REGRESSION ANALYSIS OF COLUMN OZONE AND SELECTED ATMOSPHERIC PARAMETERS IN PENINSULAR MALAYSIA FROM SCIAMACHY DATA

TAN KOK CHOOI

UNIVERSITI SAINS MALAYSIA

2015

REGRESSION ANALYSIS OF COLUMN OZONE AND SELECTED ATMOSPHERIC PARAMETERS IN PENINSULAR MALAYSIA FROM SCIAMACHY DATA

by

TAN KOK CHOOI

Thesis submitted in fulfillment of the requirements for the degree of Doctor of Philosophy

APRIL 2015

ACKNOWLEDGEMENTS

First, I want to sincerely thank and acknowledge all related parties and individuals who advised and helped me in finishing this research. Without a doubt, the most important people to whom I must convey 100 % of my respect and thanks are my supervisors, Associate Professor Lim Hwee San and Professor Mohamad Zubir Mat Jafri. They have guided, supported, and helped me throughout the course of this research, which has enabled me to develop an understanding of the subject. Thank you very much for your encouragement, attention and patience in guiding me; I can finally finish my research to fulfill the requirements for PhD study. Furthermore, the motivation that they provided contributed to my dedication to finish this research and not easily give up on solving all of the challenges that surfaced throughout my research.

In addition, I also want to thank my beloved parents and family. Their confidence and encouragement were a main factor in helping me successfully conduct and finish the research on time at Universiti Sains Malaysia. Although we live far away from each other, their love and support has driven me and given me courage to continue the PhD program and to finish this research.

Furthermore, I would like to express my gratitude to the Education Ministry for providing me financial support from MyPhD, under the MyBrain15 program, throughout my research. A special thanks to the Universiti Sains Malaysia’s Institute of Postgraduate Students for awarding a Postgraduate Research Grant (PRGS, Grant No. 1001/PFIZIK/845007) to finance my research. My appreciation also goes to the SCIAMACHY staff for their generosity in sharing some valuable satellite data with the public.

In addition, many thanks to the staff of the School of Physics, Universiti Sains Malaysia, especially for the help and cooperation of those who directly or indirectly participated in this research from the project’s start until its finish. Without their help and participation, my research could not have been conducted smoothly and finished on time.

Finally, I feel so grateful and proud to have numerous lab mates who also conducted research in our shared laboratory. I wish to thank them for their valuable experience and for sharing their knowledge throughout this research.

TAN KOK CHOOI
2015

TABLE OF CONTENTS

LIST OF TABLES

PAGE

LIST OF FIGURES

LIST OF SYMBOLS

Ar	Argon
B	Unstandardized Coefficients
β	Standardized Beta Coefficient
BrO	Bromide Monoxide
CH_4	Methane
CFCs	Chlorofluorocarbons
CO	Carbon Monoxide
CH_2O	Formaldehyde
CO_2	Carbon Dioxide
H_2CO	Formaldehyde
H_2O	Water
HCFCs	Hydrochlorofluorocarbons
InGaAs	Indium Gallium Arsenide
NO	Nitric Oxide
NO_2	Nitrogen Dioxide
N_2O	Nitrous Oxide
NMHC	Non-Methane HydroCarbon
NO_X	Nitrogen Oxides
O_3	Ozone
OCIO	Chlorine Dioxide
OH	Hydroxyl
PM_{10}	Particulate Matter
R^2	Coefficient of Determination
Si	Silicon
SO_2	Sulfur Dioxide
VOCs	Volatile Organic Compounds

LIST OF ABBREVATIONS

ADC	Analogue Digital Converter
AIRS	Atmospheric Infrared Sounder
AMC-DOAS	Air Mass Corrected-Differential Optical Absorption Spectroscopy
AMF	Air Mass Factor
ANN	Artificial Neural Network
ARL	Air Resource Laboratory
ASEAN	Association of Southeast Asian Nation
ASMC	ASEAN Specialised Meteorological Centre
DOAS	Differential Optical Absorption Spectroscopy
DOE	Department of Environment
DU	Dobson Unit
ECMWF	European Centre for Medium-Range Weather Forecasts
EDGAR	Emission Database for Global Atmospheric Research
ENVISAT	Environmental Satellite
ESA	European Space Agency
ESRL	Earth System Research Laboratory
FB	Fractional Bias
FIR	Far-Infrared
FOV	Field Of View
FTIR	Fourier-Transform InfRared
FTS	Fourier Transform Spectroscopy
GAW	Global Atmosphere Watch
GDAS	Global Data Assimilation System
GOME	Global Ozone Monitoring Experiment
HYSPLIT	HYbrid Single-Particle Lagrangian Integrated Trajectory
IA	Index of Agreement
IFOV	Instantaneous Field Of View
IPCC	Intergovernmental Panel on Climate Change
IR	Infrared
ITCZ	InterTropical Convergence Zone
LOS	Line-Of-Sight

MMD	Malaysia Meteorological Department
MRA	Multiple Regression Analysis
MR	Multiple Regressions
MS-Excel	Microsoft Excel
MSR	Multi Sensor Reanalysis
MW	Microwave
NASA	National Aeronautics and Space Administration
NCAR	National Center for Atmospheric Research
NCEP	National Centers for Environmental Prediction
NDVI	Normalized Difference Vegetation Index
NEE	Net Ecosystem Exchange
NEM	Northeast Monsoon
NIR	Near-Infrared
NN	Neural Networks
NOAA	National Oceanic and Atmospheric Administration
NRT	Near Real Time
OL	Off-Line
OMI	Ozone Monitoring Instrument
PCA	Principal Component Analysis
PCs	Principal Components
PMD	Polarization Monitoring Devices
PPM	Parts Per Million
PPBV	Parts Per Billion Volume
READY	Real-time Environmental Applications and Display sYstem
REAS	Regional Emission inventory in ASia
RMSE	Root Mean Square Errors
SCIAMACHY	SCanning Imaging Absorption spectroMeter for Atmospheric CHartographY
SCIA-VALUE	SCIAMACHY VALidation and Utilization Experiment
SCD	Slant Column Density
SPSS	Statistical Package for Social Sciences
SSM/I	Special Sensor Microwave Imager
SWIR	Short-wave Infrared

SWM	Southwest Monsoon
TOMS	Total Ozone Mapping Spectrometer
UNEP	United Nations Environment Programme
UV	Ultraviolet
VIS	Visible
VMR	Volume Mixing Ratio
WFM-DOAS	Weighting Function Modified Differential Optical Absorption Spectroscopy
WMO	World Meteorological Organization

