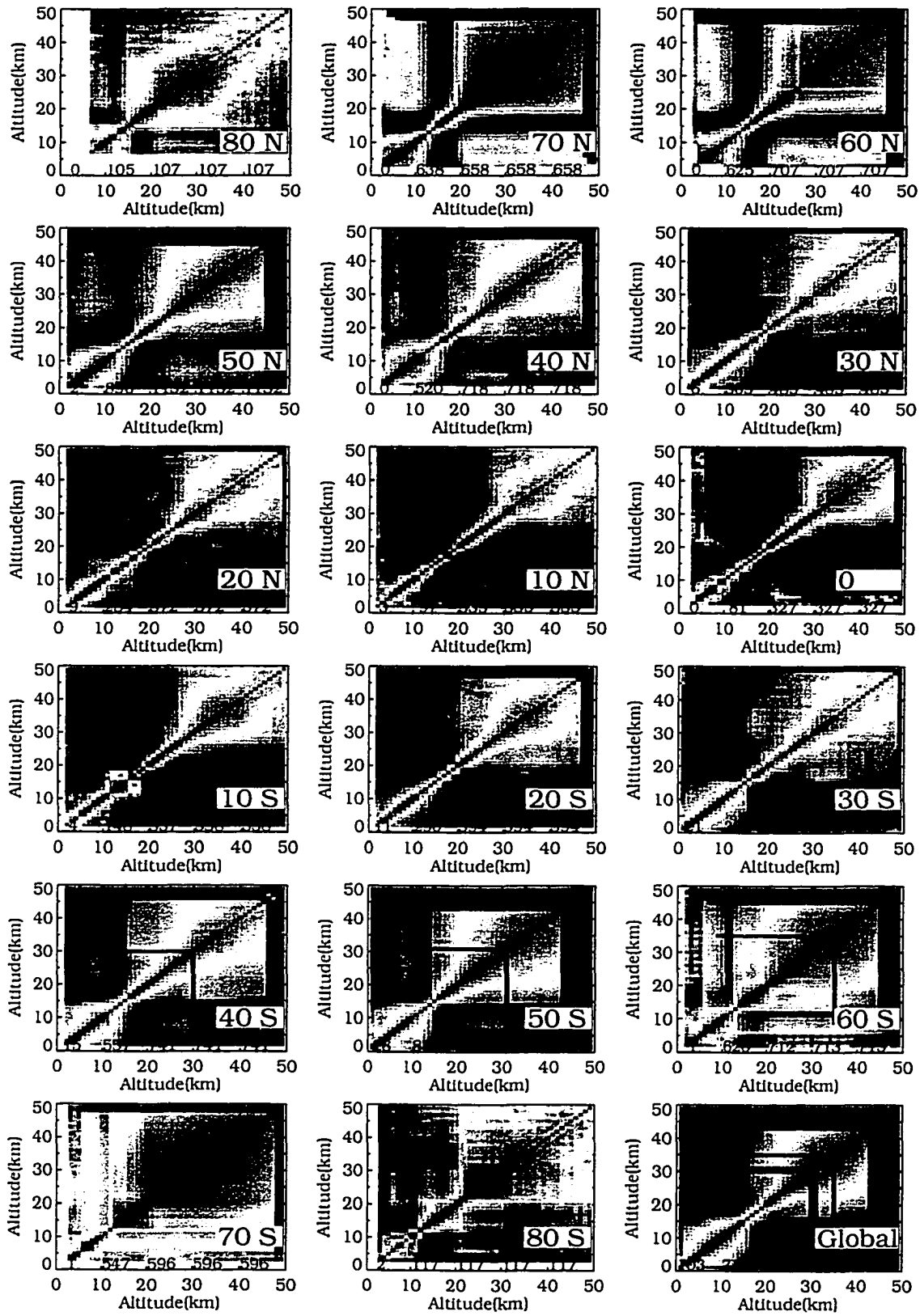


Figure C.5: Annual zonal water vapor absorption correlation coefficients for 1987.

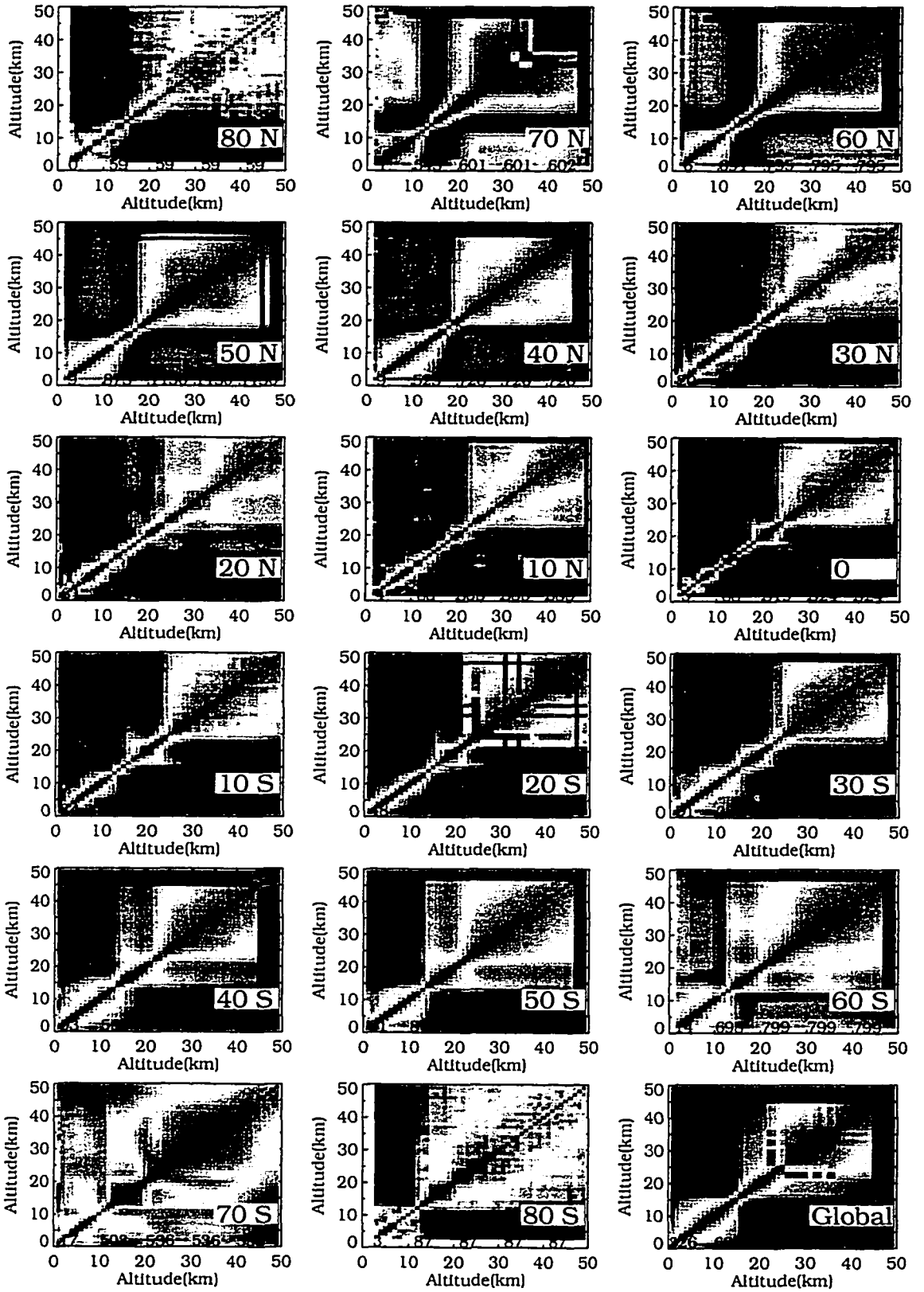


Figure C.6: Annual zonal water vapor absorption correlation coefficients for 1990.

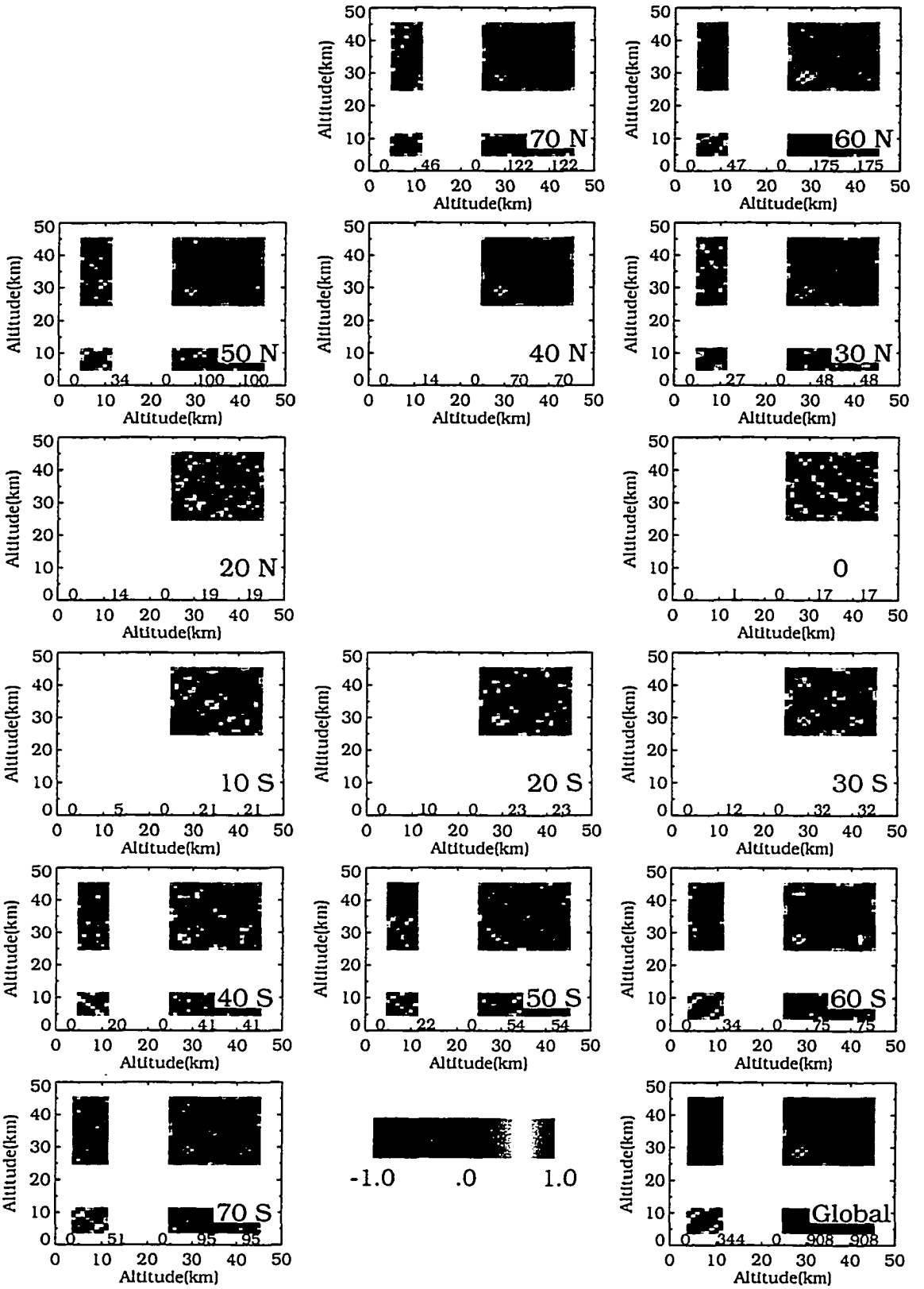


Figure C.7: Zonal water vapor absorption correlation coefficients for March 1987 calculated from two segmented fit residuals.

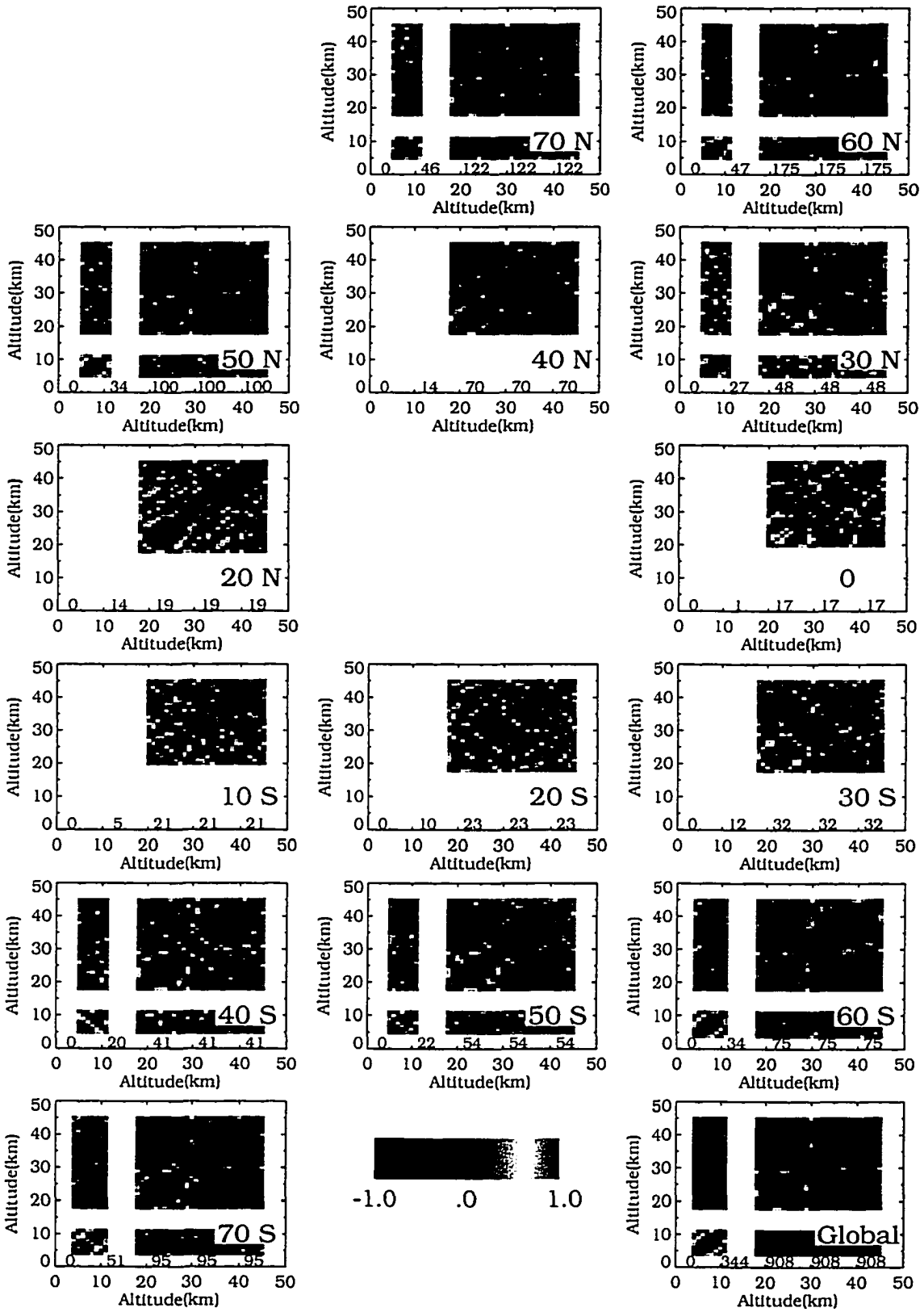


Figure C.8: Zonal water vapor absorption correlation coefficients for March 1987 calculated from three segmented fit residuals.

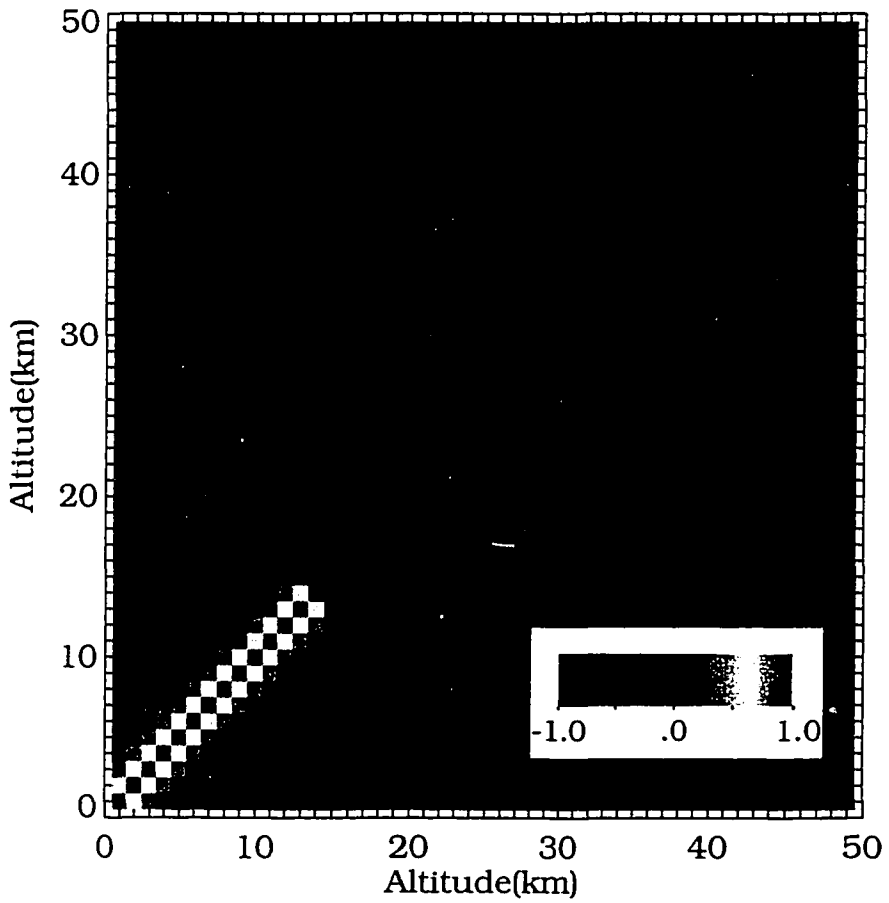


Figure C.9: "Representative" correlation coefficients for the SAGE II water vapor slant path absorption.

APPENDIX D

TOTAL EXPERIMENTAL ERROR

The monthly zonal mean total systematic, random and experimental error of the SAGE II water vapor observations, for all months of 1990, are found in this appendix. Details concerning error term formulation and application of the formal error analysis may be found in Sec. 4. The January to December error budgets are shown in Figs. D.1-D.12 for the onion peel inversion algorithm, Figs. D.13-D.24 for the operational inversion algorithm, and Figs. D.25-D.36 for the Mill-Drayson inversion algorithm. The error budget for the smoothed Mill-Drayson algorithm is nearly identical to the operational and is not shown.

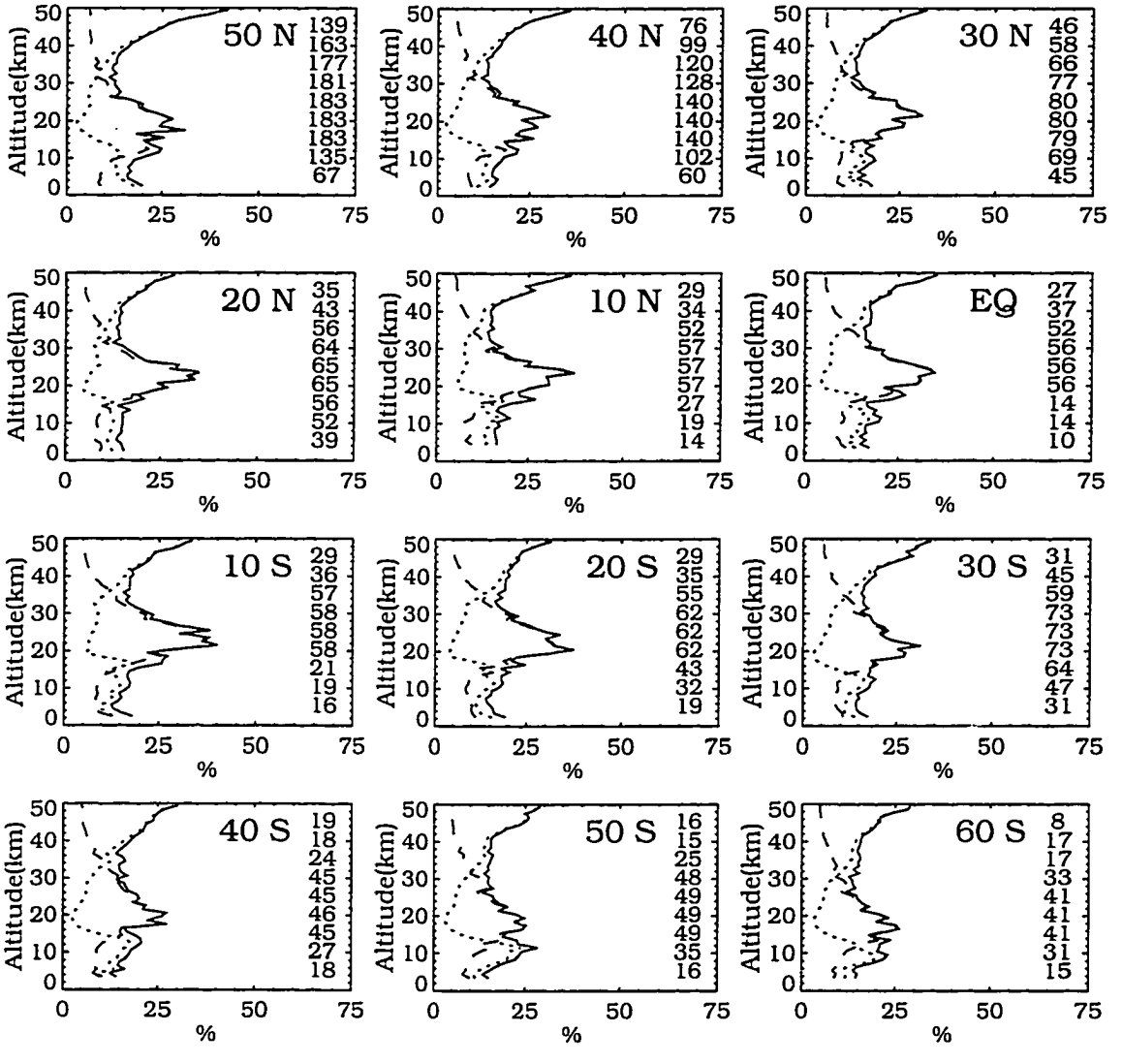


Figure D.1: Total systematic (dashed), random (dotted), and experimental uncertainty (solid) corresponding to the onion peel retrieval, January 1990.

