Università degli Studi di Ferrara

DOTTORATO DI RICERCA IN
MATEMATICA-INFORMATICA

CICLO XXIV

COORDINATORE Prof.ssa Ruggiero Valeria

Downscaling Aerosol Optical Thickness from Satellite Observations: Physics and Machine Learning Approaches

Settore Scientifico Disciplinare INF/01

Dottorando
Dott. Thi Nhat Thanh Nguyen

Tutor
Prof.ssa Eleonora Luppi
Dott. Simone Mantovani

Anni 2009-2011

Università degli Studi di Ferrara

DOTTORATO DI RICERCA IN
MATEMATICA-INFORMATICA

CICLO XXIV

COORDINATORE Prof.ssa Ruggiero Valeria

Downscaling Aerosol Optical Thickness from Satellite Observations: Physics and Machine Learning Approaches

Settore Scientifico Disciplinare INF/01

Dottorando
Dott. Thi Nhat Thanh Nguyen

.....

Tutor
Prof.ssa Eleonora Luppi

.....
Dott. Simone Mantovani

.....

Anni 2009-2011

I would like to dedicate this thesis to my beloved parents and sister.

Acknowledgements

First and foremost, I would like to show my gratitude to my supervisor, Prof. Eleonora Luppi, for her encouragements, instructions and supports to my study and research activities throughout three years of the Ph.D. course. I am sure that it would have not been possible without her constant help.

I would like to acknowledge and thank my advisor, Dr. Simone Mantovani, for his contributions of time, ideas, and research materials in my Ph.D. project. I greatly appreciate his daily guidance and supports, especially during tough times in the Ph.D. pursuit.

It is a pleasure to thank colleagues in my working group, Piero Campalani, Dr. Maurizio Bottoni, Mario Cavicchi, and Gian Piero Limone, for co-operation, work, and discussion. All results obtained in my dissertation have their contributions in some way.

I would like to thank founders of MEEO s.r.l., Dr. Stefano Natali, Dr. Marco Folegani and Dr. Simone Mantovani for finance aids to my Ph.D. and research activities.

I shoud like to thank all referees, Dr. Angela Benedetti at ECMWF, Dr. Silvia Canale, Assoc Prof. Alberto De Santis and Dr. Daniela Iacoviello at Sapienza - Università di Roma for review, correction, assessment and worthy suggestions for future work.

Finally, I owe sincere and earnest thankfulness to my dear friends, my beloved parents and sister for their unconditional love and endless spirit supports.

Abstract

In recent years, the satellite observation of aerosol properties has been greatly improved. As a result, the derivation of Aerosol Optical Thickness (AOT), one of the most popular atmospheric parameters used in air pollution monitoring, over ocean and continents from satellite observations shows comparable quality to ground-based measurements. Satellite AOT products is often applied for monitoring at global scale because of its coarse spatial resolution. However, monitoring at local scale such as over cities requires more detailed AOT information. The increase spatial resolution to suitable level has potential for applications of air pollution monitoring at global-to-local scale, detecting emission sources, deciding pollution management strategies, localizing aerosol estimation, etc. In this thesis, we investigated, proposed, implemented and validated algorithms to derive AOT maps with spatial resolution increased up to 1×1 km^2 from MODerate resolution Imaging Spectrometer (MODIS) observations provided by National Aeronautics and Space Administration (NASA), while MODIS standard aerosol products provide maps at 10×10 km^2 of spatial resolution. The solutions are considered on two perspectives: dynamical downscaling by improving the algorithm for remote sensing of tropospheric aerosol from MODIS and statistical downscaling using Support Vector Regression.

Abstract

Recentemente, gli algoritmi per l'osservazione dell'aerosol e delle sue caratteristiche da piattaforma satellitare sono stati migliorati notevolmente.

Le stime dello spesso ottico dell'aerosol - Aerosol Optical Thickness (AOT) - uno dei parametri più utilizzati per il monitoraggio dell' inquinamento atmosferico, hanno oggi un'accuratezza ed una precisione comparabile alle misurazioni effettuate con strumentazione a terra. I prodotti satellitari di AOT hanno una risoluzione spaziale adatta al monitoraggio a scala globale e non sono utilizzabili su scala locale per applicazioni su scala urbana che richiedono invece una conoscenza dettagliata della distribuzione.

La possibilità di aumentare la risoluzione consente di rendere il prodotto adeguato ad applicazioni di monitoraggio della qualit dell'aria su diverse scale (da globale fino a locale), rilevare sorgenti di emissione, definire strategie di gestione delle politiche ambientali, ecc..

In questa tesi sono stati esaminati, realizzati e validati algoritmi per la stima di mappe di AOT con risoluzione spaziale fino ad 1km utilizzando dati del sensore MODerate resolution Imaging Spectrometer (MODIS) installato a bordo dei satelliti della agenzia spaziale statunitense National Aeronautics and Space Administration (NASA). La qualit dei prodotti satellitari a maggior risoluzione è stata valutata in confronto con i prodotti MODIS Level 2 a 10km di risoluzione. Le soluzioni sono state considerate con due prospettive: downscaling dell'algoritmo nativo MODIS del prodotto AOT e downscaling mediante tecniche statistiche di Support Vector Regression.

Contents

List of Figures

Nomenclature

AATSR	Advanced Along-Track Scanning Radiometer
ACE	Aerosol Characterization Experiments
ADEOS	Advanced Earth Observation Satellite
AERONET	AErosol RObotic NETwork
AI	Aerosol Index
AOD	Aerosol Optical Depth
AOT	Aerosol Optical Thickness
AVHRR	Advanced Very High Resolution Radiometer
BAER	Bremen AErosol Retrieval
BRF	Bidirectional Reflectance Factor
CALIOP	Cloud and Aerosol LIdar with Orthogonal Polarization
CALIPSO	Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observations
CNES	le Centre National d'Etudes Spatiales
CORR	Correlation coefficient
DDV	Dense Dark Vegetation
DLR	German Aerospace Centre

ERTS	Earth Resources Technology Satellite
ESA	European Space Agency
GACP	Global Aerosol Climatology Product
GCMs	General Circulation Models
GLAS	Geoscience Laser Altimeter System
GOES	Geostationary Operational Environment Satellite
IGAC	Global Atmospheric Chemistry
ISCCP	International Satellite Cloud Climatology Project
JAXA	Japanese Aerospace Exploration Agency
KNMI	Royal Netherlands Meteorological Institute
LAMs	Limited Area Models
LC	Land Cover
LITE	Lidar In-space Technology Experiment
LUT	Look-Up Table
MERIS	MEdium Resolution Imaging Spectrometer
MISR	Multiangle Imaging SpectroRadiomete
ML	Machine Learning
MODIS	MODerate resolution Imaging Spectrometer
MRT	Minimum Reflectance Technique
MSS	Multi Spectral Scanner
NASA	National Aeronautics and Space Administration
NDBSI	Normalized Different Bare Soil Index

NDSI	Normalized Difference Snow Index
NDVI	Normalized Differential Vegetation Index
NIR	Near Infrared
NN	Neural Network
NOAA	National Ocean and Atmosphere Administration
PM	Particulate Matter concentration
POLDER	POLarization and Directionality of the Earth's Reflectances
RBF	Radial Basic Function
RCMs	Regional Climate Models
RMSE	Root Mean Square Error
SCIAMACHY	SCanning Imaging Absorption SpectroMeter for Atmospheric ChartographY
SEAWiFs	Sea-viewing Wide Field-of-view Sensors
SNR	Signal to Noise Ratio
SSA	Single Scattering Albedo
SVM	Support Vector Machine
SVR	Support Vector Regression
TARFOX	Tropospheric Aerosol Radiation Forcing Observation eXperiment
TIROS	Television Infrared Observation Satellite
TOMS	Total Ozone Mapping Spectrometer
TOMS-EP	Total Ozone Mapping Spectrometer-Earth Probe
TOP	Top Of Atmosphere
VISRR	Visible Infrared Imager Radiometer Suite

Chapter 1

Introduction

Remote Sensing allows to measure physical properties of objects without actually being in contact with them. Using devices installed on board aircrafts or satellites, Remote Sensing applied to the Earth Observation makes it possible to monitor the Earth-Atmosphere system through the analysis of the interaction of radiation with matter. The signal received is the sum of several contributions due to scattering, absorption, reflection and emission processes. Image processing techniques and specific algorithms are applied on that information to extract (direct measurement) or estimate (indirect measurement) the environmental parameters and their characteristics which are used in a large variety of applications for Earth Observation (Agriculture, Atmosphere, Forestry, Geology, Land Cover and Land Use, Mapping, Oceans and Coastal).

For Atmosphere applications focusing on the Climate Change and on the human health, the Aerosol Optical Thickness (AOT) has been recognized as one of the most important atmospheric variables to be monitored from global to local scale. AOT is representative for the amount of particulates present in a vertical column of the Earth's atmosphere. Aerosol concentration can be measured directly by ground-based filters/sensors or predicted by processing data recorded by airborne instruments or satellite-based sensors. Ground measurements have high accuracy and temporal frequency (hourly) but they are representative of a limited spatial range around ground sites. Conversely, satellite estimation provides information at global scale but lower measurement frequency (daily). Satellite aerosol measurements have fifty years history since the first visual observations

of atmospheric aerosol effects were made manually from the spacecraft Vostok by Yuri Gagarin on April 12, 1961. During this period, thousands of satellites have been launched by many governments and agencies. Nowaday, there are a large number of satellites measuring aerosols of the Earth from space.

MODerate resolution Imaging Spectrometer (MODIS) is a multi-spectral sensor on-board the two polar orbiting satellites Terra and Aqua, launched in 1999 and 2002, respectively and operated by the National Aeronautic and Space Administration (NASA). These satellites provide observations nearly the entire globe on a daily basis, and repeat orbits every 16 days. The swath about 2330 km long is divided into granules, each of which is corresponding to 5 minutes recording and hence has a length of 2030 km. The MODIS performs measurements in the solar to thermal infrared spectrum region from 0.41 to 14.235 μm, separated into 36 bands at resolutions 1 km (29 channels), 500 m (5 channels) and 250 m (2 channels) at nadir (Salomonson et al. [1989]).

The MODIS aerosol products were derived from MODIS-measured spectral radiance using physical algorithms (i.e. algorithm for remote sensing of tropospheric aerosol from MODIS) since the 90s. The first aerosol products in Collection 003 was validated globally over ocean (Remer et al. [2005]) and over land (Chu et al. [2002]) as well as used in regional validation exercises (Levy et al. [2005]; Ichoku et al. [2002]). After that, the algorithms were upgraded to create the products in Collection 004 (C004) (Kaufman and Tanré [1997]), Collection 005, Collection 051 (Remer et al. [2004]) and the newest Collection 006 released in 2004, 2006, 2008 and 2012, respectively. MODIS aerosol products are recognized as confident products hence used largely in many applications and validations as well as referred to in hundreds publications. The range of application is not limited in fields of radiation and climate (Yu et al. [2006]) but includes monitoring surface air quality for health (Chu et al. [2003]; Al-Saadi et al. [2005]), estimating iron nutrients (from dust) deposited into the ocean (Gao et al. [2000]), and assimilation for prediction of aerosol fields (Benedetti et al. [2009]).

The original methodology applied to derive AOT from MODIS follows model driven approach which concentrates on modeling of the interaction between aerosol and radiation to estimates aerosol properties. The standard MODIS aerosol product was produced using the retrieval algorithm known as Algorithm for remote

sensing of tropospheric aerosol from MODIS (Kaufman and Tanré [1997]; Remer et al. [2004]). The algorithms are designed and separated for land and ocean. Different aerosol models were simulated and their calculated parameters were stored in Look-Up Table (LUT). The algorithms assume that the aerosol properties over targeted areas were presented by proper weightings of one fine-dominated and one coarse-dominated aerosol models. Spectral reflectance from the LUT is compared with MODIS measured spectral reflectance to find the best match that is the solution to the inversion process. Beside LUT that is considered as the core technique, ancillary data and many different screening processes played an important role in this methodology. Those data and processing techniques were considered and applied on the original MODIS datasets to select appropriate data for inversion process.

Recently, machine learning approach applied in optical thickness processing has been investigated and presented in various applications ranged from classification of aerosol components (Ramakrishnan et al. [2005]), prediction based on time series data (Chen and Shao [2008]; Lu et al. [2002]; Siwek et al. [2008], Osowski and Garanty [2006]), to estimation of aerosol content and properties from different sensors (Okada et al. [2001]; Han et al. [2006]). Regarding AOT retrievals, the published results show that this approach provides a new efficient methodology in reducing processing time (Okada et al. [2001]), dealing with data uncertainties (Vucetic et al. [2008], Obradovic et al. [2010]) and improving the accuracy over specific areas (Xu et al. [2005]; Vucetic et al. [2008]). Moreover, the strong advantage of this approach is the flexibility in updating new inversion models and the simplicity to extend to data collected by new generation sensors, which would be costly if the model driven approach was applied. Following this approach, data are collected from different sources which can be satellites, ground measurements or simulated data. Because of different temporal and spatial resolutions, collected data are integrated in order to solve the differences and create samples which are used in the training phrase for creating empirical data models. Neural Networks (NNs) (Okada et al. [2001]; Xu et al. [2005]; Vucetic et al. [2008]; Obradovic et al. [2010]) and Support Vector Regression (SVR) (Lary et al. [2009]) are techniques used largely for inversion process. Related researches have shown that these approaches are competitive to physical algorithms but their limitations

are data dependence and complex modeling process requested.

Spatial resolution, representative of an area corresponding to a pixel, is an important factor of aerosol products. The best available resolution provided by MODIS standard aerosol products, up to now, is 10×10 km^2. This chosen resolution was due to a small Signal to Noise Ratio (SNR) between aerosol and noises from measurement obtained on small areas such as 0.5×0.5 km^2. However, this ratio would be bigger ten times at least if the 10×10 km^2 resolution was selected (Kaufman and Tanré [1997]). Another reason is based on validation results in which aerosol over an area of 50×50 km^2 does not vary much except over regions near major emission sources (Li et al. [2005]; Ichoku et al. [2002]). Therefore, 10×10 km^2 is considered as a suitable spatial resolution for global monitoring.

In a big context, downscaling of spatial resolution has been recognized as a popular and important methodology applied for General Circulation Models (GCMs) which are used to assess climate changes. These numerical models present various earth systems including the atmosphere, oceans, land surface and sea-ice with coarse provided spatial resolutions. Downscaling techniques are developed in order to bridge the gap between the resolution of climate models (some hundreds kilometer square in general) and regional and local scale process. Two fundamental approaches applied for the downscaling of large scale GCMs output to a finer spatial resolution are dynamical and statistical downscaling. Dynamical approach refers to the use of Regional Climate Models (RCMs) or Limited Area Models (LAMs). These focus on parameterizing atmospheric process of the large scale GCMs to produce higher resolution outputs. Statistical downscaling investigated the relationship expressed as a stochastic and/or deterministic function between large-scale atmospheric variables and local/regional climate variables. The statistical downscaling methods are generally classified into three groups: regression models, weather typing schemes and weather generators (Fowler et al. [2007]). The downscaling applications can be found in orographic/humidity/rainfall precipitation, extreme climate events and regional scale climate.

Regarding MODIS aerosol products, downscaling spatial resolution aims at providing more detailed aerosol information on observation areas, which is mean-

ingful for regional monitoring such as over countries or urban areas in particular where assumption of aerosol stability over 50×50 km^2 is not always correct. Moreover, the finer spatial resolution has great advantages in investigating aerosol distribution, detecting emission sources and deciding pollution management strategies. The challenges of this task, as mentioned above for applications using MODIS data, is high noises come from measurement instruments. In addition, large uncertainties of land surfaces over urban and cities also impact on aerosol estimation quality. Finally, appropriate algorithms as well as adaptations of complex global algorithms to derive aerosol at a finer resolution have been investigated in numerous contemporary works.

In literature, several researchers have focused on improving MODIS aerosol algorithms or proposing new methodologies for aerosol derivation. Published articles have shown achievements in which aerosol could be estimated at a finer resolution than 10×10 km^2. Grouped by regions, downscaling works were investigated and considered for New York city (Oo et al. [2008]), Sao Paolo and Beijing (Castanho et al. [2007]), Hong Kong (Li et al. [2005]; Nichol and Wong [2009]), Hong Kong and Pearl River Delta (Wong et al. [2009]). The spatial resolutions were considered at 3×3 km^2 (Oo et al. [2008]), 1.5×1.5 km^2 (Castanho et al. [2007]; Castanho et al. [2008]), 1×1 km^2 (Li et al. [2005]), and 500×500 m^2 (Nichol and Wong [2009]; Wong et al. [2009]). Following this effort but insisting on downscaling aerosol products at global scale, the MODIS team has planed to release Collection 6 in 2012 with spatial resolution increased to 3×3 km^2 (Remer et al. [2010]). However, this version was announced to have lower prediction quality than the current aerosol product at 10×10 km^2. All of the reviewed efforts follow the dynamical downscaling approach, in which surface reflectance is parameterized and regional LUTs are developed for inversion processes.

In this thesis, we investigated, proposed, implemented and validated methodologies to derive AOT maps with spatial resolution increased up to 1×1 km^2 from MODIS observations. The solutions are considered from two perspectives: dynamical downscaling by improving the algorithm for remote sensing of tropospheric aerosol from MODIS and statistical downscaling using Support Vector Regression (SVR).

Initially, the global MODIS aerosol algorithms with some adaptations are ap-

plied in smaller observation areas to derive AOT with spatial resolutionat 3×3 km^2 (Nguyen et al. [2010a]; Nguyen et al. [2010b]) and then 1×1 km^2. The algorithms perform at the global scale instead of parameterizing for specific regions. The validation carried out on data covered Europe in three years showed good correlation coefficient and acceptable errors between retrieved AOT at 1×1 km^2 with AERONET measurements (Campalani et al. [2011]). The proposed methodology was applied to develop a software package called PM MAPPER (MEEO [2010a]).

Motivated from advances of machine learning approach in AOT retrieval, the second approach investigated the usage of Support Vector Regression in deriving AOT at 1×1 km^2 of spatial resolution. This work is expected to provide a general framework which will be easy to extend to other satellite sensors. In order to evaluate the performance of SVR for AOT retrieval in comparison with the MODIS algorithms, two SVR variants were firstly applied for AOT at 10×10 km^2 (Nguyen et al. [2010c]). The evaluations for SVRs were carried out by year, by season and by land cover properties on data collected in 2006, 2007 and 2008. The experiment results show that the SVR approach outperforms the MODIS algorithm. Among two SVR variants, "instance SVR" gives better results than "aggregation SVR". SVR approach is able to improve AOT prediction quality especially over bright areas or locations with limited vegetation.

Based on validation results of SVR approach at 10×10 km^2, the "instance SVR" is selected for downscaling spatial resolutions up to 1×1 km^2. However, to apply the SVR for AOT downscaling from 10×10 to 1×1 km^2 is not a trivial task because of very large and noisy datasets to obtain 100 times more detailed maps. Moreover, the application is extended from pixel domain to map domain in which data models created by data collected on sparse locations are applied on large and continuous map areas. To deal with the above mentioned problems, the proposed methodology is based on SVR and domain knowledge (Nguyen et al. [2011]; Nguyen et al. [2012]). In this approach, the satellite-based data and ground-based measurements over areas of interest are collected and integrated using temporal and spatial constraints. After that, filtering using new integration conditions is applied to reduce the total amount of data and then clustering technique separates them into four groups having different characteristics. Then,

training datasets are selected by some strategies for each cluster and the SVR technique is applied on them to create corresponding data models. Finally, in the prediction framework, aerosol maps at spatial resolution of 1×1 km^2 are derived from MODIS L1B data using the SVR models retrieved in the previous step. Experiments were carried out on data from 2007 to 2009, covering European areas, in both pixel and map domain. Pixel domain refers to data collected around ground stations while map domain aims at all validated data in satellite maps. The validation in pixel domain shows that the SVR methodology outperforms the MODIS algorithm. Extended to map application, the proposed SVR method is robust in prediction and gives stable results in most of covering areas.

The thesis contribution is an software package to derive AOT at 1×1 km^2 from MODIS observations using the improved MODIS aerosol algorithms (Nguyen et al. [2010a]; Nguyen et al. [2010b]; Campalani et al. [2011]). The AOT maps obtained by this software have been used in SENSORE and AQUA projects in which AOTs were utilized to predict Particulate Matter concentration ($PM_{2.5/10}$) over Emilia Romagna regions in Italy and over Austria (MEEO [2010b]; MEEO [2012]). Besides, the methodology using SVR based on domain knowledge is proposed for downscaling of MODIS aerosols. The validations show promising results for the proposed approach (Nguyen et al. [2011]; Nguyen et al. [2012]).

The thesis will be organized into seven chapters. After the introduction section, fundamental overview and review of related works for AOT retrieval algorithms and downscaling techniques will be given in Chapter 2. Chapter 3 will introduce the standard methodologies for MODIS AOT retrievals. Developed on the state of the art techniques, the thesis's works will be presented in Chapter 4, 5 and 6. Finally, the conclusion section will summarize the present work and discuss future work. Details are as follows:

- *Introduction*
- *Chapter 2 (Aerosol Optical Thickness Retrieval from Satellite Observation, Physics and Machine Learning Perspectives)* presents fundamental knowledge including an overview of atmospheric aerosol, basics about satellite remote measurements and critical procedures for AOT retrievals from satellite observations. After that, the review of satellite aerosol retrieval algorithms

in physics and machine learning approaches is presented. Finally, downscaling techniques for MODIS AOT retrieval in literature are reviewed and summarized.

- *Chapter 3 (Methodologies for Aerosol Optical Thickness Retrieval from MODIS observations)* shows the state of the art of methodologies for AOT retrieval from MODIS observations. The chapter introduces theory of algorithms for remote sensing of tropospheric aerosol from MODIS in Collection 5 (model driven approach) and basics concepts of Support Vector Regression and its application for AOT retrievals from MODIS measurements (data driven approach).

- *Chapter 4 (Downscaling Spatial Resolution of Aerosol Optical Thickness from 10×10 km^2 to 1×1 km^2 using adapted MODIS aerosol algorithms)* presents the improvements of the MODIS aerosol algorithm to derive AOT with spatial resolution at 1×1 km^2 which then are used to develop a software package called PM MAPPER. This algorithm performs at global scale instead of parameterizing for specific regions. The validation for PM MAPPER AOT on data covering European areas over three years is presented.

- *Chapter 5 (Aerosol Optical Thickness Retrieval at 10×10 km^2 of Spatial Resolution using Support Vector Regression)* investigates potential of SVR in AOT retrieval at 10×10 km^2. The proposed methodology is applied in data covering European areas over three years to derive AOT at 10×10 km^2. The validations are carried out by year, by season, and by land cover class.

- *Chapter 6 (Downscaling Spatial Resolution of Aerosol Optical Thickness to 1×1 km^2 using Support Vector Regression based on Domain Knowledge)* presents the methodology replied on SVR and domain knowledge to downscale AOT up to 1×1 km^2. The proposed approach is developed and tested on real data collected over European areas over three years from 2007 to 2009. Validations show good results in both pixel and map domains.

- *Conclusions and Future Work*

Chapter 2

Aerosol Optical Thickness Retrieval from Satellite Observations, Physics and Machine Learning Perspectives

1 Fundamental Concepts

1 .1 Atmospheric Aerosol

The atmospheric aerosol is an aggregate of liquid or solid particles suspended in a gaseous medium long enough to be observed and measured. Aerosol can be produced by anthropogenic (e.g. fossil fuels burning, land use, land cover and management, fisheries, air pollution) or natural (e.g. dust storms, volcanoes, forest fires) sources. The necessity of aerosol measurements has increased dramatically in recent years in various disciplines. Many aerosol studies were originated from applications in health and environment areas in which aerosol was measured in order to ensure that the public and industrial work forces were not affected by hazardous aerosols at undesirable concentration levels. Applications for aerosol measurements can be found in chemistry, physics, biology, optics and engineering disciplines.

Atmospheric aerosols are characterized by size, shape, density, and chemical composition, all of which form a wide range of aerosol types existing in many different environmental and industrial settings. Atmospheric aerosol types with different characteristics distribute themselves in three vertical layers as shown in Figure 2.1. At each layer, aerosol has different characteristics in content, lifetime, and mixing type. The stratospheric layer contains volcanic sulfuric acid aerosols that travel at altitude of 10-15 km and make a circle around the Earth one or two year after a volcanic eruption. The tropospheric layer holds dust, smoke or industrial haze with 1-2 weeks lifetime. The boundary layer contains a large variety of aerosols from anthropogenic and natural sources. In this layer, most aerosols have a short lifetime (3-5 days) and internal mixing type (i.e. particles interact each others and grow their sizes). To fully understand aerosol effects, their characteristics must be known on local to global scales (Kaufman et al. [2002]).

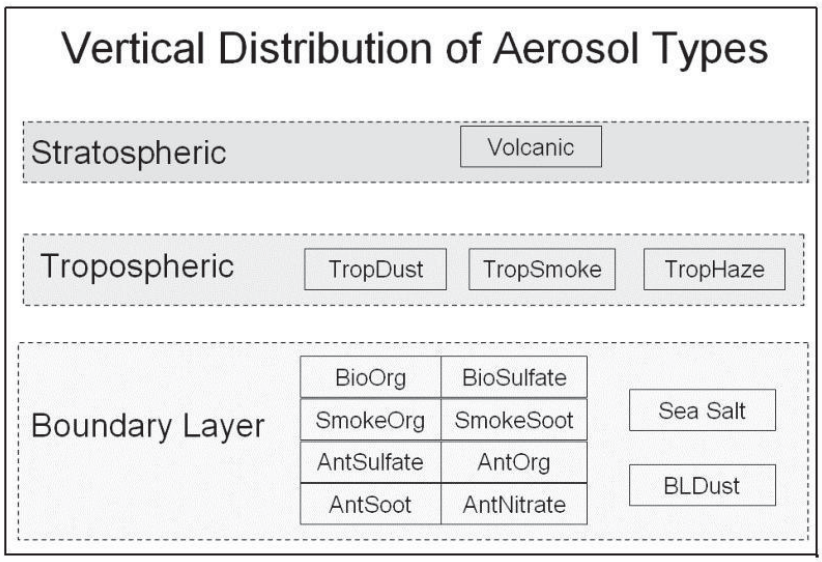


Figure 2.1: Vertical distribution of aerosol types in the atmosphere.

A variety of techniques have been used to obtain useful information about aerosol. In general, aerosol particles are able to be measured directly by sensors installed at ground based stations/aircrafts or interpreted from data recorded by airborne/satellite based sensors. Ground measurements have usually high accu-

racy and temporal frequency (from minutely to hourly) but they are representative of a limited spatial range around ground sites. Aircraft measurements are limited to a few aerosol intensive measurements campaigns (e.g. the International Global Atmospheric Chemistry (IGAC) in 1996, the Tropospheric Aerosol Radiation Forcing Observation eXperiment (TARFOX) in 1999, and three Aerosol Characterization Experiments (ACE 1-3) in 1998, 2000 and 2003, respectively). Satellite instrument, up to now, is the unique tool providing routine measurement (daily) of the Earth at the global scale.

Satellite aerosol measurements have fifty years history since the first visual observations of atmospheric aerosol effects were made manually from the spacecraft Vostok by Yuri Gagarin on April 12, 1961. The hand-held spectrophotometers were continuously used in the following missions to measure the vertical distribution of aerosol, spectrometry of the twilight and daylight, and stratospheric aerosol by cosmonauts in Vostok-2, Vostok-6, Voskhod, Soyuz-9, Apollo-Soyuz in 1961, 1963, 1965, 1970 and 1975 respectively. The first automatic image of aerosol was achieved by the Multi Spectral Scanner (MSS) on-board the Earth Resources Technology Satellite (ERTS-1) in 1975 (Griggs [1975]). However, it was with the launches of TIROS-N and Nimbus-7 satellites in 19 and 25 October 1978 that the era of satellite-based remote sensing of aerosol really began. The first aerosol products were generated from Advanced Very High Resolution Radiometer(AVHRR) on-board TIROS-N, while the longest measurements of global aerosol from space, up to now, were provided by TOMS carried by Nimbus-7. In the first 40 years of development, launched missions mainly focused on measurements of atmospheric composition, meteorology and atmospheric structure and dynamics. Since 1995, many instruments specified for air quality measurements have been launched in orbit and hence lead to opportunities of measuring columnar trace gases (e.g. ozone, nitric oxide, nitrogen dioxide, formaldehyde, glyoxal, chlorine dioxide, chlorine monoxide, and nitrate radicals) in stratosphere and troposphere.

The demand for satellite aerosol measurement is increasing in the Earth science, air quality management, and disaster management because satellite-based observations are the only available technique to measure the distribution of aerosol over the entire globe. Many biogeochemical and climate applications of aerosols

are based on the newly-derived global aerosol data. Satellite aerosols have improved understanding of global biogeochemical processes, managed aerosol source regions, global transport patterns and atmospheric lifetimes of substances. In air quality research, satellite aerosol has been used for the detection and quantification of natural and man-made air pollution events. Figure 2.2 presents the image obtain by NASA's EO-1 satellite for the Eyjafjallajökull eruption in Iceland in 2010, which led to air travel disruption in northwest Europe for six days from 15 to 21 of April and also in May 2010, including the closure of airspace over many parts of Europe.