ANALISIS REGRESI TULUS OZON DAN PARAMETER ATMOSFERA TERPILIH DI SEMENANJUNG MALAYSIA MELALUI DATA SCIAMACHY

ABSTRAK

Ozon (O_3) adalah unik di kalangan bahan-bahan pencemar kerana terhasil daripada tindak balas kimia yang kompleks di atmosfera, kerana ia tidak dilepaskan secara terus ke udara. Kedudukan ozon di atmosfera memberi kesan yang berlainan terhadap kehidupan di muka bumi (sama ada bahaya atau melindungi). Ini merupakan sebab utama O_3 menyebabkan masalah alam sekitar yang serius . Malah masalah ini adalah susah dikawal dan diramal. Objektif utama kajian ini adalah untuk membangunkan algoritma baru bagi tulus ozon di Semenanjung Malaysia dengan menggunakan kaedah statistik. Pada akhir kajian, empat persamaan regresi O_3 NEM, O_3 SWM, PCA1 (O_3 musim NEM (monsun timur laut)), dan PCA2 (O_3 musim SWM (monsun barat daya)) telah dibangunkan. Di samping itu, kajian ini turut fokus terhadap analisis dan mengkaji impak parameter-parameter atmosfera terpilih pada nilai tulus ozon di Semenanjung Malaysia. Teknik analisis regresi berganda (MRA) dan analisa komponen utama (PCA) telah digunakan untuk mencapai objektif kajian. MRA digunakan untuk menerbitkan persamaan regresi bagi O_3 NEM dan O_3 SWM. Manakala, kombinasi bagi teknik MRA dan PCA digunakan untuk menerbitkan persamaan regresi bagi PCA1 dan PCA2. Analisis statistik telah dilaksanakan untuk menguji prestasi bagi teknik MRA dan kombinasi bagi teknik MRA dan PCA, dari segi sisihan punca kuasa dua min (RMSE), persetujuan indek (IA), and pecahan kecenderungan (FB). Keputusan terbaik bagi persamaan regresi untuk tulus ozon yang diramal melalui MRA dengan

menggunakan empat pembolehubah tak bersandar didapati mempunyai korelasi yang tinggi, iaitu (R = 0.811 for SWM, R = 0.803 for NEM) dan R^2 (≈ 0.658 for SWM and ≈ 0.645 for NEM) untuk data selama enam-tahun (2003-2008). Keputusan pengesahan, R bagi NEM dan SWM monsun masing-masing menunjukkan pekali korelasi yang tinggi, iaitu 0.783 - 0.799 dan 0.752 - 0.802. Keputusan penyesuaian terbaik untuk data O_3 memberikan nilai pekali R (≈ 0.83) bagi kedua-dua musim NEM dan SWM dengan menggunakan empat pembolehubah tidak bersandar bagi kombinasi teknik MRA/PCA. Pembolehubah sepunya yang wujud dalam kedua-dua persamaan PCA1 dan PCA2 adalah H_2O vapor dan NO_2. Keputusan ini boleh dijangka kerana NO_2 adalah prapenanda bagi O_3. Pengesahan keputusan menunjukkan korelasi yang tinggi masing-masing bagi musim NEM and SWM, iaitu 0.877 hingga 0.888 dan 0.837 hingga 0.896. Nilai RMSE, IA dan FB bagi nilai O_3 yang diramal menggunakan teknik MRA adalah 15.5 DU, 0.57, dan -0.05. Manakala, nilai MRSE, IA dan FB bagi nilai O_3 yang diramal dengan menggunakan kombinasi teknik MRA/PCA adalah 10.93 DU, 0.68, dan -0.03. Nilai statistik yang diperolehi menunjukkan bahawa nilai O_3 bulanan yang diramal mencapai persetujuan yang tinggi dengan nilai O_3 yang diperoleh melalui satelit di Semenanjung Malaysia. Daripada keputusan yang diperolehi, kombinasi teknik MRA dan PCA mencapai prestasi yang lebih baik berbanding dengan teknik MRA untuk meramal nilai O_3 di Semenanjung Malaysia. Secara keseluruhannya, keputusan yang diperoleh menunjukkan bahawa kelebihan menggunakan satelit SCIAMACHY dan analisis korelasi untuk mengkaji kesan parameter-parameter atmosfera terhadap tulus ozon di Semenanjung Malaysia. Pengesahan dan perbandingan yang dilaksanakan dalam kajian ini mempamerkan keputusan penghasilan persamaan-persamaan regresi yang berketepatan tinggi dan berjaya mencapai objektif-objektif kajian.

REGRESSION ANALYSIS OF COLUMN OZONE AND SELECTED ATMOSPHERIC PARAMETERS IN PENINSULAR MALAYSIA FROM SCIAMACHY DATA

ABSTRACT

Ozone (O_3) is unique among pollutants because it is not emitted directly into the air, and its results from complex chemical reactions in the atmosphere. O_3 can bring different effects for all the living on earth (either harm or protect), depending on where O_3 resides. This is the main reason why O_3 is such a serious environmental problem that is difficult to control and predict. The aim of this study is to develop new algorithms for column O_3 in Peninsular Malaysia using statistical methods. In the end, four regression equations – denoted as O_3 NEM, O_3 SWM, PCA1 (O_3 NEM season) and PCA2 (O_3 SWM season) – were developed. In addition, the study focused on the analysis and investigation of how the selected atmospheric parameters affect column O_3 values in Peninsular Malaysia. Multiple regression analysis (MRA) and principal component analysis (PCA) methods have been utilized to achieve these study objectives. MRA was used to generate regression equations for O_3 NEM and O_3 SWM, while a combination of MRA and PCA methods were used to generate regression equations for PCA1 and PCA2. Statistical analysis has been carried out to test the performance of the MRA method and the combined MRA/PCA method, in terms of root mean square errors (RMSE), index of agreement (IA), and fractional bias (FB). The results of the best column O_3 regression equations, using MRA with four independent variables, were highly correlated (R = 0.811 for SWM, and R = 0.803 for NEM; $R^2 \approx 0.658$ for SWM, and $R^2 \approx 0.645$ for NEM for the six-year (2003-2008) data). The correlation coefficients (R) of validation for the NEM and

SWM seasons were 0.783 to 0.799 and 0.752 to 0.802, respectively. On the other hand, using the combined MRA/PCA method with four independent variables, the results of fitting the best O_3 data equations gave about the same values of R (≈ 0.83) for both the NEM and SWM seasons. The common variables that appeared in both regression equations were H_2O vapor and NO_2. This result was expected, as NO_2 is a precursor of O_3. The correlation coefficients (R) of validation for the NEM and SWM seasons were 0.877 to 0.888 and 0.837 to 0.896, respectively. Using MRA methods, the RMSE, IA, and FB for the predicted O_3 were found to be 15.5 DU, 0.57, and -0.05, respectively. On the other hand, using the combined MRA/PCA method, the RMSE, IA, and FB for the predicted O_3 were found to be 10.93 DU, 0.68, and -0.03, respectively. These statistical values indicated a very good agreement between the predicted and observed monthly O_3 for Peninsular Malaysia. The obtained results demonstrate that the combined MRA/PCA method performs slightly better than the MRA method in predicting the O_3 value in Peninsular Malaysia. Overall, these results clearly indicate the advantage of using satellite Scanning Imaging Absorption Spectrometer for Atmospheric Chartography (SCIAMACHY) and correlation analysis to investigate the impact of atmospheric parameters on total column O_3 over Peninsular Malaysia. The validation and comparison conducted in this study demonstrate the high accuracy of the regression equations and hence has successfully accomplished the objectives of this research.