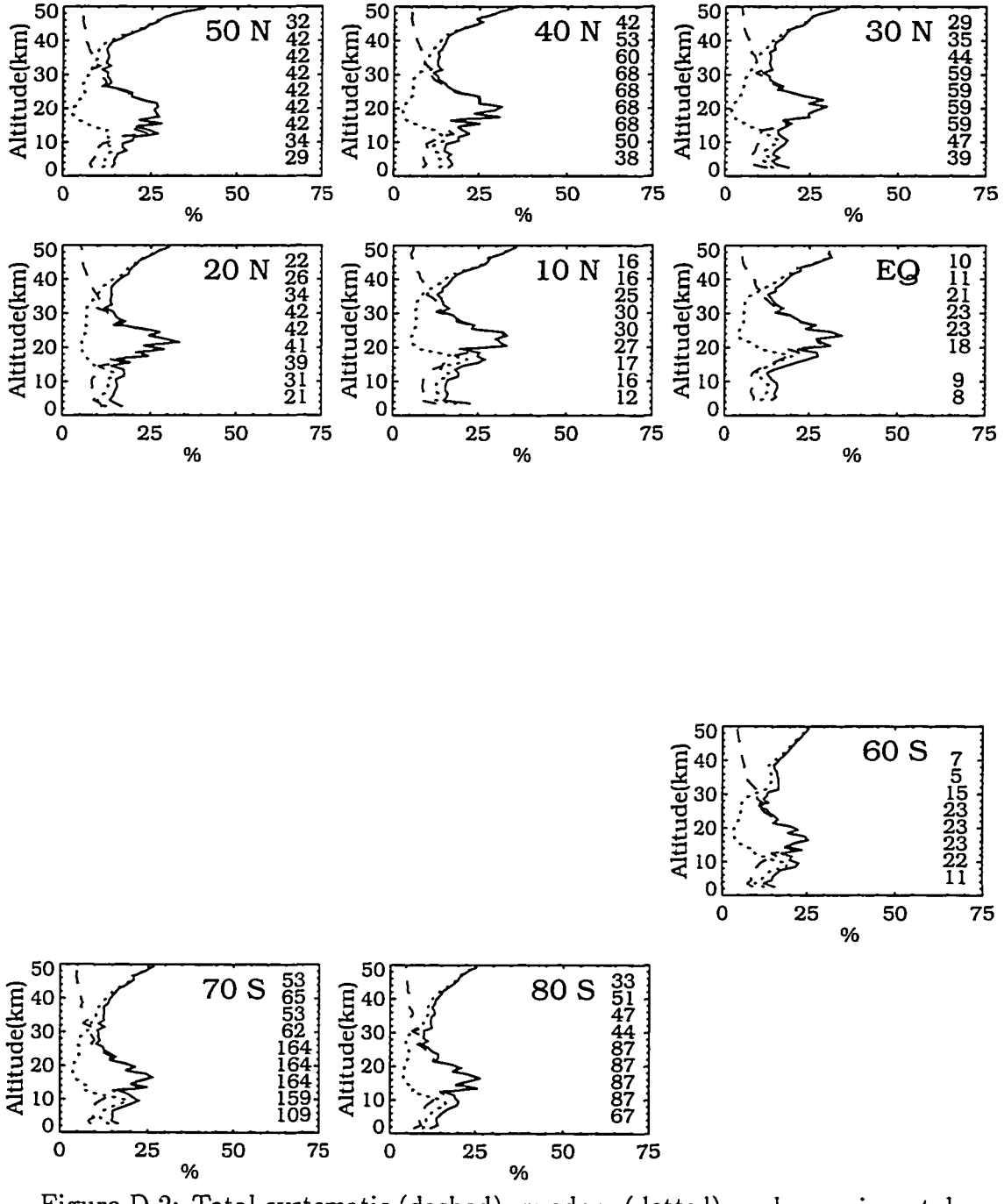


Figure D.2: Total systematic (dashed), random (dotted), and experimental uncertainty (solid) corresponding to the onion peel retrieval, February 1990.

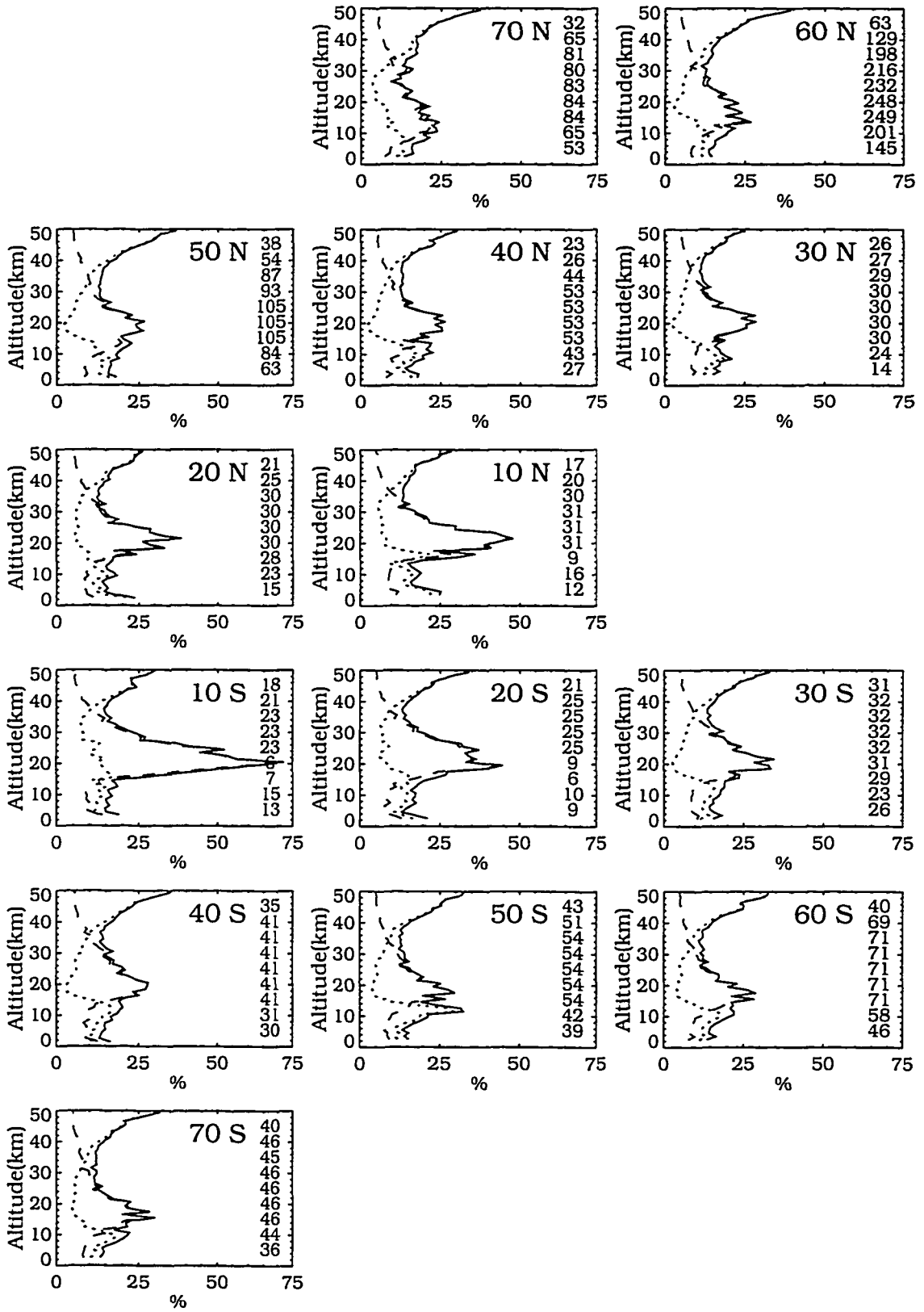


Figure D.3: Total systematic (dashed), random (dotted), and experimental uncertainty (solid) corresponding to the onion peel retrieval, March 1990.

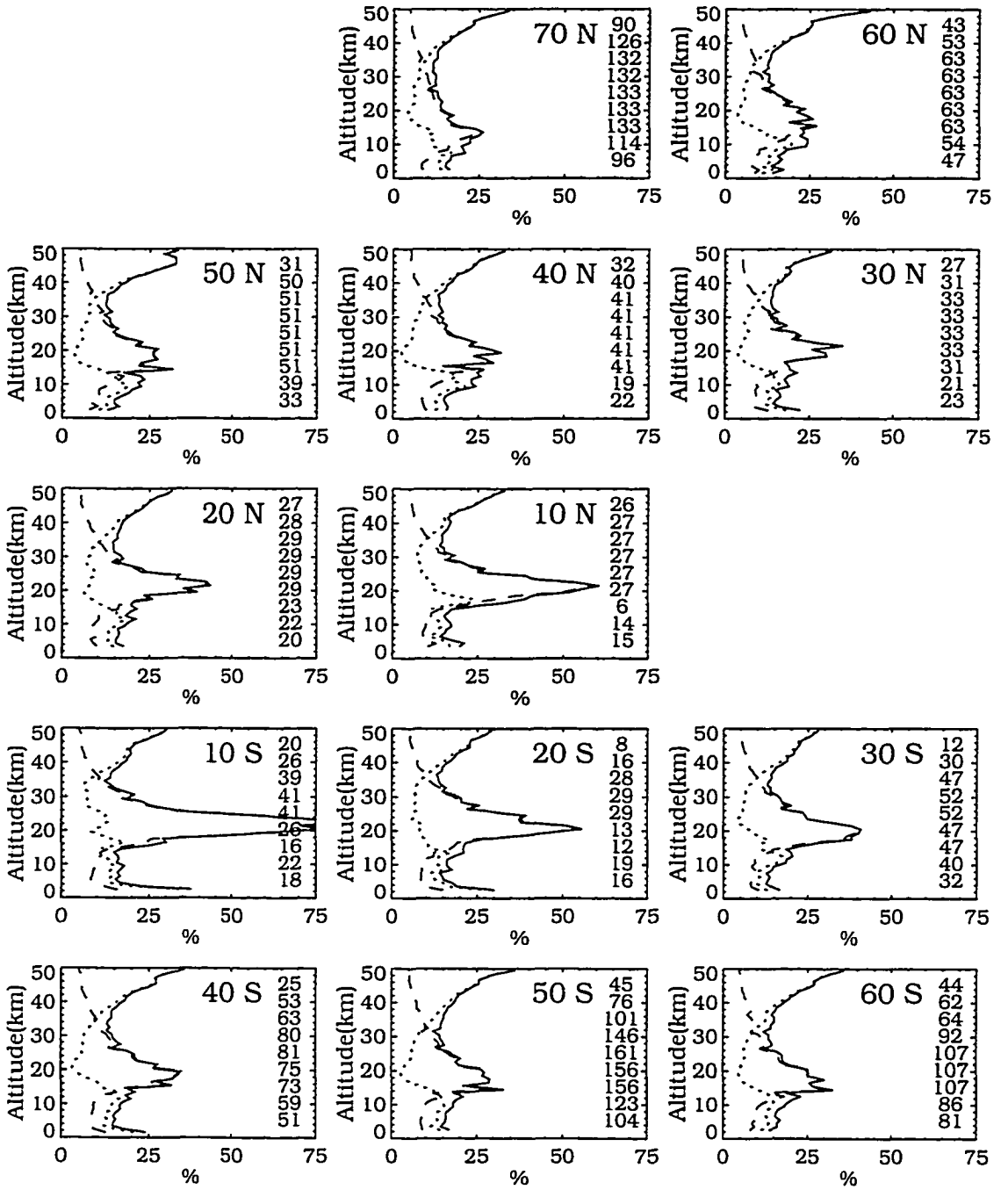


Figure D.4: Total systematic (dashed), random (dotted), and experimental uncertainty (solid) corresponding to the onion peel retrieval, April 1990.

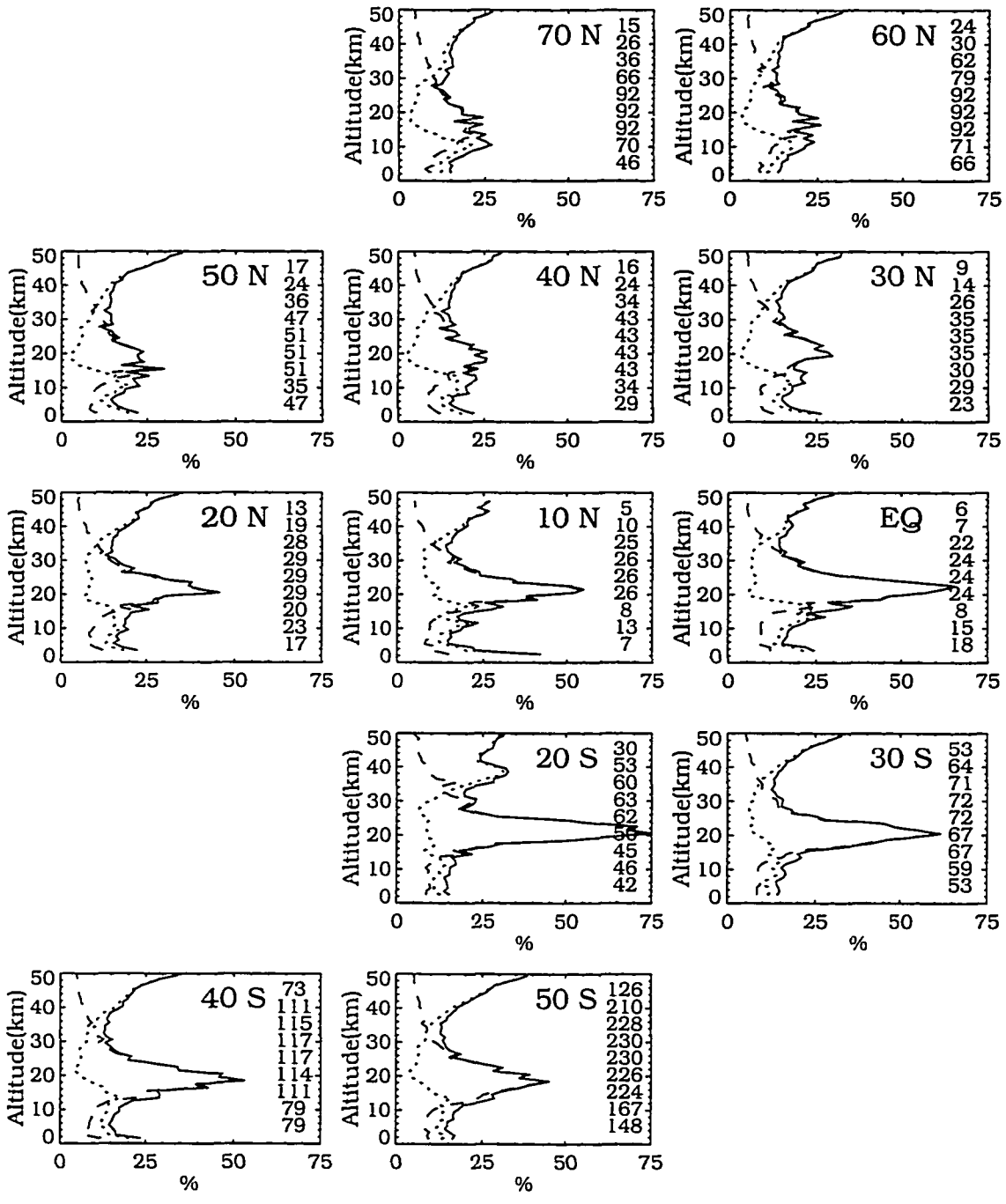


Figure D.5: Total systematic (dashed), random (dotted), and experimental uncertainty (solid) corresponding to the onion peel retrieval, May 1990.

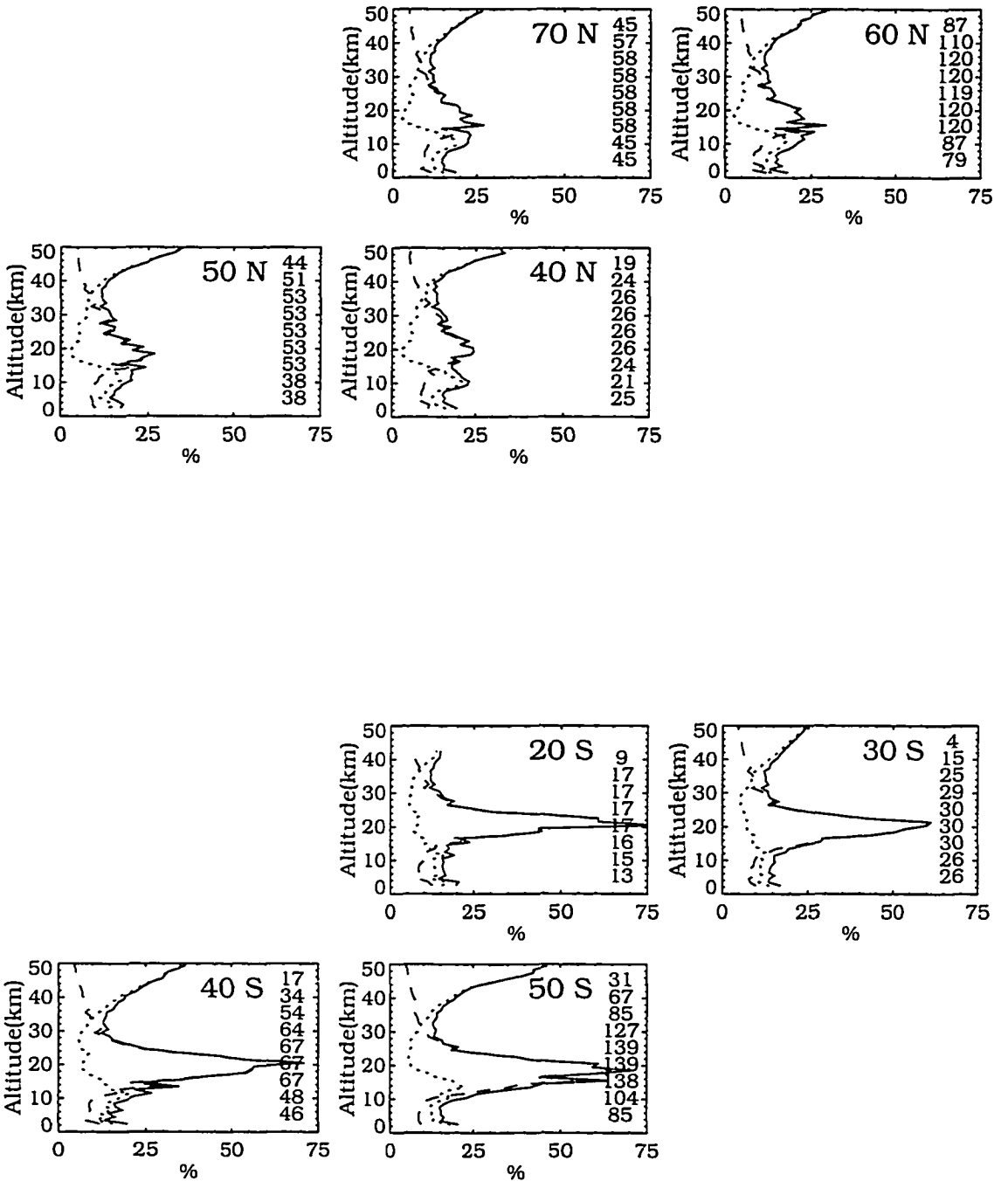


Figure D.6: Total systematic (dashed), random (dotted), and experimental uncertainty (solid) corresponding to the onion peel retrieval, June 1990.