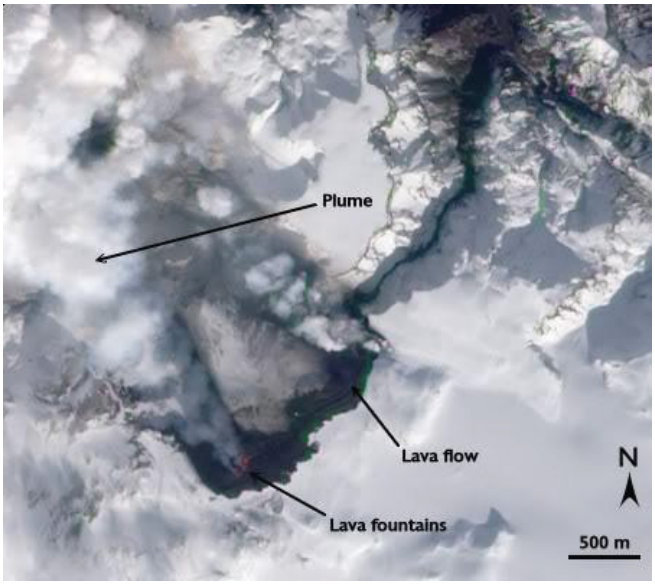


Figure 2.2: Eyjafjallajökull eruption in Iceland (Image obtained by NASAs EO-1 satellite on 24 March 2010).

1 .2 Basics about remote measurements from satellites

1 .2.1 Satellite instruments

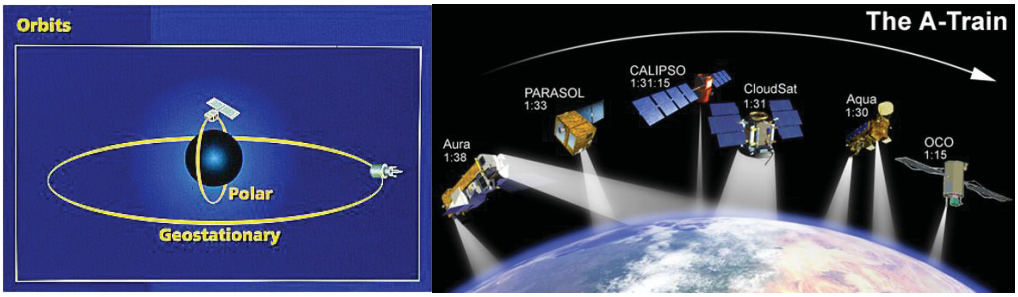
Nowadays, thousands of instruments on-board satellites have been observing the Earth. They were launched and operated by governments or space agencies such as the National Aeronautics and Space Administration (NASA), the National Ocean and Atmosphere Administration (NOAA), the European Space Agency

(ESA), le Centre National d'Etudes Spatiales (CNES) in France, the Japanese Aerospace Exploration Agency (JAXA), the China Meteorological Administration, the Royal Netherlands Meteorological Institute (KNMI), and the German Aerospace Centre (DLR). Since aerosol products obtained by TOMS and AVHRR series in 1978, aerosol information has been inferred from data obtained by later generation sensors. Aerosol required for atmospheric correction was yielded from the Sea-viewing Wide Field-of-view Sensor (SeaWiFS) thought this instrument was developed for studying marine biogeochemical processes (Gordon and Wang [1994]). With the launches of Terra and ESA EnviSAT satellites, more advanced instruments such as the MODerate resolution Imaging Spectro-radiometer (MODIS), Multiangle Imaging SpectroRadiomete (MISR), MEdium Resolution Imaging Spectrometer (MERIS), and Advanced Along-Track Scanning Radiometer (AATSR) improved significantly aerosol retrievals from satellite observations. The POLarization and Directionality of the Earth's Reflectances (POLDER) instruments on ADEOS II provided polarization measurements of backscattered solar light and hence added more capacities for aerosol retrievals. The new generation of laser satellites such as Lidar In-space Technology Experiment (LITE) (Winker et al. [1996]), the Geoscience Laser Altimeter System (GLAS) (Spinhirne et al. [2005]), and the most recently launched Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observations (CALIPSO) allows to derived more aerosol characteristics such as vertical distribution of aerosols, backscatter, extinction, and depolarization ratios. Table 2.1 summaries common instruments and satellites used to derived aerosol properties from space.

Satellites were launched in geosynchronous, medium, and low orbits which are referred to as GEO ($\sim$ 35,786 km), MEO (2,000 - 25,000 km), and LEO (<2,000 km), respectively. Aerosol measurements are often derived using data recorded by sensors on GEO and LEO satellites (Figure 2.3a). GEO satellites orbit the Earth once for every daily rotation of the Earth at a stationary position to the equator. GEO satellites can provide multiple views (as short as every 5 minutes) of a large region of the globe per day. The spatial resolution of sensors can be 1 km if a large telescope is used. LEO satellites have altitudes range from 250 to 700 km which allows to orbit the planet in approximate 90 minutes. Most of satellites relevant to atmospheric composition measurements have sun-synchronous orbit

(i.e. satellites always observe the day side of the Earth) that inclines about 96.1° to the west of the north pole at 705 km altitude. Because of the low orbit, each LEO satellite measurements only provide a swath of the Earth surface at most 3,000 km of width. Any location on the Earth is only observed once a day or less.

In order to measure the complex, dynamic and multi-dimensional aspects of the global aerosol system, multiple sensors are integrated by A-Train. A-Train is a constellation of satellites flying on near-identical orbits as a pack, therefore it provides near-simultaneous observations of the same part of the Earth through dozens of sensors on eight satellites (Figure 2.3b).



(a) (b)

Figure 2.3: (a) Polar and Geostationary satellite orbit (b) NASA A-Train satellite constellation (Aura, PARASOL, CALIPSO, CloudSat, and Aqua).

1 .2.2 Remote sensing measurements from satellites

Satellites provide remote measurements for the state of the atmosphere. Most aerosol retrieval applications have been designed and developed using theory of electromagnetic radiation scattered and/or absorbed by the atmospheric constituents and the surface target. Figure 2.4 describes the signals measured by satellite sensors. I_0 is incoming solar radiation to TOA and plays a role as an energy source. The gases and aerosols interact with the solar radiation, scatter radiation I_1 and I_2 in the atmosphere, and transfer radiation I_3 through the atmosphere to the Earth surface. I_1 is scattered radiation in the atmosphere which depends on the density of air, trace gases, and aerosols. I_1 is formulated as a function of height that presents the loss mechanism of source radiation. I_2 is

scattered radiation going back to satellite field of view and able to be measured by satellite sensors. The reflected radiation I_4 from the Earth surface is an important source used by Earth-viewing satellites to compare with I_2 to retrieve aerosol and gas features. The surface reflected radiation depends largely on the surface characteristics. I_5 is the satellite-retrieved TOA radiance from surface as a result of attenuated I_4 by scattering and absorption in the layer of interest. Generally, I_5 needs to be small for I_2 to be detected. Another source contributing to satellite measurements is upwelling IR radiance I_6 from the surface as a result of surface emissivity. I_7 is TOA IR radiance that can be measured by IR satellite sensors and often used for trace gas, water measurements, and cirrus cloud screening.

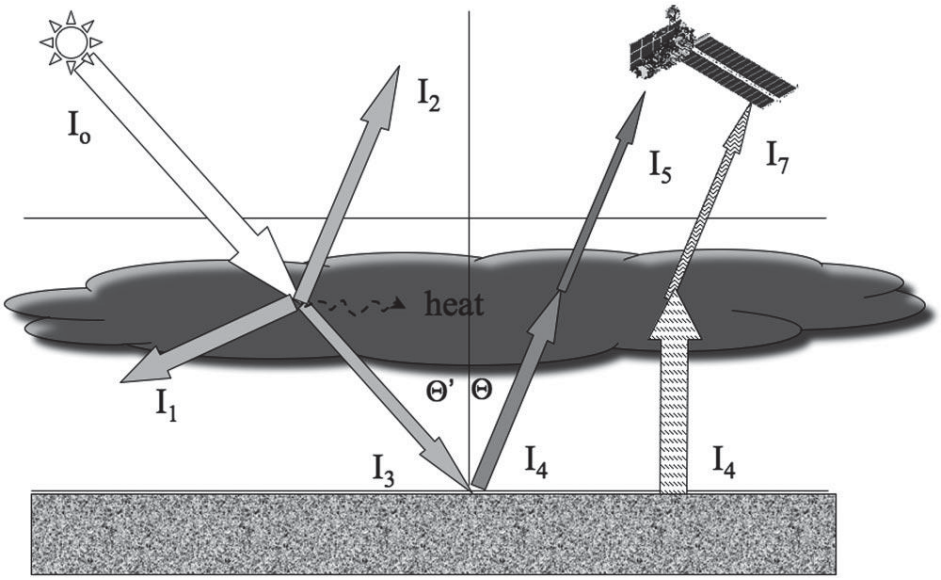


Figure 2.4: The basic scheme of radiative transfer processes for passive satellite remote sensing (Hoff and Christopher [2009]).

1 .3 Critical procedures for AOT retrievals

Aerosol Optical Thickness (AOT) or Aerosol Optical Depth (AOD) considered in this thesis refers to as aerosol measured in vertical direction of the atmosphere. The total reflectance calculated from TOA radiance is considered as contributions from atmospheric reflectance by gas molecules, aerosols and surface reflectance. If

AOT retrievals are carried out at IR wavelengths instead of visible wavelengths, the Earth emissivity will be taken in account. The aerosol radiance, referred as *"path radiance"* in (Kaufman and Tanré [1997]), is obtained from TOA measurements by separating the contributions of gas molecules (Rayleight) and the Earth surface. Retrieval algorithms are applied to path radiance to derive aerosol properties such as mass, ångström exponent, fine aerosol fraction from the total measured reflectance.

The quality of AOT retrievals is varied by satellite and algorithm because of their approaches in dealing with a significant degree of uncertainty from different sources including cloud contaminated data, the Earth surface reflectance, and inversion process.

- Cloud screening is an important process for distinguishing cloudy from clear-sky pixels. This process is carried out before application of AOT retrieval algorithms because optical signals from aerosols are indistinguished from clouds and hence it is not possible to apply AOT algorithms over cloudy pixels. Moreover, cloud contamination impacts significantly on AOT retrieval. The popular approaches for cloud screening are based on a set of static thresholds (e.g. radiance or temperature) applied on pixel basis. However, threshold techniques are not robust with sub-pixel clouds, high reflectance surfaces, illumination and observation geometry, sensor calibration, transparency of clouds, etc. Over water pixels, the cloud detection is more effective since the background reflectance is low and predictable. On the other hand, the variability of surface reflectance over land pixels makes cloud detection quite difficult and challenging.

- Surface reflectance, a part of contribution to total reflectance measured at TOA, is separated to from total TOA measurements to obtain path radiance used for AOT retrieval. Inaccuracies in calculating surface reflectance will increase uncertainties of AOT retrieval algorithms, which affects mainly the quality of satellite AOT derived over land where surfaces are complicated and variant against over ocean where surfaces are more homogeneous. The techniques to calculate surface reflectance strongly depend on satellite measurement properties. For MODIS and MERIS measurements, the assump-

tion is that the path radiance needs to dominate surface reflectance in TOA measured reflectance in order to minimize those surface uncertainties that impact on aerosol prediction from satellite measurements. Therefore, most of AOT retrieval algorithms are applied on dark surfaces whose reflectance is small. Also, surface reflectance is weak in the blue band (Hoyningen-Huene et al. [2003]) and red band (Remer et al. [2005]), hence those bands are used to derive aerosol properties. For multi-angle measurements (MISR, (A)ATSR), the ratio of surface reflectance at different viewing geometries are employed (Diner et al. [2008]) while for POLDER, polarized characteristics of surface reflection and atmospheric scattering are used (Deuzé et al. [2001]).

- The inversion process focuses on deriving aerosol and its properties (i.e. composition, size, shape ...) from satellite measurements. The inversion process is based on relationship between AOT and correlation inputs such as TOA reflectance, climate parameters (e.g. solar/sensor zenith/azimuth angle, scattering angle, geolocation, wind speed ...). Up to now, there are two main approaches for inversion process, that is, model-driven approach based on physics theory and data-driven approach using statistics or machine learning theory. Having a long history, model-driven approaches have been applied in a large range of applications. Following this approach, parameter of different aerosol models, representative for different aerosol types, are calculated using radiative transfer theory and then stored in LUT for aerosol retrievals. Based on measured parameters, different simulated aerosol models are applied to yield simulated reflectance which will be compared to measured reflectance to find the best-fit case considered as solution. Using the matched solution, AOT and its properties are derived. The aerosol retrieval algorithms following a statistical approach, on the contrary, develop empirical aerosol models from data collected in areas of interest for a long period to predict AOT information. This approach focuses on investigating the data relationship and using statistics and machine learning theory to constraint the quality of data models. The topic of this thesis is downscaling AOT from MODIS using physics and machine

learning approaches with focus on inversion process to derive aerosol information at finer spatial resolution. Both the MODIS aerosol algorithms and SVR algorithm are investigated but a big portion of this study is dedicated to the second approach.

2 Review of satellite aerosol retrieval algorithms

Today, a large number of aerosol products have been provided freely or commercially for applications in the field of atmospheric chemical modelling. They have resulted from several aerosol retrieval algorithms processing satellite data obtained by different instruments. This section reviews aerosol retrieval methodologies in physics and machine learning perspectives which details the theory summarized in previous sections with practical examples. This section aims at providing an overview of the state of the art and therefore providing the context for this thesis.

2 .1 Physics algorithms

Physics algorithms in general are based on radiative transfer theory. Most of them are tailored for specific sensors but some are general enough to apply to a wide range of satellite instruments. Common aerosol retrieval algorithms corresponding to different instrument generations are summarized in Table 2.2 and briefly described in what follows. Complete reviews can be found in (Heiberg et al. [2010]; Lee et al. [2009]).

Geostationary satellite algorithm. Although geostationary satellite provides measurements with coarse spatial resolution, it has been the unique tool for quantifying aerosol properties from space with high temporal resolution. In this algorithm, surface reflectance is obtained from "background image" acquired from composited minimum reflectance values among numerous views of the same location for a period. The AOT retrieval is based on comparison between imagery to this "background image". The AOTs obtained by VISSR onboard GOES have uncertainty in the range of ± 18-34% (Knapp et al. [2002]).

Single-channel AVHRR algorithm. The AVHRR algorithms use the wave-

length λ=0.63 μm and a LUT assuming certain types of aerosol models and calculated by radiative transfer codes for AOT derivation. Modeling aerosol types represented by aerosol models in LUT impact strongly in the quality of retrieved aerosol. In the earlier algorithm (Stowe [1991]), non-absorbing aerosols with a modified Junge size distribution are assumed. The results are validated with shipborne sunphotometer measurements made within ± 2 hours the satellite overpasses. The comparison shows a negative bias $\tau_{sat} = 0.64\tau_{sp} - 0.02$. However, in the current algorithm used to generate AVHRR aerosol products, the log-normal aerosol size distribution is employed together with application of the Fresnel model for bidirectional reflectance of calm ocean surfaces. As a result, AOT retrievals agree with ground measurements to better than 10%. The liner regression between them is $\tau_{sat} = 0.91\tau_{sp} + 0.01$ (Stowe et al. [1997]).

Dual-channel AVHRR algorithm. The two channel algorithms has been applied to the International Satellite Cloud Climatology Project (ISCCP) to generate the Global Aerosol Climatology Product (GACP)(Mishchenko et al. [1999]). Besides AOT, the ångström exponent which is a parameter that denotes aerosol size is also derived. The algorithm used both AVHRR channels (λ = 0.65, 0.85 μm). Aerosols are assumed to be spherical with the power-law size distribution and a refractive index of 1.5-0.003i. This algorithm is susceptible to calibration errors in both channels and cloud screening errors.

TOMS Algorithm. TOMS instruments onboard Nimbus-7, ADEOS, and Earth Probe originally provide data for monitoring ozone depletion. Since the ratio of two channels (λ = 0.331, 0.360 μm) is sensitive to absorbing aerosols, they can be used to derive the Aerosol Index (AI) (Herman et al. [1997]). The advantages of this technique is that the presence of subpixel clouds does not affect the detection of aerosols. Daily global TOMS AI products have been used largely to monitor the spatial and temporal variations of smoke, dust, and other types of absorbing aerosols. The "near-UV algorithm" is used to extract additional quantitative aerosol parameters such as AOT and Single Scattering Albedo (SSA). This algorithm uses two backscattered radiances at near-UV bands, a LUT assumed by three aerosol models, and variability of the relationship between the spectral contrast and the radiance at the longer wavelength. The long-term AOT (1979 to present) over land and ocean had uncertainties about 30% in comparison

with AErosol RObotic NETwork (AERONET) observations, while AOT of non-absorbing aerosols agreed to within 20% (Torres et al. [2002]). The TOMS SSA generally agrees within 0.03 of AERONET retrievals (Torres et al. [2005]). However, this retrieval is affected by the aerosol layer altitude, the single scattering albedo, and cloud contamination.

Ocean Color Algorithms (SeaWiFS). Aerosol products are by-products from the atmospheric correction for the ocean color algorithm (Gordon and Wang [1994]). The TOA reflectance over ocean is a total of air molecules scattering, aerosols, the interactions between molecular and aerosol scattering, ocean reflection, and water-leaving reflectance. The ocean color algorithm utilizes the fact that the effects of all mentioned factors excepting aerosols are negligible at Near Infrared (NIR) wavelengths. Therefore, aerosol information can be derived from NIR bands by using a set of suitable aerosol models. The current SeaWiFS and MODIS ocean color algorithms (Collection 4) use 12 aerosol models for generating the LUTs (Wang et al. [2005]). Using combination of two most appropriate aerosol models (i.e. one fine and one coarse modes), the AOT and Ångström exponent can be retrieved.

Polarization (POLDER). The POLDER instrument provides multi-band imaging radiometer and polarimeter with eight narrow spectral bands in the visible and near infrared. The spectral variation allows to derive aerosol size and AOT. The polarization provides some information on the aerosol refractive index and shape, which contributes to improve scattering phase function. This algorithm is based on LUTs from POLDER directional, spectral, and polarized measurements for several aerosol models. The accuracy in AOT retrievals was reported as 30% (Herman et al. [1997]). The POLDER Ångström exponent correlated well with AERONET data but underestimated by 30% (Goloub et al. [1999]) because it is more sensitive to fine aerosols.

Multi-channel Algorithm (SeaWiFS, MODIS, MERIS). The MODIS aerosol algorithms are designed to retrieve aerosol information from MODIS measurements. The algorithms, separated over land and ocean, generate the most comprehensive aerosol products including AOT, fine mode fraction, effective radius of aerosol particles, and mass concentration. The retrieval uncertainty of the MODIS AOT products falls within $\pm 0.03 \pm 0.05\tau_{sat}$ over ocean and $\pm 0.05 \pm 0.15\tau_{sat}$

over land (Remer et al. [2005]). However, the modified aerosol algorithms in Collection 5 products improves significantly accuracy as shown in later validation studies (Levy et al. [2007]).

BAER (Bremen Aerosol Retrieval) algorithm aims at deriving aerosol optical properties over land and ocean. It is a generic approach that have been applied to MODIS, MERIS, SCIAMACHY, and SeaWiFS. In this algorithms, the blue channel of satellite instrument is used to separate surface and atmospheric reflection, which works appropriately over dark surfaces but deserts are need individual treatments. This technique, a variant of the "dark target" method proposed by (Kaufman and Tanré [1997]) for MODIS, assumes a dynamic estimation of the surface reflectance using a linear mixing model of vegetation and soil, tuned by the Normalized Differential Vegetation Index (NDVI) of the satellite scene. The derivation of AOT and its optical properties is based on LUT pre-calculated by a radiative transfer model. Sensitivity tests have shown that the BAER algorithm can reach an agreement of 20% between retrieved and ground-based AERONET AOT at 0.421 μm. The error in AOT decreases with the zenith distance of the sun and the satellite (Hoyningen-Huene et al. [2003]).

Multi-angle, Multi-channel (MISR). MISR observations provides measurements of nine different view angles (i.e. one at nadir and eight symmetrical views at 26.1, 45.6, 60.6, and 70.5 degrees forward and after of nadir) at four spectral bands (0.443, 0.558, 0.672, and 0.866 μm). The combination of multiple bands and multi-angles allows to estimate AOT and additional properties such as the Ångström exponent, SSA, number fraction, and volume fraction. The algorithms are designed for water, Dense Dark Vegetation (DDV), and heterogeneous land (Martonchik et al. [1998]). Over dark water, water-leaving radiances at red and near-infrared wavelengths are considered as zero, therefore it is similar to the ocean color algorithm. The algorithm for DDV uses an angular shape for the surface Bidirectional Reflectance Factor (BRF) with angular measurements. In heterogeneous land, empirical orthogonal functions derived from the spectral contrast by multi-angle observations are used to determine AOT and the aerosol model. Validations of MISR AOTs using AERONET AOTs show a positive bias of 0.02 with an overestimation of 10% over southern Africa (Diner et al. [2001]), an overestimation over China (Christopher and Wang [2004]), and a linear rela-

tionship of $\tau_{sat} = 0.92\tau_{sp} + 0.02$ and a retrieval error of $00.4 \pm 0.18\tau_{sp}$ over United States (Liu et al. [2004]).

2 .2 Machine Learning algorithms

The Machine Learning (ML) approach has recently been investigated to improve quality of aerosol monitoring. ML techniques provide a more general framework for aerosol retrieval than physical techniques, hence they can be applied to data obtained by a large range of satellite instruments. ML techniques are used in the inversion process and play a role as radiative transfer model in algorithms based on physical approach. The approach is efficient in reducing processing time (Okada et al. [2001]), dealing with data uncertainties (Vucetic et al. [2008];Obradovic et al. [2010]), improving estimation accuracy (Xu et al. [2005]; Vucetic et al. [2008]; Nguyen et al. [2010c]), flexibly updating new inversion models, and easily extending to other types of sensors. However, its limitations are data dependence and complex modeling process.

In terms of aerosol estimation field from satellite observations, there are various applications using Neural Network (NN). Böttger [2000] developed an algorithm for aerosol and cloud characteristic retrievals from POLDER measurements using NNs. The data obtained by radiative transfer simulation were used to create empirical NN models for deriving aerosol, cloud optical thickness, and aerosol type (i.e. refractive index and the mode radius of the cloud droplet size distribution). Following the same approach, Okada et al. [2001] used information in LUT to train a NN system, then applied it on ADEOS/OCTS data to retrieve aerosol properties over the ocean. Li et al. [2001] developed both NN and multi-threshold techniques to detect smoke from forest fires using AVHRR measurements. Two techniques were employed separately or in combination depending on the size of an image and smoke conditions. The methodologies were validated to TOMS AI data over Canada in 1998. AVHRR and TOMS smoke detections were reasonable correspondent, but quite different and complementary. AVHRR is most sensitive to low, dense smoke plumes located near fires, whereas smoke detected by TOMS is dispersed, thin, elevated, and further away from fires. Another application for SeaWiFS AOT and Ångström exponents in (Jamet et al. [2004])

trained NN with radiative transfer computations and applied to Mediterranean images in 2000 to derive monthly mean maps. In comparison with ground-based measurement at three AERONET stations, higher agreement and improvements compared to operational SeaWiFS aerosol products were found. Regarding the MERIS instrument, NN model was created from simulated data to retrieve AOT from TOA signals recored at 13 MERIS wavelengths in Beal et al. [2007]. The validation on 61 MERIS images over AERONET sites showed good accuracy as well as competition with results obtained by 31 MODIS images. Brajard et al. [2007] estimate AOT over ocean case 1 waters using a topological NN (Kohonen map). Different from previous studies, NN was trained with real data collected from 85 MERIS images in 2003 and 2004 over the Mediterranean Sea. The algorithm was validated using 5 day MERIS images and AERONET AOT on a small island (Lampedusa) in the Mediterranean Sea. Utilizing Support Vector Machine (SVM), Ma et al. [2009] classified cloud and aerosol from the Cloud-Aerosol Lidar with Orthogonal Polarization (CALIPSO) measurements.

Regarding the AOT retrievals from MODIS measurements, many studies employed both NNs and SVR for aerosol retrieval. In Xu et al. [2005], data recorded by MISR and MODIS are integrated with AERONET measurements to create data samples. Satellite data were collected during 2002 and 2003 covering the continental USA. These data are joined spatially and temporally with AOT measurements from 34 AERONET stations over the continental USA. After that, NN is applied for inversion process. The validation results showed that the accuracy of AOT prediction was able to increase up to 10% when both MISR and MODIS data were used instead of individual sensor's measurements. Following the same approach, Vucetic et al. [2008] compared performances of model-driven approach (the MODIS algorithm) and data-driven approach (NN) in AOT retrievals from MODIS. The NN approach was applied on 3646 collocated MODIS and AERONET observations over the continental USA. The experiments show situations in which each approach presented its advantages and disadvantages. However, NN is very competitive with the MODIS algorithm and can be considered as a feasible technique for AOT retrievals. A complete summary of related studies from this research group are presented in Obradovic et al. [2010]. Besides the topic of AOT retrievals, Lary et al. [2009] investigated the use of NN

and SVR for AOT correction in order to remove bias of the standard MODIS AOT products. The proposal algorithms applied on data at the global scale. The experiment results showed SVR outperformed NN techniques and was able to improve correlation coefficient between AERONET AOT and MODIS AOT from 0.86 to 0.99 for MODIS Aqua and from 0.84 to 0.99 for MODIS Terra.

3 Review of downscaling techniques for MODIS AOT retrievals

Downscaling has been recognized as a popular and important methodology applied for General Circulation Models (GCMs) which are used to assess climate changes. These numerical models present various earth systems including the atmosphere, oceans, land surface and sea-ice with coarse spatial resolutions. Downscaling techniques are developed in order to bridge the gap between the resolution of climate models (some hundreds kilometre square in general) and regional and local scale process. Two fundamental approaches applied for the downscaling of large scale GCMs output to a finer spatial resolution are dynamical and statistical downscaling. Dynamical approach refers to the use of Regional Climate Models (RCM) or Limited Area Models (LAMs). These focus on parameterizing atmospheric process of the large scale GCMs to produce higher resolution outputs. Statistical downscaling investigated the relationship expressed as a stochastic and/or deterministic function between large-scale atmospheric variables and local/regional climate variables. The statistical downscaling methods are generally classified into three groups: regression models, weather typing schemes and weather generators. The full review and comparisons in advantages and disadvantages between two downscaling approaches are presented detailed in Fowler et al. [2007].

Regarding the downscaling aerosol products from MODIS observations, in literature, several studies focused on improving MODIS aerosol algorithms or proposing new methodologies for aerosol derivation. Published articles have shown achievements in which aerosol could be estimated at a finer resolution than 10×10 km^2. Grouped by regions, downscaling studies were investigated

and considered for New York city, Sao Paolo, Hong Kong, Hong Kong and Pearl River Delta with the spatial resolutions at 3×3 km^2, 1.5×15 km^2, 1×1 km^2, and 500×500 m^2, respectively. Oo et al. [2008] used the MODIS algorithm with proposed ratio of VIS (0.66 μm) and MIR (2.12 μm) as 0.73 instead of 0.59 over New York city. The validation showed that obtained AOT can be achieved at 3×3 km^2 or more detail 1.5×1.5 km^2 of spatial resolution and still keep good accuracy in comparison with AERONET AOT. Focusing on new aerosol models, Castanho et al. [2008] proposed the use of local surface reflectance and aerosol critical reflectance techniques to automatically estimate aerosol optical property (i.e. SSA). The methodology is applied to retrieve AOT from MODIS data collected from 2002 to 2005 and then compared with collocated data measured by AERONET in Sao Paulo. The good results could be achieved when the dynamic methodology using two aerosol optical models was applied (slope 1.06 $\pm$ 0.08, offset 0.01 $\pm$ 0.02, $R^2 \sim 0.6$). Li et al. [2005] generated a new LUT with accommodate aerosol models to HongKong city using MODIS measurements with empirical SSA around 0.91 - 0.94 instead of 0.97 as in the standard MODIS algorithms. The AOT validated to Microtops II sunphotometer observations showed retrieval errors within 15% to 20%. In comparison with Particulate Matter concentration (PM_{10}), obtained AOT data were much better correlated than MODIS AOT at 10×10 km^2. Nichol and Wong [2009] continued to downscale AOT retrieved from MODIS over Hongkong city to 0.5×0.5 km^2. In this methodology, Minimum Reflectance Technique (MRT) was used to estimate surface reflectance, while LUT was calculated using SBDART code for the AOT retrieval. The validation using MODIS data collected in 2007 at the HongKong AERONET site showed good linear fitting correlation coefficient (R^2) of 0.93. The same approach was applied and presented promising results over Pearl River Delta region, China (Wong et al. [2009]). Following this trend but deciding on downscaling aerosol products at global scale, the MODIS team has planed to release Collection 6 in 2012 with spatial resolution increased to 3×3 km^2 (Remer et al. [2010]). However, this version was announced to have lower prediction quality than the current aerosol product at 10×10 km^2. All of reviewed studies follow dynamical downscaling approach, in which surface reflectance was parameterized and regional LUTs were developed for inversion processes.

Table 2.1: Instruments and satellites used to derived aerosol properties form space.

Instrument	Satellite	LifeTime	Species	Orbit	Tem./Spa. resolutions (day/km2)	Spectral range (μm)
VISRR	GOES-1∼12	1975-pre.	Visible/IR imagery	Geo.	?/1×1-8×8, depend on wavelength	5(0.65 12.5)
AVHRR	TIROS-N	1978-pre.	H20, Cloud imagery	PS	0.5/ 1.1×1.1	4(0.58 11.5)
	NOAA6 16	1979-pre.	H20, Cloud imagery	PS	0.5/ 1.1×es1.1	5(0.58 - 12)
TOMS	Nimbus-7	1978-1993	O3	PS	?/47×47	6(0.308 - 0.36)
	ADEOS	1996-1997	-	-	-	6(0.308 - 0.36)
EP-TOMS	Earth Probe	1996-pre.	-	-	?/39×39	6(0.308 - 0.36)
POLDER	ADEOS	1996-1997	Aerosol properties	PS	?/6×6	9(0.443 - 0.910)
POLDER-2	ADEOS	2002-2003	-	-	-	9(0.443 - 0.910)
POLDER-3	PARASOL	2004-pre.	-	-	-	8(0.44 - 0.91)
SeaWiFS	OrbView-2	1997-pre.	?	PS	1/ 4.5×4.5	8(0.412 - 0.865)
MODIS	Terra	1999-pre.	H2O, clouds, aerosols	PS	1-2/ 0.25×0.25, 0.5×0.5, 1×1	36(0.4 - 14.4)
-	Aqua	2002-pre.	H2O, clouds, aerosols	PS, A-Train	-	-
MISR	Terra	1999-pre.	Aerosols	PS	9/ 0.275×0.275, 0.55×0.55, 1.1×1.1	4(0.45 - 0.87)
AATSR	ENVISAT	2002-pre.	Surface T	PS	?/1×1	7(0.55 - 12)
MERIS	-	-	AOD	-	3/1.2×1.2	15(0.4 - 1.05)
SCIAMACHY	-	-	O3, NO2, H2O, N2O, CO, CH4, CHOCHO, OClO, H2CO, SO2, aerosols, P, T	-	3/30×60	0.24 - 2.4
CALIOP	CALIPSO	2006-pre.	Lidar profiles of aerosols	PS, A-Train	?/0.03-0.06 (vertical), 0.333 (horizontal)	2(0.532 - 1.064)

Table 2.2: Aerosol retrieval algorithms for different satellite instruments (AOT - Aerosol Optical Thickness, AAI - Absorption Aerosol Index, SSA - Single Scattering Albedo, α - Ångström exponent, η - fine aerosol fraction.