CHAPTER 1

INTRODUCTION

1.1 Overview

Ozone (O_3) is a gas that is present in a very small part of our atmosphere when compared with other gases. However, O_3 is an important chemical constituent of the atmosphere and plays a key role in atmospheric energy budget and chemistry, air quality and global change (Dueñas et al., 2004; Ahammed et al., 2006; Lin et al., 2008). Furthermore, O_3 is also recognized as an important secondary air pollutant in the atmosphere and has received significant attention in the literature (Zunckel et al., 2006). In addition, O_3 also has a significant impact on the atmosphere's radiation budget and is a primary greenhouse gas (Wu and Chan, 2001). Hence, if high levels of surface O_3 and anthropogenic emissions cause abrupt changes in atmospheric O_3, a pollutant and greenhouse gas, these shifts may create environmental problems and contribute to climate change (Guicherit and Roemer, 2000; Vingarzan, 2004).

Industrialization, urbanization, rapid traffic growth and increasing levels of anthropogenic emissions have resulted in a substantial deterioration of air quality over Asia (Reddy et al., 2012). In contrast to the thinning of the O_3 layer in the stratosphere, the O_3 burden in the troposphere is generally increasing because of increasing emissions of precursors such as nitrogen oxides (NO_X) and volatile organic compounds (VOCs). NO_X and O_3 levels in urban air and amounts of nitrate in the aerosol particles over South Asia indicate that NO_X levels are not negligible (Lal et al., 2000).

The decrease of stratospheric O_3 leads to the increase of ultraviolet (UV) radiation at the earth's surface, resulting in the increased risk of several severe human diseases, such as skin cancer and eye cataracts (Kondratyev and Varotsos, 1996). In addition to its effects on mankind, UV radiation can disturb the normal genetic activity of plants and has negative impacts on the growth of plants. Furthermore, O_3 is a strong greenhouse gas in the earth's atmosphere, because it can absorbs both infrared (IR) and UV radiation. Variations of atmospheric O_3 inevitably affect global or regional climate change (IPCC, 2013). Therefore, investigating the spatial and temporal variability of the total column O_3 at the regional or global scale has caught the attention of scientists and policymakers (Bowman and Krueger, 1985).

Even though an increasing number of studies on surface O_3 in Malaysia were reported, limited studies have been conducted on total column O_3 over the vast landmass of Peninsular Malaysia. Malaysia has experienced economic growth due to urbanization, industrialization, and rapid traffic growth, which has resulted in a drastic increase in the emissions of air pollutants into the atmosphere (Streets and Waldhoff, 2000). Heavy pollution is being created by emissions that primarily result from increases in the volume of motor vehicles, industrial areas, and trans-boundary pollution. The high resolution that is associated with special satellite specifications is required to study how atmospheric parameters (including greenhouse gases) affect atmospheric O_3.

Several studies have been conducted that aim to develop tools that can achieve short-term forecasts of O_3 levels (Banja et al., 2012). These studies typically focus on examining O_3 levels to ensure it is not exceeding the threshold value. Medical and environmental authorities could use this valuable information to announce public health warnings.

Various statistical analysis of long-term total column O_3 records have been performed to examine the effect of external variables on total O_3 using ground-based measurements (Wohltmann et al., 2007) and/or satellite measurements (Brunner et al., 2006). Ground-based measurements have the advantage that they often span time periods longer than those available from satellite measurements (Brunner et al., 2006). On the other hand, satellite instruments perform measurements at a higher temporal frequency (daily) and provide global coverage (Knibbe et al., 2014).

A common method that is widely applied when developing prediction models correlates pollution data and meteorological parameters with concentrations of a particular pollutant. Furthermore, models that depend on the statistical analysis of air quality data have been intensively used to express the ozone-hydrocarbons-nitrogen oxides relationship (Peton et al., 2000). Over last decade, there have been numerous statistical methods for O_3 forecasting in the literature. A critical review of these methods (extreme value, regression, and space-time methods) can be found in Thompson et al. (2001) and Schlink et al. (2003).

1.2 Ozone Circulation and Chemistry

At ground level, O_3 is a harmful pollutant. In this state, O_3 can damage human health, vegetation, and other living systems. O_3 alters molecules and gradually destroys these entities (Shan et al., 2008). Inversely, stratospheric O_3 is considered good for humans and other living forms because it protects the biosphere from harmful UV radiation (Bracher et al., 2005). Approximately 90% of atmosphere O_3 is contained in the “ozone layer.” This layer protects us from harmful UV radiation that results from sunlight.

Stratospheric O_3 reacts with other gases especially chlorofluorocarbons (CFCs) leading to O_3 depletion (Solomon, 1999). The chemistry of O_3 depletion involves various complex catalytic reactions in which O_3 molecules are converted to atomic oxygen and molecular oxygen. This process is sometimes natural and sometimes anthropogenic, however both can occur simultaneously (Oluleye and Okogbue, 2013). Both stratospheric and tropospheric O_3 is largely produced naturally through photochemical and chemical reaction (Ogunjobi et al., 2007). The balance between production and loss determines the magnitude of O_3 surplus or deficit at any location. Natural O_3 production and depletion appeared to be less problematic since both are sufficiently slow and thus capable of removing any irregularity in either of the ways.

O_3 plays a key role in biogeochemical cycles, air quality, and global change. As a pollutant and greenhouse gas, O_3 should be brought under effective control, but the variability of its natural background level in the boundary layer makes difficult the definition of the concentration above which it becomes a pollutant (Ahammed et al., 2006). Tropospheric O_3 concentrations and growth rates exhibit large spatial and temporal variabilities (Naja and Lal, 2002). Therefore, there is an uncertainty in the estimation of its contribution to the radiative forcing and tropospheric chemistry.

Photodissociation of O_3 molecules by solar UV radiation ($\lambda < 320$ nm) produce an excited oxygen atom (O^1D radical) that in turn, through its reaction with water (H_2O) vapor, give rise to hydroxyl (OH) radicals, which are the main oxidizing agents. These radicals react with carbon monoxide (CO), methane (CH_4), non-methane hydrocarbons, nitrogen dioxide (NO_2), hydrochlorofluorocarbons (HCFCs), etc. and control their flux into the stratosphere (Hauglustaine et al., 1994). O_3 in the lower troposphere is produced mainly by photochemistry involving pollutants that are released from various industrial and other anthropogenic activities (Fishman and Crutzen, 1978). NO_X and hydrogen oxide radicals (OH and peroxy) act as catalysts in this process.