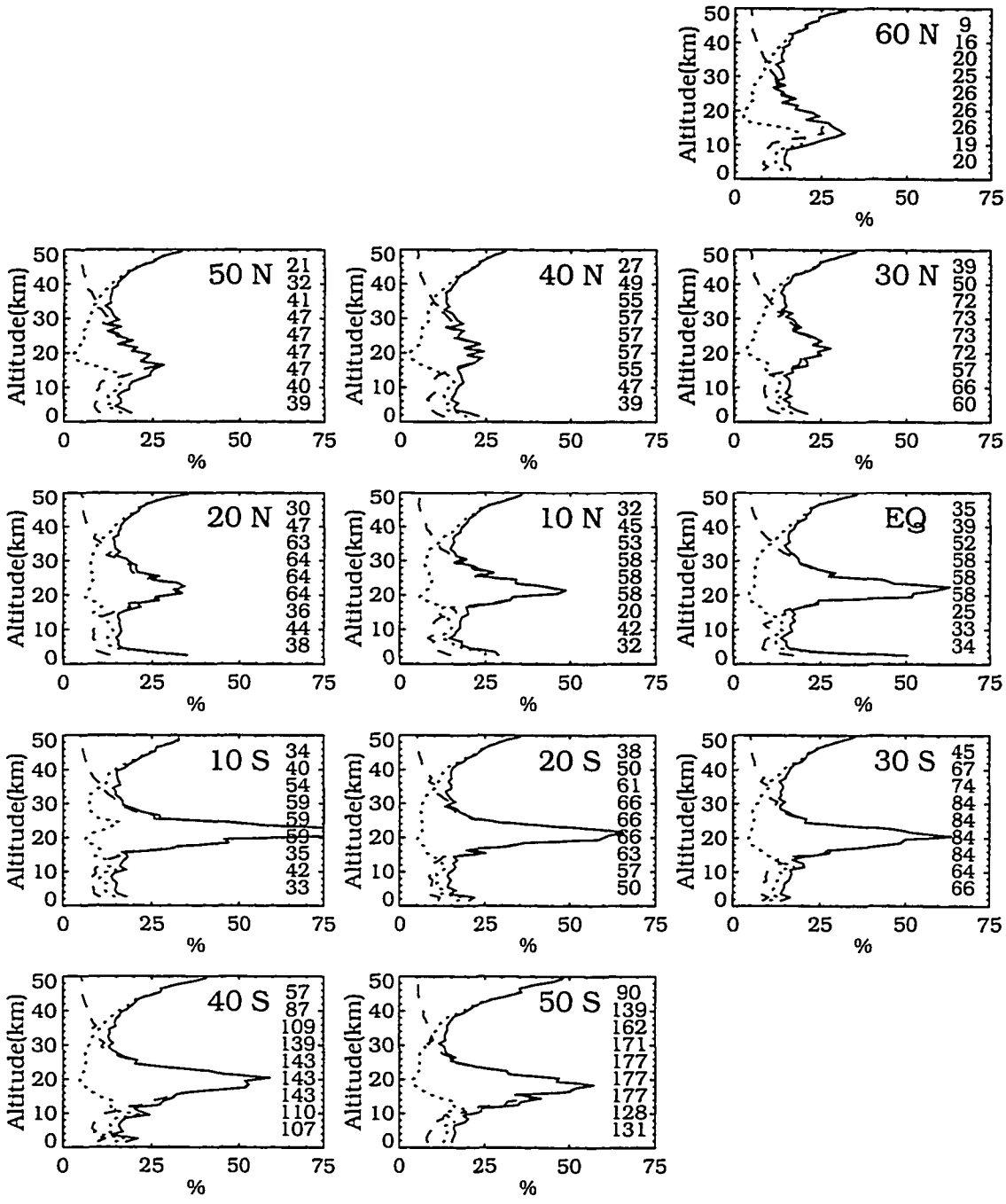


Figure D.7: Total systematic (dashed), random (dotted), and experimental uncertainty (solid) corresponding to the onion peel retrieval, July 1990.

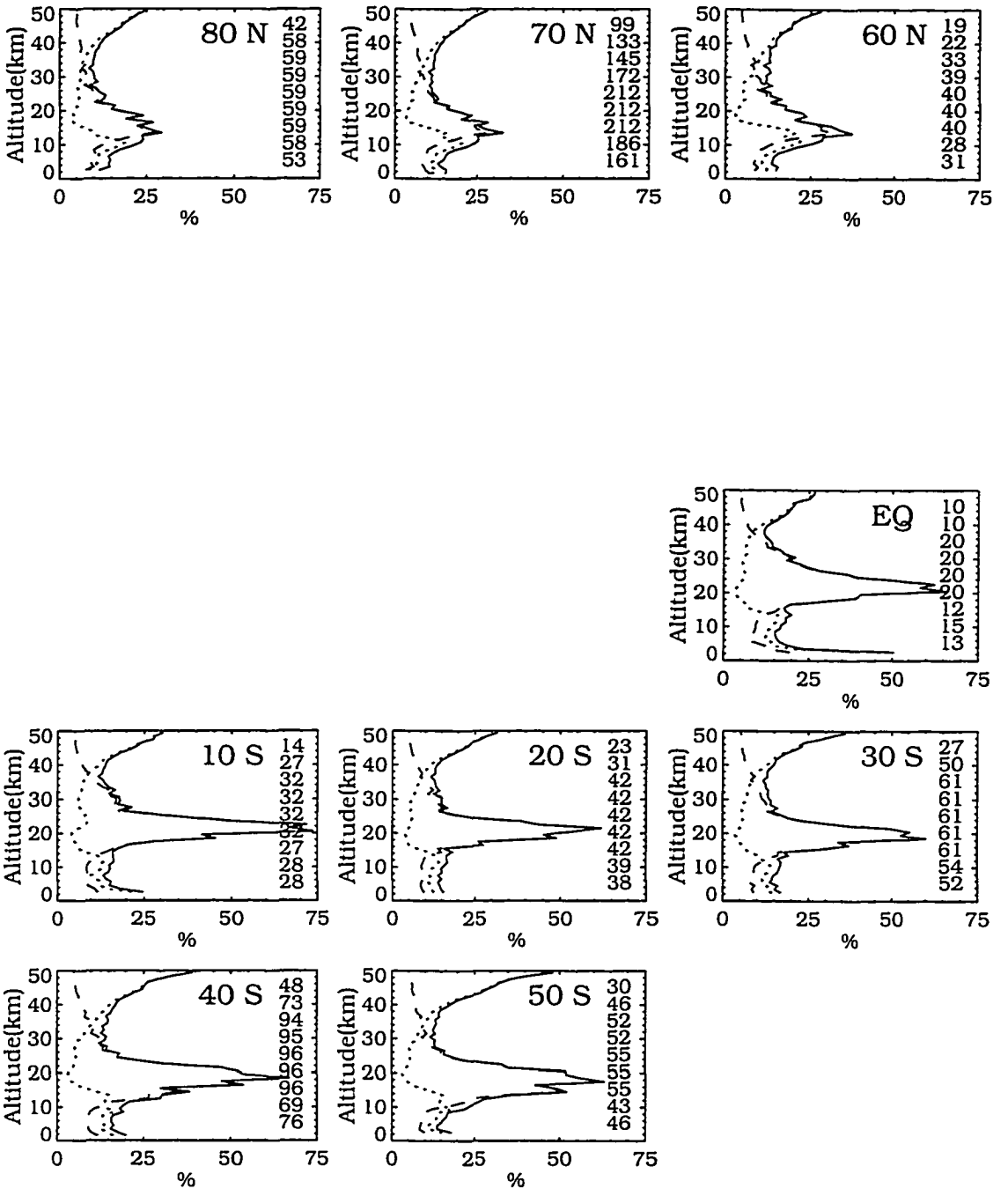


Figure D.8: Total systematic (dashed), random (dotted), and experimental uncertainty (solid) corresponding to the onion peel retrieval, August 1990.

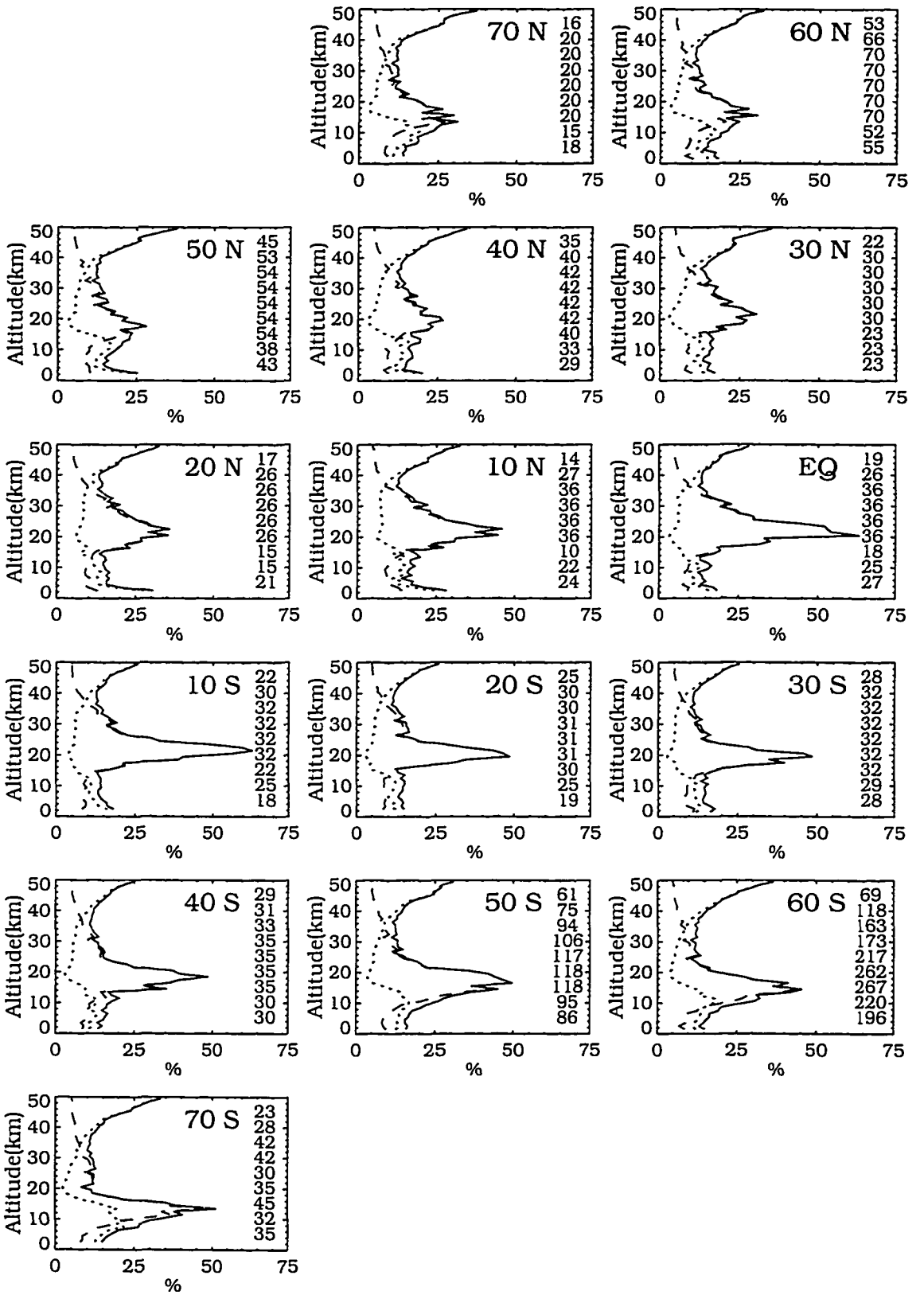


Figure D.9: Total systematic (dashed), random (dotted), and experimental uncertainty (solid) corresponding to the onion peel retrieval, September 1990.

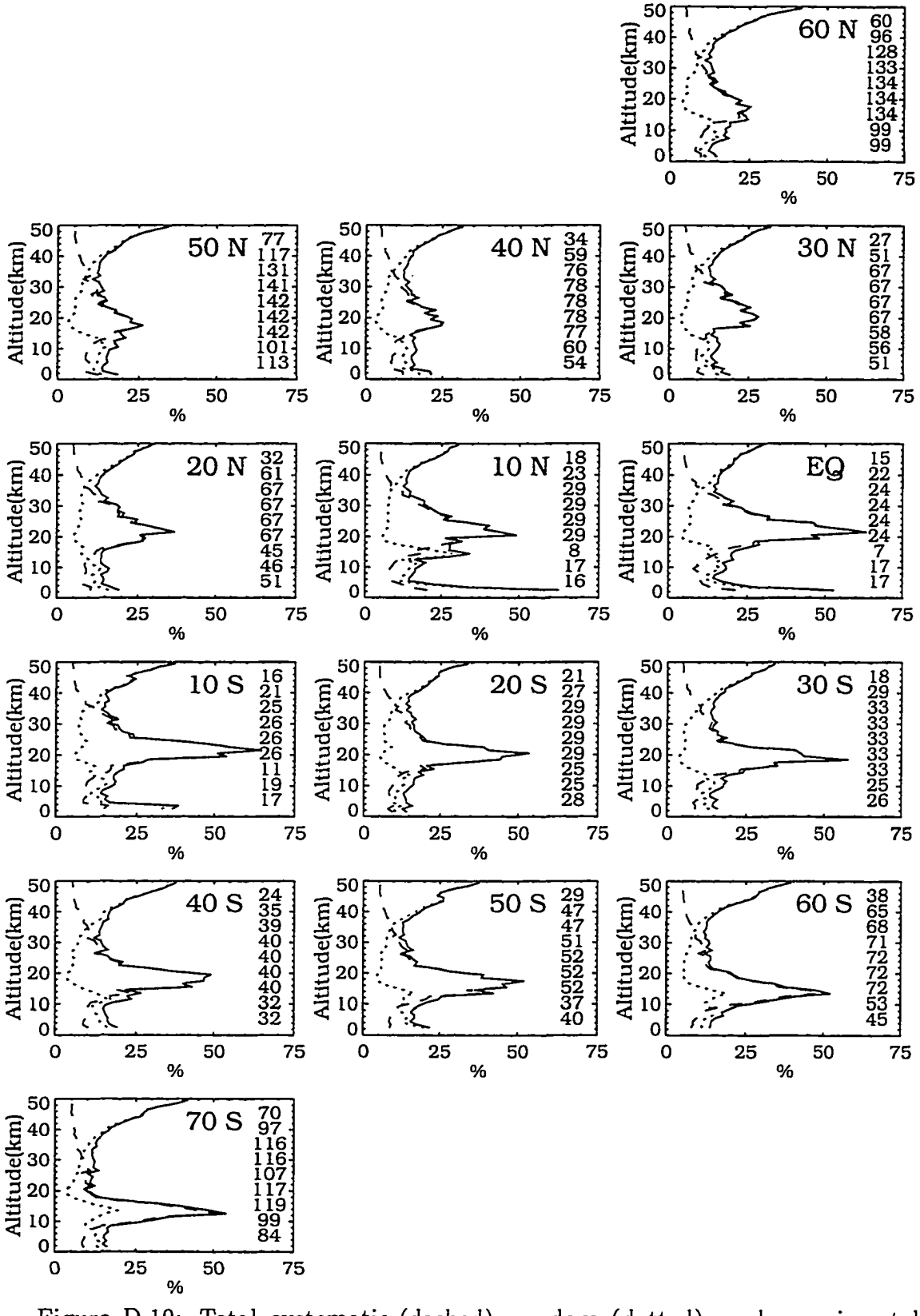


Figure D.10: Total systematic (dashed), random (dotted), and experimental uncertainty (solid) corresponding to the onion peel retrieval, October 1990.

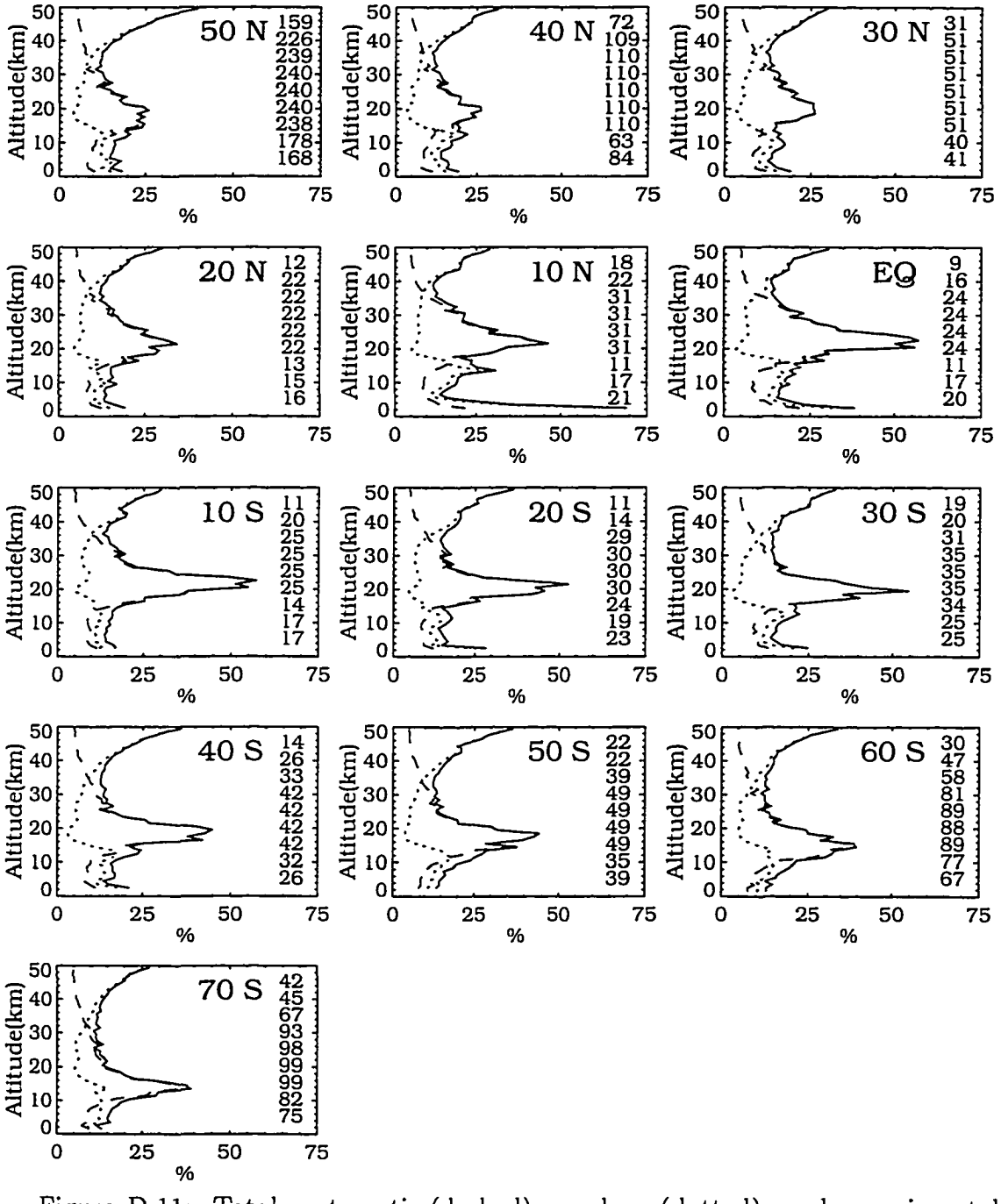


Figure D.11: Total systematic (dashed), random (dotted), and experimental uncertainty (solid) corresponding to the onion peel retrieval, November 1990.