Instrument	Satellite	Aerosol Retrieval Algorithm	Aerosol Products	Aerosol Spatial Resolution (km^2)
VISRR	GOES-1$\tilde{1}$2	Geostationary Satellite Algorithm (Knapp et al. [2002])	AOT	?
AVHRR	TIROS-N	Single-channel AVHRR algorithm (Stowe [1991]; Stowe et al. [1997])	AOT	?
-	NOAA 6$\tilde{1}$6	Dual-channel AVHRR Algorithm (Stowe et al. [1997]; Mishchenko et al. [1999])	AOT, α	?
TOMS	Nimbus-7	TOMS algorithm (Herman et al. [1997])	AI, AOT, SSA	40×40
-	Earth Probe	TOMS algorithm (Herman et al. [1997]; Torres et al. [2002])	AI, AOT, SSA	-
POLDER	ADEOS	Polarization algorithm (Herman et al. [1997])	AOT, α	20×20
POLDER-2	ADEOS	-	-	-
POLDER-3	PARASOL	-	-	-
SeaWiFS	OrbView-2	Ocean Color Algorithm (Gordon and Wang [1994])	AOT, α	?
MODIS	Terra	Multi-Channel Algorithm (Remer et al. [2005])	AOT, α, η	10×10
-	Aqua	-	-	-
MISR	Terra	Multi-Angle, Multi-Channel Algorithm (Martonchik et al. [1998]; Diner et al. [2008])	AOT, α, SSA, η, and mass concentration	17.6×17.6
AATSR	ENVISAT	(Grey et al, 2006)	?	?
MERIT	-	Multi-Channel Algorithm (von Hoyningen-Huene et al. [2006])	?	?
SCIAMACHY	-	Limb Sounding (Graaf and Stammes [2005])	AOT	60×30
CALIOP	CALIPSO	Active Sensing Algorithm	AOT, α	40×40

Chapter 3

Methodologies for Aerosol Optical Thickness Retrieval from MODIS observations

In this chapter, the methodologies for AOT retrieval from MODIS observations are presented. Firstly, background information of instruments and released data products are introduced. After that, algorithms over ocean and land for remote sensing of tropospheric aerosol from MODIS applied in Collection 5 are summarized. SVR with fundamental mathematics theory for building regression models from data will be presented in the next sections together with brief description of SVR application for AOT retrieval.

1 Background Information

1 .1 Characteristics of the MODIS and AERONET instruments

The MODerate resolution Imaging Spectrometer (MODIS) is a multi-spectral sensor on-board the two polar orbiting satellites Terra and Aqua, launched in 1999 and 2002, respectively and operated by the National Aeronautic and Space Administration (NASA). Terra and Aqua with the descending (southward) and

ascending (northward) orbit cross the equator about 10:30 and 13:30 local sun time, respectively. Each MODIS views the Earth with a swath width of 2330 km, therefore these satellites provide observations nearly the entire globe on a daily basis, and repeat orbits every 16 days. The swath is divided into granules, each of which is corresponding to 5 minutes recording. Therefore, each granule is corresponding to a map of 2030 pixels of 1 km width in the direction of satellite path and 1354 pixels of non-uniform width (i.e. The real pixel size projected on the earth far away from nadir is larger than those at nadir because of the influence of instrument scan and the Earth's curvature). The MODIS performs measurements in the solar to thermal infrared spectrum region from 0.41 to 14.235 μm, separated into 36 bands at resolutions 1 km, 500 m and 250 m at nadir. The MODIS aerosol retrievals use seven wavelength bands and other bands for cloud and other screening procedures as shown in Table 3.1.

Table 3.1: Characteristics of MODIS channels used in aerosol retrieval

Band#	Bandwidth (μm)	Weighted Central Wavelength (μm)	Resolution (m)	Ne$\Delta\rho$ ($\times 10^{-4}$)	Maxρ	Required SNR
1	0.620-0.670	0.646	250	3.39	1.38	128
2	0.841-0.876	0.855	250	3.99	0.92	201
3	0.459-0.479	0.466	500	2.35	0.96	243
4	0.545-0.565	0.553	500	2.11	0.86	228
5	1.230-1.250	1.243	500	3.12	0.47	74
6	1.628-1.652	1.632	500	3.63	0.94	275
7	2.105-2.155	2.119	500	3.06	0.75	110

The spectral reflectance ρ_λ of the wavelength λ are defined as a function of measured spectral radiance L_λ, the solar zenith angle θ_0 and the solar irradiance $F_{0,\lambda}$ as follows:

$$\rho_\lambda = L_\lambda \frac{\pi}{F_{0,\lambda} cos(\theta_0)} \tag{3.1}$$

The "Noise Equivalent Different Spectral Radiance" ($Ne\Delta L$) is a property of the instruments. Using it, Noise Equivalent Different Spectral Reflectance ($Ne\Delta\rho$) and Signal to Noise Ratio (SNR) can be calculated (see Table 3.1).

The AErosol RObotic NETwork (AERONET) is a global system of ground-based Remote Sensing aerosol network established by NASA and PHOTONS

(University of Lille 1, CNES, and CNRS-INSU) (AERONET [2011]). The network imposes standardization of instruments, calibration, processing and distribution. The main products are AOT, inversion products, and precipitable water in diverse aerosol regimes. Aerosol Optical Thickness is measured by the CIMEL Electronique 318A spectral radiometers, sun and sky scanning sun photometers in four or more wavelengths to include 0.440, 0.670, 0.870, and 1.020 μm, in every 15 minutes during midday and more often during sunrise and sunset. The global distribution of AERONET stations is illustrated in Figure 3.1

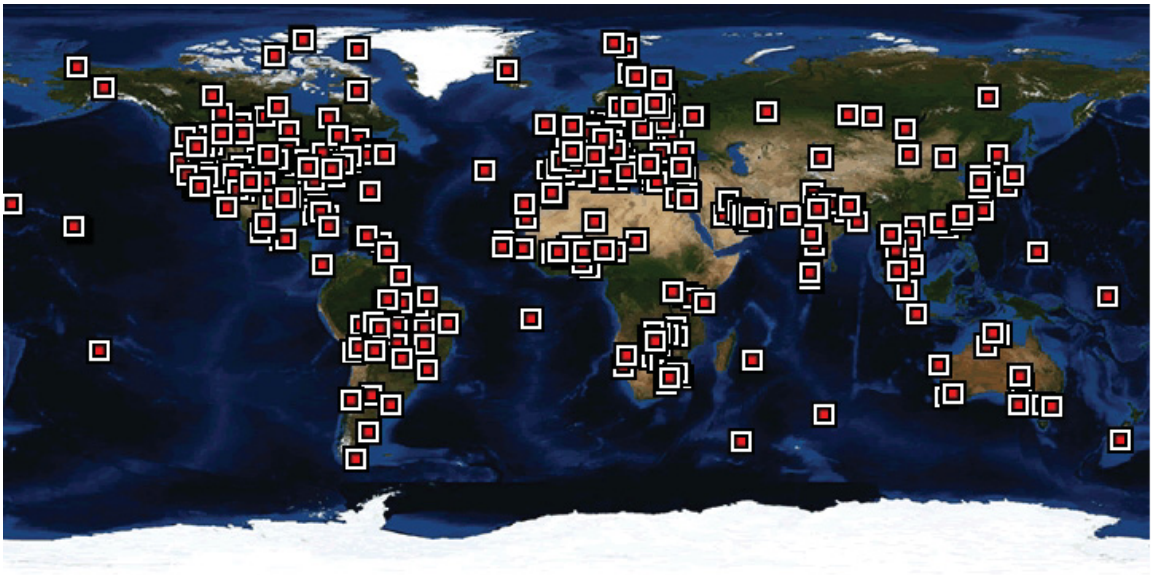


Figure 3.1: The global distribution of AERONET ground stations (AERONET [2011])

1 .2 Data Description

The MODIS Level 1B (L1B) data acquired by MODIS sensors on board the Terra and Aqua satellites represent measurements of a spectrum region from 0.415 to 14.235 μm divided into 36 channels at 1 km, 500 m, and 250 m resolution at nadir. A scene covers an area on the Earth surface of 2030 km in the direction of the satellite orbit and of 1354 km of non-uniform width. The spectral reflectance are calibrated and provided in products named MOD02/MYD02 for Terra/ Aqua. In

addition, the corresponding geo-location product containing geodetic coordinates, ground elevation, solar and satellite zenith and azimuth angles for each 1 km sample is provided together with L1B data, known as MOD03/ MYD03 for Terra/ Aqua.

The MODIS Level 2 (L2) refers to two processed products, that is, MOD035/ MYD035 Wisconsin cloud mask product and MOD07/ MYD07 atmospheric profile product. The MODIS cloud mask products provide information of cloud contamination by labeling every pixels of data as either confident clear, probably clear, uncertain, or confidently cloudy. The algorithm used 17 of 36 MODIS spectral bands for making tests of cloud contamination. The MODIS Atmospheric Profiles products (MOD07/ MYD07) include ozone information, atmospheric stability, temperature and moisture profiles, and atmospheric water vapor. In this product, all parameters are provided at 5×5 km^2 spatial resolution for Terra and 1×1 km^2 spatial resolution for Aqua when cloud-free.

MOD04 L2 is the aerosol products derived by the MODIS software package Collection 005. MOD04 L2 products characterized by spatial resolution of 10×10 km^2 provide AOT estimations at seven wavelengths (0.470, 0.550, 0.670, 0.870, 1.240, 1.630 and 2.130 μm) over ocean and three wavelengths over continental areas (0.470, 0.550 and 0.670 μm) together with respective geometry information and other various parameters.

All MODIS L1B, L2 atmospheric products, and aerosol products are written in Hierarchical Data Format (HDF) with each parameter stored as a Scientific Data Set (SDS). Besides, the ancillary data from NCEP (National Center for Environmental Prediction) including the GDAS $1^\circ \times 1^\circ$ 6 hourly meteorological analysis and TOVS/TOAST $1^\circ \times 1^\circ$ daily ozone analysis are considered in the MODIS algorithm.

AERONET collaboration provides the global AOT data computed for three data quality levels: Level 1.0 (unscreened), Level 1.5 (cloud-screened), and Level 2.0 (cloud-screened and quality-assured). Level 1.0 AOT is unscreened and may not have final calibration applied. Level 1.5 AOT has automatically cloud screening but maybe not have final calibration. Level 2.0 data are applied the pre- and post-field calibrations, cleared automatically of cloud and inspected manually.

2 Algorithm for remote sensing of tropospheric aerosol from MODIS: Collection 5

The MODIS aerosol algorithm includes two independent algorithms for deriving aerosol over ocean and land. Both algorithms use calibrated, geolocated reflectances provided by MOD02/MYD02 and MOD03/MYD03 products for Terra/Aqua. The MODIS cloud masks in MOD35/MYD35 and meteorological data are considered as inputs for algorithms as well. The theoretical basis of the algorithms has not changed from inception but some improvements have been carried out and organized by "Collection". Both algorithms output the AOT and related information at different wavelengths. This section presents theoretical basis of the latest MODIS aerosol algorithms: Collection 5.

2 .1 Algorithm over ocean

The algorithm for aerosol retrieval over the ocean in Collection 5 (C005-O) is presented in (Remer et al. [2004]). The algorithm follows the LUT approach using aerosol and surface parameters pre-computed by radiative transfer codes. The algorithm assumes that a proper weighting of one fine and one coarse lognomal aerosol modes can represent the ambient aerosol properties over the target. The simulated reflectance calculated from LUT is compared to MODIS measured spectral reflectance to find the best fit which will be considered as the solution of the inversion process. Although the inversion process is core of the algorithm, other procedures such as cloud screening, glint and sediment masking are considered and applied in order to guarantee the quality of the aerosol retrievals. Figure 3.2 presents a flowchart of the over ocean aerosol algorithm. Reflectances of the six wavelengths used in this algorithm ($\rho_{0.55}$, $\rho_{0.66}$, $\rho_{0.86}$, $\rho_{1.24}$, $\rho_{1.6}$, $\rho_{2.13}$) are calculated for 10 km boxes of 20×20 pixels at 500 m resolution and considered as inputs for the algorithm. The box sized 10 km determines spatial resolution for the final aerosol products. The ocean algorithm requires 400 pixels in the box to be ocean pixels identified by MOD35/MYD35 mask as a condition of algorithm application.

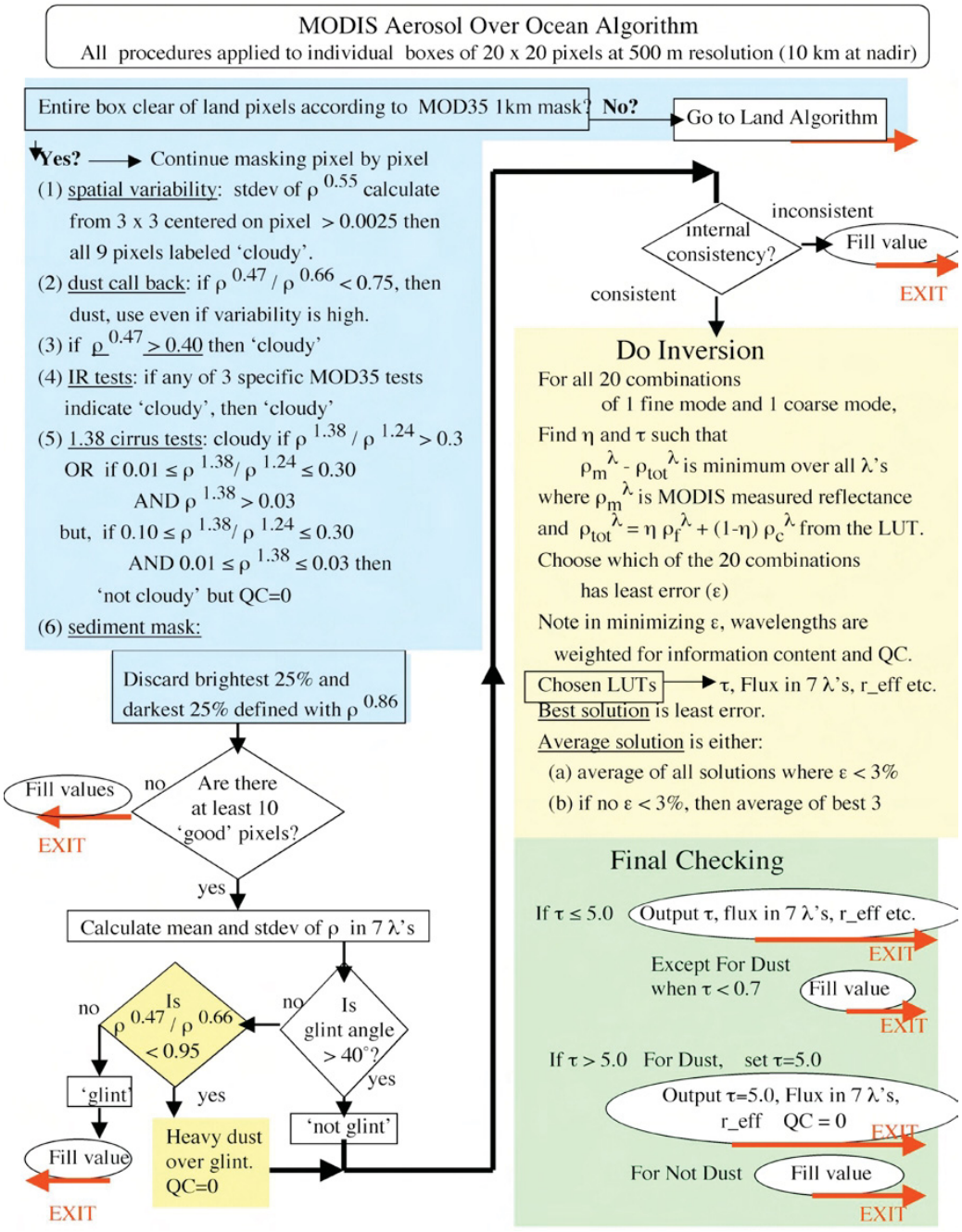


Figure 3.2: Flowchart of the over-ocean aerosol retrieval algorithm (Remer et al. [2005])

2 .1.1 Selection of pixels: cloud, glint and sediment masking

The cloud screening does not use the standard MOD35/MYD35 product as the primary cloud mask because this product cannot eliminate completely cloudy pixels, mistakes heavy aerosol as cloud, mis-retrieves important aerosol events over ocean, and allows cloud contamination. In the over-ocean aerosol algorithm, the main cloud mask is based on the difference in spatial variability between aerosols and clouds (Martins et al. [2002]). The algorithm exams the standard deviation of $\rho_{0.55}$ in a box of 3×3 500 m pixels. If this value is greater than 0.0025, all pixels are labeled as cloud and discarded. However, if heavy dust check using the ratio $\rho_{0.47}/\rho_{0.66}$ is less than 0.75, the center pixel will be considered as dust and included in the retrieval even if the spatial variability is high. The spatial variability test sometimes fails at the centers of large, thick clouds or also cirrus which can be spatially smooth. The given solution uses the test $\rho_{0.47} > 0.40$ to identify cloudy pixels, in addition. This threshold is extremely high and then, maybe discards non-absorbing aerosol and heavy aerosol loading ($\tau > 5.0$).

Cirrus clouds are identified by infrared and near-infrared tests. Three infrared tests provided by the standard MODIS cloud mask MOD35 include IR cirrus test, 6.7 μm test, and Delta IR test (Ackerman et al. [1998]). If any one of three tests fails, the pixel is labeled as cloud. The near-infrared cirrus test is based on the reflectance in the 1.38 μm channel and the ratio $\rho_{1.38}/\rho_{1.24}$ (Gao et al. [2002]).

The final mask applied to the data is the sediment mask which determines whether ocean scenes are contaminated by river sediments and discards those pixels (Li et al. [2003]). Spectral reflectances over water with suspended sediments have high values in the visible but not in the wavelengths longer than 1 μm, which distinguishes clear and sediment ocean water.

All pixels that evaded the cloud mask tests and the sediment mark are sorted according to their $\rho_{0.86}$ value in order to discard the 20% darkest and 25% brightest pixels. The filter aims at eliminating residual cloud contamination, cloud shadows, or other unusual extreme condition in the box. If there are at least 10 good pixels remained after all, they will be used to calculate reflectance mean and standard deviation for six wavelengths. Otherwise, no retrieval is attempted and all aerosol products in the 10-km box are given fill values.

The glint angle is defined as

$$\Theta_{glint} = cos^{-1}[(cos\theta_s cos\theta_v) + (sin\theta_s sin\theta_v cos\phi)] \tag{3.2}$$

where θ_s, θ_v, and ϕ are the solar zenith, the satellite zenith, and the relative azimuth angles between the sun and satellite, respectively. The ocean algorithm was designed to retrieve only over dark ocean, away from glint. If $\theta_{glint} > 40^o$, glint contamination can be avoided. Otherwise, several consistency checks of the spectral reflectances are carried out to exit the procedure or continue onto the inversion after assigning appropriate quality flags.

2 .1.2 Inversion Process

The inversion process uses the six reflectances measured from MODIS (0.55 - 2.13 μm) as inputs and derives three parameters: the AOT at one wavelength ($\tau_{0.55}^{tot}$), the reflectance weighting parameter at one wavelength ($\eta_{0.55}$), and the effective radius which is the ratio of the third and second moments of the aerosol size distribution. The inversion is based on a LUT that consists of four fine modes and five coarse modes (et. al. [108]) and is constructed using the radiative transfer code of Ahmad and Fraser [1982]. Each AOT mode is described by TOA reflectances in six wavelengths calculated for a variety of geometries, a rough ocean surface with nonzero water-leaving radiance only at 0.55 μm ($\rho_{0.55}^s = 0.005$), and several values of $\tau_{0.55}^{tot}$ for each single-mode aerosol model.

The procedure require a fine mode and a coarse mode for each retrieval. The reflectance calculated from LUT is a weighting combination of the reflectance values for an atmosphere with a pure fine mode "f" and the reflectance of an atmosphere with a pure coarse mode "c". In practice, $\eta = \tau_{0.55}^f / \tau_{0.55}^{tot}$, which represents the fraction of total optical thickness at 0.55 μm contributed by the fine mode.

$$\rho_\lambda^{LUT} = \eta \rho_\lambda^f(\tau_{0.55}^{tot}) + (1-\eta)\rho_\lambda^c(\tau_{0.55}^{tot}) \tag{3.3}$$

For each of 20 combinations of one fine mode and one coarse mode, the inver-

sion finds the pair of $\tau_{0.55}^{tot}$ and $\eta_{0.55}$ that minimizes the error (ϵ) defined as

$$\epsilon = \sqrt{\frac{\sum_{\lambda=1}^{6} N_\lambda (\frac{\rho_\lambda^m - \rho_\lambda^{LUT}}{\rho_\lambda^m + 0.01})^2}{\sum_{\lambda=1}^{6} N_\lambda}} \tag{3.4}$$

where N_λ is the sum of good pixels at wavelength λ, ρ_λ^m is the measured MODIS reflectance at wavelength λ, and p_λ^{LUT} is calculated from the combination of modes in the LUT and is defined by Equation 3.3. The wavelength 0.88 μm was chosen as the primary wavelength for accuracy consideration because it is less affected by variability in water-leaving radiances and exhibits a strong aerosol signal. The 20 solutions are then sorted by values of ϵ. The solution may not be unique. The average solution averages all solutions with $\epsilon < 3\%$. If no solution has $\epsilon < 3\%$, it averages the three best solutions. When the solution is found, related parameters can be inferred.

The final checks are employed before the final results are output. The retrieved AOT will be in a range of -0.01 to 5. Negative optical depths are possible and occur only in situations with low optical depth. They are not actually physical but are there to de-bias long term statistics. Those values are reported with lower quality flags.

2 .2 Algorithm over land

The reflectance over land, obtained at the top of the atmosphere (TOA), is a function of successive order of radiation interactions in surface-atmosphere system. The TOA angular spectral reflectance $\rho_\lambda(\theta_0, \theta, \phi)$ depends on solar zenith θ_0, view zenith θ and relative azimuth angles ϕ. It is contributed from scattering of radiation (the atmospheric path reflectance), the surface reflection directly transmited to the TOA (the surface function) and the reflection of radiation from outside the sensor's Field of View (FoV) (the environment function) that is often neglected. Therefore, an approximation is defined as:

$$\rho_\lambda^*(\theta_0, \theta, \phi) = \rho_\lambda^a(\theta_0, \theta, \phi) + \frac{F_\lambda(\theta_0) T_\lambda(\theta) \rho_\lambda^s(\theta_0, \theta, \phi)}{1 - s_\lambda \rho_\lambda^s(\theta_0, \theta, \phi)} \tag{3.5}$$

where $F_{d\lambda}$ is the normalized downward flux for zero surface reflectance, T_λ represents upward total transmission into the satellite FoV, s_λ is atmospheric backscattering ratios, and ρ^s_λ is the angular spectral surface reflectance.

In Equation 3.5, except for the surface reflectance, each term in the right hand represents the aerosol constribution in the cloud free conditions. Therefore, the global aerosol can be described by a set of aerosol types and loading whose conditions are pre-calculated and stored in LUT.

The algorithm takes observations of MODIS spectral reflectances to retrieval aerosol properties (AOT at 0.55 μm, Fine model Weighting at 0.55 μm (FW or $\eta_{0.55}$), and the surface reflectance at 2.1 μm (ρ^s_{212}). Similar to the over-ocean aerosol algorithm, the land algorithm assumes that one fine-dominated aerosol model and one coarse-dominate aerosol model can be combined with proper weightings to represent the ambient aerosol properties over the target. Using LUT, the algorithm determines the conditions that best similarity to the MODIS-observed spectral reflectance ρ^m_λ, that is the solution to the inversion. The flowchart of the over-land aerosol algorithm is presented in Figure 3.3.

2 .2.1 Selection of Pixels

The spectral data include the 0.66 and 0.86 μm channels (MODIS channels 1 and 2 at 250 m resolution), the 0.47, 0.55, 1.2, 1.6 and 2.1 μm channels (channels 3, 4, 5, 6 and 7 at 500 m), and the 1.38 m channel (channel 26 at 1 km). The geo-location data are at 1 km and include angles, latitude, longitude, elevation and date. The L1B reflectance values are corrected for water vapor, ozone, and carbon dioxide before proceeding. The measured reflectances are organized into 10-km boxes of 20x20 or 40x40 pixels, depending on the channel.

All pixels in the considered box are evaluated pixel by pixel to identify whether the pixel is cloudy, snow/ice, or water. The land algorithm is applied to coastal boxes that include both land and water pixels but retrieval aerosol quality will decrease. The standard MODIS cloud mask product (MOD/MYD35) provide all masking information. Besides, spatial variability cloud mask was added to remove thin and cirrus cloudy pixels. Enhanced snow/ice mask using a window of 8 contiguous pixels to label the center pixel as snow/ice if all neighbors are

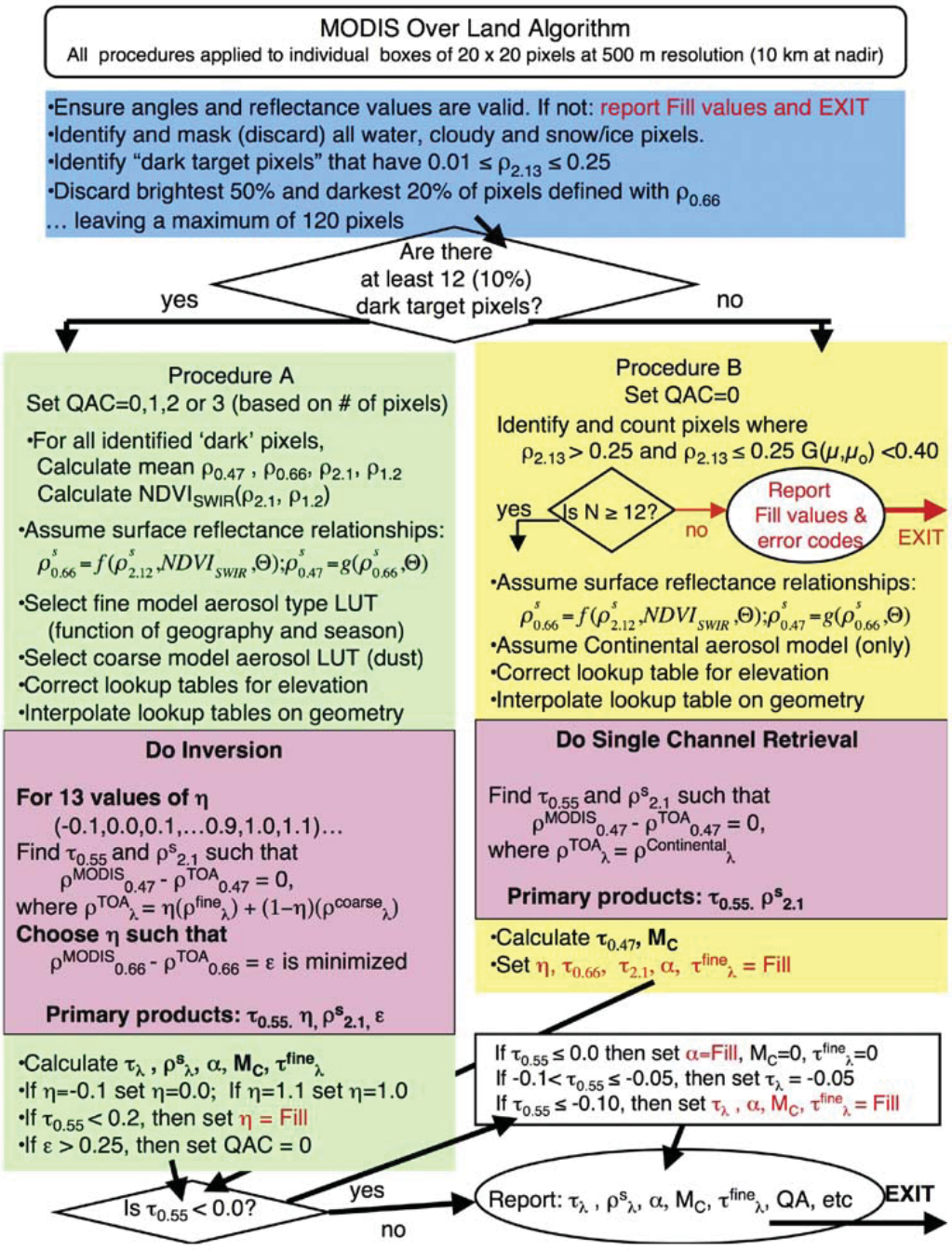


Figure 3.3: Flowchart of the over-land aerosol retrieval algorithm (Remer et al. [2004])

snow/ice. The subpixel water is identified, in the second time, if its Normalized Difference Vegetation Index (NDVI) is smaller than the threshold 0.10.