The first instrument for routine monitoring of total column O_3 was developed by Gordon M. B. Dobson in the 1920s. The instrument called a Dobson spectrophotometer, measures the intensity of sunlight at two wavelengths, where one Dobson Unit (DU) is defined as 0.01 mm thickness at standard temperature and pressure. O_3 layer thickness is expressed as DU, by measuring its physical thickness when compressed it in the earth's atmosphere (Jasim et al., 2010). It originated, and continues to be widely used, as a measure of total column O_3, which is dominated by O_3 in the stratospheric O_3 layer (Jasim et al., 2013).

1.3 Introduction to SCIAMACHY

Scanning Imaging Absorption Spectrometer for Atmospheric Chartography (SCIAMACHY) is a passive remote sensing, moderate-resolution imaging spectrometer onboard the European Space Agency's (ESA) Environmental Satellite (ENVISAT) satellite. It was launched in March 2002 as a multi-national contribution to ENVISAT by scientists from Belgium, Germany, and the Netherlands (Burrows et al., 1988).

As one of ten instruments onboard ENVISAT, SCIAMACHY is designed to measure solar radiation that is scattered, reflected and transmitted at a relatively moderate resolution (0.2 nm to 1.5 nm) from the earth's atmosphere through eight channels over a spectral range in the UV, visible (VIS) and near-infrared (NIR) region (214 nm–2386 nm) (Gottwald et al., 2006). Thus, it can be used to study land surfaces, oceans, and the atmosphere, especially to provide information about the dynamics, radiation balance, and composition of the atmosphere. In addition, SCIAMACHY consists of seven polarization monitoring devices (PMD), which are used to measure upwelling radiation at certain wavelengths.

SCIAMACHY is a grating spectrometer, whereas it is sensitive to polarization. A combination of spectral channels and PMD measurements allows instrument polarization characteristics to be enhanced. Thus, it can provide useful information with moderate spatial resolution that is utilized in the field of view (FOV), such as the interpretation of the aerosols and clouds.

Additionally, SCIAMACHY is capable of measuring both earthshine radiation and extraterrestrial irradiance, as it performs measurements in nadir, limb, and solar/lunar occultation viewing geometries. SCIAMACHY is extraordinary because the distribution profiles of trace gases (in the troposphere and stratosphere) and the total column can be retrieved through the combination of all three viewing geometries.

1.4 SCIAMACHY's Goal

SCIAMACHY is designed to operate synchronously, and its scientific objective is to understand global environmental problems related to atmospheric chemistry and physics and to acquire more information about global atmospheric distribution. Thus, SCIAMACHY can measure different trace gases, such as O_3, bromide monoxide (BrO), chlorine dioxide (ClO_2), sulfur dioxide (SO_2), formaldehyde (H_2CO), nitrogen dioxide (NO_2), carbon monoxide (CO), carbon dioxide (CO_2), methane (CH_4), H_2O vapor, nitrous oxide (N_2O), chlorine dioxide (OCIO) and aerosols, and their distributions, along with radiation, cloud cover, and cloud top height (Buchwitz et al., 2004).

SCIAMACHY retrieves the reflection, scattering, and absorption characteristics in the troposphere and stratosphere based on the comparison of the upwelling radiation and extraterrestrial solar irradiance that is observed in the different viewing geometries. These data will be used to support climate-related studies and provides information on some phenomena that affect the atmospheric chemistry.

1.5 SCIAMACHY Instrument Description

ENVISAT, i.e., SCIAMACHY's platform, is an advanced sun synchronous, polar-orbiting earth observation satellite that passes the equator at the same local solar time (10:00 AM) and can take measurements for the ocean, land, ice, and atmosphere. This environmental satellite has an orbit period of approximately 100 minutes, always passing close to the poles and moving against the earth's rotation due to an inclination angle of 98.55 °(Khlystova, 2010). The orbit has a 35-day repeat coverage that is associated with the position of the earth's surface.

SCIAMACHY consists of three basic parts: the radiant cooler, the electronic units, and the optical unit. The optical configuration of SCIAMACHY is shown in Figure 1.1. It comprises a spectrometer, a telescope, a mirror system, several detectors, gratings and electronics and thermal subsystems (Afe, 2005). The instrument consists of two scan mirrors: the nadir mirror and limb mirror, which allow incoming light to enter the instrument. In the nadir mode, the nadir mirror is utilized for the scan mode in the east-west direction (across track). While in the limb mode, the elevation scan (up-down direction) and azimuth scan (east-west direction) are carried out using both mirrors.

A telescope mirror with a 31-mm diameter is used to focus on the spectrometer's entrance slit (8 mm high and 180 microns wide) to produce the instrument's FOV at 0.045 °x 1.88 °. When passing through a pre-disperser prism, the collimated light is separated into 8 spectral channels in the spectral range. In each channel, a grating is placed to disperse the light and ensure that it is focusing on eight arrays of detectors. To a large extent, the optical components are designed to avoid stray light along the light paths.

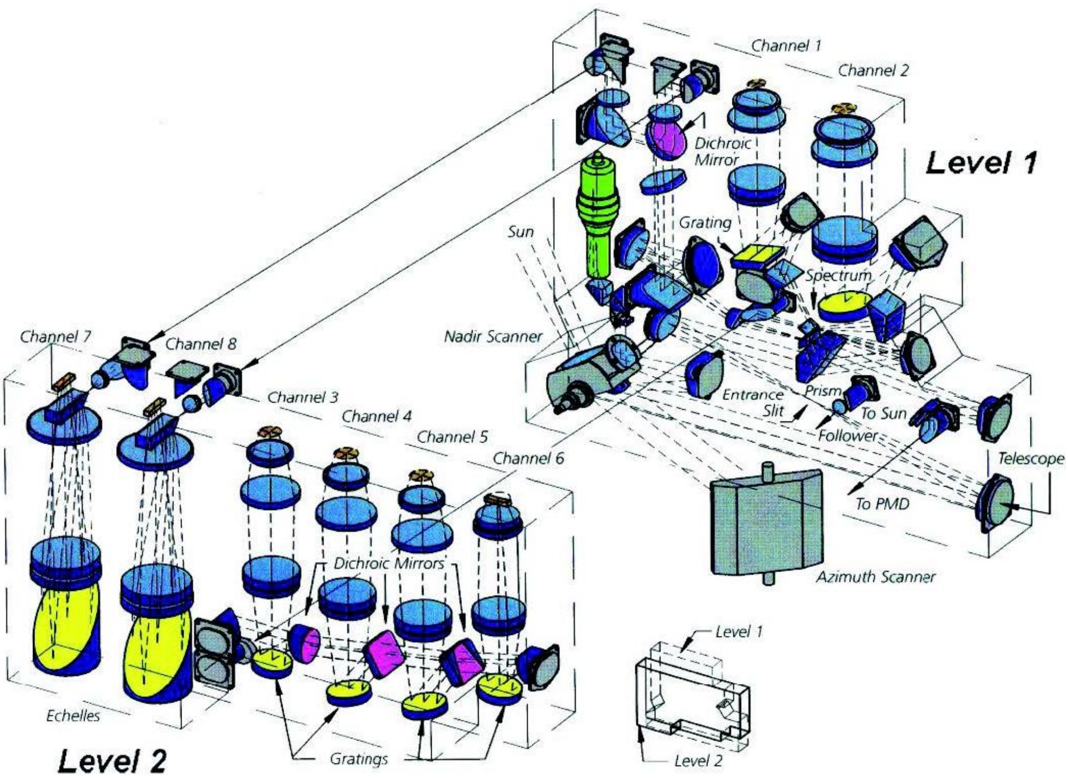

Figure 1.1. The optical configuration of SCIAMACHY, levels 1 and 2 (Source: Khlystova, 2010).