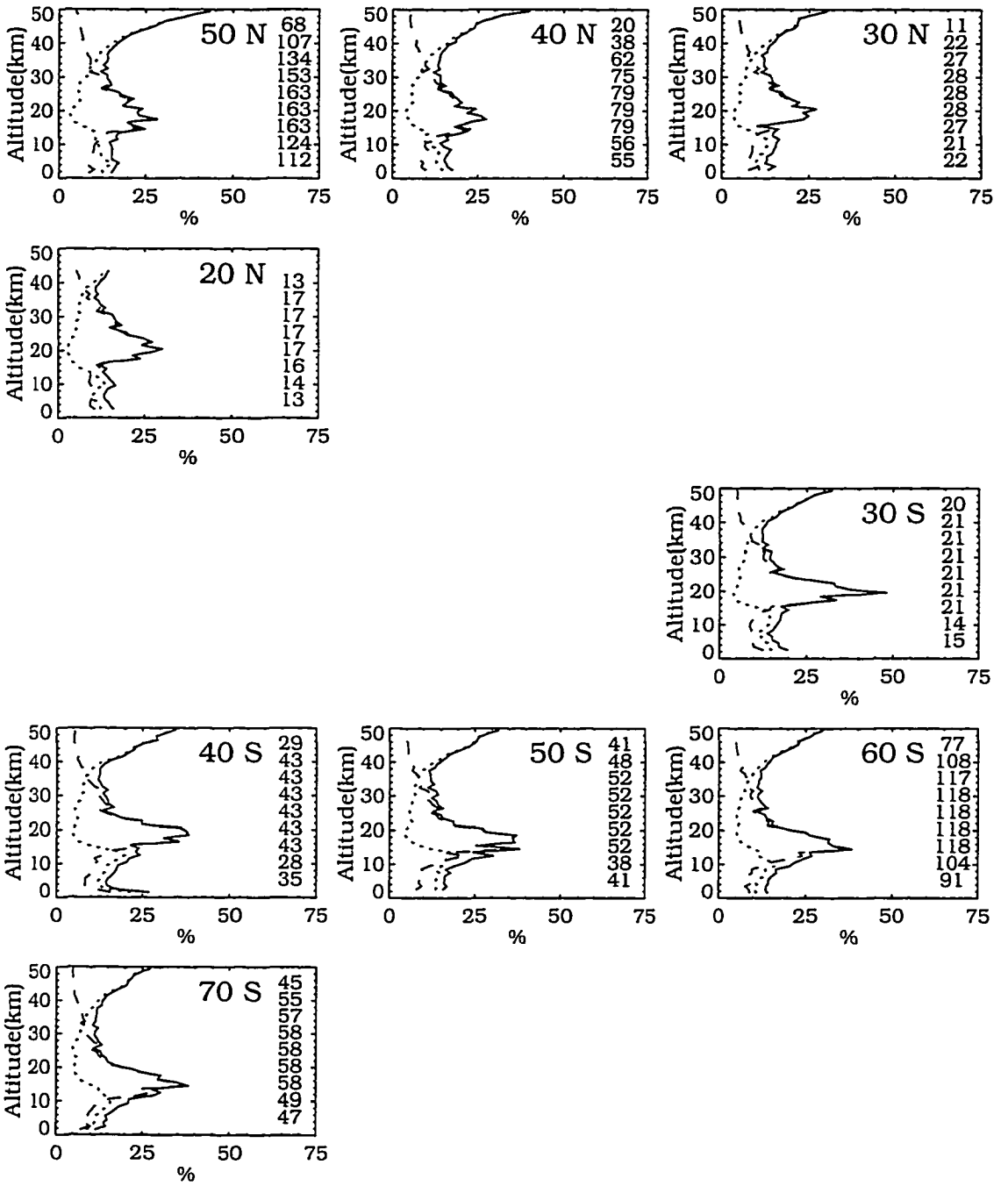


Figure D.12: Total systematic (dashed), random (dotted), and experimental uncertainty (solid) corresponding to the onion peel retrieval, December 1990.

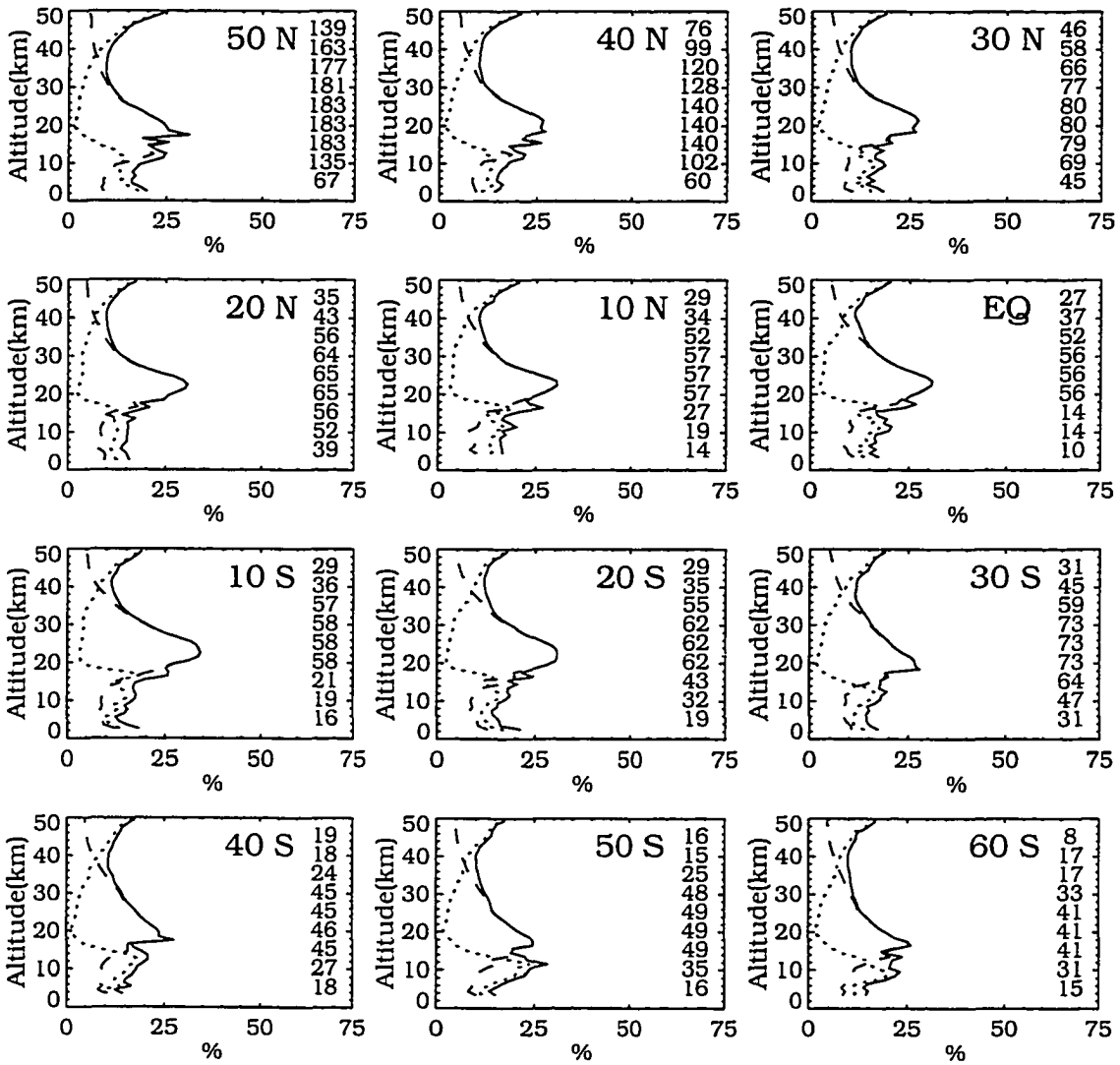


Figure D.13: Total systematic (dashed), random (dotted), and experimental uncertainty (solid) corresponding to the operational retrieval, January 1990.

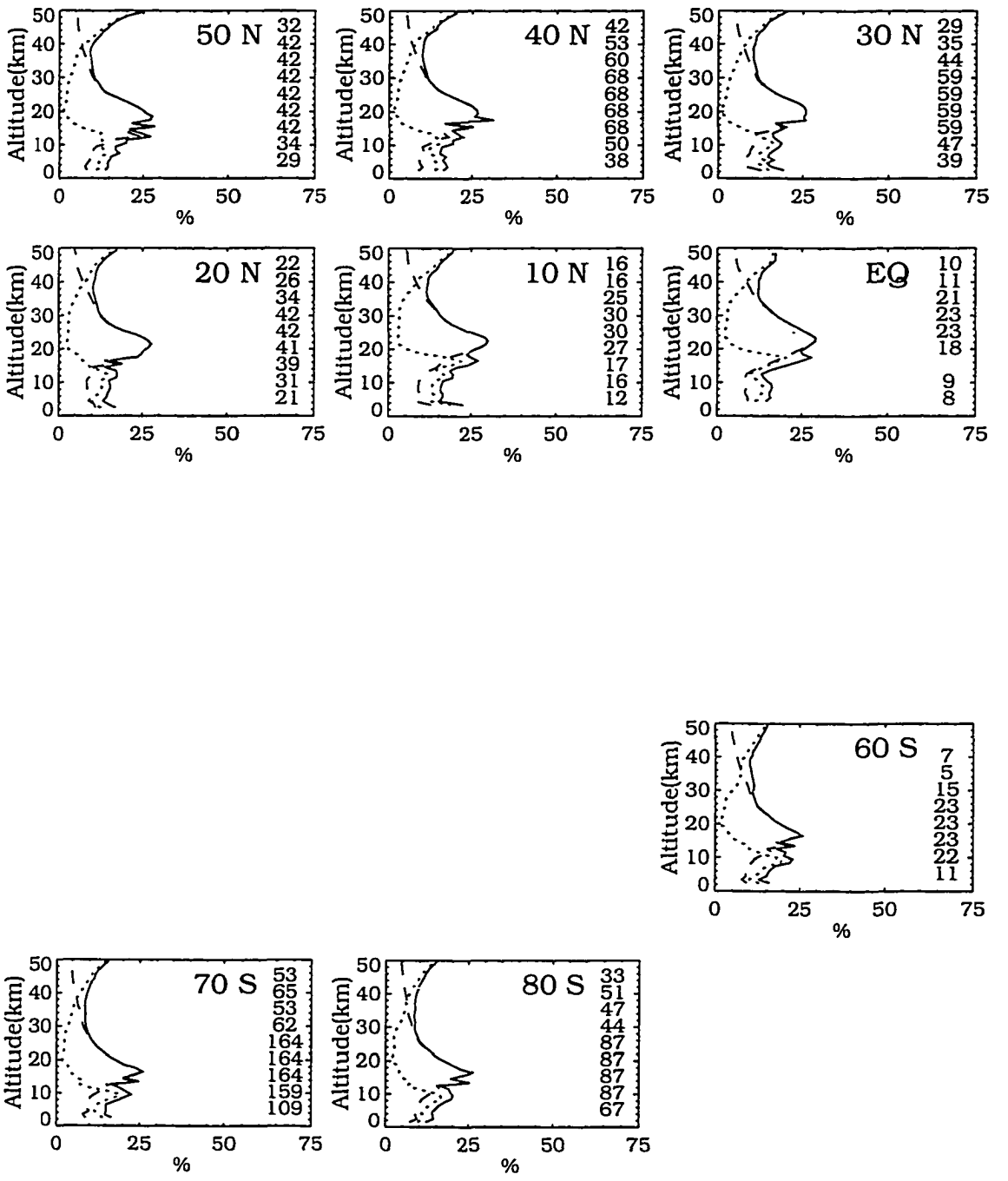


Figure D.14: Total systematic (dashed), random (dotted), and experimental uncertainty (solid) corresponding to the operational retrieval, February 1990.

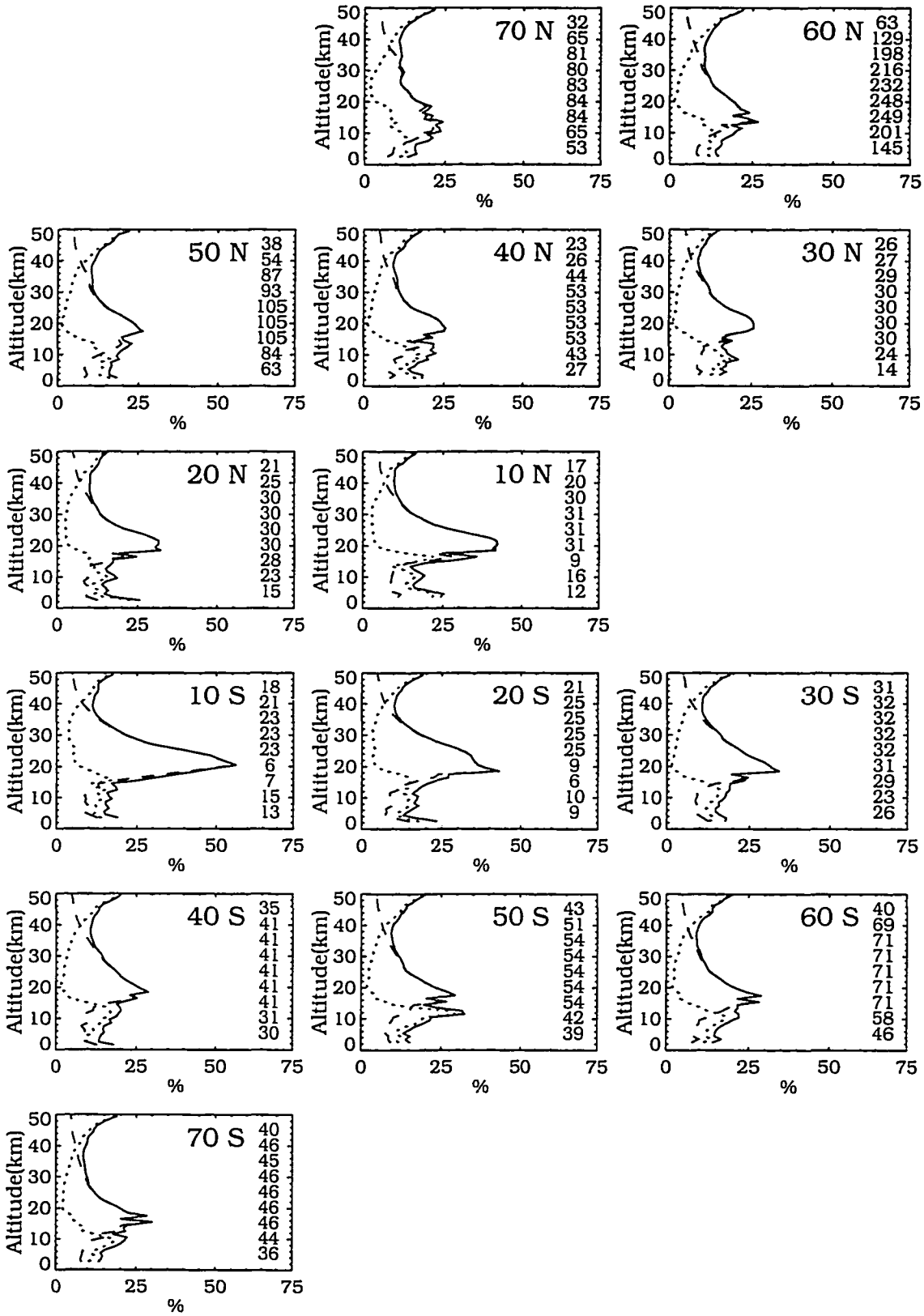


Figure D.15: Total systematic (dashed), random (dotted), and experimental uncertainty (solid) corresponding to the operational retrieval, March 1990.

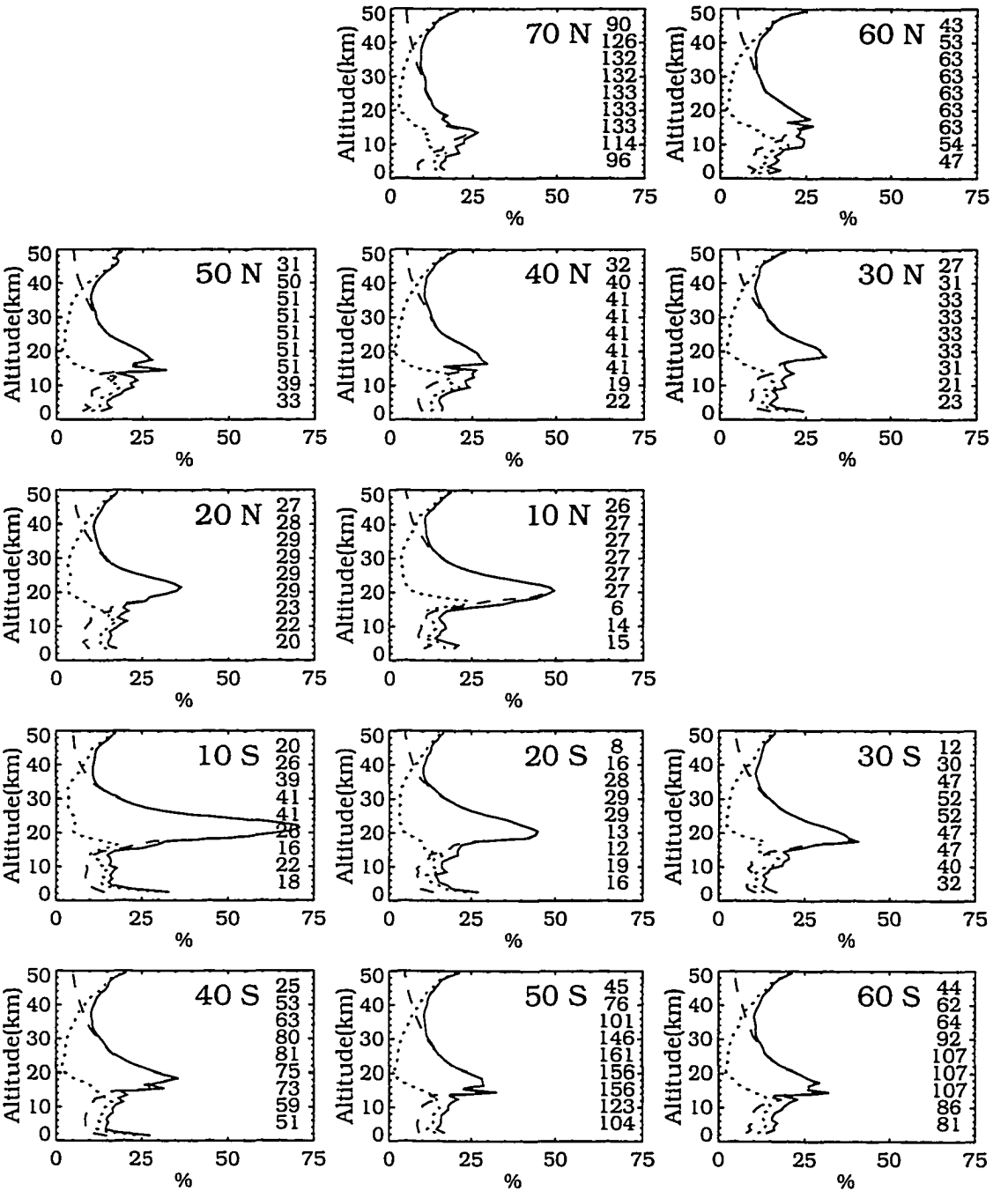


Figure D.16: Total systematic (dashed), random (dotted), and experimental uncertainty (solid) corresponding to the operational retrieval, April 1990.