The over land aerosol algorithm follows Dense Dark Vegetation (DDV) approach that applies aerosol retrievals over "dark pixel". Dark pixels are selected using their reflectance at 2.13 μm ($0.01 \leq \rho_{2.13} \leq 0.25$). The pixels remained after masking and and dark target selection are sorted in term of their visible reflectance $\rho_{0.66}$. The pixels with the darkest 20% and brightest 50% of $\rho_{0.66}$ are discarded in order to remove pixels possibly contaminated by cloud shadows, odd surfaces at the dark or bright end. The remaining 30% of the pixels will be used in the procedure A if there are at least 12 of these pixels remaining from the original 400 in the 20x20 box. The mean measured reflectance is calculated from these selected pixels in four wavelengths ($\rho_{0.47}$, $\rho_{0.66}$, $\rho_{2.1}$, $\rho_{1.2}$). Depending on whether the number of dark target pixels is bigger than 12, Procedure A, the inversion for dark surfaces, or Procedure B, alternative retrieval for brighter surface, will be selected.

2 .2.2 VIS/SWIR surface reflectance assumptions

Over land, the surface reflectance at $\rho^s_{0.66}$ and $\rho^s_{0.47}$ are derived using VIS/SWIR relationship which depends on the scattering angle Θ and the Normalized Difference Vegetation Index at SWIR $NDVI_{SWIR}$ defined in Equation 3.6 and Equation 3.7.

$$\Theta = \cos^{-1}(-\cos\theta_0 \cos\theta + \sin\theta_0 \sin\theta \cos\phi) \tag{3.6}$$

where θ_0, θ, and ϕ are the solar zenith, sensor view zenith, and relative azimuth angles, respectively.

$$NDVI_{SWIR} = (\rho^m_{1.24} - \rho^m_{2.12})/(\rho^m_{1.24} + \rho^m_{2.12}) \tag{3.7}$$

where $\rho_{1.24}$ and $\rho_{2.12}$ are the MODIS measured reflectances of the 1.24 and 2.1 μm, respectively.

The surface reflectance at 0.66 μm are derived directly from 2.12 μm, while 0.47 μm surface reflectance is estimated indirectly from 0.66 μm because the relationship of 0.47 to 0.66 μm is stronger than to 2.12 μm.

$$\rho^s_{0.66} = f(\rho^s_{2.12}) = \rho^s_{2.12} * a_{0.66/2.12} + b_{0.66/2.12}$$
$$\rho^s_{0.47} = g(\rho^s_{0.66}) = \rho^s_{0.66} * a_{0.47/0.66} + b_{0.47/0.66} \tag{3.8}$$

where

$$\begin{aligned} a_{0.66/2.12} &= a^{NDVI_{SWIR}}_{0.66/2.12} + 0.002\Theta = 0.27, \\ b_{0.66/2.12} &= 0.00025\Theta + 0.033, \\ a_{0.47/0.66} &= 0.49, and \\ b_{0.47/0.66} &= 0.005 \end{aligned} \tag{3.9}$$

where in turn

$$\begin{aligned} a^{NDVI_{SWIR}}_{0.66/2.12} &= 0.48; NDVI_{SWIR} < 0.25, \\ a^{NDVI_{SWIR}}_{0.66/2.12} &= 0.58; NDVI_{SWIR} < 0.75, \\ a^{NDVI_{SWIR}}_{0.66/2.12} &= 0.48 + 0.2(NDVI_{SWIR} - 0.25); 0.25 \leq NDVI_{SWIR} \leq 0.75 \end{aligned} \tag{3.10}$$

2 .2.3 Aerosol Models and LUT

The AOT at nearly 14,000 AERONET sky retrievals (both spherical and spheroid assumptions) satisfying AERONET team recommended threshold of $\tau_{0.44} > 0.4$ are collected. Retrievals are based on assumption that spheres generally described sites dominated by fine (radius < 0.6 μm) aerosols, whereas spheroids generally represented sites dominated by coarse (radius > 0.6 μm) aerosols. The cluster analysis upon the spherical retrievals are performed, hence three aerosol types representing the global fine-dominated aerosol regimes were pointed out. These are a "nonabsorbing" aerosol model (Single Scattering Albedo - SSA or $\omega_0 \sim 0.95$) corresponding to urban/industrial aerosol in the industrialized northern hemisphere, and "absorbing" aerosol model ($\omega_0 \sim 0.85$) found in sooty and/or savanna-burning regions of South America and Africa, and a "moderately absorbing" aerosol model representative of biomass burning and incomplete fossil fuel burning in the developing world. Performed in a similar way, analysis of spheroid retrievals showed that a single model represented global dust aerosol. Each aerosol "model" is comprised of two lognormal modes, either dominated by the fine mode (the three spherical models) or the coarse mode (the spheroid model).

The average aerosol properties of each aerosol type were used to calculate

scattering/extinction properties using a Mie code or T-matrix code that were integrated over size distribution. From these properties, the spectral dependence of τ and phase functions were characterized.

Based on dominant aerosol type resulted from clustering analysis, an aerosol type was assigned to each AERONET site (as a function of season). This information was extrapolated to all regions and mapped onto a 1^o longitude x 1^o latitude grid such that a fine aerosol type is assigned for each grid point, globally.

The algorithm over land performs an inversion of three channels (0.47, 0.66, and 2.12 μm) to retrieve τ, η, and the surface reflectance. The inversion process is based on LUT. In the C005 algorithm, the LUT is indexed in relation to the channel 0.55 μm and computed at the four central wavelengths (0.466, 0.553, 0.644, and 2.119 μm). The aerosol model-dependent parameters of equation 3.5 are calculated for several values of aerosol total loadings and for a variety of geometry. The LUT represent spherical aerosol models (Continental, moderately absorbing, absorbing, and nonabsorbing) and the one spheroid model (dust).

The scattering and reflectance parameters are calculated for seven aerosol loadings ($\tau_{0.55}$ = 0.0, 0.25, 0.5, 1.0, 2.0, 3.0, and 5.0). TOA reflectance is calculated for nine solar zenith angles (θ_0 = 0.0, 6.0, 12.0, 24.0, 36.0, 48.0, 54.0, 60.0, and 66.0), 16 sensor zenith angles (θ = 0.0 to 65.8, approximate increments of 6.0), and 16 relative azimuth angles (ϕ = 0.0 to 180.0 increments of 12.0). All of these parameters are calculated assuming a surface reflectance of zero.

When surface reflectance is present, the second term in equation 3.5 is nonzero. The flux is a function of the atmosphere, while the atmospheric backscattering term s and the transmission term T are functions of both atmosphere and the surface. The radiative transfer code is run two additional times with two distinct positive values of surface reflectance to calculate s and T. Then, these values of F_d, s, and T are included within the LUT for each τ index, wavelength, and aerosol model.

2 .2.4 Inversion for dark surfaces and brighter surfaces

The algorithm assigns the fine aerosol mode using location and time as mentioned above. From the LUT, ρ^a, F, T and s (for the fine and coarse mode) are

interpolated for angle, resulting in six values for each parameter, corresponding to aerosol loading. For different values of FW between -0.1 and 1.1 (intervals of 0.1), the algorithm attempts to find the AOT at 0.55 μm and the surface reflectance at 2.12 μm matching the MODIS measured reflectance at 0.47 μm. The solution is the one where the error ϵ at 0.66 μm is minimized. In formula,

$$\begin{aligned} \rho^m_{0.47} - \rho^*_{0.47} &= 0 \\ \rho^m_{0.66} - \rho^*_{0.66} &= \epsilon \\ \rho^m_{2.12} - \rho^*_{2.12} &= 0 \end{aligned} \tag{3.11}$$

where

$$\rho^*_\lambda = \eta(p^{fa}_\lambda + F^f_{d,\lambda} T^f_\lambda \rho^f_\lambda/(1 - s^f_\lambda \rho^s_\lambda)) + (1-\eta)(\rho^{ca}_\lambda + F^c_{d,\lambda} T^c_\lambda \rho^c_\lambda/(1 - s^c_\lambda \rho^s_\lambda)) \tag{3.12}$$

where in turn, λ = 2.12, 0.66 and 0.47 μm, $\rho^a = \rho^a(\tau)$, $F = F(\tau)$, $T = T(\tau)$, $s = s(\tau)$ are functions of τ indices in the LUT, and $\rho^s_{0.66}$, $\rho^s_{0.47}$ are described in Equation 3.8. The primary products are AOT ($\tau_{0.55}$), FW ($\eta_{0.55}$), and the surface reflectance ($\rho^s_{2.12}$), and the error ϵ.

The Procedure B is applied to derived AOT at pixels having 2.12 μm reflectance brighter than 0.25. The Continental aerosol is assumed, therefore the LUT reflectance is calculated using one aerosol mode with $\eta = 1.0$. The primary products for Procedure B are AOT ($\tau_{0.55}$) and the surface reflectance ($\rho^s_{2.12}$)

2 .2.5 Derivation of Fine Mode AOD, Mass Concentration and other secondary parameters

After inversion processes, besides primary products ($\tau_{0.55}$, $\eta_{0.55}$, and $\rho^s_{2.12}$), the secondary products can be calculated. These include the fine and coarse model optical depths ($\tau^f_{0.55}$, $\tau^c_{0.55}$), mass concentration M, the spectral total, fine, and coarse model optical thickness (τ_λ, τ^f_λ, τ^c_λ) and Ångström exponent α. They are defined as:

$$\begin{aligned} \tau^f_{0.55} &= \tau_{0.55}\eta_{0.55} \\ \tau^c_{0.55} &= \tau_{0.55}(1 - \eta_{0.55}) \end{aligned} \tag{3.13}$$

$$M = M_c^f \tau_{0.55}^f + M_c^c \tau_{0.55}^c \tag{3.14}$$

where M_c^f and M_c^c are mass concentration coefficients for the fine and coarse models.

$$\begin{aligned} \tau_\lambda &= \tau_\lambda^f + \tau_\lambda^c \\ \tau_\lambda^f &= \tau_{0.55}^f (Q_\lambda^f / Q_{0.55}^f) \\ \tau_\lambda^c &= \tau_{0.55}^c (Q_\lambda^c / Q_{0.55}^c) \end{aligned} \tag{3.15}$$

where Q_λ^f and Q_λ^c represent model extinction coefficients at wavelength λ

$$\alpha = ln(\tau_{0.47}/\tau_{0.66})/ln(0.466/0.644) \tag{3.16}$$

2 .2.6 Low and Negative Optical Depth Retrievals

The C005 algorithm allows negative τ retrievals. As result of positive and negative noise in the MODIS observations, the underestimation and overestimation of surface reflectance and aerosol properties, retrievals of negative τ is statistically imperative in order to avoid bias. However, a large negative retrieval is not acceptable. The strategy for negative values of τ is presented in Figure 3.2

3 Aerosol Retrieval Using Support Vector Regression

The basic idea underlying the data analysis approach followed by SVR is to use a set of preliminary data, characterized by already known target values, to derive regression criteria to be applied on a new set of items. In the prediction procedure applied to new datasets, prediction values are assigned to them conforming to the analogies with the preliminary known data. In the last decade numerous applications of the SVR methods have been made in several domains of chemistry and physics, among the latter ones several are in the domain of signal processing in which statistical methods are well suited. In the following, after recalling briefly the mathematical basis of the method, we will explain how to apply the

SVR technique to retrieve AOT information from MODIS observations.

3 .1 Basic concepts of Support Vector Regression

SVR was proposed using the idea of Support Vector Machine (SVM) developed for classification application. A SVM constructs hyperplanes in a high- or infinite-dimensional space to separate training datasets. The optimal hyperplanes will have the largest distance to the nearest training data point of any class (i.e. maximum margin hyperplanes), which reduces the generalization error of the classifier. In the similar way, SVR constructs maximum margin hyperplanes in a high- or infinite- dimensional space for the regression purpose. The optimization problem for hyperplanes can be solved more easily in its dual formulation, therefore a standard dualization method utilizing Langrange multipliers is applied. This method provides a Support Vector expansion in which hyperplanes are described by a linear combination of a specific subset of the training patterns called Support Vectors (SVs). The nonlinear SVR is achieved by simply preprocessing the training patterns by a map Φ into some high dimension space called feature spaces F using kernel functions $k(x_i, x)$ and then linear SVR algorithm can be applied. In the next sections, the mathematics and solution for ϵ-SVR are presented, together with brief description of its implementation in practice.

3 .1.1 The framework

Given training data $\{(x_1, y_1), \ldots, (x_l, y_l)\} \subset X$ where X denotes the space of the input patterns (i.e. $X \subset \Re^d$). The $\epsilon - SVR$ (Vapnik [1995]) is to find a function $f(x)$ that has at most ϵ deviation from the actually obtained target y_i from the training data and is as 'flat' as possible in order to minimize the expected risk. In the case of linear function $f(x)$, it is taken in the form:

$$f(x) = \langle w, x \rangle + b \quad with \quad w \subset X, b \in \Re \tag{3.17}$$

where $\langle ., . \rangle$ denotes the dot product in X. In the case of 3.17, *flatness* is to find the function $f(x)$ that presents an optimal regression hyperplane with minimum

w. The problem can be written as a convex optimization problem:

$$\begin{aligned} &minimize \quad \frac{1}{2}\|w\|^2 \\ &subject \quad to \quad \begin{cases} y_i - \langle w, x\rangle - b \leq \epsilon \\ \langle w, x\rangle + b - y_i \leq \epsilon \end{cases} \end{aligned} \tag{3.18}$$

The assumption in 3.18 is that a function f actually exists and approximates all pairs (x_i, y_i) with ϵ precision. The convex optimization problem is feasible. However, in the case of infeasible constraints of the optimization problem or allowing of some errors, the slack variables ξ_i, ξ_i^* are introduced. Then, the problem can be formulated as follow:

$$\begin{aligned} &minimize \quad \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{l}(\xi_i + \xi_i^*) \\ &subject \quad to \quad \begin{cases} y_i - \langle w, x\rangle - b \leq \epsilon + \xi_i \\ \langle w, x\rangle + b - y_i \leq \epsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0 \end{cases} \end{aligned} \tag{3.19}$$

The constant $C > 0$ determines the trade-off between the flatness of f and the amount up to which deviations larger than ϵ are tolerated. This is corresponding to ϵ-insensitive loss function $|\xi|_\epsilon$ described by

$$|\xi|_\epsilon := \begin{cases} 0 & if \quad |\xi| \leq \epsilon \\ |\xi| - \epsilon & otherwise \end{cases} \tag{3.20}$$

The situation is presented graphically in Figure 3.4. Only the points outside the shaded tube are penalized and contribute to the cost function.

3 .1.2 The dual problem and quadratic program

The SVR problem can be solved by classical Lagrangian optimization techniques. A Lagrangian function L is constructed from objective function and corresponding constraints in 3.19 by introducing a dual set of variables. The solution is to find a saddle point which minimizes L with respect to the primal variables w and b, and maximizes L with respect to the dual variables. This is known as *the Lagrangian*

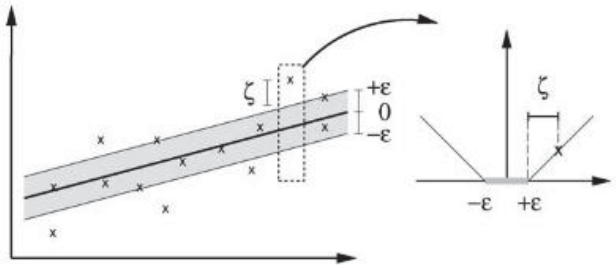


Figure 3.4: The soft margin loss setting for a linear SVM (Scholkopf and Smola [2002])

primal problem. The Lagrangian function is constructed and formulated as follow:

$$\begin{aligned} L :=& \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{l}(\xi_i + \xi_i^*) - \sum_{i=1}^{l}(\eta_i\xi_i + \eta_i^*\xi_i^*) \\ & - \sum_{i=1}^{l}\alpha_i(\epsilon + \xi_i - y_i + \langle w, x_i\rangle + b) \\ & - \sum_{i=1}^{l}\alpha_i^*(\epsilon + \xi_i^* + y_i - \langle w, x_i\rangle - b) \end{aligned} \tag{3.21}$$

where η_i, η_i^*, α_i, α_i^* are Lagrangian multipliers. The dual variables in 3.21 have to satisfy positivity constraints

$$\alpha_i^*, \eta_i^* \geq 0 \tag{3.22}$$

in which α_i^* refers to α_i and α_i^*.

From the saddle point condition, the partial derivatives of L with respect to the primal variables (w, b, ξ_i, ξ_i^*) are set to zero for the condition of optimality.

$$\partial_b L = \sum_{i=1}^{l}(\alpha_i^* - \alpha_i) = 0 \tag{3.23}$$

$$\partial_w L = w - \sum_{i=1}^{l}(\alpha_i - \alpha_i^*)x_i = 0 \tag{3.24}$$

$$\partial_{\xi_i^*} L = C - \alpha_i^{(*)} - \eta_i^{(*)} = 0 \tag{3.25}$$

Substituting 3.23, 3.24, and 3.25 into 3.21 yields the dual optimization problem known as *the Lagrangian dual problem*. The primal problem of finding a saddle point for $L(w, b)$ is transformed into the easier one of maximizing $L(\alpha_i^*)$ that only depends on Lagrange multipliers.

$$\begin{aligned} &maximize \quad \begin{cases} -\frac{1}{2}\sum_{i,j=1}^{l}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)\langle x_i, x_j\rangle \\ -\epsilon\sum_{i=1}^{l}(\alpha_i + \alpha_i^*) + \sum_{i=1}^{l} y_i(\alpha_i - \alpha_i^*) \end{cases} \\ &subject \quad to \quad \sum_{i=1}^{l}(\alpha_i - \alpha_i^*) = 0 \quad and \quad \alpha_i, \alpha_i^* \in [0, C] \end{aligned} \tag{3.26}$$

In the formula 3.26, the dual variables ξ_i, ξ_i^* are eliminated through the condition 3.25. From the Equation 3.24, the w can be calculated as

$$w = \sum_{i=1}^{l}(\alpha_i - \alpha_i^*)x_i \quad thus \quad f(x) = \sum_{i=1}^{l}(\alpha_i - \alpha_i^*)\langle x_i, x\rangle + b \tag{3.27}$$

The w is described as a linear combination of the training pattern x_i, therefore the complexity of the function's representation by SVs is independent of the dimensionality of the input space X, and depends only on the number of SVs.

The Karush-Kuhn-Tucker (KKT) conditions (Karush [1939]; Kuhn and Tucker [1959]) are used to calculate b. Following these, at the point of solution the product between dual variables and constraints has to be zero.

$$\begin{aligned} \alpha_i(\epsilon + \xi_i - y_i + \langle w, x_i\rangle + b) = 0 \\ \alpha_i^*(\epsilon + \xi_i^* - y_i + \langle w, x_i\rangle + b) = 0 \end{aligned} \tag{3.28}$$

and

$$\begin{aligned} (C - \alpha_i)\xi_i = 0 \\ (C - \alpha_i^*)\xi_i^* = 0 \end{aligned} \tag{3.29}$$

Therefore,

$$\epsilon - y_i + \langle w, x_i\rangle + b \leq 0 \quad and \quad \xi_i = 0 \quad if \quad \alpha_i < C \tag{3.30}$$

$$\epsilon - y_i + \langle w, x_i \rangle + b \geq 0 \quad if \quad \alpha_i > 0 \tag{3.31}$$

In conjunction with an analogous analysis on α_i^*, b is identified with the following constrain

$$\begin{aligned} &max\{-\epsilon + y_i - \langle w, x_i \rangle | \alpha_i < C \quad or \quad \alpha_i^* > 0\} \leq b \leq \\ &min\{-\epsilon + y_i - \langle w, x_i \rangle | \alpha_i > 0 \quad or \quad \alpha_i^* < C\} \end{aligned} \tag{3.32}$$

In summary, the solution for the optimal hyperplanes (w, b) can be given by Lagrangian approach (3.27 and 3.32). The hyperplanes are described by training pattern x_i where dual variables α_i^* are not zero. Those examples are called Support Vector as illustrated in Figure 3.5. The expansion of w in term of x_i is sparse, which means we do not need all x_i to describe w.

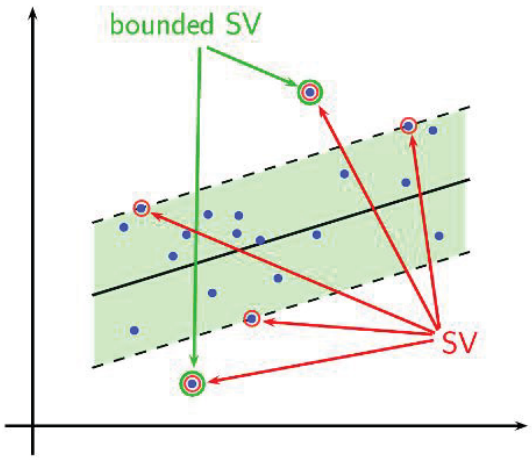


Figure 3.5: SVs and bounded SVs

3 .1.3 Non-linear Support Vector Regression and Kernel Method

The SV algorithm nonlinear is based on the principle of the Cover's theorem for separability of patterns which states that a classification problem cast in high-dimensional nonlinear space is more likely to be linearly separable. Therefore, non-linear SVR, similar to nonlinear SVM, is designed to operate in two stages:

- performance of a map $\Phi : X \to F$ for the feature vector x_i from input space X into a high dimensional feature space F.
- construction of optimal separating hyperplanes in the high-dimensional space.

As described in the previous section, the SV algorithm only depends on dot product between pattern x_i. Hence, the operation in high dimensional space do not have to be performed explicitly if a function $k(x, x')$ is found such that $k(x, x') := \langle \Phi(x), \Phi(x') \rangle$. $k(x, x' \rangle$ is called a kernel function. The SV optimization problem in 3.26 is restated as follows:

$$\begin{aligned} &maximize && \begin{cases} -\frac{1}{2}\sum_{i,j=1}^{l}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)k(x_i, x_j) \\ -\epsilon\sum_{i=1}^{l}(\alpha_i + \alpha_i^*) + \sum_{i=1}^{l} y_i(\alpha_i - \alpha_i^*) \end{cases} \\ &subject \quad to && \sum_{i=1}^{l}(\alpha_i - \alpha_i^*) = 0 \quad and \quad \alpha_i, \alpha_i^* \in [0, C] \end{aligned} \tag{3.33}$$

Likewise the expansion of f in 3.27 can be written as

$$w = \sum_{i=1}^{l}(\alpha_i - \alpha_i^*)\Phi(x_i) \quad and \quad f(x) = \sum_{i=1}^{l}(\alpha_i - \alpha_i^*)k(x_i, x) + b \tag{3.34}$$

Some common kernel functions are listed here:

- The homogeneous polynomial kernels k with $p \in N$: $k(x, x') = \langle x, x' \rangle^p$
- The inhomogeneous polynomial kernels k with $p \in N, c \geq 0$: $k(x, x') = (\langle x, x' \rangle + c)^p$
- The hyperbolic tangent kernels k with some $\kappa > 0$ and $c < 0$: $k(x, x') = tanh(\kappa\langle x, x' \rangle + c)$
- The Gaussian radial basis function kernels k with σ: $k(x, x') = e^{-\frac{\|x - x'\|^2}{2\sigma^2}}$

3 .1.4 Implementation

The convex programming algorithms described above can be used directly on moderately sized sample datasets (up to 3000). However, it is difficult on larger

datasets because of computer's memory and CPU limitation in computation and storage of the dot product matrix $k(x_i, x_j)$. Therefore, the implementation requires specific techniques, mostly relying on heuristics for breaking the big problem down into smaller chucks.

The first solution was introduced by Vapnik [1982]. The method is to start with an arbitrary subset (a chunk) that fits into memory, train the SVR algorithm on it, keep the SVs and fill the chuck up with data whom the current estimator would make errors on (i.e. data lying outside the ϵ-tube of the current regression). The system is retrained and the iteration is continued until all KKT-conditions are satisfied. This chucking algorithm just postponed the underlying problem of dealing with large datasets but this problem is not completely avoided. A variation of this solution was proposed to use only a subset of the variables as a working set and optimize the problem with respect to them while freezing the other variables (Osuna et al. [1997]).

Recently, the Sequential Minimal Optimization (SMO) algorithm was proposed in which subsets with size of 2 were selected iteratively and the target function was optimized with respect to them (Platt [1999]). The key point is that for a working dataset of 2, the optimization subproblem can be solved analytically without explicitly invoking a quadratic optimizer. This method has been reported to have good convergence properties and easy implementation.

3 .2 Application of Support Vector Regression in Aerosol Optical Thickness Retrievals

In application of AOT retrievals, SVR is applied in the inversion process to construct models from training datasets and then to use them for predicting AOT information of new datasets. The SVR models represent the relationship between satellite observations and AOT information. Other procedures such as cloud screening, pixel selection, or map prediction application need to be carried out independently. The SVR application in AOT retrievals can be summarized in four main steps:

- Collection of data covering interested areas for a long period in order to construct SVR data models for future prediction. Data can be obtained

from simulation systems or measurements by ground/satellite-based instruments. Depending on application, data often consist of reflectances/radiances, geolocation, meteorological factors, and target AOT information. The selected data should have strong correlations with target information.

- Processing data by application of certain screening processes (e.g. cloud screening, cloud tests, glint mask, sediment mask, etc) to select appropriate pixels, integration to combine data recorded by different instruments, and feature extraction to design input for the regression. This step aims at creating sample datasets $\{(x_1, y_1), \ldots, (x_l, y_l)\}$, in which x_i refer to as data features and y_i represent target AOTs, for the SVR algorithms.

- Construction of SVR models by application of the SVR algorithm on training datasets. It includes choosing SVR types, implementation of the SVR algorithm, setting SVR configurations for selected training datasets, and validation of developed SVR models.

- Application of SVR constructed models for prediction of AOT information on new datasets.

4 Summary

The multi-spectral sensor MODIS on-board the two polar orbiting satellites Terra and Aqua provides observations nearly the entire globe on a daily basis. The MODIS performs measurements in the solar to thermal infrared spectrum region from 0.41 to 14.235 μm, separated into 36 bands at resolutions 1 km (29 channels), 500 m (5 channels) and 250 m (2 channels) at nadir (Salomonson et al. [1989]). MODIS observations are divided into granules each of which covers an area of 2030×1354 km^2. Various data products derived from MODIS measurements are provided (e.g. MODIS L1B data, MODIS L2 products, MOD04 L2).

The global system of ground-based Remote Sensing aerosol network AERONET provides AOT information, inversion products, and precipitable water in diverse aerosol regimes. AOT is measured in four or more wavelengths to include 0.440, 0.670, 0.870, and 1.020 μm, in every 15 minutes during midday and more often

during sunrise and sunset. AERONET AOT products have three data quality levels: Level 1.0 (unscreened), Level 1.5 (cloud-screened), and Level 2.0 (cloud-screened and quality-assured).

Following model driven approach, "algorithm for remote sensing of tropospheric aerosol from MODIS" (Kaufman and Tanré [1997]; Remer et al. [2004]) is applied to derive the standard MODIS aerosol products. The algorithms are designed separately for land and ocean. Different aerosol models were simulated and their calculated parameters were stored in LUT. The algorithms assume that the aerosol properties over targeted areas were presented by proper weightings of one fine-dominated and one coarse-dominated aerosol models. Spectral reflectance from the LUT is compared with MODIS measured spectral reflectance to find the best match that is the solution to the inversion process. Beside the LUT approach that is considered as the core technique, ancillary data and many different screening processes played an important role in this methodology. Those data and processing techniques were considered and applied on the original MODIS datasets to select appropriate data for inversion process.

Conversely, data driven approach for AOT retrieval from MODIS is carried out by machine learning technique such as SVR. The basic idea underlying the data analysis approach followed by SVR is to use a set of preliminary data, characterized by already known target values, to derive regression criteria to be applied on a new set of items. In the prediction procedure applied to new datasets, prediction values are assigned to them conforming to the analogies with the preliminary known data. The linear SVR technology constructs maximum margin hyperplanes in a high- or infinite- dimensional space for the regression purpose. The nonlinear SVR is achieved by preprocessing the training patterns by a map into some high dimension space called "feature space" using kernel functions and then linear SVR algorithm can be applied. In AOT retrieval application, SVR plays a role as radiative transfer model in algorithms based on model driven approach.

In the next chapters, our work based on standard methodologies mentioned above will be presented. Downscaling techniques using an adapted MODIS aerosol algorithms and its validation will be described in the next chapter and then it will be followed by methodologies for AOT retrieval at 10×10 and 1×1

km^2 using SVR.

Chapter 4

Downscaling Spatial Resolution of Aerosol Optical Thickness from 10×10 km^2 to 1×1 km^2 using adapted MODIS aerosol algorithms

This chapter presents the improvements of the MODIS aerosol algorithm to derive AOT with spatial resolution at 1×1 km^2 which is then used to develop a software package called PM MAPPER (Campalani et al. [2011]; MEEO [2010a]). This algorithm performs at global scale instead of parameterizing for specific regions. The algorithm is an extension version of former methodology applied to derive MODIS AOT maps at 3×3 km^2 (Nguyen et al. [2010a]; Nguyen et al. [2010b]). The validation was carried out on data covering European areas from 2007 to 2009.

1 Methodology

The proposed idea is simple and straightforward in which the MODIS aerosol algorithms described in Chapter 2 were applied directly on smaller data boxes in

order to obtain AOT information at increased spatial resolutions. Besides, the first cloud masking process was carried out by a software program called SOIL MAPPER instead of the MODIS cloud mask in MOD35/MYD35 products.