1.5.1 Spectral Channels Characteristics

SCIAMACHY consists of eight spectral channels, and each channel has a 1024-pixel monolithic diode array. The detector material for channels 1 through 5 is made of silicon (Si), while indium gallium arsenide (InGaAs) is used for channels 6 through 8, as this material is suitable for a larger spectral range. The range of wavelengths between 214 and 1063 nm was continuously measured with channels 1 through 6. Additionally, channels 7 and 8 provided measurements in the spectral ranges of 1934 to 2044 nm and 2259 to 2386 nm, respectively. Table 1.1 summarizes the characteristics of the eight spectral channels.

Table 1.1. Characteristics of SCIAMACHY's science channels (Source: Schneising, 2008).

Channel	Spectral range [nm]	Resolution [nm]	Detector material	Temperature range [K]
1	214-334	0.24	Si	204.5-210.5
2	300-412	0.26	Si	204.0-210.0
3	383-628	0.44	Si	221.8-227.8
4	595-812	0.48	Si	222.9-224.3
5	773-1063	0.54	Si	221.4-222.4
6	971-1773	1.48	InGaAs	197.0-203.8
7	1934-2044	0.22	InGaAs	145.9-155.9
8	2259-2386	0.26	InGaAs	143.5-150.0

1.5.2 Polarization Measurement Devices Characteristics

The polarization state of the incoming solar radiation is the main factor that affects the spectrometer's measurement sensitivity. Thus, SCIAMACHY is equipped with PMD, which enables the instrument to measure polarized light that is perpendicular to the SCIAMACHY optical plane by assuming that extraterrestrial solar light is unpolarized. This polarized beam is split into six different spectral bands (Table 1.2).

Polarization is measured in the direction that is parallel to the spectrometer's entrance slit and is associated with the overlap between spectral channels 2 to 6 and 8 with the first six general PMD. Meanwhile, the third PMD overlaps the seventh PMD 45° sensor in order to measure the polarization in a direction at 45° to the spectrometer slit. In addition, the PMD detectors need to be cooled in order to reduce noise.

Table 1.2. PMD characteristics (Source: Gottwald et al., 2006).

PMD Sensor	Spectral Range [nm]	Detector Temperature [℃]	Detector Material
PMD 0	310-365	-18	Si
PMD 1	455-515	-18	Si
PMD 2	610-690	-18	Si
PMD 3	800-900	-18	Si
PMD 4	1500-1365	-18	Si
PMD 5	2280-2400	-18	Si
45 ° Sensor	800-900	-18	Si

1.5.3 Field of View and Instantaneous Field of View Characteristics

SCIAMACHY's FOV is defined as the across-track path length of data acquisition and is determined by the swath width. The instantaneous field of view (IFOV) fixes an area on the earth's surface, which is observed from an operating altitude at one particular moment in time. Table 1.3 shows the swath width that depends on the FOV and IFOV for different viewing geometries.

Table 1.3. Field of view characteristics (Source: Yan, 2005).

Geometry	IFOV	The Swath Width
Nadir	0.045 ° x 1.8 ° (across track/along track)	+/- 480 km across track
Limb	0.045 ° x 1.8 ° (elevation/azimuth)	+/- 480 km in azimuth, 0-150 km in elevation
Solar Occultation	0.045 ° x 0.72 ° (elevation/azimuth)	0-150 km in elevation
Moon Occultation	0.045 ° x 1.8 ° (elevation/azimuth)	

1.5.4 Viewing Geometries

SCIAMACHY measurements are conducted in nadir, limb, and solar/lunar occultation geometries when the incoming radiation enters the instrument through one of three ports from an altitude of approximately 800 km (Figure 1.2). The maximum swath width of 30 km (along-track) by 960 km (across-track) can be achieved in the nadir mode if the nadir mirror scans an area on the ground and the atmospheric volume beneath the spacecraft is observed.

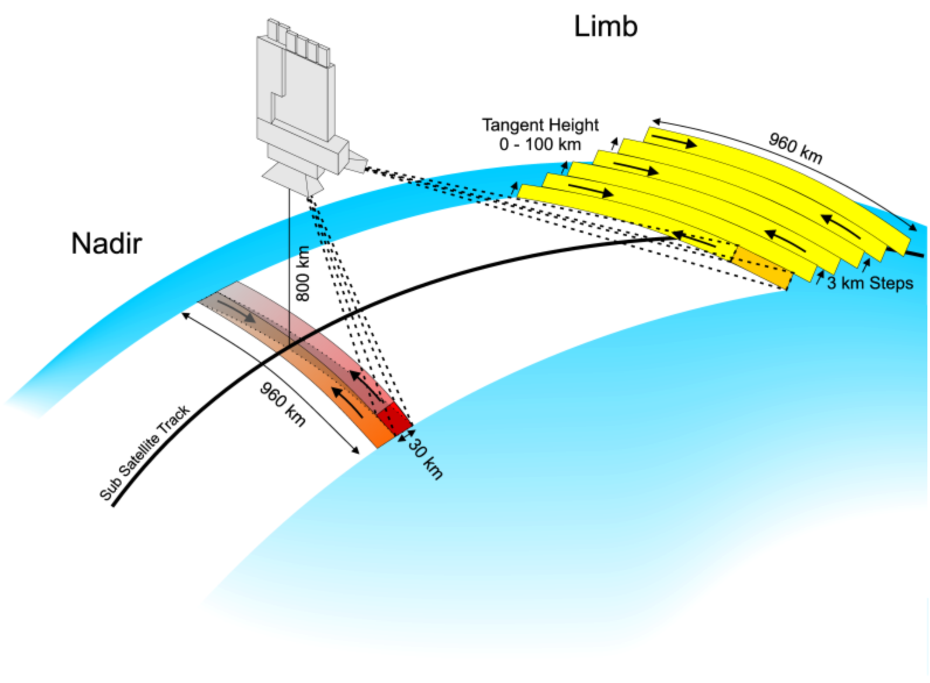

Figure 1.2. Measurement modes for SCIAMACHY (Source: Schneising, 2008).