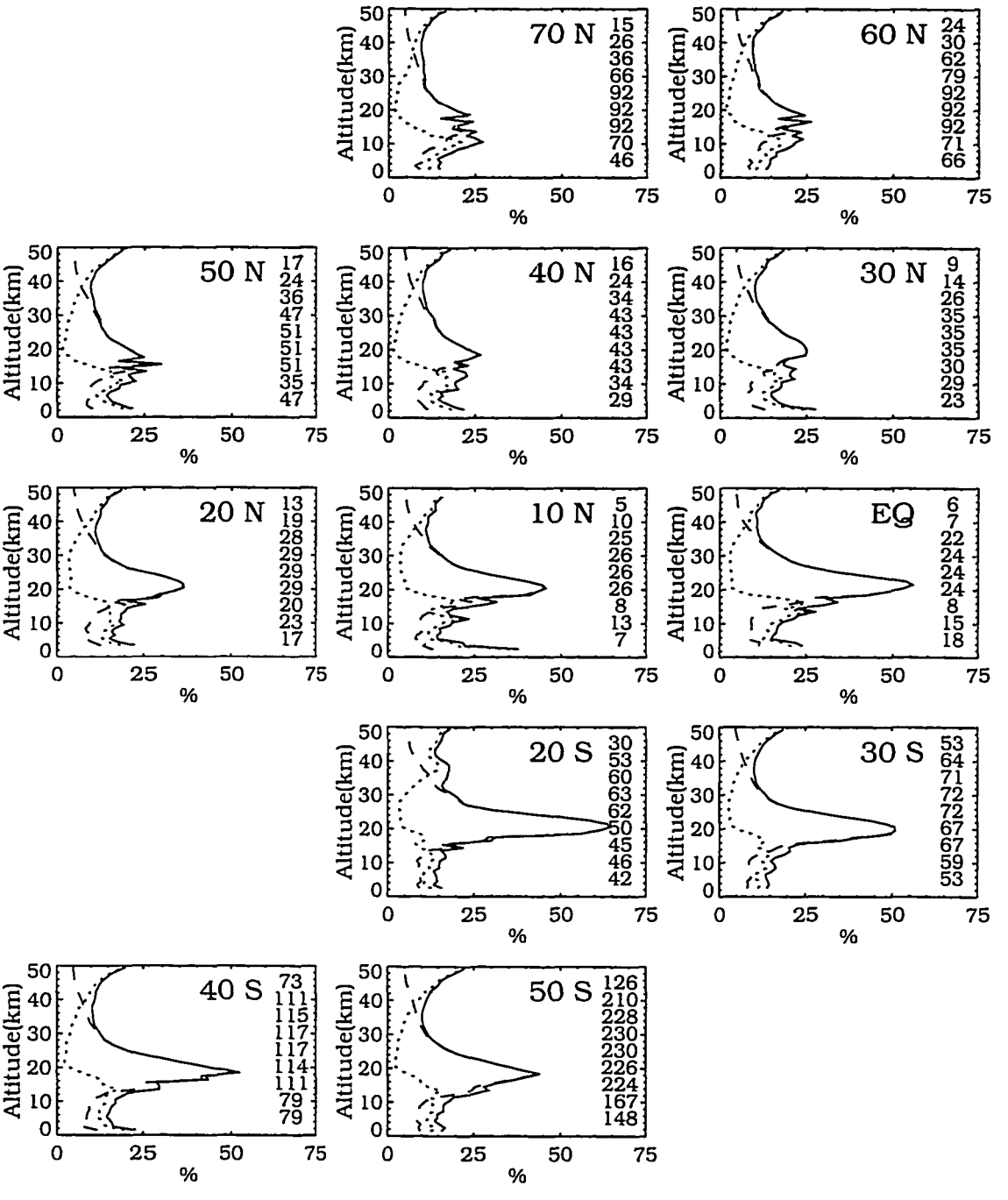


Figure D.17: Total systematic (dashed), random (dotted), and experimental uncertainty (solid) corresponding to the operational retrieval, May 1990.

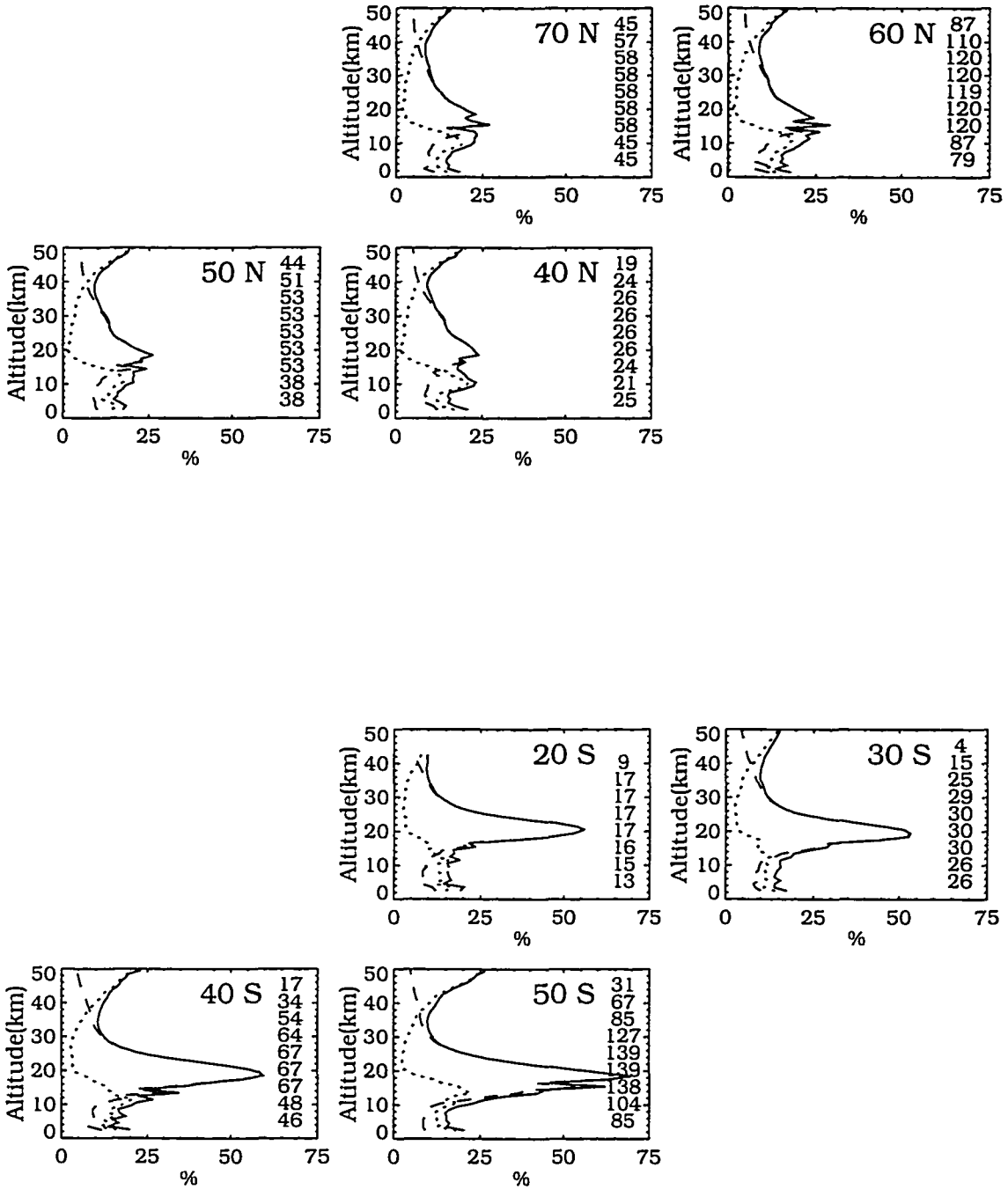


Figure D.18: Total systematic (dashed), random (dotted), and experimental uncertainty (solid) corresponding to the operational retrieval, June 1990.

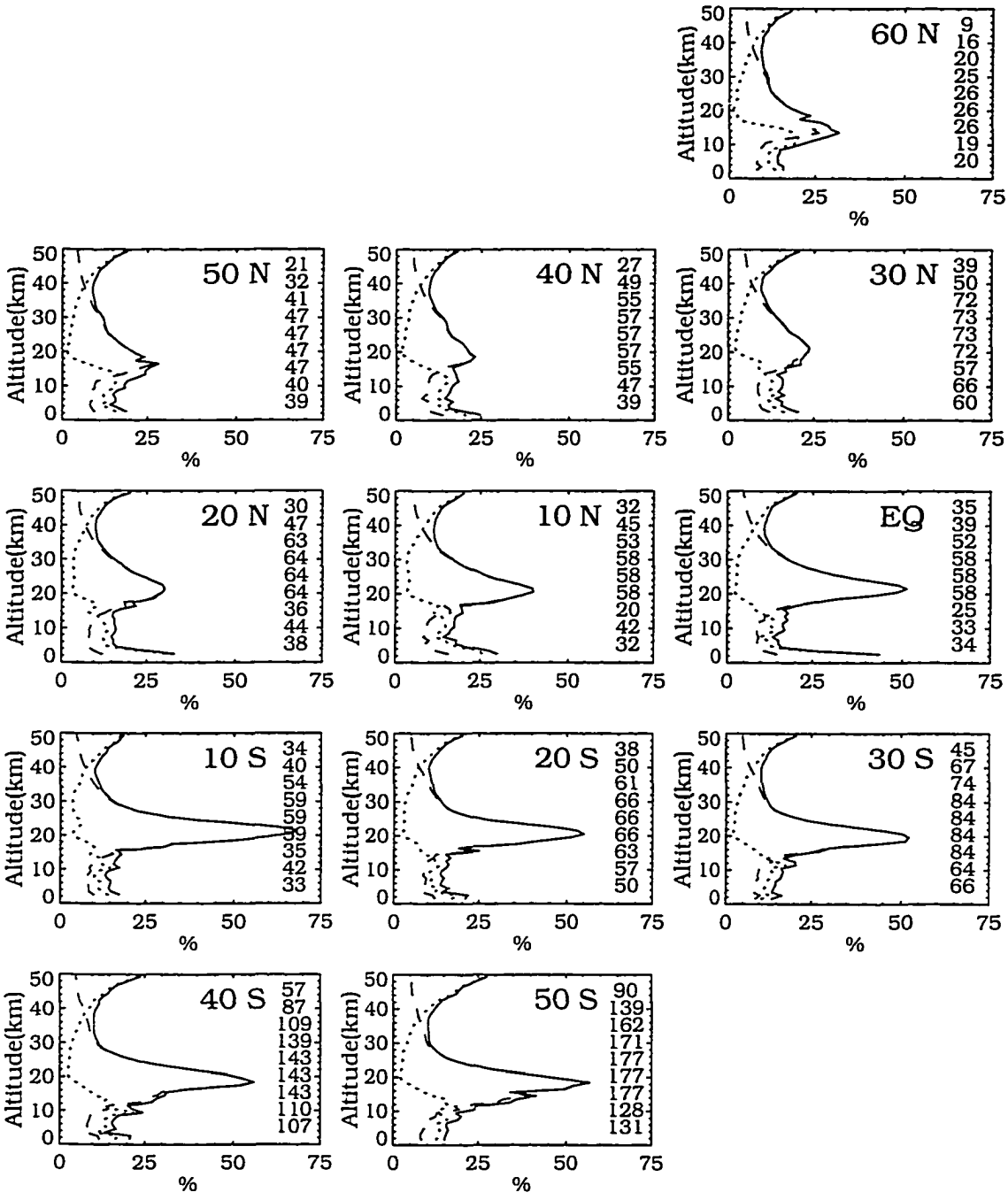


Figure D.19: Total systematic (dashed), random (dotted), and experimental uncertainty (solid) corresponding to the operational retrieval, July 1990.

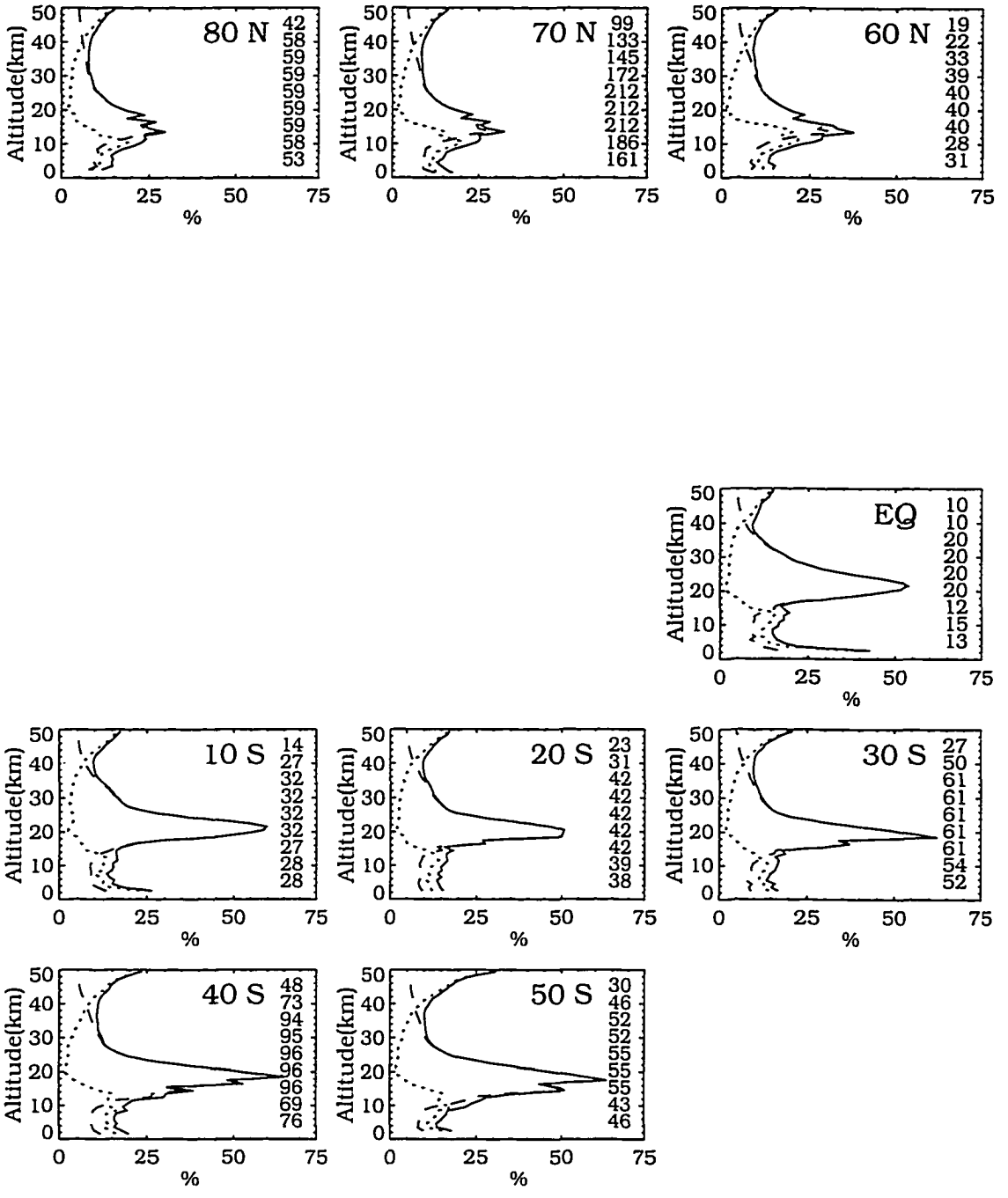


Figure D.20: Total systematic (dashed), random (dotted), and experimental uncertainty (solid) corresponding to the operational retrieval, August 1990.

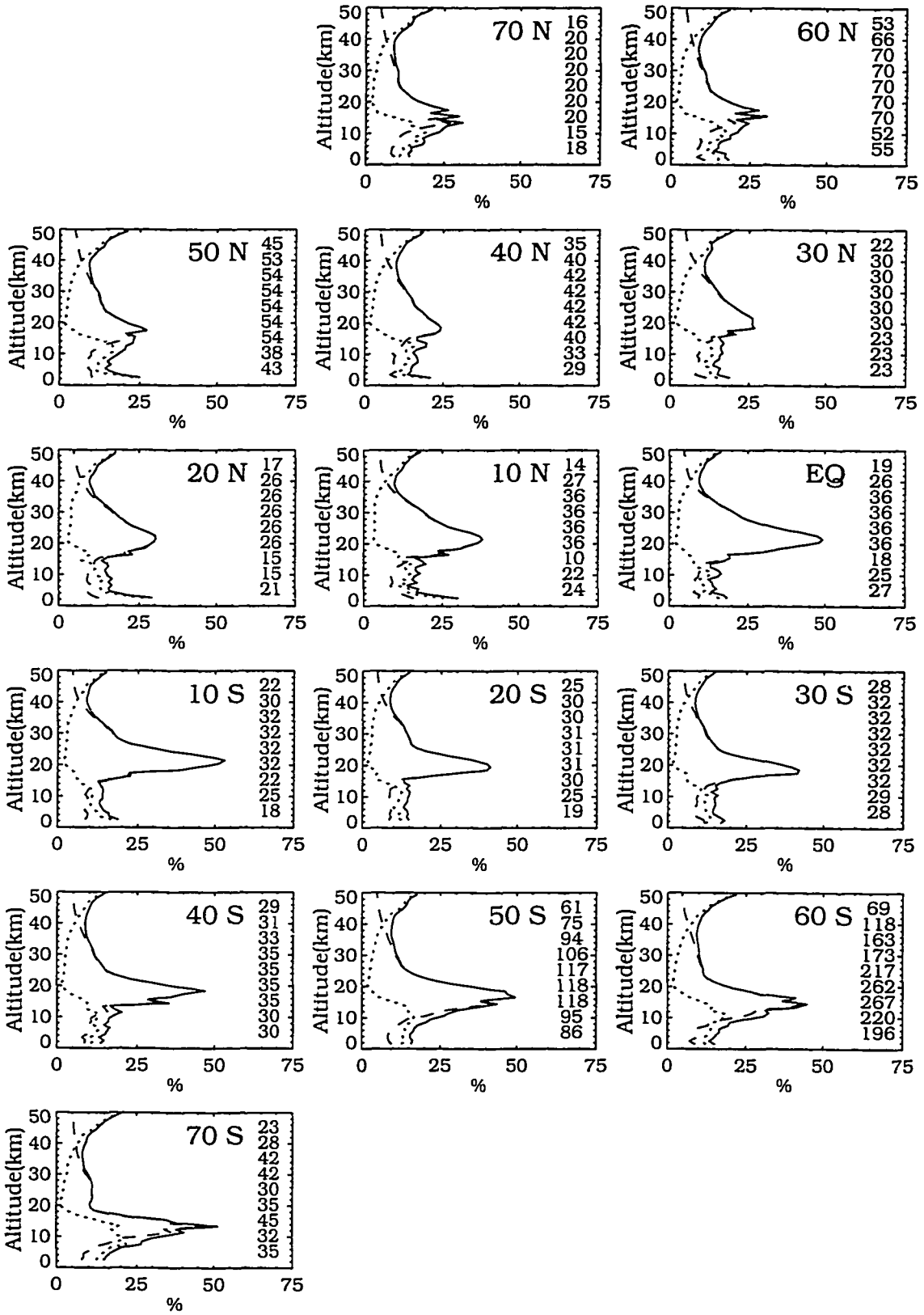


Figure D.21: Total systematic (dashed), random (dotted), and experimental uncertainty (solid) corresponding to the operational retrieval, September 1990.

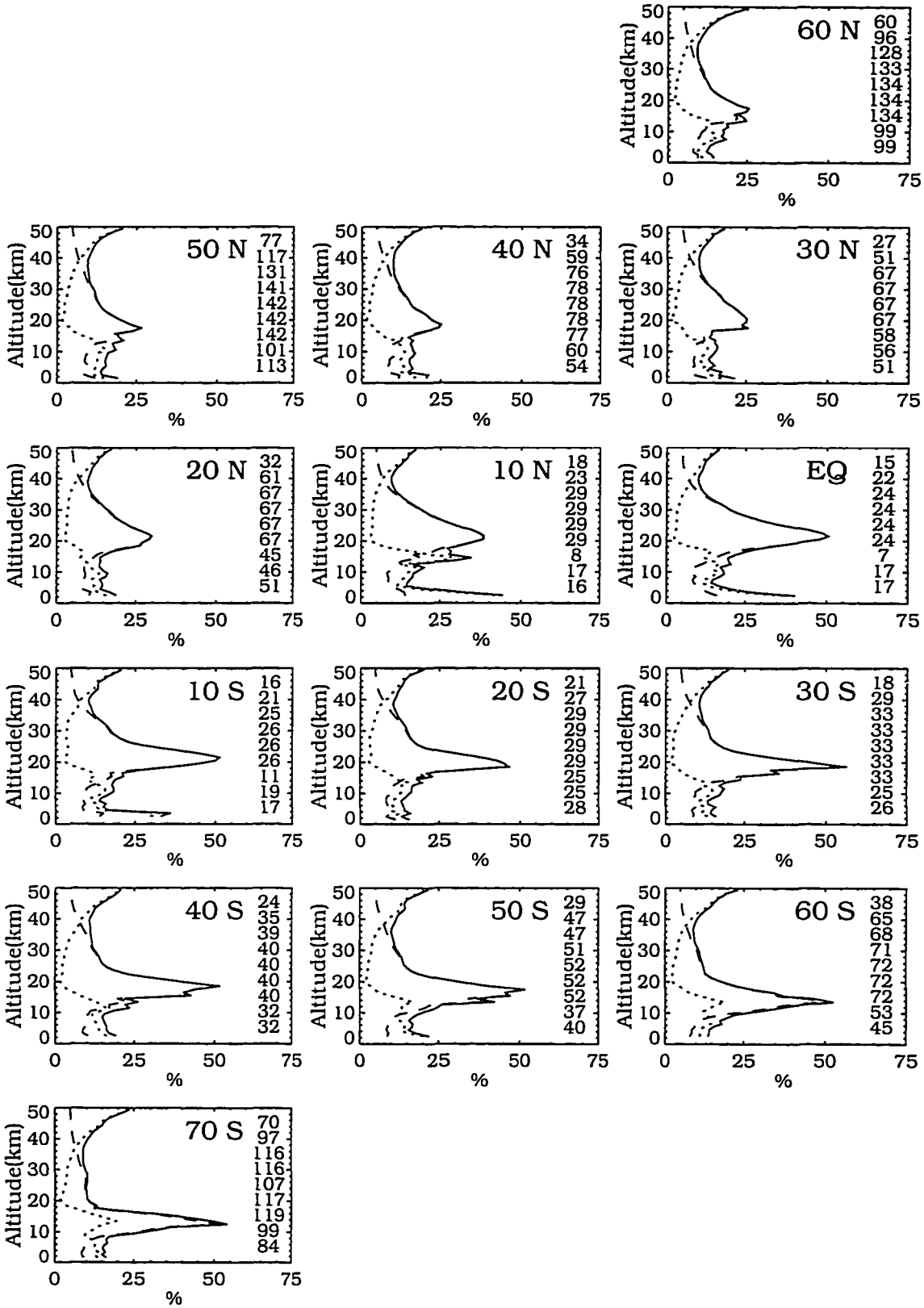


Figure D.22: Total systematic (dashed), random (dotted), and experimental uncertainty (solid) corresponding to the operational retrieval, October 1990.