1 .1 Selection of pixels by Land Cover Classification

The recently developed SOIL MAPPER is a fully automatic, unsupervised, software package which, from the analysis of multispectral optical satellites data, allows to generate Land Cover (LC) classification maps. The algorithm is based on spectral fuzzy rule-based per-pixel classification method, originally presented and discussed in Baraldi et al. [2006], consisting of two levels of processing. Firstly, input reflectances are used to calculate features including Brightness (Bright), Normalized Difference Vegetation Index (NDVI), Normalized Different Bare Soil Index (NDBSI), Normalized Difference Snow Index (NDSI). After that, linguistic labels (low, middle, high) are assigned to fuzzy sets providing a complete partition of feature space consisting of the scalar variables mentioned above. In this step, kernel spectral rules are computed flatly in order to group data into different raw clusters. The second step is based on a built-in hierarchy of values of spectral rules and feature fuzzy sets to stratify data to a list of kernel spectral categories. Based on this list, we are able to determine a set of 57 different classes, out of which 40 refer to different land types, from dense vegetation to bare soil, and the remaining classes refer to cloud, water, ice, snow etc. Most recent upgrading of the SOIL MAPPER consisted in extending its capabilities to generate standardized outputs from various optical satellite sensors, thus allowing comparison of land classifications deduced from different satellite-flown sensor images. When applied to MODIS data, the SOIL MAPPER software uses reflectances recorded in seven wavelengths (0.66, 0.87, 0.47, 0.55, 1.64, 11.03, and 12.02 μm) as input to process land cover maps which identify land, water, and cloudy-free pixels in order to apply aerosol retrieval algorithms, and provide land surface information.

1 .2 Modification of the MODIS Aerosol Algorithm

In order to retrieve AOT concentration maps at higher spatial resolution than 10×10 km^2, the original aerosol algorithms over both land and ocean have been

modified in three points. The first is a major change while the second and third are minor changes.

- The algorithms are applied on smaller boxes sized 1×1 and 2×2 pixels of the MODIS reflectances at 1 km and 500 m resolution respectively, instead of those sized 10×10 and 20×20, as presented in the original MODIS aerosol algorithm. Consistently with this modification, we adapt all thresholds by the linear reduction. One of the important thresholds used by the MODIS algorithm over land is the number of good pixels in a box which must be identified before applying the aerosol derivation procedure. Good pixels are used to estimate aerosol and they are identified when 50% of brightest and 20% of darkest pixels of the total cloud-free pixels in a box are discarded. This procedure is intended to eliminate noise and bright pixels that are not suitable to derive aerosol. In the original method of elaboration of the MODIS data, the threshold number is 12 pixels for each box sized 20×20 pixels at 500 m resolution. In our approach, with smaller boxes sized 2×2 pixels for the case of 1×1 km^2 of resolution, the new threshold for a box is 1 pixel which may decrease the quality of the aerosol retrieval in later steps.

- The second cloud masking over land and ocean for MODIS algorithms is based on the methodology proposed by Martins et al. [2002]. In our approach, the cloud mask is calculated over whole map instead of for each block sized 20×20 500-m pixels as used in the MODIS algorithm. This adaptation will provide a homogeneous cloud mask which is convenient to process data to obtain AOT maps with a fine spatial resolution.

- The condition to apply ocean algorithm or land algorithm on boxes is changed. In the original version, if all pixels in the considered box belong to the water class, then the ocean algorithm is applied, otherwise land algorithm is performed. Our work uses a condition that if the land aerosol derivation procedure fails at the first time but the percentage of the water pixels in the box is bigger than 50%, then the aerosol algorithm over ocean is applied again. This improvement focuses on retrieving more information along coastlines when the methodology was firstly applied to obtain AOT at 3×3 km^2 (Nguyen et al. [2010a]; Nguyen et al. [2010b]). However with

small processing data boxes (i.e. a box sized 1×1 km^2), this rule is less important.

2 Implementation and PM MAPPER application

The software of the MODIS aerosol algorithms over land and ocean at 10×10 km^2 of resolution are implemented and provided in the International MODIS/AIRS Processing Package (IMAPP) by University of Wisconsin-Madison (CIMSS [2011]). The IMAPP MODIS Level 2 software package provides a complete atmosphere processing system processing MODIS data acquired by direct broadcast from Terra and Aqua spacecrafts and deriving cloud mask (MOD35/MYD35), cloud top properties and cloud phase (MOD06CT/MYD06CT), atmospheric profiles (MOD07/MYD07), aerosol (MOD04/MYD04), sea surface temperatures and nea-infrared water vapor. Regarding to aerosol retrievals, the following main modules are involved

- The IMAPP MODIS flat file extractors (FLATFILE) create binary flat files and header files in ENVI readable format from input L1B 1km, 500m, 250m and geolocation MODIS Direct Broadcast or DAAC formatted HDF files.
- The cloud mask module (CLOUDMASK) creates the MODIS cloud mask 48 bit array for each pixel in a given scene.
- The IMAPP MODIS aerosol module (AEROSOL) creates the MODIS aerosol product, which consists of these 6 arrays at a resolution of 10×10 1 km MODIS pixels for daytime only scenes.
- The IMAPP MODIS binary to HDF file module (BINTOHDF) converts the flat binary files that are the standard IMAPP output format into HDF files for the aerosol products.

The required platforms are g77 gnu (gcc 3.2.1) compiler and Linux operating systems. In addition, the aerosol module refers to a set of static files describing LUT and ancillary data of NCEP GDAS1 (i.e. model profiles of temperature,

humidity, etc) and TOVS/TOAST ozone files. Most of main modules are written in Fortran programming language and run in sequence by bash script programs. This release requires $\sim$ 2.7 GB of disk space for the software, ancillary data files and the input/output test datasets.

The IMAPP MODIS software package were re-designed and modified in order to obtain AOT products at 1×1 km^2, which is called the PM MAPPER software package. The cloud mask module is replaced by the SOIL MAPPER software. Other modules are modified to work on data boxes sized 1×1 km^2. The modification of spatial variability cloud mask and aerosol retrieval conditions are carried out in the aerosol module. Figure 4.1 presents the flowchart of MODIS aerosol algorithm in comparison with the PM MAPPER application. Meanwhile, an example of the MODIS AOT map at 10×10 km^2 and corresponding PM MAPPER AOT map at 1×1 km^2 is illustrated in Figure 4.2.

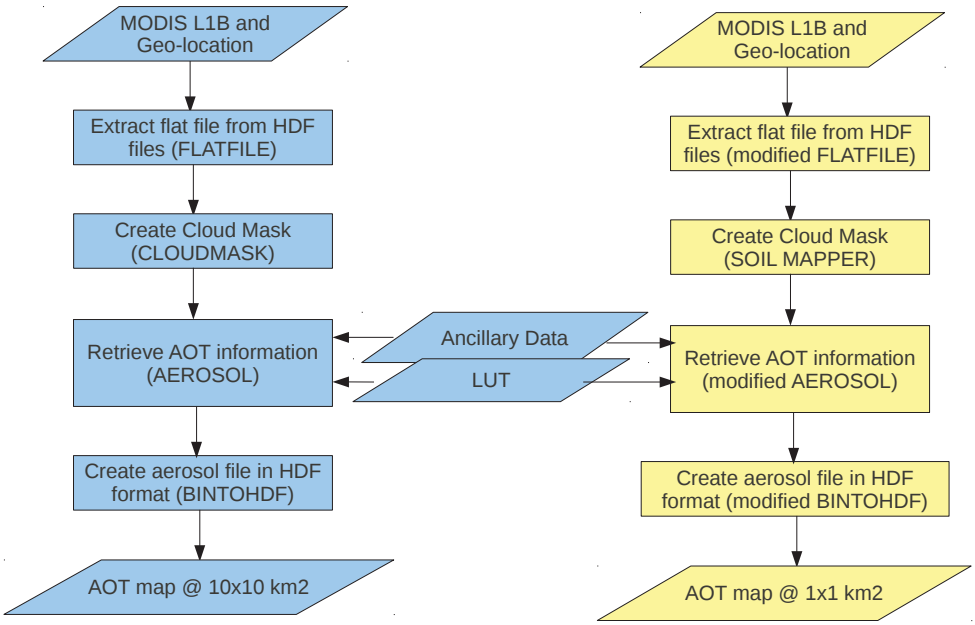
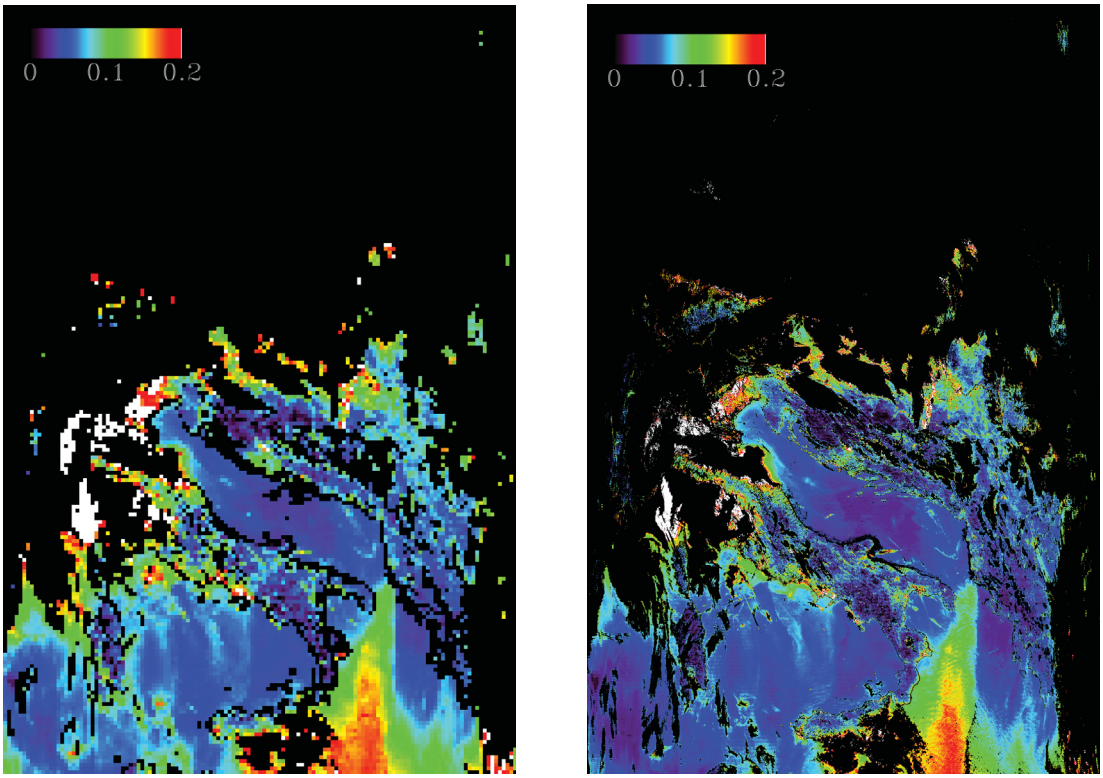


Figure 4.1: The flowchart for aerosol retrieval of the IMAPP MODIS package (left) and the PM MAPPER software (right).



(a) (b)

Figure 4.2: (a) The MOD02 L2 AOT map at 10×10 km^2 and (b) The PM MAPPER map at 1×1 km^2 of MODIS image acquired on January 29, 2008 at 09:55 UTC.

3 Validations

In this section the validation results of the 1-km AOT retrievals of PM MAPPER which takes MODIS Level 1B data as input and yields AOT information at increased resolution. PM MAPPER has passed through several intermediate steps that went from 10×10 to the current 1×1 km^2 of spatial resolution. In (Nguyen et al. [2010c]), AOT products at 3×3 km^2 have been validated by direct comparison with MODIS retrievals and showed a higher ability to retrieve pixels over land and coastlines. The validation of the new 1 km products has given the occasion to investigate the process of the validation itself, which in fact implicates the nontrivial comparison of spatially varying satellite data with temporally varying ground measurements. This is in itself ongoing work.

3 .1 Datasets

Validation data consist of AOT measured by AERONET sites, MODIS AOT at 10×10 km^2, and AOT at 1×1 km^2 obtained by the proposed method. The considered data covering European areas were collected in three years from 2007 to 2009. The Level 2.0 AOT measurements of 105 AERONET sites over Europe have been used. The location of the selected sites can be observed in Figure 4.3. The proposed methodology was applied on MODIS data involving a total of ca. 5500 granules to derive AOT maps at 1×1 km^2. Corresponding MODIS AOT at 10×10 km^2 of resolution in MOD04 L2 products were collected.

3 .2 Validation Method

As pointed out by Ichoku et al. [2002], the comparison of a geolocated raster map against time-series of measurements of a point in the ground cannot be achieved by matching the single pixels over the points. In the satellite maps, AOT represents for an area and may not be well matched with a sunphotometer point measurement. In case of an extremely high spatial resolution the pixel may well approximate a point, but sometimes clouds may obscure a pixel directly over a sunphotometer site instead of nearby pixels. Moreover the satellite overpass times rarely coincide with the AERONET measurement, several minutes separates the

Figure 4.3: The distribution of AERONET ground stations used to validate the satellite products.

two acquisitions on average. For these reasons it is far more appropriate and meaningful to compare spatial averages from MODIS with temporal averages from AERONET. For a proper validation then a spatio-temporal window should be used to extract the data from which statistics are collected and compared (see Figure 4.4).

In general, the same amount of airmass should be captured in a shorter time period by a sunphotometer in areas where the motion of the atmosphere is faster. Several window sizes have been used in literature to compare MODIS and AERONET data. In Ichoku et al. [2002], 50×50 km^2 boxes are chosen to match a 1-hour sunphotometer data segment. In Lary et al. [2009], 30 minutes of AERONET measurements are paired with MODIS pixels within a great circle distance of 0.25 ° and within a solar zenith angle of 0.1^o, while in Vucetic et al. [2008] a 30×30 km^2 box is used to match 1 hour of AERONET samples. The effect of different spatial windows from 30×30 up to 90×90 km^2 has been tested in Ichoku et al. [2002], but the increased resolution can now allow tests on smaller areas. The effects of the temporal window applied on the AERONET measurements has not been tested yet, instead. For all these reasons, the maps

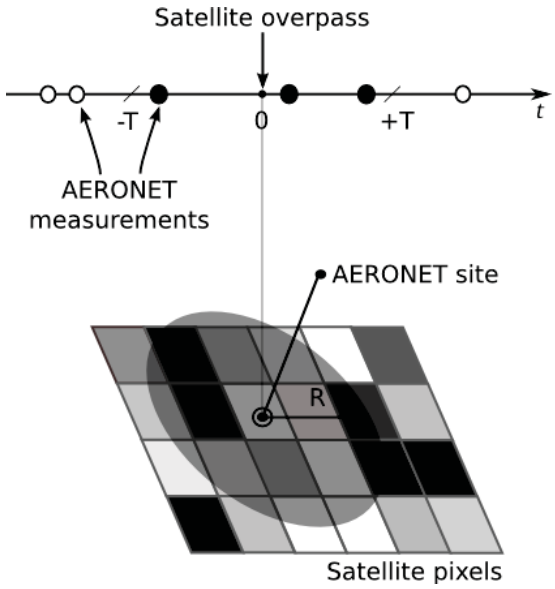


Figure 4.4: Matching a satellite map with ground data.

are validated by varying the size of the window on both the temporal and spatial axis.

For each satellite imagery of validation dataset, the pixels within 30 km were retrieved and stored in a database along with the relative AERONET ground measurements and other ancillary data as the distance of each pixel, timestamps and land cover information. The database stores about 20×10^6 satellite pixels, relative to about 50,000 satellite/AERONET matches. Customized queries to this database were then the mean for the validation. A minimum of 20% ca. of the available pixels was set for a valid match because a lower presence of satellite pixels was not considered enough to represent the whole area.

The validation of the satellite products has been iterated over different spatio-temporal windows, defined by the radius (R) and time semi-interval (T). Each match compares the average of the satellite pixels within a radius of R around an AERONET site and the average of the corresponding AERONET samples within $\pm$T minutes from the satellite overpass (see Figure 4.4).

To measure the association between the satellite and AERONET retrievals of AOT, following parameters were extracted

- Pearson's Correlation Coefficient r. It is quite sensitive to outliers and non-normality, e.g. the typically positively skewed AOT distribution, however still a good indicator with large sample sizes (between 10^2 and 10^4).

- Root Mean Square Error (RMSE). It is to evaluate the precision of satellite retrievals (r alone may not detect biased correlations);

- Mean Error (ME). It is to distinguish between under- and over-estimations of satellite retrievals.

AERONET measurements are interpolated to 550 nm to match the band of the satellite retrievals. In Lary et al. [2009] the log-log interpolation is used, but after some accuracy test on an AERONET data sample, log-linear interpolation in the frequency axis was used. The transformation formula is:

$$AOT_{550} = e^{\log(AOT_{f_1})+(550-f_1)\cdot\frac{\log(AOT_{f_2})-\log(AOT_{f_1})}{f_2-f_1}} \tag{4.1}$$

where $f1$ and $f2$ respectively the nearest lower and higher frequencies with available AERONET measurements to 550 nm.

3 .3 Results

Figures 4.5, 4.6, and 4.7 show the general results for the PM MAPPER 1×1 km^2 maps along with the results of original MODIS 10×10 km^2 maps. Samples over a discrete grid made of T=±{10, 20, 30, 40, 60, 120} min. and R={3, 6, 10, 15, 20, 25, 30} km have been smoothed via thin plate spline to create the continuous surfaces. The surface shows that no relevant trend exists on the temporal axis, although there is a slight systematic decrease of correlation of less than 0.05% every 10 minutes. As the radius of the area which selects the pixels gets smaller instead, PM MAPPER products suffer the more noisy nature of its pixels, whereas MODIS original products are more stable due to their coarser spatial resolution. 1-km AOT retrievals lose about 15% of correlation on the shorter radii (see Fig. 4.5). The same behaviors can be observed for RMSE and ME (see Figures 4.6 and 4.7).

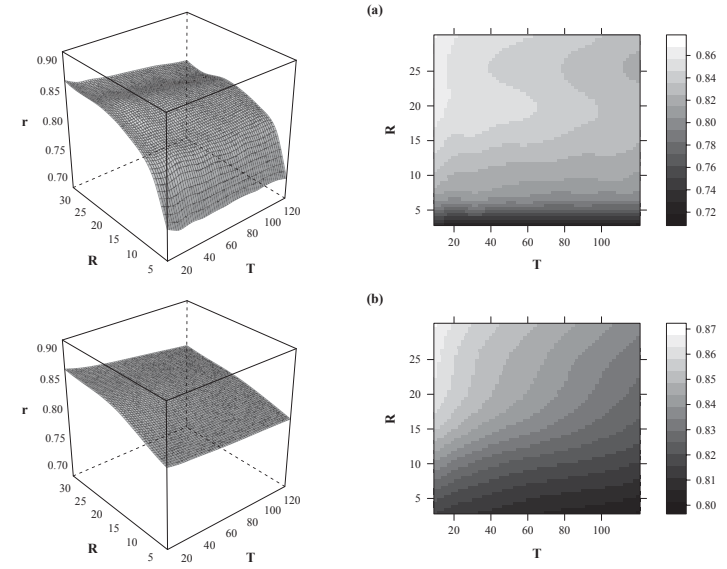


Figure 4.5: Interpolated Pearson's r surfaces and relative contour maps of AERONET AOT against (a) PM MAPPER 1×1 km^2 AOT and (b) MODIS 10×10 km^2 resampled AOT.

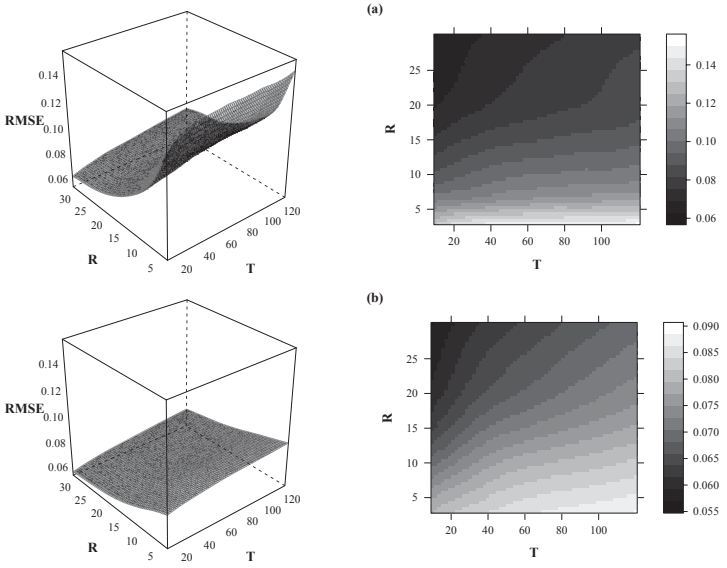


Figure 4.6: Interpolated RMSE surfaces and relative contour maps of AERONET AOT against (a) PM MAPPER 1×1 km^2 AOT and (b) MODIS 10×10 km^2 resampled AOT.

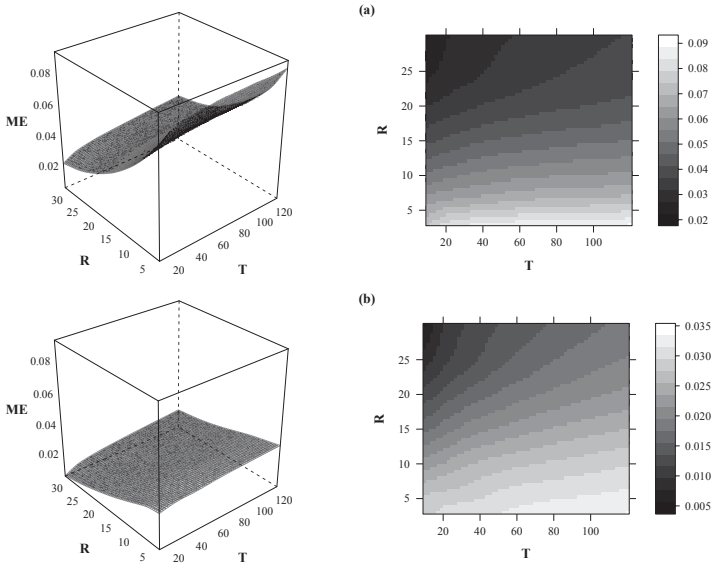


Figure 4.7: Interpolated ME r surfaces and relative contour maps of AERONET AOT against (a) PM MAPPER 1×1 km^2 AOT and (b) MODIS 10×10 km^2 resampled AOT.

Each discrete point in the 3D space is the statistic obtained from a dataset of the satellite and AERONET arithmetic means matched by the corresponding spatial and temporal constraints. The details for the chosen best case (T=10 min, R=20 km) can be observer in Figure 4.8. Both PM MAPPER and MODIS maps show a very good correlation (> 0.86) with the AERONET measurements. The Quantile-Quantile plot - which shows quantiles of the AERONET dataset against the corresponding quantiles of the satellite dataset - points out how AOT tends to be overestimated by MODIS for high values (> 0.5), while PM MAPPER seem to introduce a small positive bias for a wider range of AOT values (> 0.1). There is no highly remarkable difference between the two scatterplots, but the regression line equations show that the PM MAPPER retrievals are slightly more biased than MODIS (0.036 to 0.024), while the slope (0.97) is quite closer to unity than the original MOD04 products (0.927).

Table 4.1 shows the results for a fixed spatio-temporal window of ±10 min. and 20 km, which has been chosen as best case for both correlation with AERONET and stability along the temporal axis. The overall results are compared with

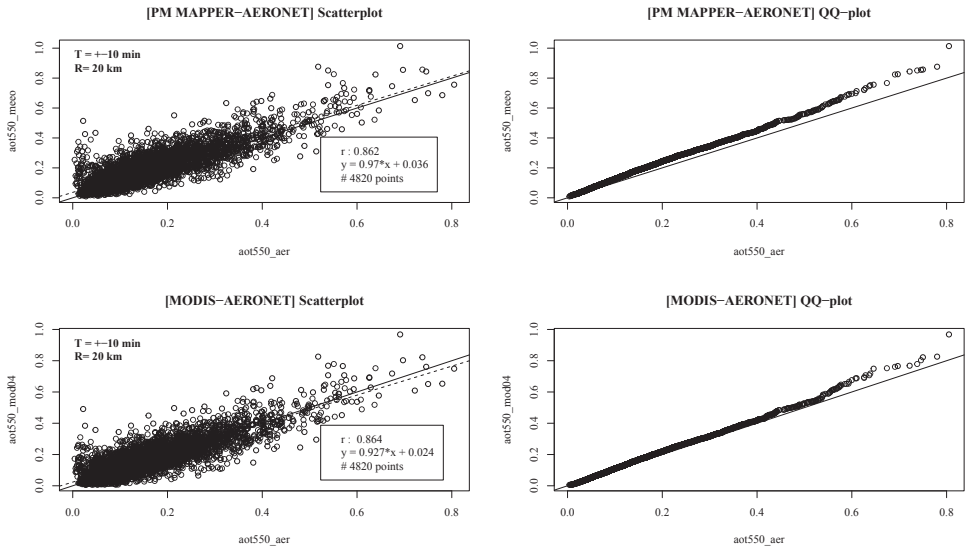


Figure 4.8: Scatter plot and Quantile-Quantile plot of AERONET against both PM MAPPER 1×1 km^2 AOT (above) and MODIS resampled 10×10 km^2 AOT (below) for a spatio-temporal window of ±10 min/20 km.

several filtered ones: by year (which are not showed here since not particularly relevant), season and land cover respectively. Looking at the table, both MODIS and PM MAPPER satellite products seem to better reproduce AOT during cold months (e.g. Winter period in comparison with Summer period). During Fall the PM MAPPER retrievals are curiously almost unbiased with respect to the other seasons. Trends have been searched by selecting the pixels over different land cover classes. Four macro-classes were used for this purpose and the results show excellent satellite retrievals over vegetation and rangeland, whereas 1-km^2 products performed worse on barren lands and particularly built-up areas. This trend is more remarkable on the PM MAPPER imagery.

4 Conclusion

The chapter presents the improvements of the MODIS aerosol algorithm to derive AOT with spatial resolution at 1×1 km^2 which then are used to develop a software package called PM MAPPER (Nguyen et al. [2010a]; Nguyen et al. [2010b];

Table 4.1: Overall and filtered validation scores for both PM MAPPER 1×1 km^2 and MODIS 10×10 km^2 AOT (R=20 km from AERONET site, T=±10 minutes from satellite overpass).

	PM MAP.	MODIS	
OVERALL # 4820	0.862 0.068 0.032	0.864 0.060 0.012	***r*** **RMSE** **ME**
Winter # 732	0.894 0.054 0.027	0.894 0.048 0.014	
Spring # 1444	0.855 0.075 0.037	0.858 0.064 0.014	
Summer # 1891	0.839 0.076 0.039	0.841 0.065 0.016	
Fall # 753	0.881 0.045 0.009	0.879 0.044 -0.001	
Vegetation # 3203	0.882 0.052 0.012	0.874 0.052 0.001	
Rangeland # 1110	0.868 0.076 0.032	0.875 0.066 0.012	
Dark Barren Land # 431	0.776 0.127 0.080	0.836 0.091 0.045	
Barren/Built-Up # 384	0.681 0.097 0.018	0.759 0.079 0.003	

MEEO [2010a]). The adaptation is replied on three main points which are application of the standard MODIS aerosol algorithms on smaller data boxes sized 1×1 km^2, modification of spatial variability cloud mask, and change of condition to apply ocean algorithm or land algorithm. Besides, the land cover classification maps are used for the first cloud screening process instead of the MODIS MOD35 cloud products. The implementation of PM MAPPER is based on the IMAPP software package provided by University of Wisconsin-Madison (CIMSS [2011]).

The validation of the 1×1 km^2 AOT products of the PM MAPPER software

package was carried out over Europe for years from 2007 to 2009 (Campalani et al. [2011]). Both the comparisons with the ground sunphotometers of the AERONET network and the MODIS 10 km AOT products have shown a highly satisfactory level of correlation. PM MAPPER 1×1 km^2 AOT products are hence validated. The satellite retrievals show a stronger association with the ground measurements during Fall and Winter. Some trend has been found also for different land use classes: e.g. the AOT retrieval over vegetation outperforms the retrieval over barren land and built-up areas. The method to validate satellite products via ground truth comparison has been investigated, showing no important sensibility of the result to the time interval which defines the ground data segment, whereas a stronger trend is shown as the radius that selects the satellite pixels approaches zero. After that, a radius of 20 km and a time semi-interval of ±10 minutes are then assessed as the best choice for satellite validation with AERONET ground stations.

Chapter 5

Aerosol Optical Thickness Retrieval at 10×10 km^2 Spatial Resolution using Support Vector Regression

This chapter presents the application of SVR for AOT retrieval at 10×10 km^2 from MODIS (Nguyen et al. [2010c]). The proposal is motivated by the better performance of SVR with respect to NNs in finding a global solution instead of a local one, and in coping with huge and high dimensional satellite data. In this thesis, a SVR method was investigated, applied, and then validated on Collection 005 datasets covering Europe in a period from 2006 to 2008.

1 Methodology

The methodology applied for AOT retrievals based on SVR technique consists of three main steps: (i) collection and processing of satellite-based data (MODIS) and ground-based sensor measurements (AERONET) over Europe for a period of three years, (ii) integration and combination of data from two sources having different temporal and spatial resolutions, and (iii) application of SVR technique for aerosol estimation.

1 .1 Datasets

The MODIS aerosol products MOD04 L2 covering European areas from 2006 and 2008 were collected. These data represent radiance/reflectance of 36 wavelengths averaged at 10×10 km^2 of resolution at nadir. Besides, it provides corresponding geolocation data including geodetic coordinates, ground elevation, solar and satellite zenith and azimuth angles for each 10×10 km^2 sample. AOT are estimated in seven wavelengths (0.470, 0.550, 0.670, 0.870, 1.240, 1.630 and 2.130 μm) over ocean and three wavelengths over continental areas (0.470, 0.550 and 0.670 μm) at 10×10 km^2 of spatial resolution.

AERONET data level 2.0, cloud-screened and quality-assured, of 105 sites distributed in Europe in the period of three years, 2006, 2007, and 2008, were collected. AOT at 0.500 μm, the closest to MODIS AOT at 0.550 μm, was used to generate SVR aerosol predictors.