1.5.5 Nadir Viewing

Two main criteria have to be fulfilled for SCIAMACHY's nadir viewing mode: the portion of the atmosphere directly below the satellite is observed, and the nadir/elevation mirror is utilized. In nadir viewing mode, global coverage is achieved in 6 days at the equator, as the maximum swath width of 960 km (across-track) (Richter et al., 2005).

The nadir mirror's scan angle affects the size of swath width. The swath width has a line-of-sight (LOS) variation, with the edge varying from 0° to a maximum of approximately 30.9°. Nadir observations are acquired for the typical footprints of 30 km (along-track) x 60 km (across-track). However, the integration time and scan speed can have an impact on the nadir observation and achieve 30 km x 30 km of the swath width.

1.5.6 Limb Viewing

In the limb mode, nadir and limb mirrors are utilized to obtain the horizon in-flight direction through the spectrometer slit's projection. In this viewing mode, azimuth and elevation mirrors are used to observe the atmospheric spectra, which are tangential to the earth's surface. Each horizontal scan of the atmosphere in limb mode covers a 960-km swath, starting at nearly 2 km below the earth's horizon and climbing to a maximum altitude of 100 km (Yan, 2005). The vertical resolution is approximately 3 km and is controlled by the geometrical FOV. Thus, the nominal integration time at the tangent height is 1.5 seconds and is associated with a spatial resolution of 960 km x 3 km.

1.5.7 Solar/Lunar Occultation Viewing

In this mode, occultation measurements are conducted in the same manner as in the limb mode, with the additional condition that SCIAMACHY is pointed toward the sun or moon in the full view of the instrument (Schneising, 2008). SCIAMACHY chooses a target, the moon or the sun, tracking it from 17.2 km above the horizon until the line of sight reaches a maximum tangent height of 100 km. Vertical distribution profiles of aerosols and trace gases can be determined by SCIAMACHY's occultation viewing.

1.5.8 Measurement Orbit

SCIAMACHY observes the solar occultation by measuring orbits from above the Northern Hemisphere in combination with preceding limb measurements. The measurements follow the solar occultation that is associated with the alternating limb/nadir with durations of 59 seconds for a limb and 80 seconds for a nadir. In the Southern Hemisphere, the monthly lunar visibility period can be observed. Only nadir observations are performed at the end of the illuminated part, while the atmosphere in flight directions is already in eclipse.

A typical sequence of SCIAMACHY measurements during one orbit is shown in Figure 1.3. The blue lines separate the measurement into two portions. In the ascending mode, the satellite travels toward the North Pole in the left portion of the orbit. Inversely, the right portion of the orbit is the descending pass, where the satellite travels southward. The instrument takes approximately 100 minutes to complete an orbit and makes 14 orbits per day.

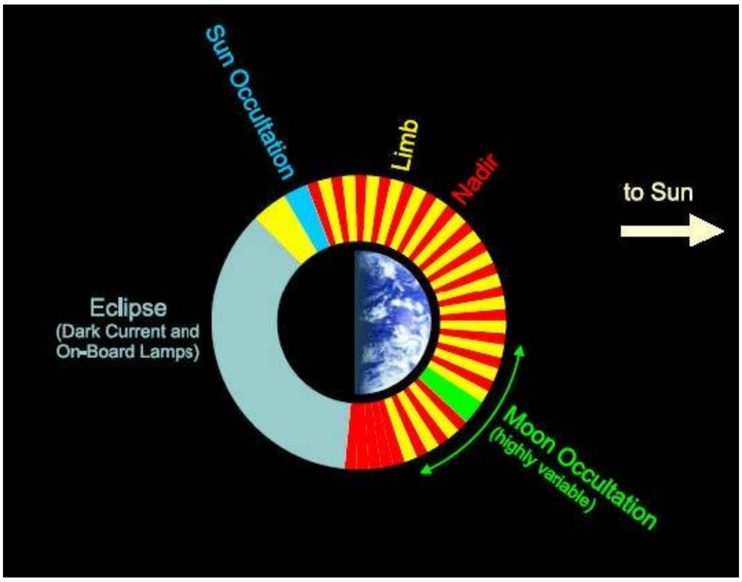

Figure 1.3. A typical measurement of SCIAMACHY's orbit (Source: Yan, 2005).

1.6 SCIAMACHY Science Processing and Products

SCIAMACHY data products consist of level 0, level 1b, level 1c, and level 2 (near real-time (NRT) and off-line products). Table 1.4 shows the main scientific products from SCIAMACHY. The level 0 product is reformatted data, which are delivered in binary analogue-to-digital converter (ADC) units and contain raw detector signals. Main input data from level 1c are from the raw data of level 0, which are the data streams for limb, nadir and occultation states, PMD data, monitoring and calibration states.

Level 1b products consist of raw data (level 0 products) combined with calibration constants to support SCIAMACHY monitoring and calibration activities (Gottwald et al., 2006). Level 2 products consist of level 2 off-line (OL) and NRT products. Level 2 NRT products are available three hours after the data acquisition. The OL processor processes limb, nadir and occultation for level 2 OL products. Meanwhile, only nadir measurements are used for level 2 NRT products.

Table 1.4. Level 2 NRT and OL SCIAMACHY operational products for the nadir and limb viewing modes that cover SCIAMACHY's entire spectral range (Source: Afe, 2005).

	Nadir			Limb		
	UV/Vis	IR	UV/Vis/IR	UV/Vis	IR	UV/Vis /IR
NRT	O_3, NO_2, SO_2, OCIO, H_2CO	H_2O, CO, N_2O, CH_4	Clouds, Aerosols			
Off-Line	O_3, NO_2, BrO, SO_2, OCIO, H_2CO, UV-index	H_2O, CO, CO_2, N_2O, CH_4, Pressure, Temperature	Clouds, Aerosols	O_3, NO_2, BrO	H_2O, CO, O_2, N_2O, CH_4, P, T	Aerosol s

1.7 Problem Statement

O_3 is unique among pollutants because it is not emitted directly into the air, and its results from complex chemical reactions in the atmosphere (Jasim et al., 2013). Besides, it has dramatically different effects depending on where O_3 resides. It can harm or protect life on earth (Struijs et al., 2010). This is the main reason why O_3 is such a serious environmental problem that is difficult to predict and control.

O_3 has been identified as one of the secondary pollutants that degrade air quality, and it is naturally present in our atmosphere (Sadanaga et al., 2012). O_3 has the ability to absorb IR radiation and it's considered as an essential and important greenhouse gas, which is presence both at the ground-level and in the earth's upper atmosphere (Freijer et al., 2002).