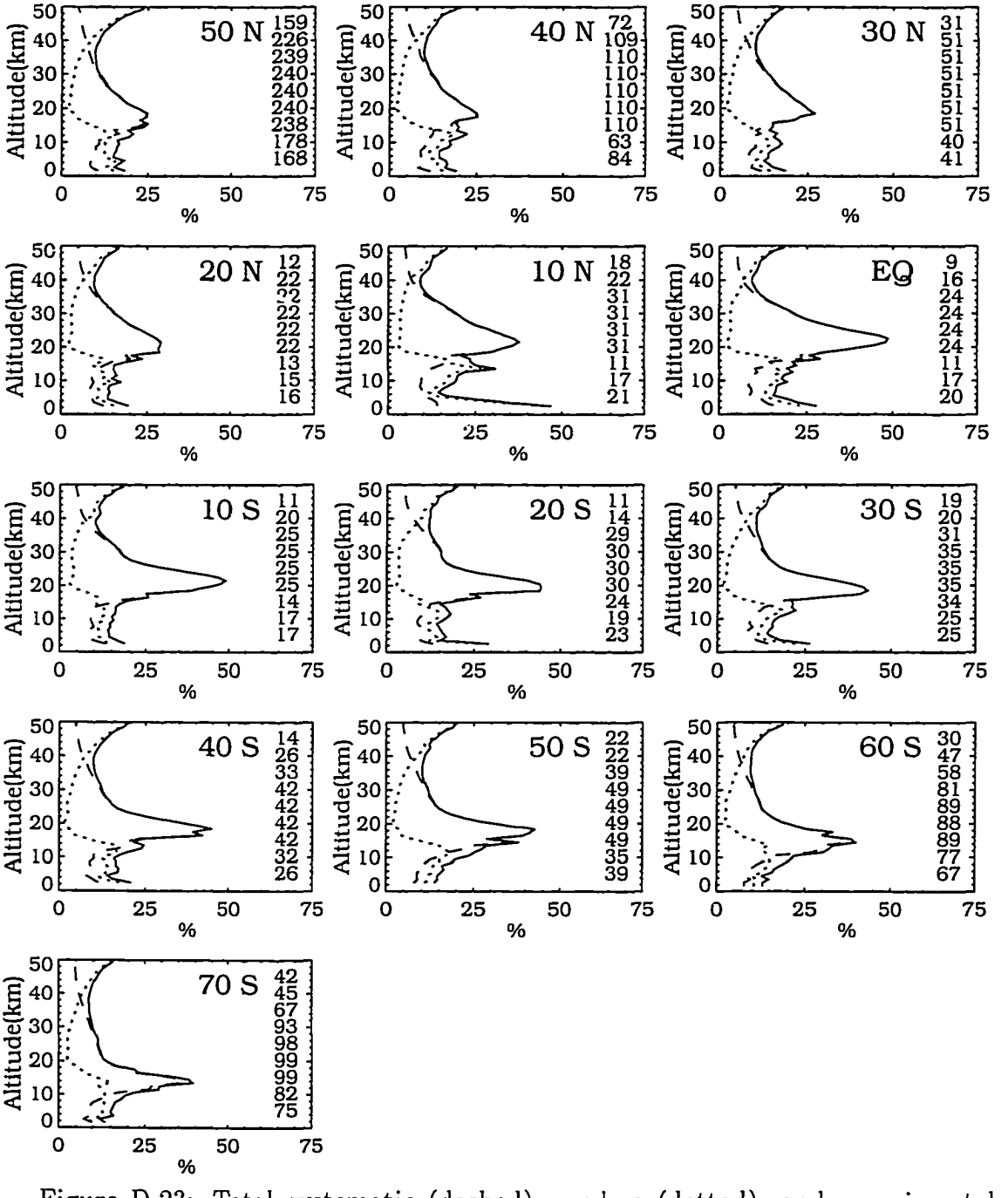


Figure D.23: Total systematic (dashed), random (dotted), and experimental uncertainty (solid) corresponding to the operational retrieval, November 1990.

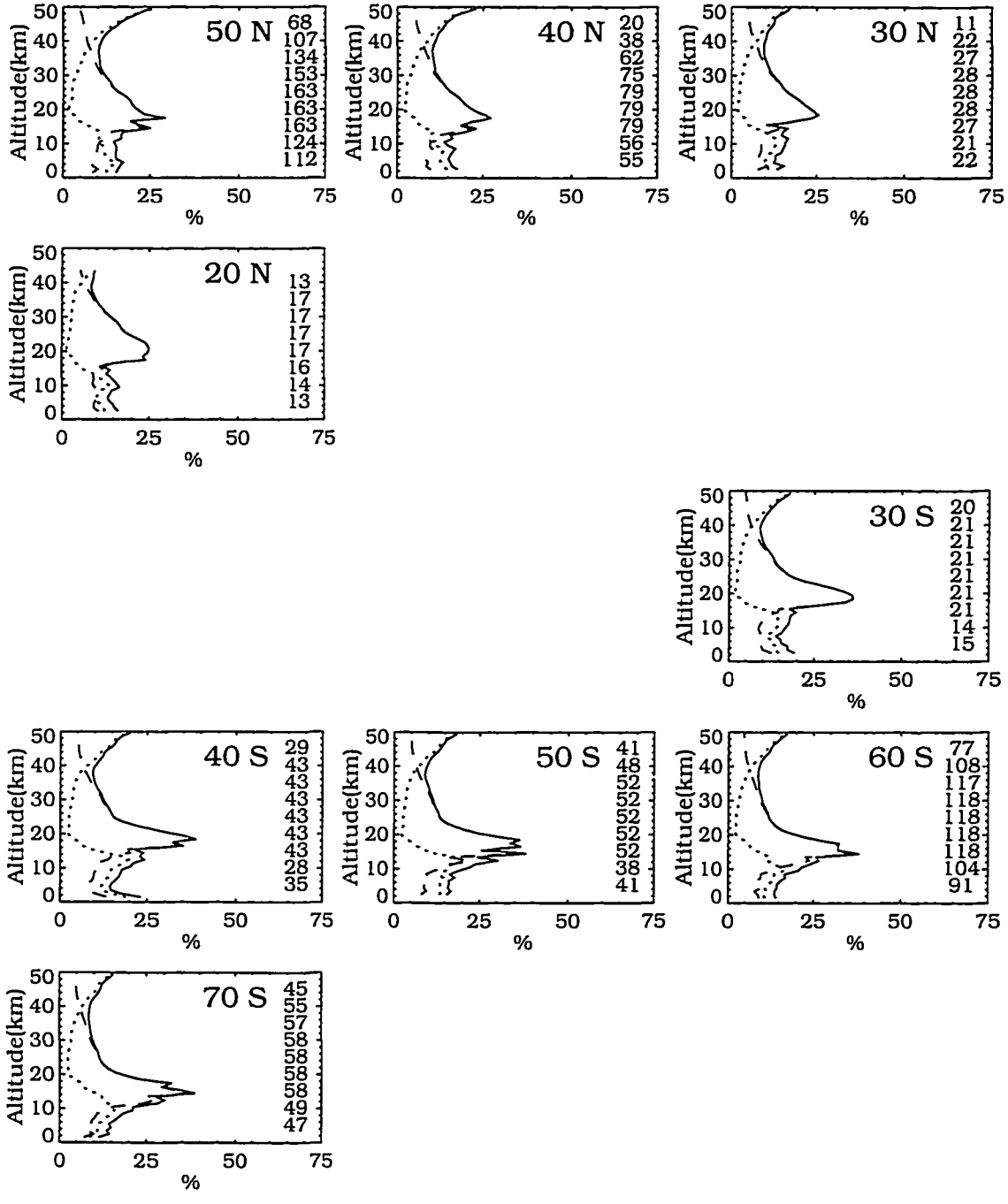


Figure D.24: Total systematic (dashed), random (dotted), and experimental uncertainty (solid) corresponding to the operational retrieval, December 1990.

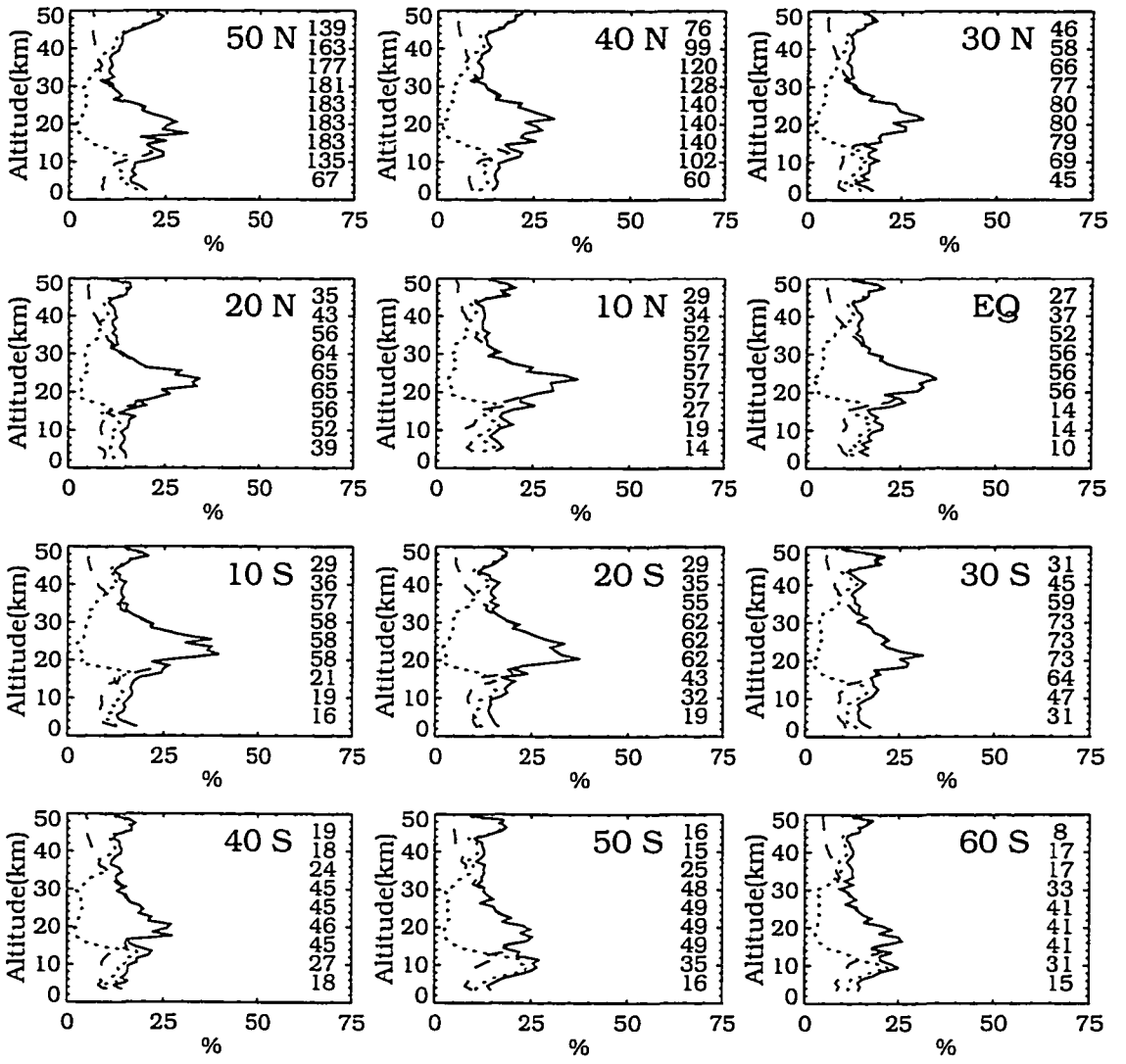


Figure D.25: Total systematic (dashed), random (dotted), and experimental uncertainty (solid) corresponding to the Mill-Drayson retrieval, January 1990.

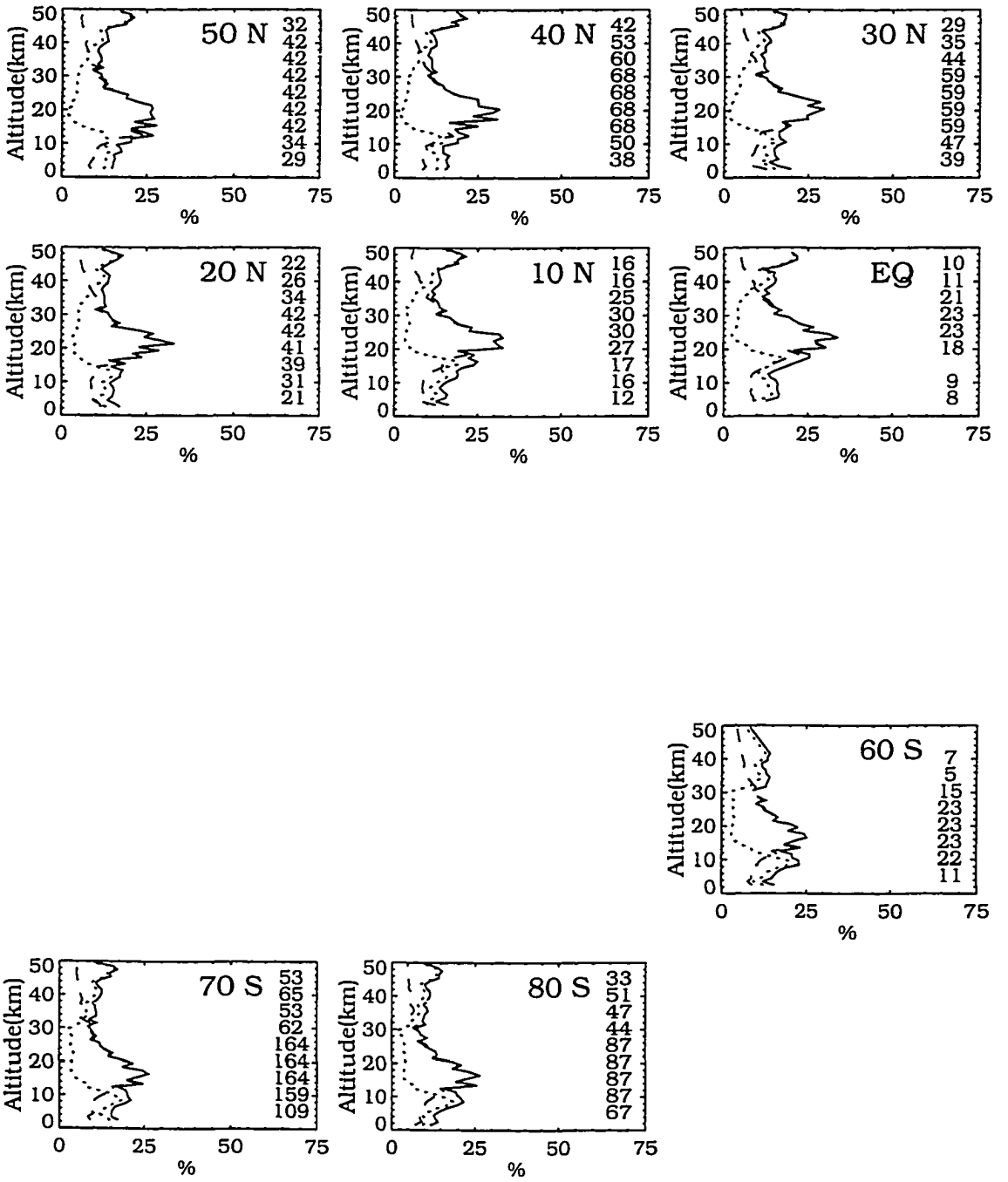


Figure D.26: Total systematic (dashed), random (dotted), and experimental uncertainty (solid) corresponding to the Mill-Drayson retrieval, February 1990.

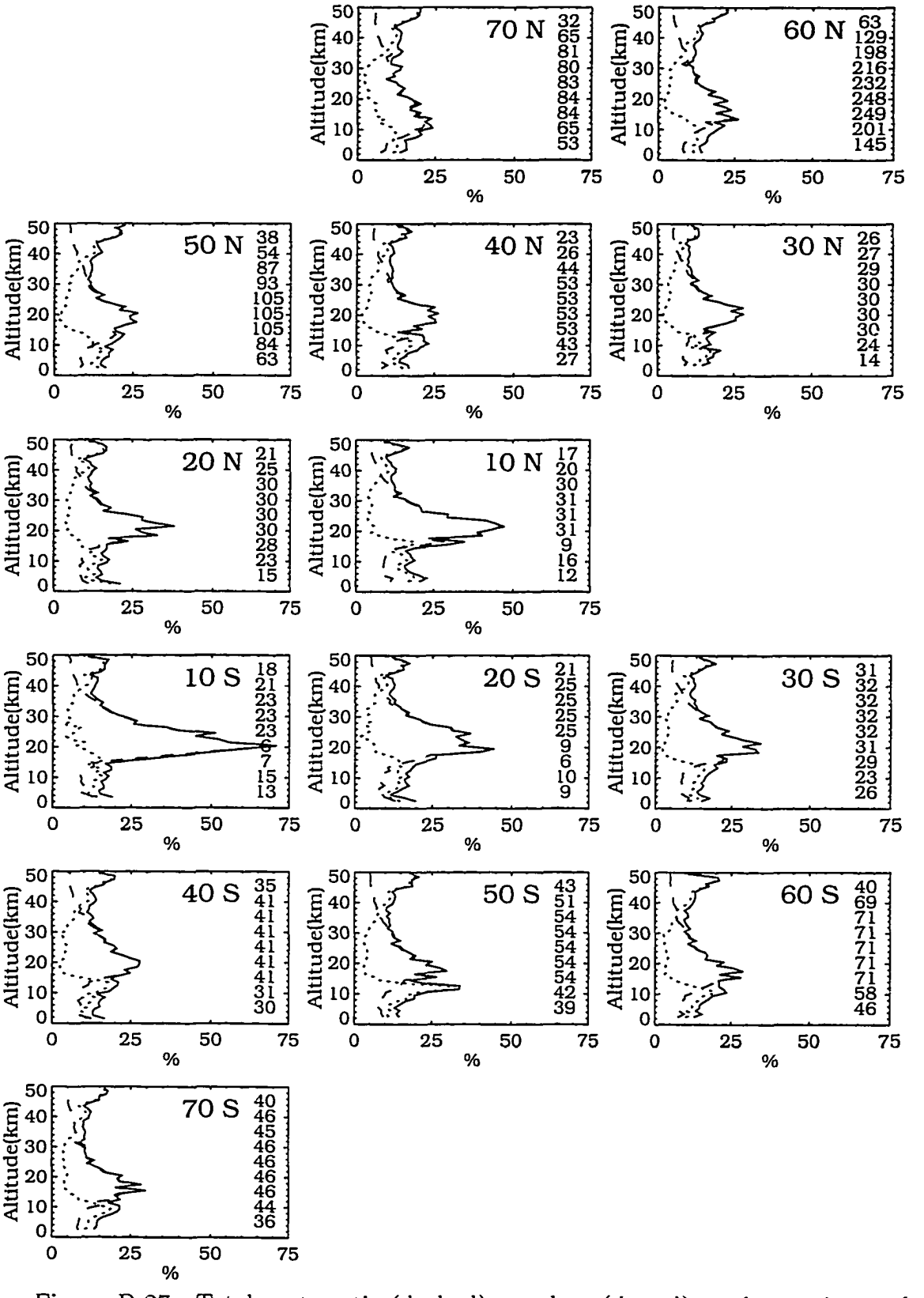


Figure D.27: Total systematic (dashed), random (dotted), and experimental uncertainty (solid) corresponding to the Mill-Drayson retrieval, March 1990.