1 .2 Data Integration

The satellite data and ground based measurements have differences of temporal and spatial resolution. Hences, data combination aims at obtaining data collocated in space and synchronized in time. The combination methodology is carried out as mentioned in Section 4.4, Chapter 4. The condition follows suggestion in Ichoku et al. [2002] for the best fit of MODIS data and AERONET at 10×10 km^2. The MODIS data are considered if their distances to AERONET sites are within a radius of 30 km, while the contemporaneous measurements of AERONET instruments are selected and averaged within a temporal window of 60 minutes around the satellite overpasses. The satellite data are collected if they are cloud-free and number of good pixels is bigger than 20% out of total pixels legal to integration conditions.

The AERONET-MODIS combinations are separated into two datasets: *instance dataset* and *aggregate dataset* (see Figure 4.4). The first one consists of 66,225 samples, each of which is a combination of measurements on a single MODIS pixel with an averaged AERONET AOT value. In instance case, many MODIS pixels collected around an AERONET at same acquisition time are matched to an AERONET AOT value. One sample is represented as a vector

that includes the following information: AERONET AOT at 0.500 μm, MODIS geometric data (solar zenith angle, solar azimuth angle, sensor zenith angle, sensor azimuth angle, scattering angle) and seven MODIS reflectances (0.646, 0.855, 0.466, 0.553, 1.243, 1.632, and 2.119 μm). The aggregate dataset contains 5,289 samples. A sample is created by combining an averaged AERONET AOT with averaged MODIS geometric data and averaged MODIS reflectances calculated on all cloud-free pixels acquired at the same time around this AERONET site. These vectors are stored in the same format as the ones in the instance dataset.

1 .3 Support Vector Regression for Inversion Process

SVR was applied to instance dataset and aggregate dataset in order to create different data models for AOT retrievals, called *instance SVR* and *aggregate SVR*, respectively. The SVR with epsilon loss function and Radial Basic Function (RBF) kernel provided by LIBSVM (Chang and Lin [2011]) was employed. The accuracy was measured on three year data cross validation in which we repeated selections of two year data for training and one year data for testing. Root Mean Square Error (RMSE) and CORRelation coefficient (CORR) were calculated from SVR AOT and AERONET AOT. SVR regularizations (C, γ, ϵ) were searched in appropriate range with exponentially growing sequences. For each case, cross validation was applied on a training data set and the best accuracy was picked. At the end of searching process, the chosen regularizations minimized mean square error in the training phase. Both instance and aggregate SVRs were used to bring out data models for AOT prediction at pixels of 10×10 km^2. The aggregate SVR data model was made by the smaller dataset (aggregate dataset), while the instance data models dealt with a large training data and difficulties in searching regularization. Experiments on both instance and aggregate SVRs were made to investigate their accuracy and computing time.

Moreover, SVR AOT were separated by different land cover types in order to investigate surface effect on aerosol retrievals. Concerning the land cover analysis, a spectral rule based software system, introduced as the SOIL MAPPER (MEEO [2011]), was used to distinguish surface types. In this experiment, a compact classification mode with 12 land cover classes was used. Cloud, snow,

and unclassified pixels were discarded, whereas the nine remaining classes (see Table 5.4) were utilized to evaluate the SVR prediction on a land cover basis. A land cover class for each pixel sized 10×10 km^2 was determined when the classification system was applied on reflectances averaged from all cloudy-free pixels of 1×1 km^2 in this box.

2 Experiments and Results

The experiments focused on assessing accuracy of SVR AOT in comparison with AERONET AOT. The AOTs obtained by aggregate SVR, instance SVR, and MODIS aerosol algorithm were considered at different conditions: by year, by season, and by surface type. The accuracies of both instance and aggregate SVR estimators are slightly better than those of the MODIS algorithm, as summarized in Table 5.1. Based on RMSE and CORR between predicted AOTs and AERONET AOTs, averaged in 3 year data, instance SVR achieves the highest accuracy, then aggregate SVR follows and finally the MODIS algorithm. This order is justified by the increase of RMSE (0.077, 0.084, and 0.090, respectively) and the decrease of CORR (0.835, 0.812, and 0.807, respectively). The MODIS and SVRs AOT data in 2008 seem to have low quality as shown by the lowest correlation with AERONET AOT. However, instance SVR in this case still outperforms (CORR=0.802) the aggregate SVR (CORR=0.758) and MODIS algorithm (CORR=0.764).

Table 5.1: MODIS algorithm, Aggregate SVR, and Instance SVR accuracy by year

Year	# Obs.	MODIS		Aggregate SVR		Instance SVR	
		RMSE	CORR	RMSE	CORR	RMSE	CORR
2006	21,555	0.095	0.831	0.087	0.847	0.086	0.850
2007	24,251	0.087	0.827	0.081	0.831	0.074	0.853
2008	20,455	0.087	0.764	0.084	0.758	0.072	0.802
Total	**66,225**	**0.090**	**0.807**	**0.084**	**0.812**	**0.077**	**0.835**

Table 5.2 shows in detail the consuming time of aggregate SVR and instance SVR for the above experiment. Executions were tested on a computer with Intel

(R) Core(TM)2 CPU 6400 @2.13 GHz, 2Gb RAM and Ubuntu 8.10 platform. Instance SVR spent about 240 seconds to predict 66,255 data, while aggregate SVR uses much smaller amount of time, 26 seconds. This difference is mainly due to the number of samples in aggregate dataset used for training in aggregate SVR less than instance dataset used in instance SVR (132,522 data compared to 10,778), which induces data models with smaller sizes. The performance time will be meaningful for further SVR applications that aim at increasing spatial resolution of aerosol retrievals.

Table 5.2: Aggregate SVR vs. Instance SVR in consuming time performance

Year	# Obs.	Aggregate SVR		Instance SVR	
		# Training	Time(s)	# Training	Time(s)
2006	21,555	3,549	7.4563	44,706	66.59
2007	24,251	3,378	8.9790	42,010	106.69
2008	20,455	3,851	9.4843	45,806	70.94
Total	**66,225**	**10,778**	**25.9196**	**132,552**	**244.22**

We carried out the same further experiments on data sets separated by seasons and surface types to consider effects of meteorological conditions and surface reflectance on aerosol retrievals. Data in pairs of years were used for training SVRs, while data on the remaining year were classified by seasons and surface classes for testing purposes.

Table 5.3 present the validation results by season. Three months (i.e. December-January-February (DJF), March-April-May (MAM), June-July-August (JJA) and September-October-November (SON)) are grouped to be presentative for a season. Moreover, relative error (RERR) defined in the equation 5.1 is calculated in order to measure error percentage between predicted and target values.

$$rerr = \frac{1}{n}\sum_{i=1}^{n}\frac{|predictedAOT_i - aeronetAOT_i|}{|aeronetAOT_i|} \tag{5.1}$$

In the coldest period (ie. DJF), aerosol retrievals obtained by all algorithms have the lowest RMSE but the highest RERR in comparison with aerosol predited in other months. Because RMSE presents absolute error while RERR shows error percentage, AOT in DJF should have low values. For all algorithms, the

performance of aerosol retrievals is worst in the period DJF but similar over all warmer seasons (ie. MAM, JJA and SON). Based on RMSE and RERR, a relationship between aerosol retrievals and aerosol values can be infered. We calculated averages of AOT (MEAN) for all seasons. Over the coldest season DJF, AOT is often low (MEAN $\sim$ 0.115) while AOT is quite similar over other seasons (MEAN $\sim$ 0.223, 0.212 and 0.175 for MAM, JJA and SON, respectively). Therefore, aerosol prediction have the highest error in DJF and similar errors in MAM, JJA and SON. The seasonal performance in fact presents the dependence of AOT prediction on AOT magnitude for considered algorithms. The instance SVR has the most competitive accuracies that are better than those of aggregate SVR and MODIS algorithm in all seasons.

Table 5.3: Instance SVR, Aggregate SVR and MODIS algorithm accuracy by season

Season	# Obs.	Instance SVR			Aggregate SVR			MODIS		
		CORR	RMSE	RERR	CORR	RMSE	RERR	CORR	RMSE	RERR
DJF	7125	0.693	0.059	0.514	0.629	0.066	0.583	0.677	0.061	0.503
MAM	17932	0.863	0.078	0.344	0.839	0.084	0.367	0.840	0.091	0.406
JJA	23859	0.808	0.082	0.385	0.771	0.089	0.404	0.762	0.099	0.461
SON	15406	0.794	0.080	0.378	0.778	0.084	0.416	0.800	0.086	0.437

MODIS used two algorithms for land and ocean because of different physical interactions between aerosol and matters. Among all surface types listed in Table 5.4, only the water class refers to water pixels while remains represent land pixels. MODIS ocean algorithm gained high accuracy (RMS=0.067, CORR=0.822), but it can be further improved by instance SVR (RMS=0.062, CORR=0.850). Out of land surface types, four classes Peat Bog, Evergreen Forest, Agricultural Areas and/or Artificial non Agricultural, Areas Scrub/Herbaceous Vegetation have a small number of samples, so their results should not be considered. In all remaining cases, instance SVR is more accurate than the MODIS algorithm. The clear improvement can be observed at Artificial Surfaces and/or Open Spaces with little or no Vegetation surface, which is consistent with results of previous studies that showed the poor performance of the MODIS algorithm on bright surfaces (Nguyen et al. [2010a]). Aggregate SVR has the worst accuracy on water pixels. It can be explained as result of the small contribution of water pixels on aver-

aged data used for training aggregate SVR model, that did not occur to instance SVR. This phenomenon influences pixels belonging to other surface types except Deciduous Forest and/or Agriculture Area class that has a large data set and therefore can be represented well by averaged values.

Table 5.4: MODIS algorithm, Aggregate SVR, and Instance SVR accuracy by land cover class

LC Class	# Obs.	MODIS		Aggregate SVR		Instance SVR	
		RMSE	CORR	RMSE	CORR	RMSE	CORR
Water	2,981	0.067	0.822	0.071	0.799	0.062	0.850
Peat Bogs	91	0.112	0.622	0.151	0.527	0.129	0.550
Deciduous Forest	2,734	0.086	0.692	0.072	0.681	0.065	0.700
Evergreen Forest	19	0.054	0.489	0.065	0.584	0.053	0.714
Deciduous Forest and/or Agricultural Area	34,316	0.080	0.824	0.075	0.824	0.702	0.833
Agricultural Areas and/or Artificial non Agricultural Areas	25	0.103	0.895	0.086	0.926	0.080	0.950
Scrub/Herbaceous Vegetation and/or Agricultural Areas	5,302	0.082	0.825	0.083	0.806	0.075	0.829
Artificial Surface and/or Open Spaces with little or no Vegetation	5,961	0.096	0.746	0.085	0.769	0.078	0.808
Scrub/Herbaceous Vegetation	134	0.060	0.892	0.075	0.871	0.066	0.882

3 Conclusion

In this chapter, a data driven approach that applies the SVR technique on MODIS and AERONET data to predict AOT information has been presented. Satellite

and ground-based data covering the European areas from 2006 to 2008 were collected and then integrated to solve spatial and temporal differences. After that, SVRs were applied on instance dataset and aggregate dataset to make different non-linear regressions for aerosol retrievals. The experiment results show that SVR approach is competitive with respect to the MODIS algorithm and able to improve prediction accuracy over areas having no or little vegetation. Out of two SVR models, instance SVR outperforms the aggregate SVR, but the time execution is longer. In fact, with 10×10 km^2 of spatial resolution, each MODIS image consists of 135×203 pixels. Increasing spatial resolution up to 1×1 km^2 (see the next chapter), more than two million pixels in an image would need to be processed. Also, the slow performance of instance SVRs hints at the need for further investigations of data selection and application of pruning techniques in the training phase.

Chapter 6

Downscaling Spatial Resolution of Aerosol Optical Thickness to 1×1 km^2 using Support Vector Regression based on Domain Knowledge

1 Introduction

In this chapter, the methodology to derive aerosol optical thickness at 1×1 km^2 over land from MODIS Level 1B data is presented. This work aims at providing more detail AOT information for local monitoring. Different from related works in literature (Oo et al. [2008]; Castanho et al. [2007]; Li et al. [2005]; Nichol and Wong [2009]; Wong et al. [2009]) as well as our downscaling work described in Chapter 3 using a physical approach (Nguyen et al. [2010a]; Nguyen et al. [2010b]), the proposed methodology exploits SVR technique and domain knowledge for aerosol optical thickness retrievals. The investigation of SVR on AOT retrival from MODIS data at 10×10 km^2 shows promissing results in the previous chapter and hence motivates the usage of SVR to deal with downscaling problem. However, the application of SVR for AOT from 10×10 to 1×1 km^2 of spatial res-

olution is not a trivial task because of a very large and noisy dataset as a result of increasing the resolution 100 times. Moreover, the application is extended from pixel domain to map domain in which data models created by data collected on sparse locations are applied on large and continuous map areas.

The proposed approach is developed and tested on real data collected over European areas in three years from 2007 to 2009. Validations show good results in both pixel and map domains. Although the former versions of this approach are published in Nguyen et al. [2011] and Nguyen et al. [2012], some changes are applied in the methodology and validation is re-done in a larger dataset in order to provide uniform results.

2 Methodology

In this section, the methodologies to create SVR models and to predict AOT maps from MODIS data are presented. Firstly, satellite-based data and ground-based measurements in the areas of interest are collected. Secondly, data from difference sources are integrated to solve the differences of temporal and spatial resolution. After that, filtering and clustering techniques exploiting physical aspects of data are applied in order to reduce the total amount of data, and to separate them into groups having different characteristics. In the fourth step, data are selected for training process using some strategies. Then, SVR is applied on training datasets to create data models for different clusters. The flowchart of model generation is presented in Figure 6.1. Finally, in the map prediction framework, aerosol maps at 1×1 km^2 of spatial resolution are derived from MODIS Level 1B data using SVR models.

2 .1 Data Collection

The collected data consist of MODIS L1B data, MOD04 L2, Land Cover (LC) map, and AERONET data Level 2.0 covering Europe areas in three years from 2007 to 2009. Datasets are used to develop empirical data models as well as to input for the map prediction framework. The description of datasets can be found in Section 3.1.2 (Chapter 3) and Section 5.2 (Chapter 5).

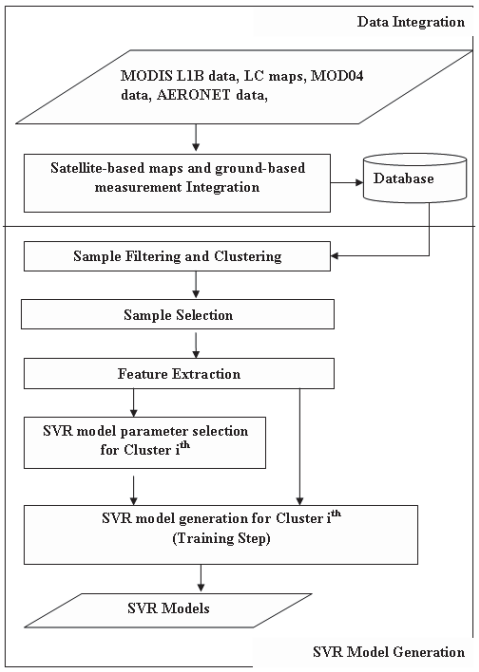


Figure 6.1: SVR approach for the AOT inversion problem.

2 .2 Data Integration

Satellite- and ground-based data are collected from different sources have different temporal and spatial resolutions which can be solved by the integration process. Satellite data include MODIS L1B data (MOD02 and MOD03) and LC maps at 1×1 km^2 of resolution, MODIS aerosol products (MOD04 L2) at 10×10 km^2 of resolution. Ground-based data are obtained from AERONET distributed sites.

All satellite maps are acquired at the same time and location, thus only re-sampling process is applied to refine MOD04 L2 products to 1 km^2 of spatial resolution. However, satellite-based and ground-based data have different temporal resolution (every day versus every 15 minutes, respectively) and different spatial resolution (1354 by 2030 of 1-km-pixel maps in comparison with site points). Therefore, time and location constraints are applied for data integration. As proposed in Ichoku et al. [2002] for MODIS and AERONET AOT integration, satellite data are considered if pixels are located over land, cloudy-free and their distances from AERONET sites are within radius R of 30 km. Meanwhile, the

contemporaneous measurements of AERONET instruments are selected and averaged within a temporal window T of 30 minutes around the satellite overpasses. All pixels are populated in a database with 1×1 km^2 of spatial resolution. This database refers to the same one used in PM MAPPER validations mentioned in Chapter 4.

Satellite-based and ground-based integration is applied to create data samples for data modeling process. The usage of integrated data aims at improving the aerosol retrieval quality by utilizing the high accuracy of ground measurements as validated in Xu et al. [2005]Vucetic et al. [2008], Lary et al. [2009], Obradovic et al. [2010] and Nguyen et al. [2010c]. Following the instance SVR approach (see Chapter 5), a sample is a combination of a satellite pixel's attributes and an arithmetic mean of AERONET AOT values that satisfied collocation and time synchronization constraints. A sample's features consist of the AERONET AOT at 0.553 μm, latitude, longitude, sensor zenith angle, solar zenith angle, relative azimuth angle, scattering angle, four reflectances at 0.646. 0.466, 1.243, and 2.119 μm, and land cover class. The feature selection is based on inputs of LUT in the MODIS algorithm.

AERONET AOT at 0.553 μm (AOT_{553}) is not measured directly from AERONET sites and it is calculated using log-linear interpolation from two AOT values of the closest channels 0.500 and 0.670 μm (AOT_{500} and AOT_{670} , respectively), as follows:

$$AOT_{553} = e^{\log(AOT_{500})+(553-500)\frac{\log(AOT_{670})-\log(AOT_{500})}{670-500}} \tag{6.1}$$

The scattering angle Θ was defined as:

$$\Theta = \cos^{-1}(-\cos\theta_0\cos\theta + \sin\theta_0\sin\theta\cos\phi) \tag{6.2}$$

where θ_0, θ and ϕ are the solar zenith, sensor view zenith and relative azimuth angles, respectively.

2 .3 Data Filtering

Using the proposed integration in Ichoku et al. [2002] for 1 km pixels, the resulting dataset becomes huge and has lower quality. This poses difficulties for

the modelling step. In order to deal with this problem, the validation results described in Chapter 4 are utilized to select more suitable datasets. The validation over data covering European areas in three years from 2007 to 2009 (see Chapter 4) investigates different spatial and temporal windows for integration of satellite data and ground-based measurements. The objective is to find intergration constraints to obtain the best quality datasets and to reduce as much as possible the number of samples. The optimal parameters are a spatial window radius of 20 km and temporal window of 10 minutes (see Figures 4.5 and 4.6). The resulting datasets have better quality and smaller amount of samples than those obtained by the original conditions in Ichoku et al. [2002]. Therefore, these conditions are proposed to select data for aerosol applications at 1×1 km^2 of spatial resolution.

2 .4 Data Clustering

The clustering technique is considered to divide and conquer big datasets and to classify aerosol prediction quality. Data are separated into specific groups and then used for modeling. The technique is based on priority of criteria applied over land surfaces excluding water, clouds, ice and snow to choose pixels for aerosol derivation in the physical approach (Kaufman and Tanré [1997]). The priority is originated from the fact that in the MODIS aerosol algorithm Collection 004 over land, the surface reflectance is determined using the relationship of surface reflectance at 0.49, 0.66, and 2.1 μm. Figure 6.2 presents the practical relationship between surface reflectance at 0.49 μm (full symbols) and 0.66 μm (empty symbols) to that at 2.1 μ over different land cover types. Depending on surface types, the variability of surface reflectance is different (e.g. vegetation surface is less variable than soil areas) and hence affects the estimation of surface reflectance for aerosol retrieval process. The quality of AOT retrieval will decrease in order as follows:

$$\begin{array}{ll} \text{first priority for} & 0.01 \leqslant \rho^*_{2.1} \leqslant 0.05 \\ \text{second priority for} & \rho^*_{3.8} \leqslant 0.025 \\ \text{third priority for} & 0.01 \leqslant \rho^*_{2.1} \leqslant 0.10 \\ \text{fouth priority for} & 0.01 \leqslant \rho^*_{2.1} \leqslant 0.15 \end{array} \tag{6.3}$$

where $\rho^*_{2.1}$ and $\rho^*_{3.8}$ are TOA reflectance at wavelength 2.1 and 3.8 μm.

The clustering technique is proposed using the first, third, and fourth priorities. Samples are separated into four groups based on thresholds in the mid-IR band 2.13 μm (from 0.01 to 0.05, from 0.05 to 0.10, from 0.10 to 0.15, and larger than 0.15). It aims at specializing SVR models for particular data groups.

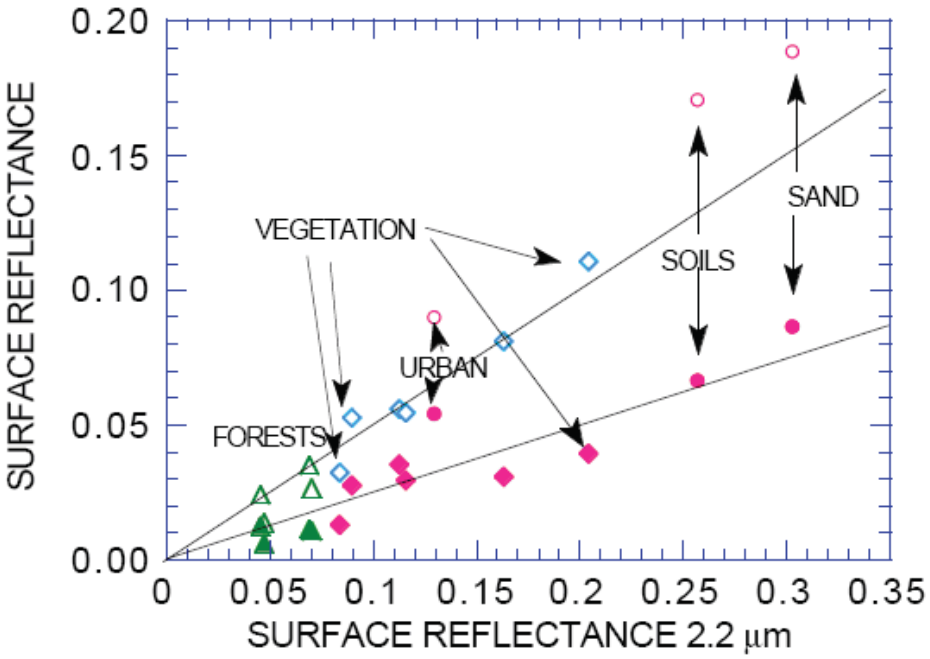


Figure 6.2: Scatter diagram between the surface reflectance at 0.49 μm (full symbols) and 0.66 μm (empty symbols) to that at 2.2 μm for several surface type (Kaufman and Tanré [1997])

2 .5 Selection of Training and Testing Datasets

For each cluster, data in two years are selected to create training datasets while data in the year which is not included are used for testing models. This process is repeated for three year of data. The big integration datasets (about 400 thousands samples of each year) will pose difficulties in the modeling step using SVR if all data are employed for training process. Moreover, these datasets consist of samples whose AERONET AOT values just range from 0 to 0.8 and are

dominated by small values (see Table 6.1). Imbalance in the datasets will lead to bias in the modelling process. Therefore, the selection of samples for training SVR models is based on two crucial factors: AERONET location and AOT size. Samples at each AERONET site are divided into four groups, called AOT groups, with AOT value thresholds 0.2, 0.4, and 0.6. After that, the same data amount is selected randomly from all available AERONET sites in each AOT group so that new training datasets still keep the ratio of data in each AOT group to total as in the original datasets. All selections used random selection. The maximum number of samples for each year in training datasets are limited at about 25000 since LIBSVM (Chang and Lin [2011]) is used for implementation. However, testing datasets includes all data in a year.

Table 6.1: Statistics of AOT values in integration datasets.

Year	# Total	% (0.0 - 0.2)	% (0.2 - 0.4)	% (0.4 - 0.6)	% (> 0.6)
2007	550,690	69.20	24.50	5.94	0.37
2008	492,640	75.45	21.71	2.49	0.35
2009	387,392	73.99	22.71	3.16	0.15

2 .6 Support Vector Regression for Inversion Process

SVR is applied on training dataset of each cluster to create a corresponding data model. This takes advantages of the divide-and-conquer strategy and therefore, it is easier to control, improve, and evaluate the SVR performance on each cluster. The inversion problem is stated as follows. Given a training dataset including l samples:

$$\{(x_1, y_1), \ldots, (x_l, y_l)\} \subset X \times \Re \tag{6.4}$$

where X denotes the space of the input patterns (i.e. $X \subset \Re^d$), the target y_i refers to as AERONET AOT at 0.553 μm. The input is expressed as a record of latitude, longitude, sensor zenith angle, solar zenith angle, relative azimuth angle, scattering angle, reflectance at 0.646 μm, reflectance at 0.466 μm, reflectance at 1.243 μm, reflectance at 2.119 μm, and land cover class. The ϵ-SVR, introduced in Chapter 2, is to find the optimal function $f(x)$ that has at most ϵ deviation from the actually obtained target y_i from the training data. The ϵ-SVR with ep-

silon loss function and Radial Basic Function (RBF) kernel provided by LIBSVM (Chang and Lin [2011]) is used in this work.

The SVR algorithm is well known by generation performance which can be achieved by good settings of the ϵ-SVR parameters (i.e. regularization C, ϵ of the lost function, and p in the kernel function RBF). Because of high cost in cross validation for parameter selection on large datasets, those parameters are estimated using a practical approach proposed in Cherkassky and Ma [2004].

Following this method, the parameter C can be chosen equal to the range of output y_i values of training data. In order to limit the sensitivity of C to possible outliers in the training data, C is proposed as

$$C = \max(|\bar{y} + 3\sigma_y|, |\bar{y} - 3\sigma_y|) \tag{6.5}$$

where $\bar{y}$ and σ are the mean and the standard deviation of the y values of training data.

Parameter ϵ is estimated using the assumption that the value of ϵ should be proportional to the input noise variance. Based on the empirical results, the practical ϵ is proposed as:

$$\epsilon = t\sigma\sqrt{\frac{\ln l}{l}} \tag{6.6}$$

where t, l and σ are the empirical dependency on the number of training data (proposed as 3), the number of samples in the training data and the variance of additive noise δ, respectively. δ is described by:

$$y = f(x) + \delta \tag{6.7}$$

where δ is independent and identically distributed (i.i.d) zero mean random noise, x is a multivariate input and y is scalar output, $f(x)$ is regression function.

$\bar{\sigma}$ is denoted as the practical noise variance estimated from the training data which will be used as σ in 6.6 for ϵ selection:

$$\bar{\sigma} = \frac{l^{\frac{1}{5}}k}{l^{\frac{1}{5}}k - 1}\frac{1}{l}\sum_{i=1}^{l}(y_i - \bar{y}_i)^2 \tag{6.8}$$

where k is window size, proposed in the 2 - 6 range, of k-nearest-neighbours regression, $\bar{y}_i$ is a local average of training data estimated from k nearest neighbours.

The width parameter p in RBF kernel is presented as follows:

$$K(x_i, x_j) = e^{-\frac{\|x_i - x_j\|^2}{2p^2}} \tag{6.9}$$

where x_i is a training data.
p is appropriately selected to reflect the input range of the training/test data. For the multivariate d-dimensional problem, p is proposed to be calculated as $p^d \sim (0.1 - 0.5)$ where d input variables are pre-scaled to $[0, 1]$ range.

Based on mentioned theory, the SVR parameter selection is carried out in two steps:

- Initializing values of C, ϵ and p from training data using the methodology described above.

- Tuning parameter ϵ by changing empirical dependency parameter t in 6.6 which was proposed as 3. The ϵ calculated by this methodology in our training dataset is very small, which makes most training data become support vectors. In this case, data models become complex and overfitting to training datasets. The small ϵ is due to a large training dataset, very small target values, and repeated target values on many samples as a result of integration process in which many satellite pixels are matched to one AERONET sites. The changing reduced number of support vectors to approximately 50% - 60% of total number of training data. Moreover, the new ϵ did not increase Mean Square Error (MSE) much with respect to the old one in cross-validation tests.

2 .7 Map Prediction Framework

The map prediction framework is developed to derive AOT maps from MODIS L1B data using generated SVR models. The detail of software development can be found in Limone [2011]. The algorithm flowchart is presented in Figure 6.3.

As mentioned in Section 2.1.2 in Chapter 2, cloud screening process is a crucial task to guarantee aerosol retrieval quality. In the map prediction framework, LC maps distinguish types of pixels and perform the first cloud screening. This is due to the fact that the aerosol estimation algorithm over land is applied on pixels of land instead of cloud, water, ice, snow. Because the AOT estimation on cloud contamination or bright pixels from satellite reflectance is not correct, we apply a second cloud screening process using the cloud masking procedure developed for retrieval of aerosol properties by MODIS.

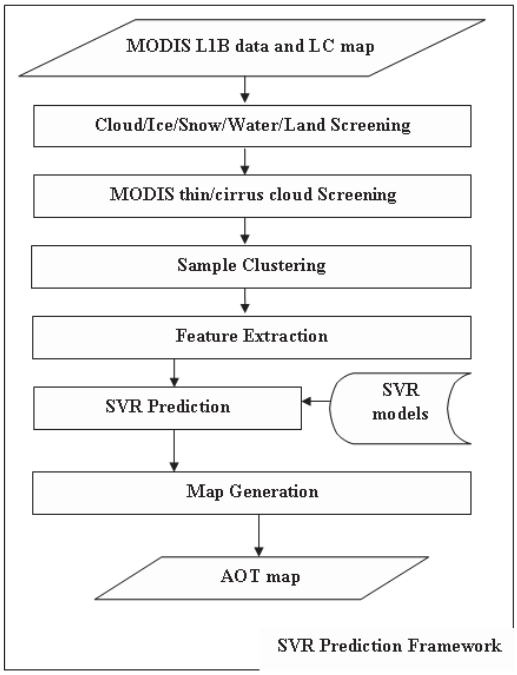


Figure 6.3: The map prediction framework.