Due to the lack of observational studies of greenhouse gases in Malaysia, most studies have depended on ground station data. Malaysia, one of the tropical countries in Southeast Asia, has undergone rapid economic development and urban expansion, and transportation facilities have led to increased fossil fuel consumption over the past few decades in Malaysia, which has resulted in the increased emission of air pollutants, especially in industrial areas and cities. Therefore, there is a need to focus the study on the regression analysis between selected atmospheric parameters and column O_3.

Malaysia has very limited atmospheric data (especially total column O_3) from ground stations. Thus, satellite data has provided a good method to acquire and analyze the temporal and spatial atmospheric gases in the Malaysian region. In addition, because it is located in Southeast Asia, Malaysia's climatology is dominated by a strong northeast monsoon (NEM) and southwest monsoon (SWM) (Jasim et al., 2014). These monsoons have different influences on the atmospheric parameters in terms of how they affect climate or the number of pollutants that they bring to Malaysia, along with the contributions of many regional pollutant sources.

The most frequent methodology that is used for the meteorological adjustment of O_3 is multiple linear regression analysis. Multiple linear regression analysis is one of the most widely used methodologies for expressing a response variable's dependence on several independent variables in order to obtain a linear input-output model for a given data set (Al-Alawi et al., 2008). Therefore, satellite remote sensing is one of the most effective approaches for monitoring the distributions of atmospheric gases on a global scale with moderate spatial-temporal resolution (Baker et al., 2010).

The free downloaded data from SCIAMACHY onboard ENVISAT make this spectrometer a useful space instrument for observing the earth's atmospheric gas concentrations. In addition, the processed data of atmospheric pollutant gases (CO_2, CH_4, NO_2 and O_3) were available from January 2003 to December 2009, which were developed by the Institute of Environmental Physics (IUP) at the University of

Bremen in Germany. To generate a model for O_3 prediction, multiple regression analysis (MRA) and principal component analysis (PCA) were utilized in this study with the SCIAMACHY data.

Asia has experienced rapid economic and industrial development, which has increased its air pollutants. Hence, there is great interest in the tropospheric O_3 chemistry in Asia. In the case of Southeast Asia and other tropical countries, biomass burning, especially through forest fires, is also expected to contribute to the amount of tropospheric O_3 (Pochanart and Kreasuwun, 2001). Although an increasing number of O_3 studies in Asia have been reported, few studies have been conducted on the trend of atmospheric gases over Peninsular Malaysia. Thus, backward trajectory analysis is used to investigate the long-range transport of atmospheric O_3 and precursors during NEM and SWM.

1.8 Objectives

There are a few aims or objectives for this research:

(1) To develop a new algorithm of the column ozone (O_3) in Peninsular Malaysia by predicting the regression equations using the retrieved atmospheric parameters from SCIAMACHY satellite data for the period from 2003 to 2008.

(2) To compare the performance of the MRA method and the combined MRA/PCA method to predict the column ozone (O_3). In addition, using statistical methods to investigate and analyze the effects of the selected atmospheric parameters on column ozone (O_3) values.

(3) To analyze the trend of column ozone (O_3) over Peninsular Malaysia from 2003 to 2009. Furthermore, to use backward trajectory analysis during NEM

and SWM (with the HYSPLIT model) to study the long-range transportation of atmospheric O_3 and precursors.

(4) To validate the newly generated column ozone (O_3) algorithm with ground-truth, SCIAMACHY and AIRS satellite data.

1.9 Novelty of the Study

In this study, the satellite data are retrieved from SCIAMACHY and employed for the prediction of O_3 through regression equations. Statistical methods are utilized to analyze the atmospheric data and generate new regression equations for column ozone O_3. Overall, this study focused on atmospheric gas modeling and is the first study to use SCIAMACHY data to analyze the effects of atmospheric parameters on column O_3 and to develop the regression equations in Peninsular Malaysia. In addition, this study also tests the performance of column O_3 modeling for two different methods (the MRA and combined MRA/PCA methods). Backward trajectory analysis is also used to identify patterns for long-range transportation of atmospheric O_3 and precursors that arrived in Peninsular Malaysia.

1.10 Scope of the Study

This study primarily focuses on the development of new algorithms for column O_3 in Peninsular Malaysia. The algorithms were predicted through regression analysis using the retrieved atmospheric parameters (processed and available) from SCIAMACHY satellite data from 2003 to 2008. Thus, two different methods (MRA and MRA/PCA) were utilized to compare the performance of the MRA method and the combined MRA/PCA method. In addition, this investigation also been carried out to analyze the effects of selected atmospheric parameters on column O_3 values using statistical methods. Furthermore, five study sites across Peninsular Malaysia have been selected to analyze the column O_3 trend from 2003 to 2009. Besides, the HYSPLIT model was used to examine the long-range transportation of atmospheric O_3 through backward trajectory analysis.

1.11 Thesis Structure and Contents

This dissertation consists of six chapters, which are briefly described in the following paragraph.

In Chapter 1, a detailed research-related review of this study is provided. The content briefly describes the SCIAMACHY instrument and explains the operation and processing of satellite data. In addition, this chapter also discusses O_3 circulation and chemistry in the atmosphere. Overall, this chapter seeks to provide a scientific background for the O_3 chemistry and physics and the description for SCIAMACHY's instrument. This chapter also discusses the problem statement, originality, and the objectives of this study.

Chapter 2 seeks to discuss a literature review based on related work that uses the SCIAMACHY data. Therefore, it gives an overview that is associated with the present outlook and status for greenhouse gases and the application of statistical analysis in atmospheric remote sensing. In addition, this chapter also discusses the introduction and study of the earth's atmosphere. Chapter 3 elaborates further on this research's methodology and study area. The contents include a description of the research procedure, the standard SCIAMACHY dataset description and selection, and the data processing method. In addition, this chapter also includes the software and tools utilized in this study. The in-situ data obtained from Malaysia Meteorological Department (MMD) are used for validation purposes. Furthermore, the mechanism of the long-range transport of backward trajectory analysis, including a backward trajectory model and automatic circulation type scheme, is also discussed.

In Chapter 4, an overall analysis and discussion of all obtained results is thoroughly discussed. The contents include a discussion of the O_3 prediction algorithm using the MRA and the combined MRA/PCA methods. In this chapter, the results also describe the O_3 distribution in Peninsular Malaysia with mapping from 2009. Additionally, backward trajectory analysis undertakes a supporting investigation of the atmospheric O_3 through its long-range transportation over Peninsular Malaysia, which is associated with hotspot counts data for Sumatera, Borneo and Peninsular Malaysia.