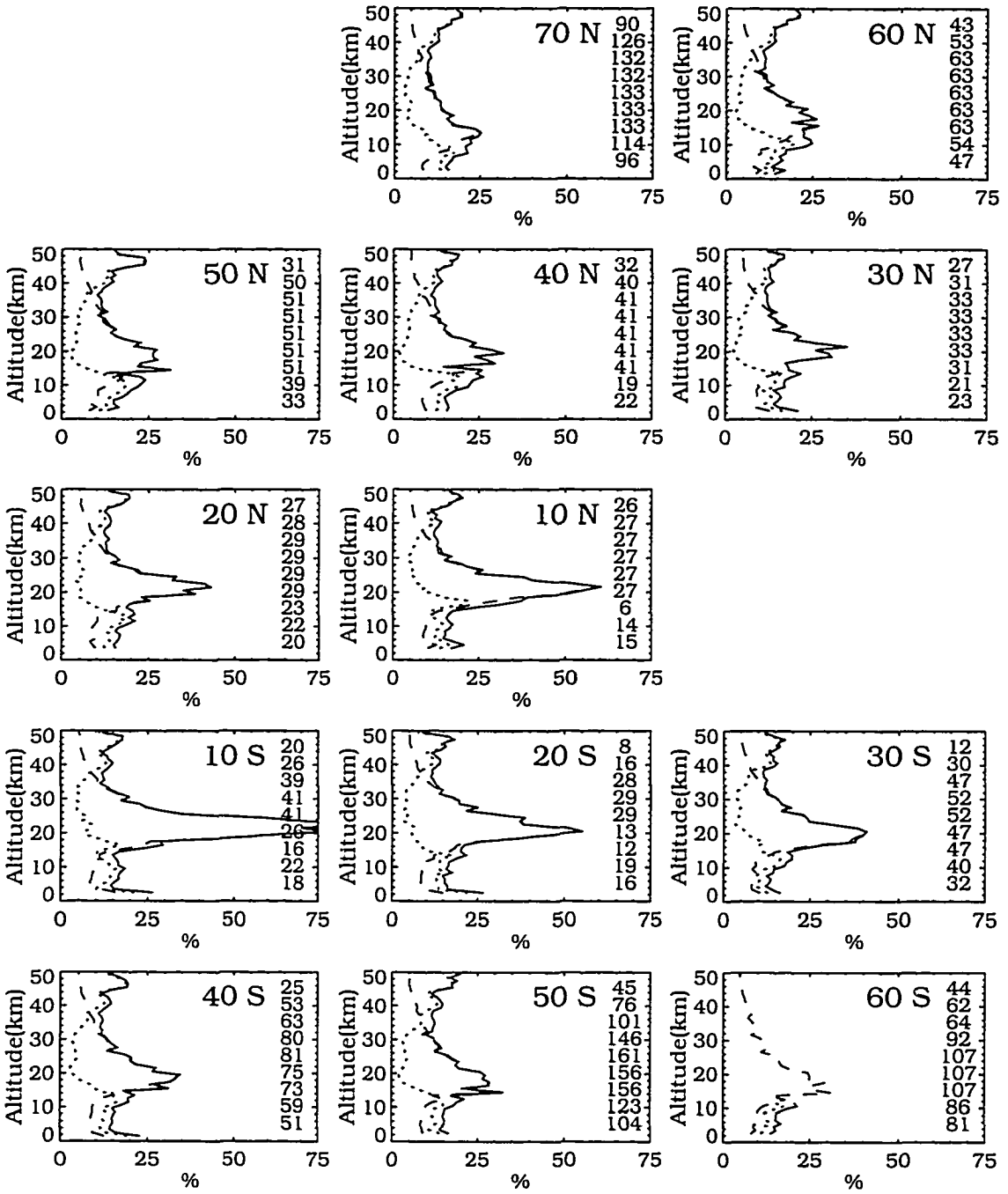


Figure D.28: Total systematic (dashed), random (dotted), and experimental uncertainty (solid) corresponding to the Mill-Drayson retrieval, April 1990.

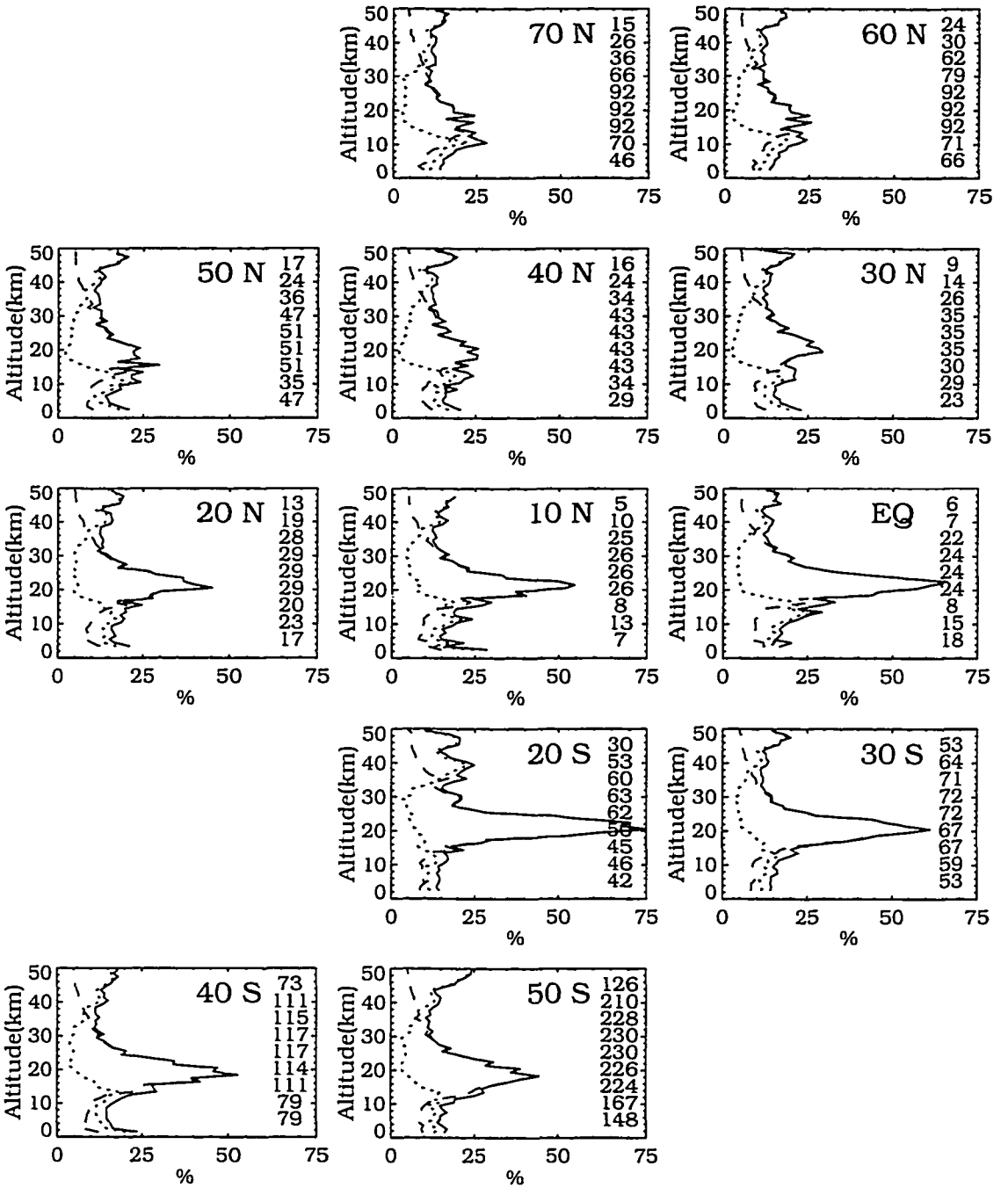


Figure D.29: Total systematic (dashed), random (dotted), and experimental uncertainty (solid) corresponding to the Mill-Drayson retrieval, May 1990.

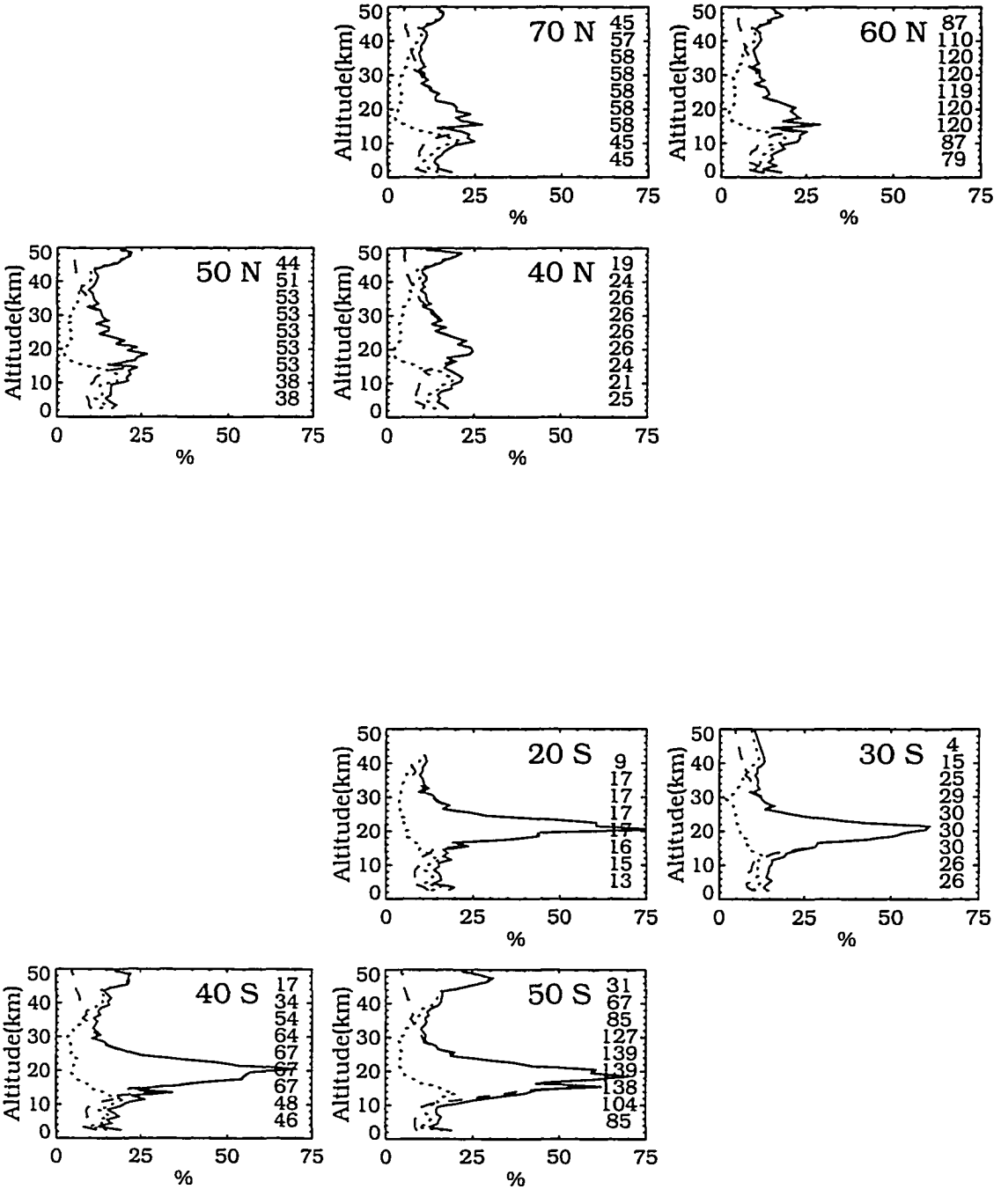


Figure D.30: Total systematic (dashed), random (dotted), and experimental uncertainty (solid) corresponding to the Mill-Drayson retrieval, June 1990.

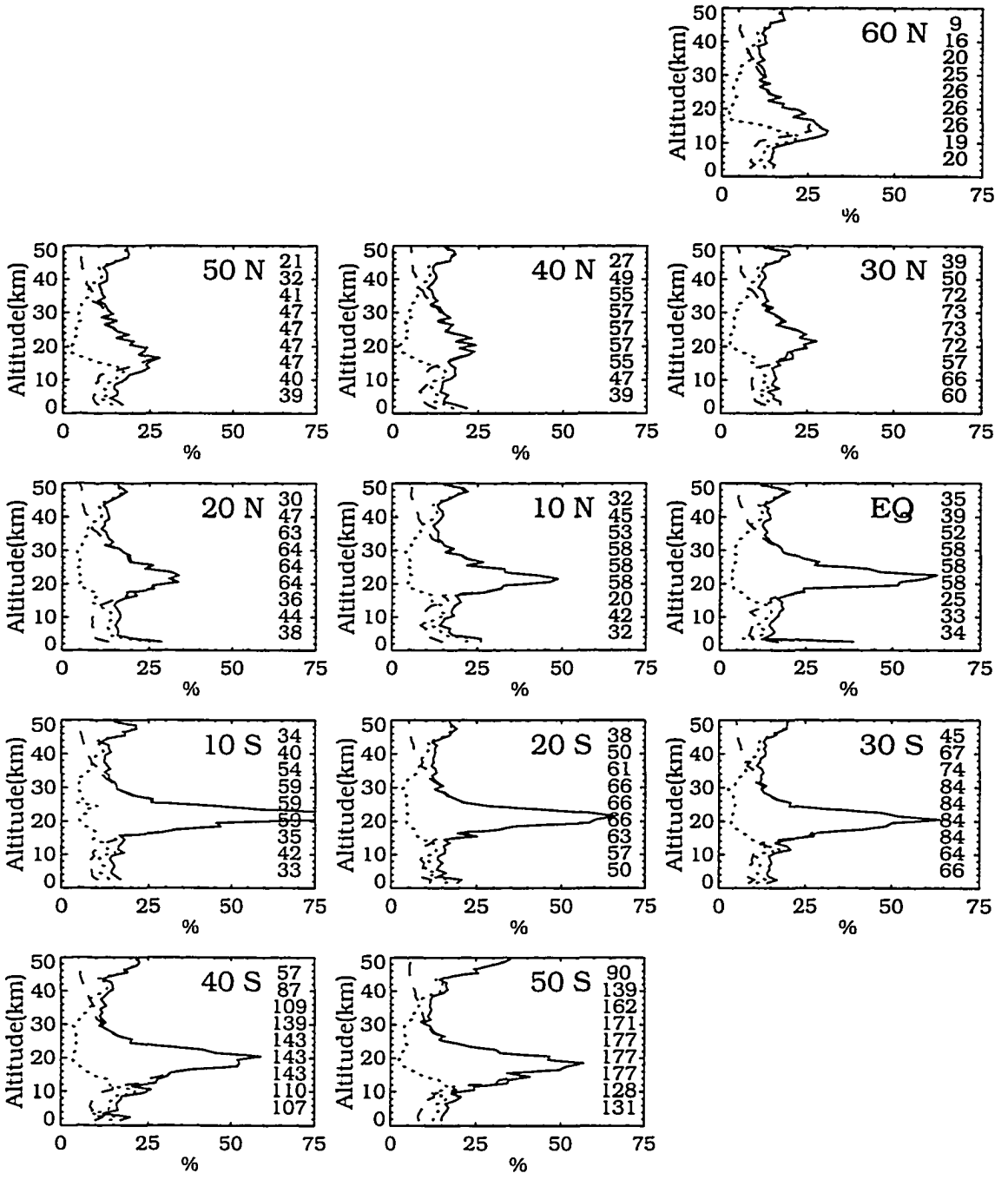


Figure D.31: Total systematic (dashed), random (dotted), and experimental uncertainty (solid) corresponding to the Mill-Drayson retrieval, July 1990.

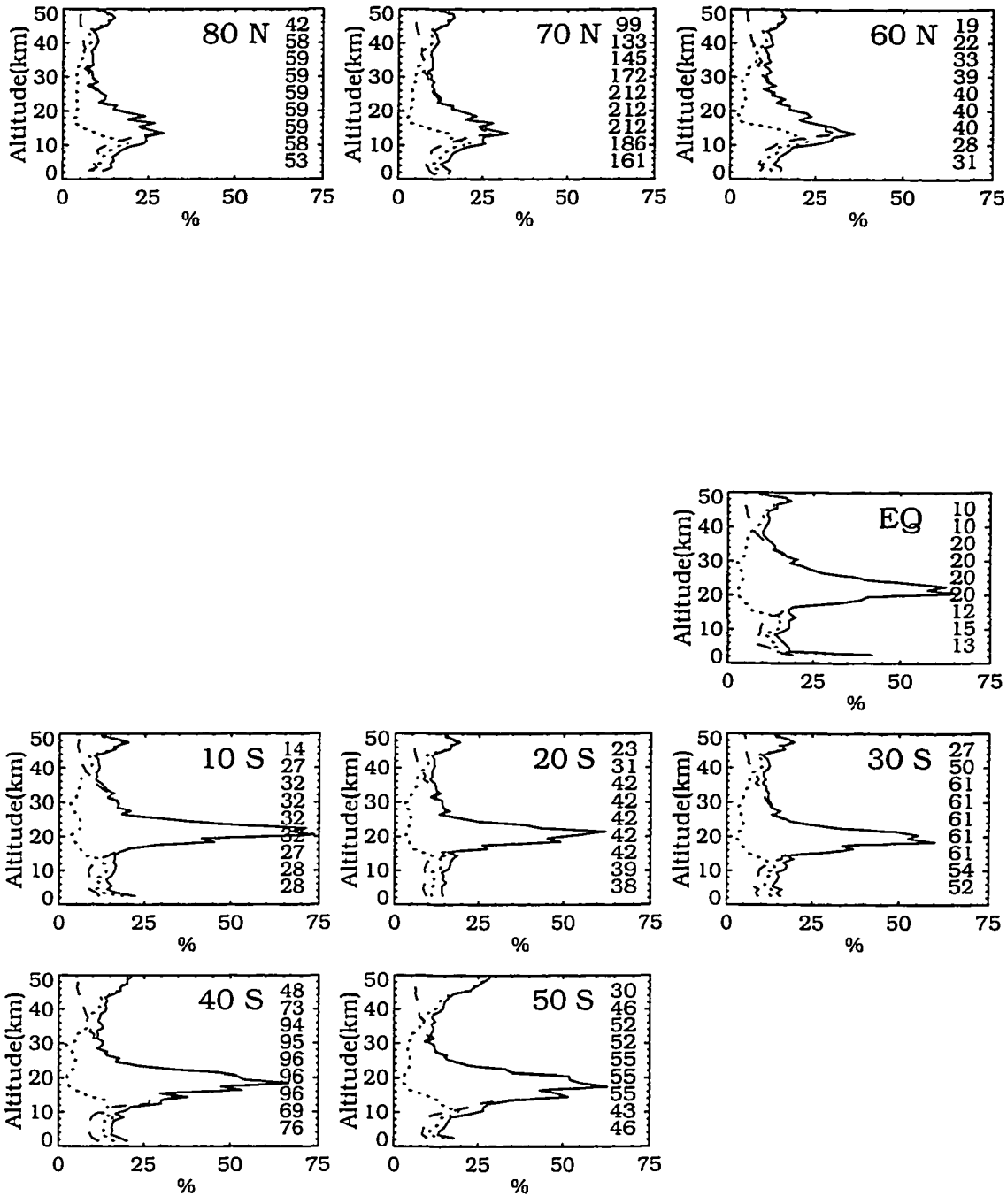


Figure D.32: Total systematic (dashed), random (dotted), and experimental uncertainty (solid) corresponding to the Mill-Drayson retrieval, August 1990.