The second cloud screening algorithm is based on spatial variability of reflectances on TOA in the visible wavelengths. Clouds show high spatial variability in the range from hundred meters to few kilometers, while aerosol in general is very homogeneous. The original algorithm is proposed in Martins et al. [2002] for cloud masking over ocean but this procedure has been extended to land and applied in both aerosol algorithms in Collection 005 (Remer et al. [2004]). The land algorithm generates a cloud mask using spatial variability of the 0.47 and 1.38 μm channels with thresholds 0.0025 and 0.003, respectively. If the standard

deviation calculated for each group of 3×3 pixels is greater than the corresponding threshold, then the area of the entire 3×3 pixel box is considered as clouds. In addition, tests on visible channel reflectance thresholds are carried out. If the reflectance at 0.47 μm and 1.38 μm are greater than 0.4 and 0.025, respectively, the pixel is considered as a cloudy pixel. In our approach, all calculations are applied at 1 km pixels for both 0.47 μm and 1.38 μm channels instead of 500 m and 1 km pixels, respectively as in the Collection 5 algorithm.

After cloud scanning processes, selected pixels are grouped into four clusters in order to apply the corresponding SVR data model to predict aerosol optical thickness. The final process collects predicted pixels, integrates with geo-information and then generates the AOT map.

3 Experiments and Results

3 .1 Pixel Domain

Pixel domain refers to as pixels collected in areas around AERONET sites and used to create and test SVR models. Three experiments are designed in order to validate the proposed methodology, compare performance at 1×1 km^2 of spatial resolution between SVR and MODIS algorithm, and investigate the relationship between SVR AOT and MODIS AOT.

The collected data, covering Europe in three year from 2007 to 2009, consist of MODIS L1B data and LC map at 1×1 km^2 of resolution, MOD04 L2 at 10×10 km^2 of resolution, and AERONET data Level 2.0. After integrating satellite-based and ground-based measurements, there are samples at 31, 36 and 33 AERONET sites for 2007, 2008, and 2009, respectively. The sites distribution is presented in Figure 6.4.

The statistics on the total dataset and clusters are presented in Table 6.2. There are 1,425,733 samples for three years. In the clustering step, those samples are grouped into four groups based on proposed thresholds of the mid-infrared band 2.13 μm. The clusters 1, 2, and 3 which are considered as having good pixels for AOT estimation hold most of the data, i.e. 19.63%, 48.34% and 23.14% of the total, respectively.

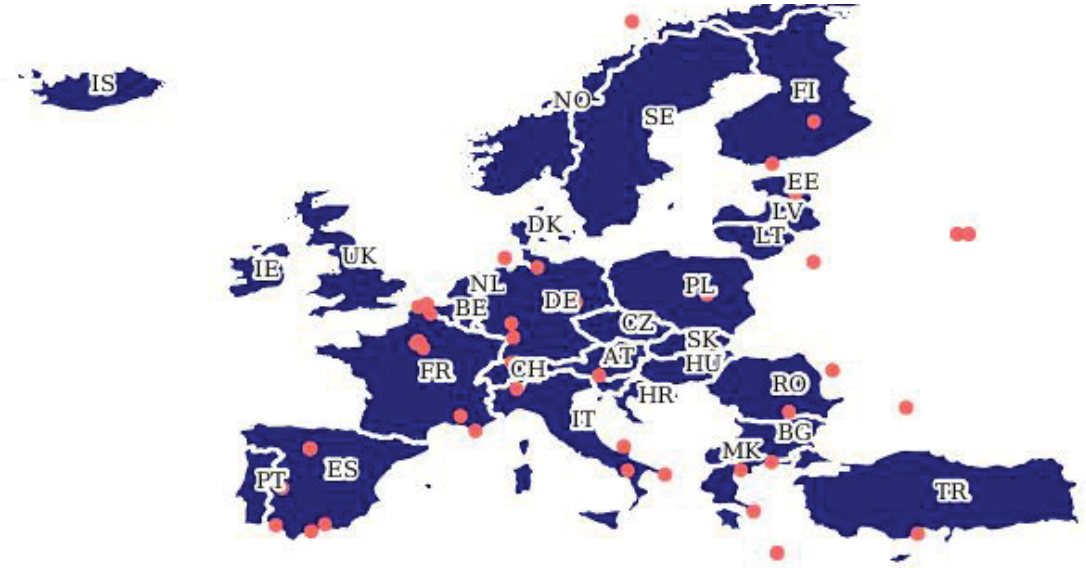


Figure 6.4: Distribution of AERONET sites over the Europe area used in data modeling.

For each cluster, selection of training and testing datasets follows the methodology mentioned in Section 6.2.5. After that, SVR is applied on training datasets to create data models and then use them to predict AOT in testing datasets. The evaluation was carried out on each cluster using Mean Error (ME), Root Mean Square Error (RMSE) and CORRelation coefficient (CORR), all of which are calculated from AOT values obtained by the SVR method and AERONET measurements.

Instance and aggregate validations were carried out. In "instance validation", all AOTs predicted by SVRs around every AERONET site are matched directly to a corresponding AERONET AOT and validated. In "aggregate validation", all SVR AOTs are aggregated by acquisition time and AERONET site, then averaged and validated to corresponding AERONET AOT values. The aggregate validation gives more stable results than instance validation when data at different spatial resolutions are compared. Results for each clusters are averaged from results of three years.

Table 6.3 and 6.4 show number of samples used in training phase (#Training), averaged amount of Support Vectors by year (#SVs/Year), number of samples

Table 6.2: Statistics on total dataset.(#: Number of samples, %: Percentage of the total)

Year	# AER site	# Raw data	# Clus.1	# Clus.2	# Clus.3	# Clus.4
2007	31	549,441	98,760	264,693	125,272	60,770
2008	36	491,323	113,347	238,651	104,334	34,991
2009	33	384,969	67,875	185,911	100,364	30,819
#		1,425,733	279,928	689,255	329,970	126,580
%			19.63	48.34	23.14	8.88

in testing dataset (#Ins./#Agg), and accuracy of SVR AOT in comparison with AERONET measurements on the pixel domain (CORR, RMSE, ME). Using the proposed approach, four clusters achieve acceptable accuracy. For instance validation, the prediction error increases gradually from cluster 1 to 4 (i.e. RMSE $\sim$ 0.0550, 0.0621, 0.0751, 0.0851, respectively), which shows decrease trend of AOT retrieval quality by group as mentioned in theory. SVR models underestimate AOT values, which is represented by negative ME. The general results on CORR, RMSE, and ME, calculated on overall pixels, are 0.7855, 0.0672 and -0.0029, respectively. For aggregate validation, the same results and conclusion can be observed. The final CORR, RMSE and ME obtained by the second validation are 0.8518, 0.058, and -0.0023 which are better than the first validation results since aggregated AOT values are used. These results are considered as good for AOT estimation at 1×1 km^2 of spatial resolution where inputs are very variant and noisy.

Table 6.3: Instance validation between SVR AOT and AERONET AOT on different clusters.

Clus.	# Training	# SVs/Year	# Ins.	CORR	RMSE	ME
1	58,721	28,606	279,928	0.8069	0.0550	-0.0039
2	70,246	31,372	689,255	0.8041	0.0621	-0.0033
3	46,169	26,779	329,970	0.7291	0.0751	-0.0003
4	48,413	21,211	126,580	0.6746	0.0851	-0.0102
All			1,422,349	0.7855	0.672	-0.0029

In order to compare SVR performance with the MODIS algorithm, instance and aggregate validations on the same dataset were carried out for MODIS AOT

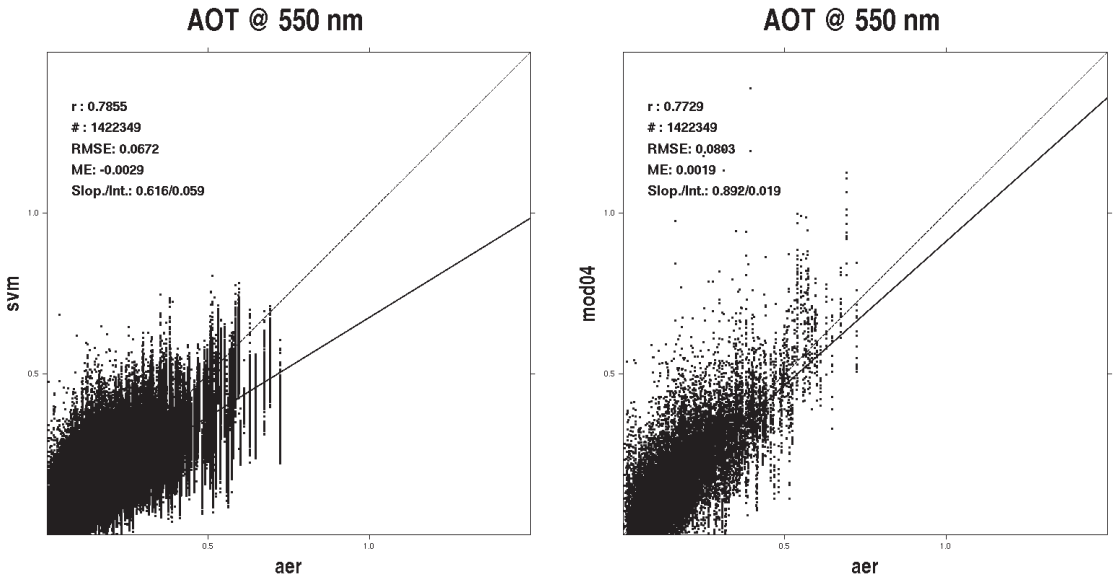
Table 6.4: Aggregate validation between SVR AOT and AERONET AOT on different clusters.

Clus.	# Training	# SVs/Year	# Agg.	CORR	RMSE	ME
1	58,721	28,606	1,803	0.8330	0.0567	-0.0032
2	70,246	31,372	2,169	0.8459	0.0571	-0.0019
3	46,169	26,779	1,747	0.7976	0.0647	0.0001
4	48,413	21,211	1,004	0.6574	0.0825	-0.0090
All			2,213	0.8518	0.058	-0.0023

and AERONET AOT. MODIS AOT obtained from MOD04 L2 products are resampled at 1×1 km^2 of spatial resolution for comparison. The results of both validations are shown in Table 6.5. Based on comparison with AERONET AOTs, SVR AOTs are more accurate than MODIS AOTs. In instance validation, general CORR/RMSE are 0.786/ 0.672 for SVR and 0.773/ 0.080 for MODIS, while in aggregate validation, these values are 0.852/0.058 versus 0.835/0.063. The SVR models underestimate AERONET AOTs while MODIS overestimates them (i.e. Instance/Aggregate ME $\sim$ -0.0029/-0.0023 versus 0.0019/0.0031 for SVR and MODIS, respectively). The overestimation of the MODIS alogrithm for small AOT values was stated in Remer et al. [2005] as well.

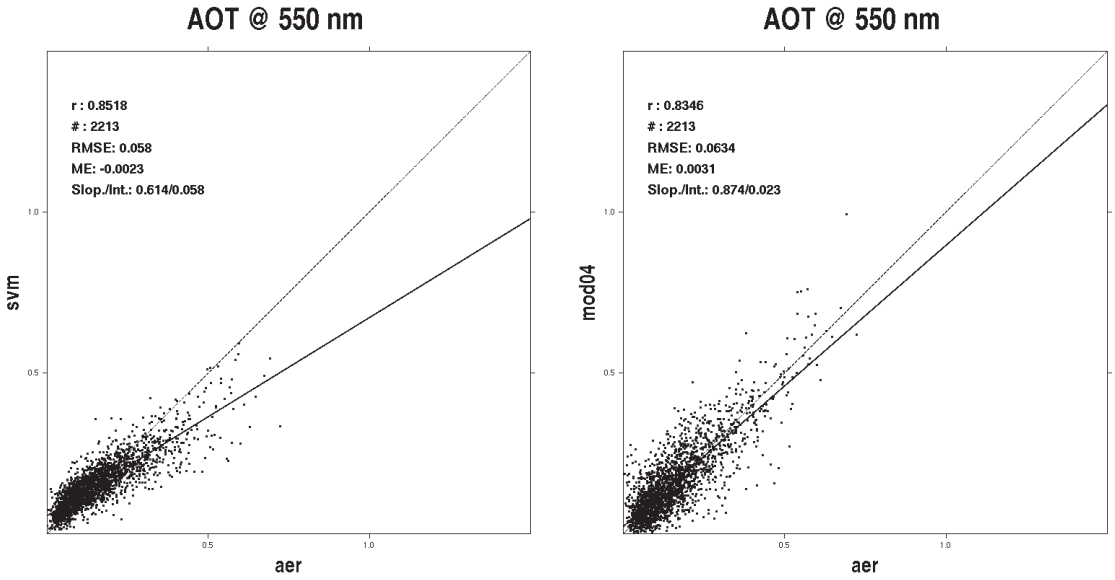
Figures 6.5 and 6.6 present scatterplots of SVR AOT - AERONET AOT and MODIS AOT - AERONET AOT in instance and agrregate cases. The bias can be observed in SVR models much more then MODIS models. The bias in SVR may be explained by imbalance datasets used for modelling (i.e. the samples of small AOT dominate in datasets) while in MODIS algorithm, the much bigger AOT values can be expressed by coarse aerosol modes hence the MODIS algorithm is less biased in these AOT ranges. The solution for SVR in this case will need further investigation.

The last experiment aims at investigating the relationship between SVR AOT and MODIS AOT around all AERONET sites. These results will be milestones to explain the results of the next validation in which SVR AOT maps are compared directly with MODIS AOT maps. Because AERONET AOT is considered as ground truth to validate both SVR and MODIS algorithms, this experiment just imply the relative difference between SVR AOT and MODIS AOT. The difference of AOT increases from cluster 1 to 4 but bigger difference can be observed at pixels



(a) SVR - AERONET (b) MODIS - AERONET

Figure 6.5: Instance scatterlots



(a) SVR - AERONET (b) MODIS - AERONET

Figure 6.6: Aggregate scatterlots

Table 6.5: Instance and Aggregate Validation between MODIS AOT and AERONET AOT on different clusters.

Clus.	Instance Validation				Aggregate Validation			
	# Ins.	CORR	RMSE	ME	# Agg.	CORR	RMSE	ME
1	279,928	0.8164	0.0588	-0.0147	1803	0.8169	0.0641	-0.0098
2	689,255	0.8010	0.0714	0.0010	2169	0.8266	0.0632	0.0029
3	329,970	0.7131	0.0931	0.0123	1747	0.8030	0.0711	0.01137
4	126,580	0.5790	0.1172	0.0156	1004	0.7127	0.0873	0.0149
All	1,422,349	0.7729	0.0803	0.0019	2213	0.8346	0.0634	0.0031

of clusters 3 and 4. The SVR AOT has underestimation trend in comparison with MODIS AOT, which is represented by negative ME (i.e. Instance/ Aggregate ME $\sim$ -0.0048/-0.0054). The general CORR/ RMSE calculated on overall pixels are 0.766/ 0.081 and 0.819/ 0.066 for instance and aggregate cases, respectively. However, only instance validation can be carried out over map domain (see details in the next section).

Table 6.6: Instance and Aggregate Validation between SVR AOT and MODIS AOT on different clusters.

Clus.	Instance Validation				Aggregate Validation			
	# Ins.	CORR	RMSE	ME	# Agg.	CORR	RMSE	ME
1	279,928	0.8421	0.0531	0.0108	1803	0.8065	0.0640	0.0066
2	689,255	0.8149	0.0705	-0.0043	2169	0.8192	0.0642	-0.0048
3	329,970	0.6906	0.0951	-0.0126	1747	0.7762	0.0749	-0.0112
4	126,580	0.5150	0.1221	-0.0258	1004	0.5768	0.1002	-0.0238
All	1,422,349	0.7664	0.0813	-0.0048	2213	0.8194	0.0663	-0.0054

3 .2 Map Domain

Map domain refers to all cloud-free pixels on images recorded by MODIS. The experiment carried out in map domain aims at evaluating quality of SVR models when they are used to derive AOT maps from MODIS L1B data. The validation of algorithms performing on map domain is challenging because there is no confident target for comparison. MOD04 L2 product, which provides aerosol monitoring at the global scale, is used in these experiments. However, as shown in the previous section, re-sampled MODIS AOT also presents low quality in comparison with

ground truth AERONET AOT for some cases (e.g. pixels of cluster 4). In this section, two experiments are considered. The first one validates directly SVR AOT maps with MODIS AOT maps, while the second considers this relationship by country in order to investigate spatial performance of SVR methodology.

3 .2.1 Validation by Map

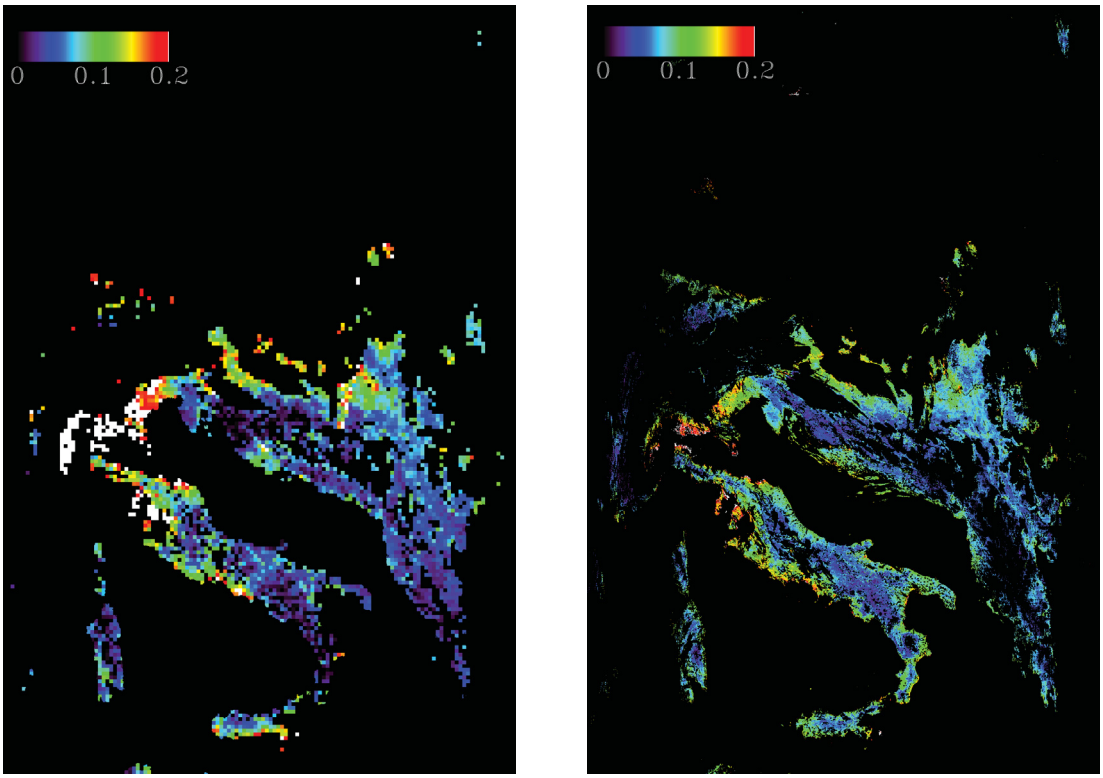
The MODIS L1B data in three years from 2007 to 2009 covering European areas are collected. The validation dataset consists of 163 images spreading over months of a year since AOT estimation strongly depends on season (see Table 6.7). After applying the map prediction framework, we received 163 AOT maps at 1×1 km^2 of spatial resolution. Corresponding MOD04 L2 maps are collected and resampled into 1 km maps by simply dividing one 10×10 km^2 pixel to one hundred of 1×1 km^2 pixels with same AOT values.

In the first experiment, every two corresponding maps are compared directly. Since the algorithms work on different spatial resolutions and use different methodologies for scanning good pixels, the two AOT maps are not completely overlapped. Therefore, the CORR and RMSE are calculated only on match pixels which have both SVR AOT and MODIS AOT. Moreover, only AOT maps with matching percentages bigger than 10% are considered in validation. An illustration of AOT map estimated by our SVR and MODIS algorithm is shown in Figure 6.7.

Table 6.7: Statistics of validation data.

Years	# Map	# Jan-Apr	# May-Aug	# Sep-Dec
2007	55	18	20	17
2008	50	17	22	11
2009	58	17	25	16
Total	163	52	67	44

Table 6.8 presents the numerical results of the experiment on validation datasets. SVR AOT of clusters 1 and 2, occupying a big quantity of data (45.42% and 44.01%, respectively), have good correlation coefficient and small error in comparison with MODIS AOT (CORR/RMSE $\sim$ 0.753/0.054 and 0.750/0.063, respectively). The worst case happens to cluster 4 with COR $\sim$ 0.505 and RMSE



(a) (b)

Figure 6.7: (a) The MOD04 L2 over-land AOT map at 10×10 km^2 and (b) The SVR over-land AOT map at 1×1 km^2 MODIS image acquired on January 29, 2008 at 09:55 UTC.

$\sim$ 0.1158. Regarding the validation between SVR AOT and MODIS AOT in pixel domain (see Table 6.6), the results obtained are consistent. In details, the decrease of CORR can be observed in clusters 1, 2, 3, and 4 while RMSE slightly increases. The overall CORR/RMSE are 0.7523/0.0614 in map domain versus 0.7664/0.0813 in pixel domain. However, ME over map domain is bigger than ME in pixel domain. The domination of pixels at clusters 1 and 2 can explain for this phenomenon. These results are generally compable between the two validations, which implies the stability of the proposed methodology when it is extended from pixel domain to map domain in a real application.

Table 6.8: SVR - MODIS validation for different clusters on the map domain (Clus.: cluster, #T: total number of pixels, %T: percentage of cluster pixels to total, #M: number of matched pixels, %M: percentage of matches to total number of cluster)

Clus.	# T	%T	#M	%M	COR	RMSE	ME
1	19,433,481	45.42%	16,266,684	0.842275	0.753316	0.054772	0.020683
2	18,832,321	44.01%	18,217,527	0.942983	0.750185	0.063028	0.014063
3	3,871,651	9.05%	3,773,566	0.924000	0.691041	0.081333	-0.001737
4	651,759	1.52%	370,262	0.776989	0.504973	0.115876	-0.008600
All	42,802,004		38,629,156	0.876382	0.752327	0.061459	0.016858

Moreover, additional comments shoud be given for ratio of pixels of different clusters. As shown in Table 6.8, clusters 1 and 2 hold most of pixels (45.42% and 44.01% of total, respectively). It is a result of strictly scanning process of good pixels in the map prediction framework. However, this process is necessary when estimation is carried out directly on values of 1 km pixel instead of averaged values of all good pixels at 500 m selected in a box sized 10×10 km^2 as in the MODIS algorithm.

Figure 6.8 presents the validation results of all datasets sorted by time. Season trends can be observed all three years from the AOT chart. SVR AOT maps in cold seasons (i.e. Jan-Mar/ Sep-Dec) have high correlation with MODIS AOT maps than in warm seasons (i.e. Apr-Jun, Jul-Sep). The biggest difference can be seen on maps in summer, which is represented by high RMSE and low CORR between SVR and MODIS AOT maps.

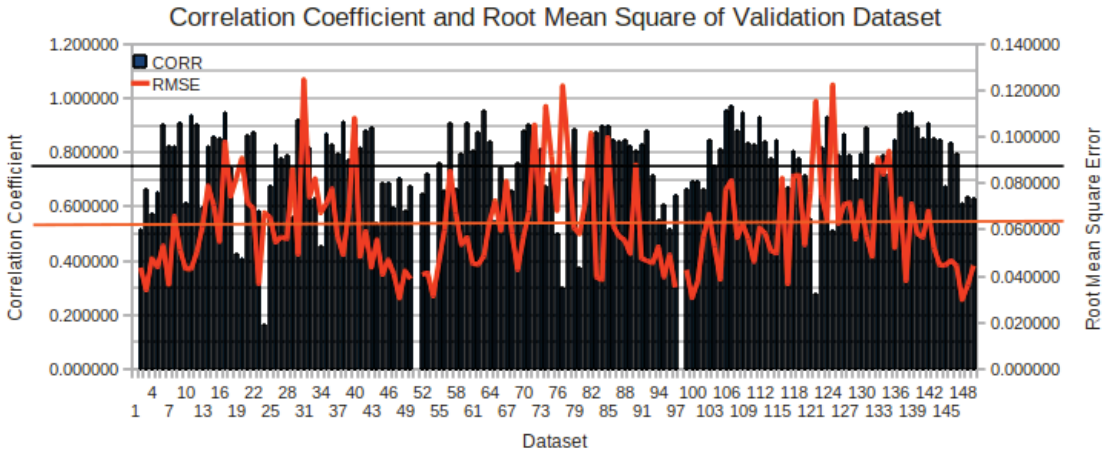


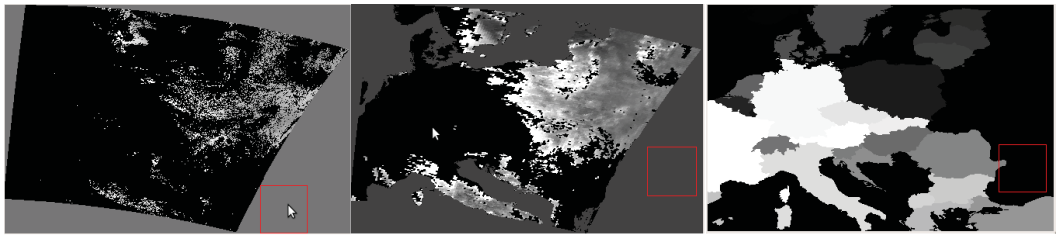
Figure 6.8: CORR and RMSE of Validation Datasets on Map Domain

3 .2.2 Validation by Country

In the second experiment on map domain, the validation is separated by country in order to investigate performance of the proposed SVR methodology in different spatial areas. Considered datasets consist of SVR AOT maps at 1×1 km^2 and MOD04 L2 maps at 10×10 km^2 of spatial resolutions. Moreover, a shapefile which is a geospatial vector data format for storing geometric location and associated attribute information is included in our validation datasets. The shapefile presenting European country's borders is applied to filter data over Countries. The shapefile is a geospatial vector data format for storing geometric location and associated attribute information. The validation methodology includes the following steps.

- Collection of validation datasets. The experiment uses the same datasets as in the previous one, detailed in Table 6.7.
- The SVR map prediction framework is applied on selected validation datasets and yields AOT maps at 1×1 km^2. Corresponding MOD04 L2 maps at 10×10 km^2 are selected and re-sampled into 1×1 km^2 spatial resolution.
- All satellite maps and European shapefile are projected on the Earth surface grid using the same projection Lat/Lon WGS84.

- All maps and the shapefile are overlapped. Then, corresponding data are recorded by Country. Figure 6.9 presents an example of SVR AOT map, MODIS AOT map, and corresponding overlaid shapefile that is created for two specific maps after they are projected on the Earth grid.
- For each Country, validation is carried out by calculating number of overlapped pixels (# Pixel), number of cover maps (# Map), CORR, RMSE and ME between corresponding SVR AOT and MODIS AOT.



(a) SVR AOT map @ 1×1 km^2 (b) MODIS AOT map @ 10×10 km^2 (c) Projected shape file

Figure 6.9: AOT maps and shapefile of the MODIS image acquired on April 25, 2008 at 10:00 UTC.

Table 6.9 presents validations results by Country in which country's validation parameters are calculated for each year and then averaged in three years. Out of 34 European countries considered, validation datasets cover 31 countries except Iceland (IS), Malta (MT), and Cyprus (CY). Over validation datasets, the number of overlapped pixels of 1 km is about 28 millions in which the biggest proportions are hold by areas of Poland (PL), France (FR), Germany (DE), Romania (RO), Italy (IT), Sweden (SE), Hungary (HU), and Finland (FI) (more than one million pixels for each country). Meanwhile, the smallest amounts are in areas such as Liechtenstein (LI), Luxembourg (LU), and Ireland (IE) (445, 23677, and 33257, respectively), which will be hence skipped in the following discussion.

In term of correlation coefficient between SVR AOT and MODIS AOT, values range from 0.503 to 0.921 and can be divided into two groups. The first one which includes Norway (NO), Portugal (PT), and Finland (FI) has low CORR (0.503, 0.684 and 0.695). The second group consists of 25 remained countries in which

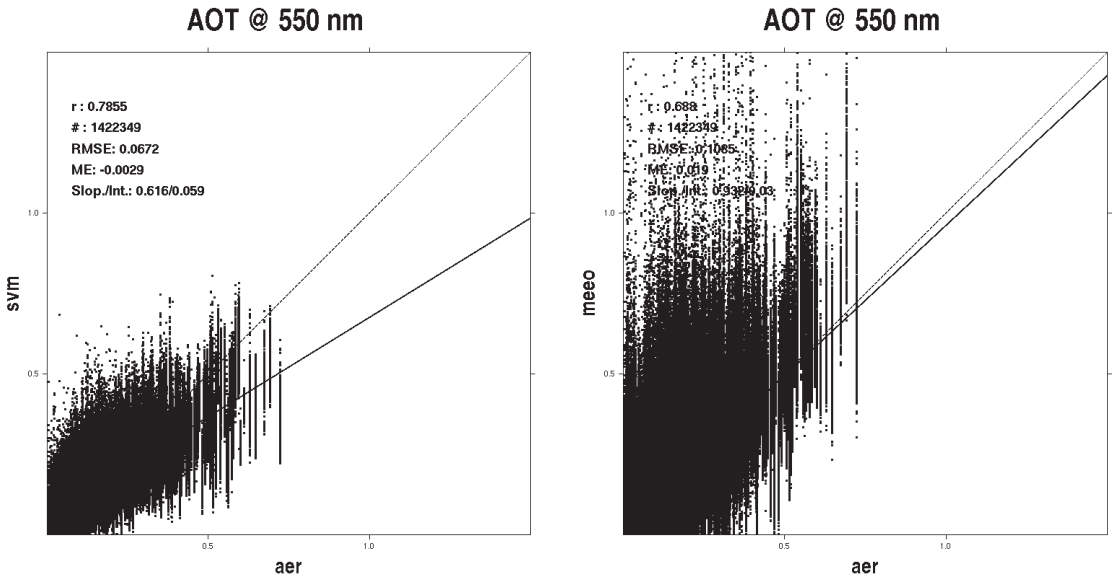
22 countries have CORR bigger than 0.800. Low correlations can be explained by rare AERONET sites where data are collected to produce data models over NO, PT and FI (see Figure 6.4). However, in most of the considered regions, the SVR models achieve good results in map domain. RMSE varies by Country and ranges from 0.0485 to 0.111. In table 6.9, results are presented in ascending order of RMSE. The average CORR and RMSE over considered countries are 0.831 and 0.064. Positive ME in average (0.005) presents the overestimation of SVR AOT with respect to MODIS AOT.