Chapter 5 focuses on the comparison and validation of the O_3 regression model. The predicted algorithms were validated with the observed SCIAMACHY and Atmospheric Infrared Sounder (AIRS) data for 2009. In addition, statistical methods have been used to compare the performance of the MRA method and the combined MRA/PCA method in predicting O_3. Furthermore, the predicted algorithms are also compared with the in-situ and AIRS observed data. In Chapter 6, the dissertation ends with conclusions from the findings and a presentation of the limitations of this study. This chapter also proposes recommendations for possible future work.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

Scientists now regard anthropogenic (human-caused) global climate change to be the most important environmental issue of our time. The possibility that humans might alter world climate is not a new idea. In 1895, Svante Arrhenius, who subsequently received a Nobel Prize for his work in chemistry, predicted that the carbon dioxide (CO_2) that was released through coal burning would cause global warming. Since the early 1980s, climate change has been a topic of important discussion within the scientific community and among international and general political communities (Cheang, 1993).

The Industrial Revolution brought more extensive agriculture, new industrial processes, and a rapid increase in the world's population. The atmospheric concentration of greenhouse gases has increased rapidly due to the increasing of this human activity. In 1988, the United Nations Environment Programme (UNEP) and World Meteorological Organization (WMO) formed the Intergovernmental Panel on Climate Change (IPCC). This panel brings together scientists from many nations and a wide variety of fields to assess the current state of knowledge about climate change.

Analyzing and forecasting atmospheric parameters has recently become an important topic of atmospheric and environmental research (Amita, 2010). In fact, studies have been conducted that aim to develop short-term forecasts in atmospheric variables, especially in the ozone (O_3) level (Banja et al., 2012). Thus, in forecasting or prediction studies, researchers predict estimated or known observations from the past and utilize them as input for the predicted model. Seeing how well the output matches the known observations can test the model's performance.

2.2 The Atmosphere and Climate

The earth's atmosphere consists of 78 percent nitrogen and almost 21 percent oxygen, with the remaining 1 percent composed of a variety of trace gases, including CO_2 and argon (Ar) (Lutgens et al., 2006). In addition, an aerosol, i.e., a liquid droplet, is collectively suspended in the air. Atmospheric aerosols play an important role in rain production and the energy budget. Meanwhile, the atmospheric concentrations of water (H_2O) vapor vary from approximately 0 to 4 percent and are mainly controlled by available moisture and air temperature.

With respect to altitude, the atmosphere has four distinct layers with contrasting temperatures (Figure 2.1). The troposphere is the layer of air immediately adjacent to the earth's surface (Wallace and Peter, 2006). It extends from approximately 18 km over the equator to approximately 8 km over the poles, where the air is cold and dense. The tropopause is a distinct region between the troposphere and stratosphere, and the temperature decreases up to this region. The troposphere contains approximately 75% of the total mass of the atmosphere, as gravity holds most air molecules close to the earth's surface.

The earth's surface experiences the highest pressure and decreases with the height of the atmosphere associated with the barometric formula. The stratosphere is the second lowest layer of the earth's atmosphere and extends up approximately 50 km from the tropopause. The increasing temperature within this layer is due to the stratospheric O_3's absorption of ultraviolet (UV) solar radiation between 290 to 330 nm. This absorption thus makes the atmosphere warmer toward the top of the stratosphere. The mesosphere, or the middle layer, exists above the stratosphere. At approximately 50 km above the mesosphere, the thermosphere begins. It is a highly ionized region, where gases are heated by a steady flow of cosmic radiation and high-energy solar.

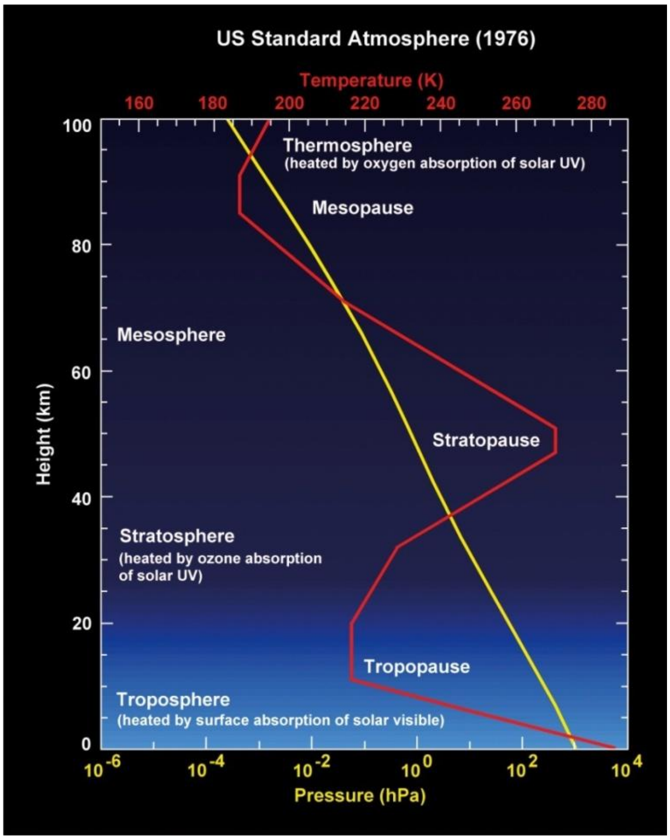

Figure 2.1. Atmospheric temperature and pressure profiles for the middle latitudes (US Standard Atmosphere) (Source: Gottwald et al., 2006).

Meanwhile, the amount and distribution of incoming solar radiation determines the weather and climate on earth. To achieve an equilibrium climate, the incoming absorbed solar radiation must balance the outgoing radiation (Trenberth and Fasullo, 2009). Of the solar energy that reaches the outer atmosphere, clouds and atmospheric gases reflect about one quarter, while CO_2, H_2O vapor, O_3, methane (CH_4), and few other gases absorb another quarter. About half of the incoming solar radiation (i.e., insolation) reaches the earth's surface (Trenberth et al., 2009). Most of this energy is in the form of light or infrared (IR) energy. The earth's surface and H_2O absorb most of this energy. Bright surfaces (e.g., ice, sand, and snow) reflect the rest (Figure 2.2).

Most solar energy comes in the form of intense, high-energy light or near-infrared (NIR) wavelengths. This short-wavelength energy passes relatively easily through the atmosphere to reach the earth's surface. The energy re-released from the earth's warmed surface is lower-intensity, longer-wavelength energy in the far-infrared (FIR) part of the spectrum. Atmospheric gases, especially CO_2 and H_2O vapor, absorb much of this long-wavelength energy and then re-release it into the lower atmosphere, letting it slowly leak out into space. This re-irradiated energy provides most of the lower atmosphere's heat. If the atmosphere were as transparent to IR radiation as it is to visible (VIS) light, the earth's average surface temperature would be approximately 20 ℃ colder than it is now (Trenberth and Fasullo, 2013).

This phenomenon is called the greenhouse effect because the atmosphere, loosely comparable to the glass of a greenhouse, transmits sunlight while trapping heat inside. The greenhouse effect is a natural atmospheric process that is necessary for life as we know it. However, too much greenhouse effect, caused by the burning of fossil fuels and deforestation, may cause harmful environmental change.