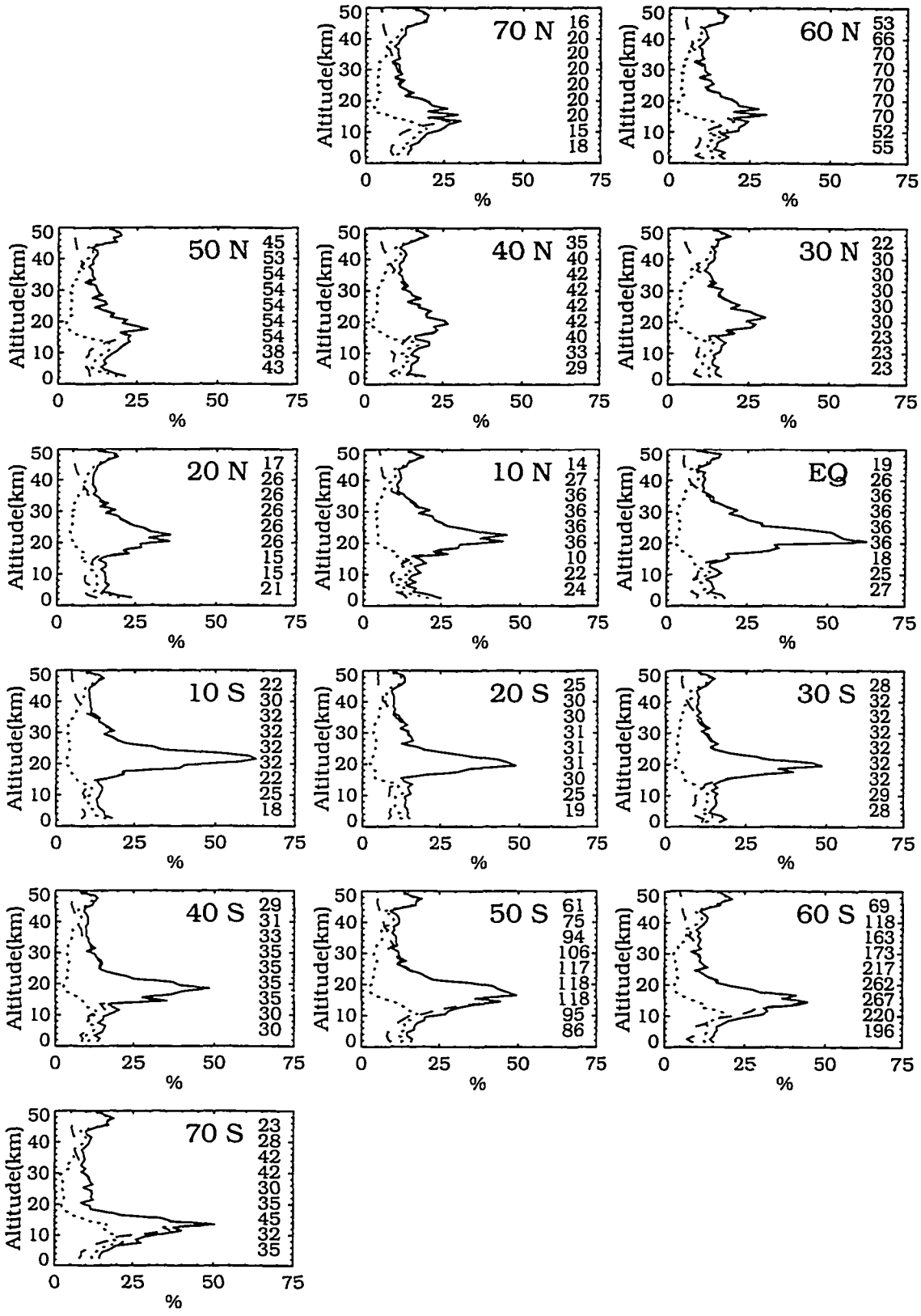


Figure D.33: Total systematic (dashed), random (dotted), and experimental uncertainty (solid) corresponding to the Mill-Drayson retrieval, September 1990.

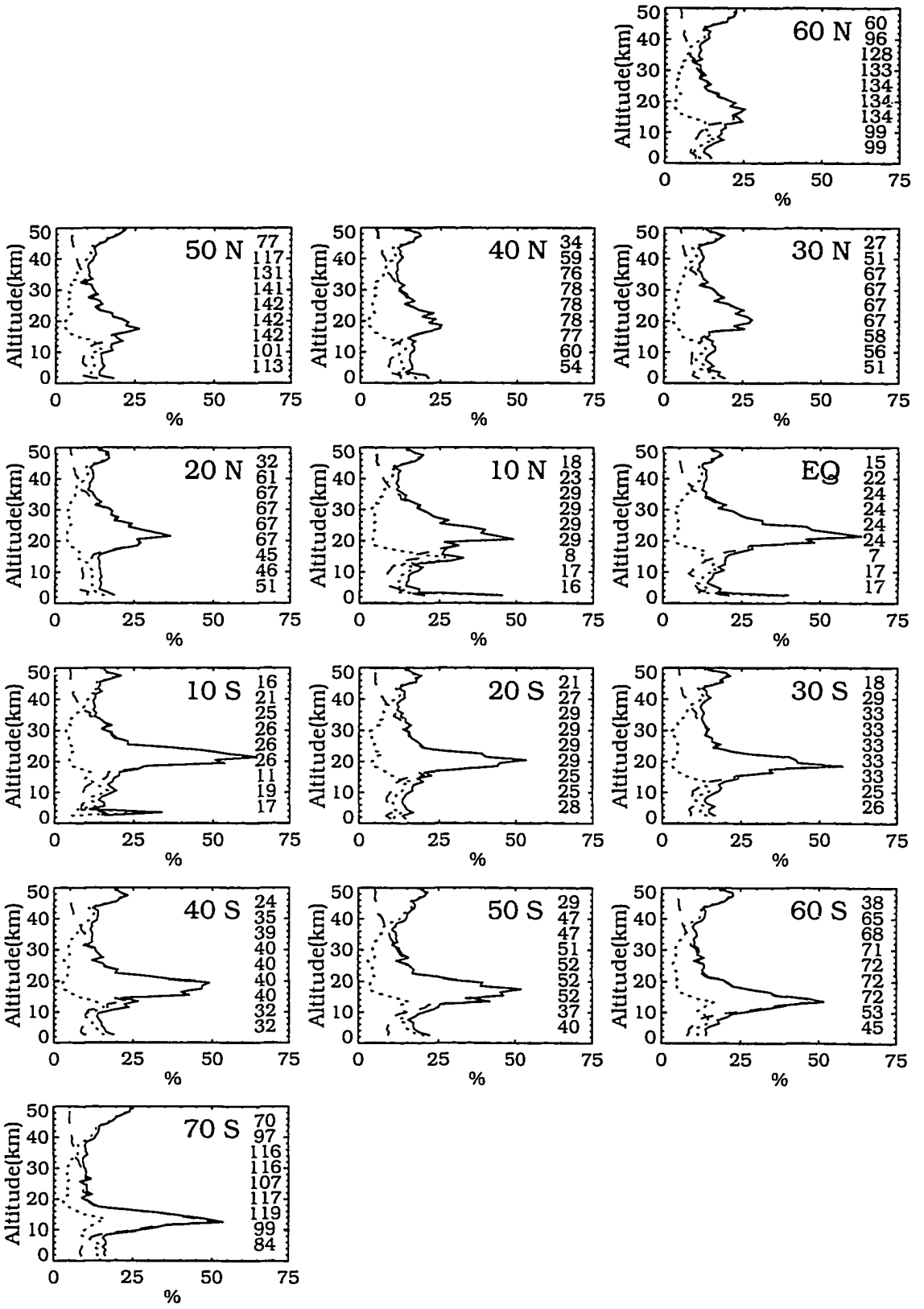


Figure D.34: Total systematic (dashed), random (dotted), and experimental uncertainty (solid) corresponding to the Mill-Drayson retrieval, October 1990.

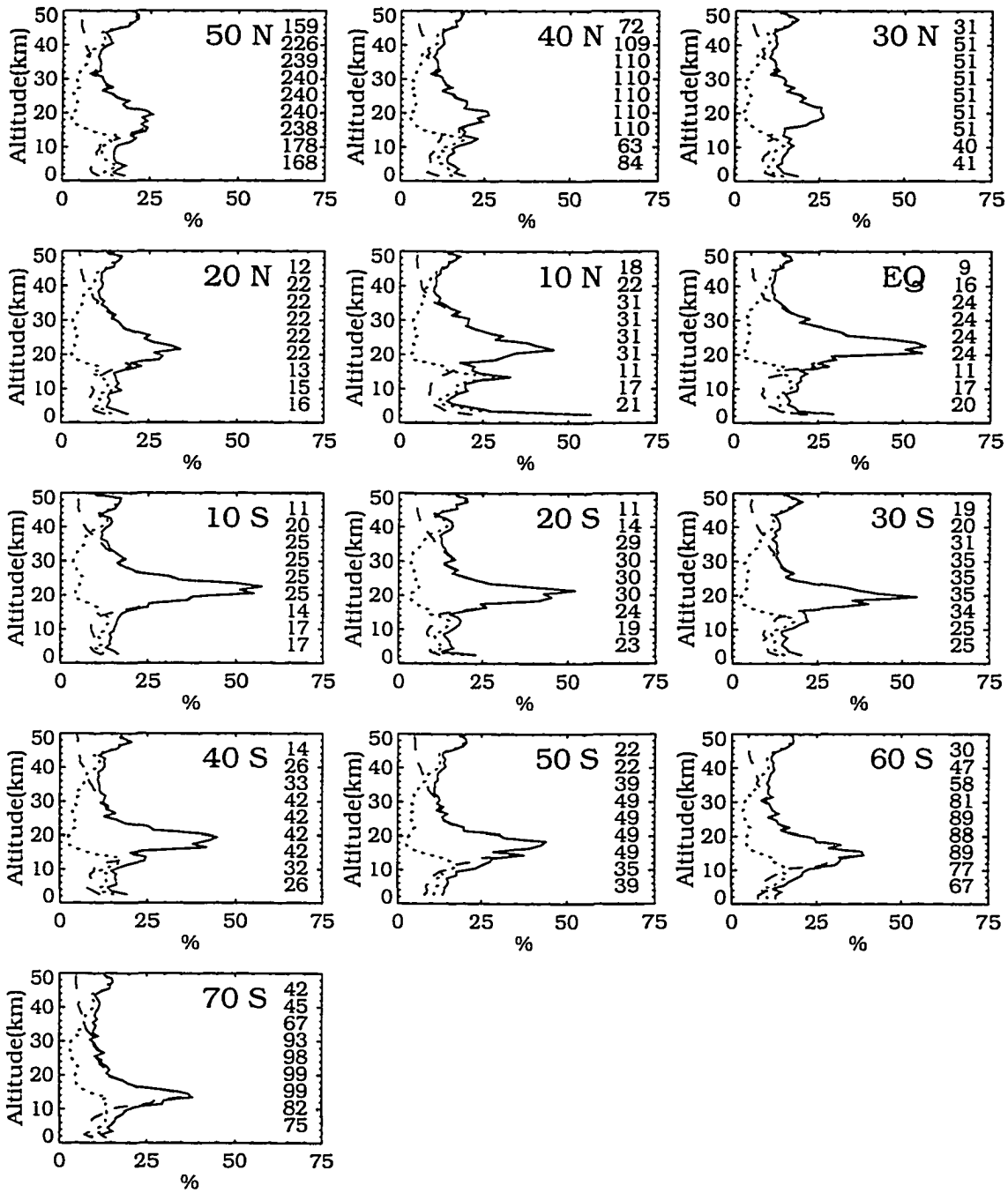


Figure D.35: Total systematic (dashed), random (dotted), and experimental uncertainty (solid) corresponding to the Mill-Drayson retrieval, November 1990.

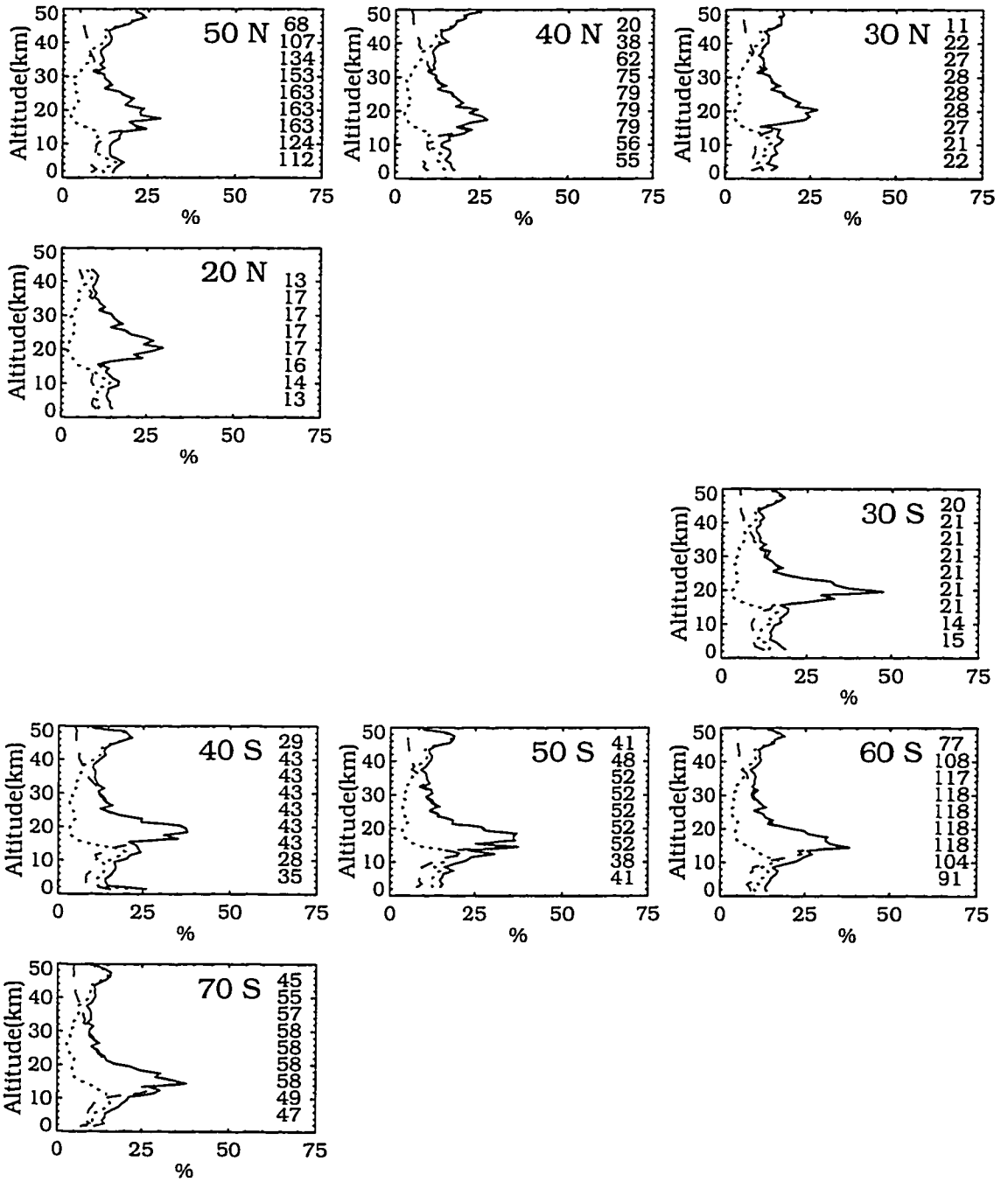


Figure D.36: Total systematic (dashed), random (dotted), and experimental uncertainty (solid) corresponding to the Mill-Drayson retrieval, December 1990.

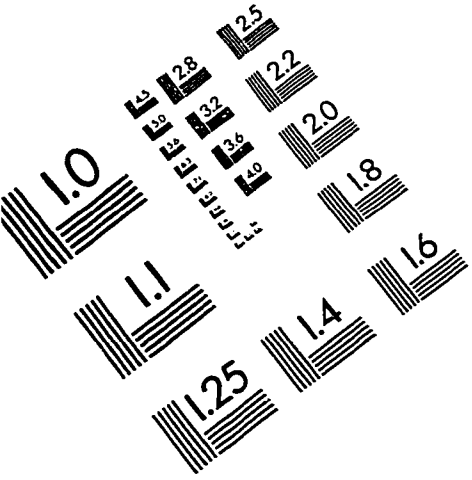
VITA

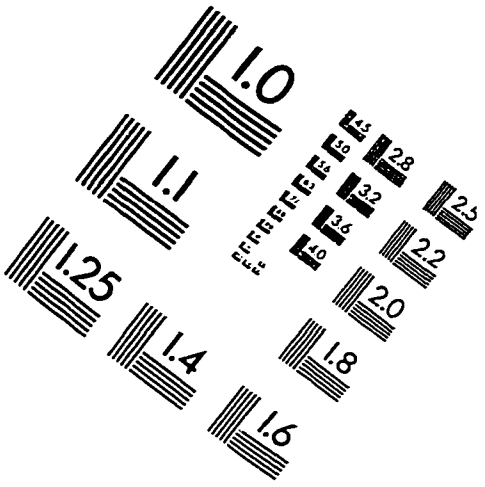
Jack Cutler Larsen was born on January 29, 1952, in Sunbury, Pa., and resided in the Cleveland, Ohio, area during his early years where he graduated from Berea High School in 1970. He enrolled at Ohio State University where he received the Bachelor of Science degree (with honors) in Electrical Engineering in March 1975. In the Fall of 1975 he began his graduate studies at The George Washington University in the Environmental Modeling program through the Joint Institute for the Advancement of Flight Sciences (JIAFS) at NASA Langley Research Center. He received the Master of Science degree from The George Washington University in September of 1977. He taught mathematics for a semester at The Old Dominion University for the PACE program and then joined the staff of Systems and Applied Sciences (now known as Raytheon STX) in Hampton, Va.

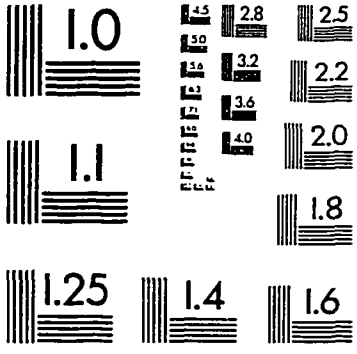
From the Fall of 1977 to the Spring of 1992 Mr. Larsen held the position of atmospheric scientist with STX where he worked on several atmospheric remote sensing projects. These include the Laser Heterodyne Spectrometer experiment, analysis of high resolution infrared atmospheric absorption interferometer spectra, and the Stratospheric Aerosol and Gas Experiment II. From 1979 to 1984, Mr. Larsen was enrolled in the Doctoral Program at Old Dominion University where he completed many of his academic requirements. In June of 1992 he returned full time to complete the degree. Mr. Larsen rejoined Raytheon STX in Lanham, Md., as a principal scientist working on the next generation design of ozone profiling instruments.

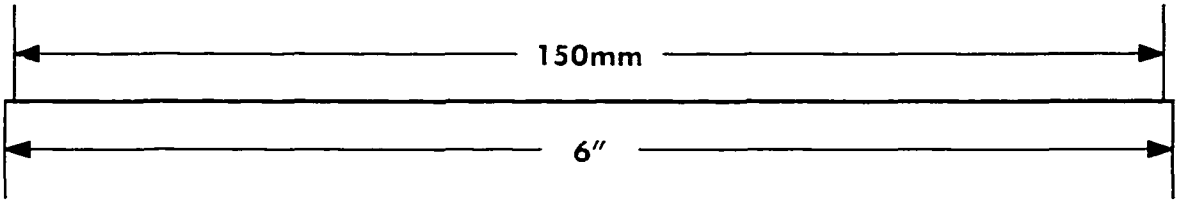
Mr. Larsen is author/co-author of 20 papers and over 30 presentations related to atmospheric remote sensing, inversion methods, and radiative transfer. He is a co-recipient of numerous NASA awards. Mr. Larsen is a member of the American Meteorological Society and the American Institute of Aeronautics and Astronautics.

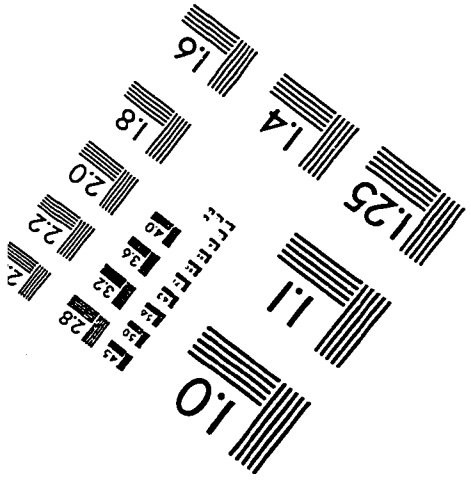
IMAGE EVALUATION TEST TARGET (QA-3)

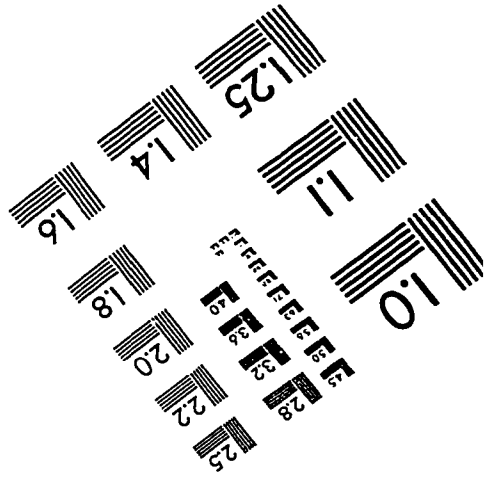












APPLIED IMAGE, Inc
1653 East Main Street
Rochester, NY 14609 USA
Phone: 716/482-0300
Fax: 716/288-5989