4 PM MAPPER and SVR Aerosol Optical Thickness at 1×1 km^2, some comparisons

In order to make a comparison between PM MAPPER AOT and SVR AOT at 1×1 km^2, additional validation for PM MAPPER AOT was carried out. The overall validation for PM MAPPER AOT was already presented in Chapter 4, while the evaluation for SVR AOT at 1×1 km^2 was described in the previous sections. Therefore, this additional validation evaluates PM MAPPER AOT over both pixel and map domains using the validation datasets and methodologies used in Section 6.3.

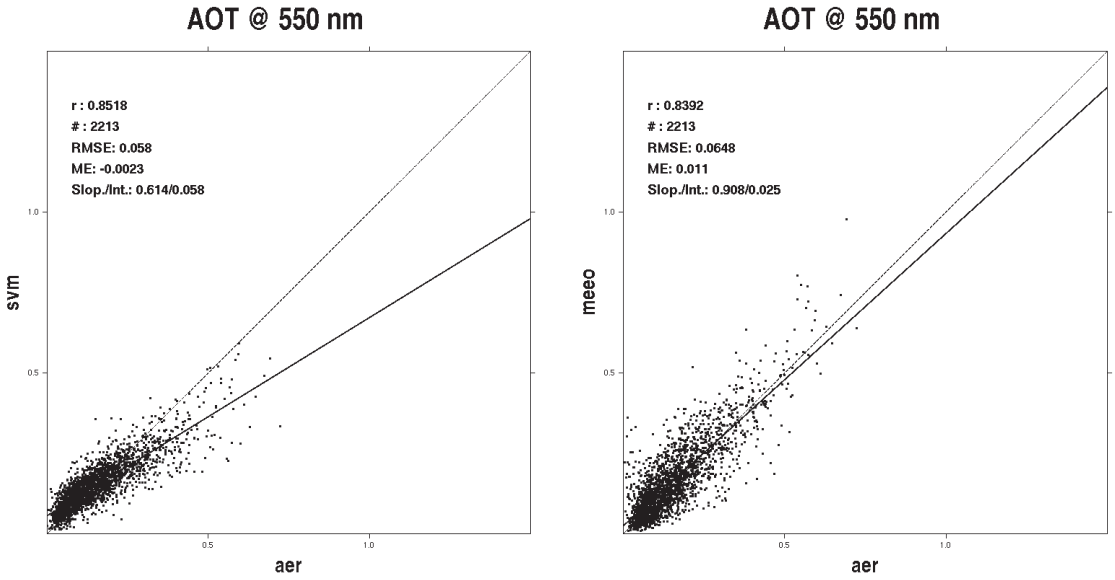
Figures 6.10 and 6.11 present validation results on pixel domain between PM MAPPER AOT and SVR AOT at 1×1 km^2 with respect to AERONET AOT. For instance validation, SVR AOT is more accurate than PM MAPPER AOT (CORR/RMSE $\sim$ 0.786/0.067 and 0.069/0.109, respectively). However, the PM MAPPER AOT is less biased than SVR AOT, although PM MAPPER values have more spread. Aggregate validation results imply similar conclusions with CORR/RMSE $\sim$ 0.852/0.058 and 0.839/0.065 for SVR AOT and PM MAPPER AOT, respectively.

Figure 6.12 presents the validation results on map domain for PM MAPPER AOT. It can be used for comparison with a similar validation between SVR AOT and MODIS AOT shown in Figure 6.8. Since PM MAPPER AOT is derived from the improved MODIS aerosol algorithms, it is well correlated with MODIS AOT. The overall average CORR/RMSE between PM MAPPER AOT and MODIS



(a) SVR - AERONET (b) PM MAPPER - AERONET

Figure 6.10: Instance scatterlots



(a) SVR - AERONET (b) PM MAPPER - AERONET

Figure 6.11: Aggregate scatterlots

AOT are 0.909/0.035, while those between SVR AOT and MODIS AOT are 0.752/0.061 (i.e. SVR AOT is less close to MODIS AOT two times in comparison with PM MAPPER AOT).

Observing the RMSE and CORR trend on all datasets, there is not much difference by dataset for the pair of PM MAPPER AOT and MODIS AOT maps, while strong variability can be observed between SVR AOT and MODIS AOT maps (see Figure 6.8). It can be explained by two points: (i) PM MAPPER is developed from the MODIS aerosol alogrithm and (ii) PM MAPPER and MODIS algorithms provide global models while SVR algorithm creates a regional model (for European areas). Therefore, besides the reason of using different techniques for AOT downscaling, the variability between SVR AOT and MODIS AOT also reflects effects of local measurements in SVR with respect to global AOT retrieval represented by MODIS AOT maps.

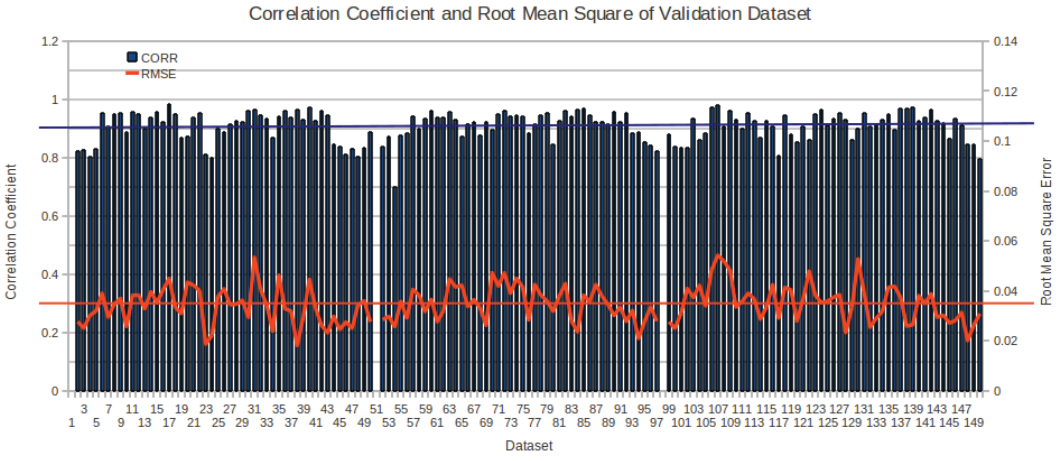


Figure 6.12: CORR and RMSE of Validation Datasets on Map Domain for PM MAPPER AOT

5 Conclusion

In this chapter, the methodology to estimate aerosol optical thickness at 1×1 km^2 from MODIS L1B data using SVR based on domain knowledge is presented. In

the proposed approach, the satellite-based data and ground-based measurements over areas of interest are collected and integrated using temporal and spatial constraints. After that, filtering using new integration constraints is applied to reduce the total amount of data and then a clustering technique is ised to separate them into four groups having different characteristics. Training datasets are then selected by some strategies for each cluster and the SVR technique is applied on each cluster to create corresponding data models. Finally, in the prediction framework, aerosol maps at spatial resolution of 1×1 km^2 are derived from MODIS L1B data using the SVR models.

Experiments were carried out on data from 2007 to 2009, covering European areas, in both pixel and map domain. The evaluation results show that the proposed approach deals well with two problems: (i) very large and noisy datasets and (ii) the going from pixel domain to map domain. The validation results in pixel domain show that the SVR methodology performs better than the MODIS algorithm. Extended to map application, the proposed SVR method is robust in prediction, which is justified by the similarity of CORR and RMSE in map and pixel domain. In the validation by country which looks at the spatial performance of SVR, the SVR prediction is stable in most of the covered areas as well.

Finally, some comparisons between PM MAPPER AOT and SVR AOT were carried out on the mentioned validation datasets. The validation results show that on pixel domain, SVR AOT is more accurate but more biased than PM MAPPER AOT in comparion with AERONET AOT. On map domain, PM MAPPER AOT maps are two times closer to MODIS AOT maps than SVR AOT maps to MODIS AOT maps. The validation between PM MAPPER and SVR AOT with another aerosol products such MISR AOT in the map domain are suggested because it will give more independent and objective results.

Table 6.9: Validation Results by Country

Code	Name	# Pixel	# Map	COR	RMSE	ME
EE	Estonia	261,462	99	0.868396	0.048515	0.019100
PT	Portugal	138,970	24	0.683860	0.048736	0.013472
SE	Sweden	1,897,810	130	0.803948	0.051939	0.004872
LV	Latvia	451,078	108	0.859885	0.054171	0.021641
LU	Luxemburg	23,677	113	0.948417	0.055878	0.022818
ES	Spain	756,017	86	0.800025	0.056285	-0.007739
CZ	Czech Republic	630,426	149	0.885758	0.057078	-0.005350
CH	Switzerland, Helvetia	160,721	131	0.855811	0.057325	-0.007667
DE	Germany	2,989,088	160	0.898717	0.057624	0.003829
FR	France	3,681,009	141	0.850485	0.057637	0.011961
SK	Slovakia	455,353	134	0.822694	0.059075	0.010408
FI	Finland, Suomi	1,162,321	78	0.695238	0.059263	0.032473
DK	Denmark	159,357	145	0.781215	0.059312	0.006878
AT	Austria	708,399	139	0.825595	0.059605	0.008826
PL	Poland	3,770,511	159	0.899658	0.060347	0.001062
RO	Romania	2,576,598	121	0.888492	0.060698	0.017723
LT	Lithuania	738,917	128	0.894415	0.062264	0.009214
NL	Netherlands	196,872	129	0.921203	0.062742	0.016241
BE	Belgium	256,262	121	0.883091	0.065659	0.001484
BG	Bulgaria	767,033	92	0.889166	0.065968	0.005143
HU	Hungary	1,373,091	131	0.852861	0.068059	0.021919
MK	Macedonia	146,586	91	0.911175	0.068472	-0.008624
NO	Norway	359,867	106	0.503442	0.070029	-0.019578
IT	Italy	2,274,898	125	0.875706	0.070847	-0.000461
GR	Greek	318,130	100	0.890871	0.071205	-0.017578
UK	United Kingdom	427,402	76	0.902495	0.071457	-0.024617
HR	Croatia	690,709	118	0.773915	0.071615	0.036812
LI	Liechtenstein	445	122	0.778836	0.071988	-0.021929
SI	Slovenia	290,881	121	0.803320	0.072788	0.032711
IE	Ireland	33,257	38	0.852035	0.102796	-0.058150
TR	Turkey, West	216,173	74	0.741390	0.110665	-0.048890
All		27,855,941	163	0.830815	0.063549	0.004831

Chapter 7

Conclusions and Future Work

This thesis aims at downscaling aerosol optical thickness from 10×10 km^2 to 1×1 km^2 from MODIS observations using physics and machine learning approaches. The finer spatial resolution brings great advantages for investigating aerosol distributions, detecting emission sources and making pollution management strategies. However, the challenges of this task are high noise coming from the measuring instruments, large uncertainties related to land surfaces over urban areas and cities, and finding appropriate downscaling algorithms which are still being investigated. In this thesis, the solutions are considered under two perspectives: dynamical downscaling by improving the algorithm for remote sensing of tropospheric aerosol from MODIS data and statistical downscaling using support vector regression.

Initially, the global MODIS aerosol algorithms including some adaptations are applied over smaller observation areas to derive AOT with spatial resolution at 3×3 km^2 (Nguyen et al. [2010a]; Nguyen et al. [2010b]) and 1×1 km^2. The downscaling algorithms perform at the global scale instead of over specific regions. The validation carried out on data covering Europe in three years showed good correlation coefficient and acceptable errors between achieved AOT at 1×1 km^2 with AERONET measurements (Campalani et al. [2011]). The proposed methodology was applied to develop a PM MAPPER software package (MEEO [2010a]). The AOT maps obtained by this software have been used in the SENSORE and AQUA projects in which AOTs were utilized to predict Particulate Matter concentration ($PM_{2.5/10}$) over the Emilia Romagna region in Italy and

over Austria (MEEO [2010b]; MEEO [2012]).

Since the proposed downscaling techniques follow the model driven approach, strong knowledge from domain experts will be required to improve and update new models or extend to other satellite sensor's measurements. On the other hand, the data driven approach using statistics/machine learning techniques in AOT retrieval provides a more flexible framework. Moreover, this approach has the advantage to reduce processing time, deals with data uncertainties and improves the accuracy in specific areas. Therefore, the second approach investigated the usage of SVR in deriving AOT at 1×1 km^2 of spatial resolution. In order to evaluate the performance of SVR for AOT retrieval in comparison with the MODIS algorithms, two SVR variants were firstly applied for AOT at 10×10 km^2 (Nguyen et al. [2010c]). The evaluations for SVRs were carried out by year, by season and by land cover properties. The experiment results showed that in some situations, the SVR approach outperformed the MODIS algorithm. Among two SVR variants, instance SVR gave better results than aggregate SVR but had higher computational time. The SVR approach improved AOT prediction quality especially over bright or scarce vegetation areas.

Based on the validation results of the SVR approach at 10×10 km^2, the instance SVR is selected for downscaling spatial resolutions down to 1×1 km^2. However, the application of SVR for AOT from 10×10 to 1×1 km^2 is not a trivial task because of very large and noisy datasets as a result of pursuing 100 times more detailed maps. Moreover, the application is extended from pixel domain to map domain in which data models created by data collected on sparse locations are applied on large and continuous map areas. To deal with the above mentioned problems, the proposed methodology is used SVR and domain knowledge (Nguyen et al. [2011]; Nguyen et al. [2012]). In this approach, the satellite-based data and ground-based measurements over areas of interest are collected and integrated using temporal and spatial constraints. After that, filtering using new integration conditions is applied to reduce total amount of data and clustering technique separates obtained data into four groups having different characteristics in order to "divide and conquer" a big dataset. Then, training datasets are selected by some strategies for each cluster and the SVR technique is applied on them to create corresponding data models. Finally, in the prediction framework,

aerosol maps at spatial resolution of 1×1 km^2 are derived from MODIS L1B data using the SVR models retrieved in the previous step. Experiments were carried out on data covering European areas from 2007 to 2009, in both pixel and map domain. Pixel domain refers to data collected around ground stations while map domain aims at all validated data in satellite maps. In pixel domain validation, the SVR methodology outperformed the MODIS algorithm. Extended to map application, the proposed SVR method was robust in prediction and gave stable results in most of areas covered.

Two proposed downscaling approaches are able to derive AOT maps at 1×1 km^2 from MODIS measurements with an acceptable quality. However, each approach has both advantages and disadvantages. In the first approach, AOT maps at 1×1 km^2 is well correlated to MODIS AOT at 10×10 km^2. Since the downscaling algorithms perform at global scale and did not apply any regional adaptation, the quality of PM MAPPER AOT at 1×1 km^2 is acceptable but worse than MODIS AOT at 10×10 km^2 in a validation which compares both PM MAPPER AOT and MODIS AOT with AERONET AOT measurements. The overall CORR/RMSE/ME are 0.862/0.068/0.032 and 0.864/0.060/0.012 for PM MAPPER AOT and MODIS AOT (see Table 4.1, Chapter 4). It can be explained by effects of MODIS instrument noise when AOT at finer spatial resolution is derived. However, PM MAPPER AOT is little less biased than MODIS AOT. The slope/intercept of the linear regression line are 0.970/0.036 and 0.927/0.024 for aggregate validation between PM MAPPER - AERONET AOT and MODIS AOT - AERONET AOT, respectively (see Figure 4.8, Chapter 4).

In the second approach using SVR replying on domain knowledge, the SVR AOT is more accurate than MODIS AOT in comparison with AERONET AOT. The overall CORR/RMSE/ME are 0.786/0.068/-0.0029 and 0.773/0.089/0.0019 for instance validation (0.852/0.058/-0.0023 and 0.835/0.063/0.0031 for aggregate validation) of SVR - AERONET AOT and MODIS - AERONET AOT, respectively (see Figures 6.5 and 6.6, Chapter 6). However, the SVR AOT is more biased than MODIS AOT (i.e. slope/intercept $\sim$ 0.616/0.059 and 0.892/0.019 for instance validation (0.614/0.058 and 0.874/0.023 for aggregate validation) in comparison with AERONET AOT (see Figures 6.5 and 6.6, Chapter 6).

The downscaling technique using the adapted MODIS aerosol algorithms pro-

vides AOT maps at 1×1 km^2 at global scale. In order to improve the accuracy of PM MAPPER AOT, some regional adaptations should be investigated and embedded into the current global algorithms. In this way, regional models are created and customized for specific areas. In the downscaling approach using SVR and domain knowledge, the bias problems of data models appear because samples with small AOT values dominate the training datasets. Therefore, the bias can be reduced if more samples with large AOT values were collected and used in the training phase.

In future, another aerosol product such as MISR AOT will be used for validations of PM MAPPER and SVR AOT in order to obtain more independent and objective results. Further investigations on the modeling process will be carried out to improve the quality of AOT retrievals using the SVR approach. Besides the usage of SVRs, additional techniques for spatial data will be considered to improve the prediction quality over all satellite maps. From the view point of applicability, the performance time will be considered more in depth and the proposed methodologies will be applied on data recorded by different satellite sensors.

Appendix

1 Satellite Datasets

Satellite datasets populated in database for 1×1 km^2 include 2600 PM MAPPER AOT maps over Europe ranging from 01/01/2007 to 31/12/2008 and 7500 MOD04 L2 maps obtained by Collection 005 over Europe from 01/01/2006 to 31/12/2009.

2 List of AERONET ground stations over Europe

Table 1: List of 105 AERONET sites over Europe

Name	(Lon, Lat)	Elev.	Description
Abisko	(18.816999,68.349998)	390	Abisko-Sweden
Ahi-De-Cara	(-3.22972,37.116669)	2103	AhideCara-Spain
Amsterdam-Island	(77.573334,-37.810001)	30	AmsterdamIsland-Netherlands
Andenes	(16.00861,69.278328)	379	Andenes-Norway
Arcachon	(-1.16322,44.663528)	11	Arcachon-France
Armilla	(-3.24222,37.133331)	691	Armilla-Spain
ATHENS-NOA	(23.775,37.987999)	130	Athens-Greece
Autilla	(-4.60306,41.997219)	873	AutilladelPino-Spain

Avignon	(4.87807,43.932751)	320	Avignon-France
Barcelona	(2.11697,41.38567)	125	Barcelona-Spain
Belsk	(20.79167,51.83667)	190	Belsk-Poland
Birkenes	(8.25231,58.388451)	230	Birkenes-Norway
BORDEAUX	(-0.57917,44.788029)	40	Bordeaux-France
Bucarest	(26.525,44.450001)	44	Bucarest-Romania
Bucharest-Inoe	(26.02972,44.348061)	93	Bucharest-Romania
Caceres	(-6.34347,39.47858)	397	Caceres-Spain
Carpentras	(5.05833,44.083328)	100	Carpentras-France
Clermont-Ferrand	(2.96194,45.759998)	1464	ClermontFerrand-France
CRETE	(25.66783,35.337669)	140	Crete-Greece
Creteil	(2.44278,48.788502)	57	Cretil-France
Davos	(9.8438,46.81292)	1596	Davos-Switzerland
Dunkerque	(2.36812,51.035351)	0	Dunkerque-France
Eforie	(28.632219,44.075001)	40	Eforie-Romania
El-Arenosillo	(-6.73347,37.105)	0	ElArenosillo-Spain
Ersa	(9.35929,43.00367)	80	Ersa-France
ETNA	(15.01943,37.613499)	736	Mt.Etna-Italy
EVK2-CNR	(86.813332,27.958611)	5050	Nepal-Italy
Fontainebleau	(2.68028,48.40667)	85	Fontainebleau-France
FORTH-CRETE	(25.282419,35.332691)	20	ForthCrete-Greece
Gerlitzen	(13.90667,46.678329)	1900	Gerlitzen-Austria
Gotland	(18.950001,57.916672)	10	GotlandIsland-Sweden
Granada	(-3.605,37.164001)	680	Granada-Spain
Hamburg	(9.97333,53.568329)	105	Hamburg-Germany
Helgoland	(7.88736,54.17786)	33	HelgolandIsland-Germany
Helsinki	(24.960649,60.203732)	52	Helsinki-Finland
Helsinki-Lighthouse	(24.926359,59.948971)	20	HelsinkiLighthouse-Finland

Hornsund	(15.55,77)	0	Hornsund-Norway
Hyytiala	(24.295891,61.846451)	191	Hyytiala-Finland
IFT-Leipzig	(12.43528,51.352501)	125	Leipzig-Germany
IMAA-Potenza	(15.72,40.599998)	820	Potenza-Italy
IMC-Oristano	(8.5,39.91)	10	IMCOristano,Sardinia-Italy
IMS-METU-ERDEMLI	(34.255001,36.564999)	3	Turkey
Irkutsk	(103.086639,51.799801)	670	Irkutsk-Russia
ISDGM-CNR	(12.33198,45.436981)	20	ISDGM-CNR,Venezia-Italy
Ispra	(8.6267,45.803051)	235	Ispra-Italy
Izana	(-16.49906,28.30932)	2391	Izana-Spain
Kanzelhohe-Obs	(13.907,46.678001)	1526	Kanzelhoehe-Austria
Karlsruhe	(8.4279,49.0933)	140	Karlsruhe-Germany
Kuopio	(27.63361,62.89241)	105	Kuopio-Finland
Kyiv	(30.49667,50.363609)	200	Kyiv-Ukraine
Laegeren	(8.35139,47.480282)	735	Laegeren-Switzerland
Lampedusa	(12.63167,35.51667)	45	Lampedusa-Italy
Lannion	(-3.46194,48.730831)	15	Lannion-France
Lecce-University	(18.111389,40.33511)	30	Lecce-Italy
Le-Fauga	(1.2846,43.384232)	193	LeFauga-France
Lille	(3.14167,50.611671)	60	Lille-France
London-UCL-UAO	(-0.1311,51.524269)	46	London-UnitedKingdom
Longyearbyen	(15.649,78.222832)	30	Longyearbyen-Norway
Mace-Head	(-9.9,53.330002)	20	Galway-Ireland
Mainz	(8.3,49.999001)	150	Mainz-Germany
Malaga	(-4.4775,36.715)	40	Malaga-Spain
Messina	(15.56683,38.197498)	15	Messina-Italy
Minsk	(27.601,53.919998)	200	Minsk-Belarus
Modena	(10.94528,44.631672)	56	Modena-Italy

Moldova	(28.815599,47.000099)	205	Kishinev-Moldova
Moscow-MSU-MO	(37.509998,55.700001)	192	Moscow-Russia
Munich-Maisach	(11.258,48.209)	520	Munich-Germany
OHP-OBSERVATOIRE	(5.71,43.935001)	680	OHPObservatoire-France
Oostende	(2.925,51.224998)	23	Oostende-Belgium
Palaiseau	(2.20833,48.700001)	156	Palaiseau-France
Palencia	(-4.51569,41.988571)	750	Palencia-Spain
Palgrunden	(13.1515,58.755329)	49	Granvik-Sweden
Paris	(2.33333,48.866669)	50	Paris-France
Pic-du-midi	(0.1413,42.937222)	2898	PicduMidi-France
Pitres	(-3.22222,36.933331)	1252	Pitres-Spain
Porquerolles	(6.16139,43.001389)	10	PorquerollesIsland-France
REUNION-ST-DENIS	(55.48333,-20.883329)	0	LaReunion-France
Rome-Tor-Vergata	(12.64733,41.83955)	130	Rome-Tor-Vergata
Rossfeld	(7.62475,48.33514)	167	Rossfeld-France
SAGRES	(-8.87352,37.04771)	26	SAGRES-Italy
Saint-Mandrier	(5.94417,43.06694)	44	SaintMandrier-France
Salon-de-Provence	(5.12028,43.60556)	60	SalondeProvence-France
Sevastopol	(33.51733,44.615829)	80	Sevastopol-Ukraine
Seysses	(1.25972,43.50333)	179	Seysses-France
SMHI	(16.15,58.580002)	0	Norrkoping-Sweden
Sodankyla	(26.62957,67.366623)	184	Sodankyla-Finland
Tarbes	(0.08333,43.25)	350	Tarbes-France
Tenerife	(-16.633329,28.033331)	10	Tenerife-Spain
The-Hague	(4.32682,52.110481)	18	TheHague-Netherlands
Thessaloniki	(22.959999,40.630001)	60	Thessaloniki-Greece

Tomsk	(85.046997,56.477329)	130	Tomsk-Russia
Toravere	(26.459999,58.255001)	70	Travere-Estonia
Toulon	(6.00944,43.135559)	50	Toulon-France
Toulouse	(1.37389,43.574718)	150	Toulouse-France
Tremiti	(15.49011,42.117962)	4	Tremiti-Italy
TUBITAK-UZAY-Ankara	(32.778,39.890999)	924	Ankara-Turkey
Tuz-Golu	(33.336109,38.749439)	907	Tuz-Golu
Ussuriysk	(132.163498,43.700401)	280	Ussuriisk-Russia
Valladolid-Sci	(-4.7148,41.6562)	701	Valladolid-Spain
Venise	(12.5083,45.3139)	10	Venice-Italy
Villefranche	(7.32889,43.683891)	130	Villefranche-France
Xanthi	(24.918949,41.14677)	54	Xanthi-Greece
Yakutsk	(129.366669,61.661671)	118	Yakutsk-Russia
Yekaterinburg	(59.544998,57.03833)	300	Yekaterinburg-Russia
Zvenigorod	(36.775002,55.695)	200	Zvenigorod-Russia

References

S. A. Ackerman, K. I. Strabala, W. P. Menzel, R. A. Frey, C. C. Moeller, and L. E. Gumley. Discriminating clear sky from clouds with MODIS. *J. Geophys. Res.*, 103:32 139–32 140, 1998. 34

AERONET. AErosol Robotic Network (AERONET), October 2011. URL `http://aeronet.gsfc.nasa.gov/`. ix, 30

Z. Ahmad and R. S. Fraser. An iterative radiative transfer code for the oceanatmosphere system. *J. Atmos. Sci.*, 39:656665, 1982. 35

J. Al-Saadi, J. Szykman, R. B. Pierce, C. Kittaka, D. Neil, D. A. Chu, L. Remer, L. Gumley, E. Prins, L. Weinstock, C. MacDonald, R. Wayland, F. Dimmick, and J. Fishman. Improving national air quality forecasts with satellite aerosol observations. *Bull. Am. Met. Soc.*, 86(9):1249–1261, 2005. 2

A. Baraldi, V. Puzzolo, P. Blonda, L. Bruzzone, and C. Tarantino. Automatic spectral rule-based preliminary mapping of calibrated landsat tm and etm+ images. *IEEE Transactions on Geoscience and Remote Sensing*, 44(9):2563–2586, 2006. 55

D. Beal, F. Baret, C. Bacour, and X-F. Gu. A method for aerosol correction from the spectral variation in the visible and near infrared: application to the meris sensor. *Int. J. Remote Sens*, 28:761–779, 2007. 23

A. Benedetti, J.-J. Morcrette, O. Boucher, A. Dethof, R. J. Engelen, M. Fisher, H. Flentje, N. Huneeus, L. Jones, J. W. Kaiser, S. Kinne, A. Mangold,

M. Razinger, A. J. Simmons, and M. Suttie. Aerosol analysis and forecast in the European Centre for Medium-Range Weather Forecasts Integrated Forecast System: 2. Data assimilation. *J. Geophys. Res*, 144(D13205): doi:10.1029/2008JD011115, 2009. 2

U. Böttger. Retrieval of Aerosol Characteristics from Bidirectional, Spectral and Polarization Measurements of POLDER/ADEOS. *Retrieval of Aerosol Characteristics from Bidirectional, Spectral and Polarization Measurements of POLDER/ADEOS . ALPS , 1999 , Meribel (France)*, 2000. 22

J. Brajard, A. Niang, S. Sawadogo, F. Fell, R. Santer, and S. Thiria. Estimating aerosol parameters above the ocean from meris observations using topological maps. *Int. J. Remote Sens.*, 28:781–795, 2007. 23

P. Campalani, T. N. T. Nguyen, S. Mantovani, P. Campalani, M. Bottoni, and G. Mazzini. Validation of PM MAPPER aerosol optical thickness retrieval at 1 x 1 km^2 of spatial resolution. *Proceedings of the 19th International Conference on Software, Telecommunications and Computer Networks*, 2011. 6, 7, 54, 68, 103

A. D. Castanho, L. T. Molina, R. Prinn, and P. Artaxo. Comparison of Aerosol Optical Properties and High Spatial Resolution Retrievals with MODIS in Sao Paulo, Mexico City and Beijing Megacities. *Ame. Geo. Uni.*, 2007. 5, 77

A. D. Castanho, J. V. Martins, and P. Artoxo. Modis aerosol optical depth retrievals with high spatial resolution over an urban area using the critical reflectance. *J. Geo. Res.*, 113, 2008. 5, 25

C. C. Chang and C. J. Lin. LIBSVM - A Library for Support Vector Machines, 2011. URL `http://www.csie.ntu.edu.tw/~cjlin/libsvm/`. 71, 83, 84

Q. Chen and Y. Shao. The Application of Improved BP Neural Network Algorithm in Urban Air Quality Prediction: Evidence from China. *Proc. 2008 IEEE Pacific-Asia Workshop on Computational Intelligence and Industrial Application (PACIIA 2008)*, pages 160–163, 2008. 3

V. Cherkassky and Y. Ma. Practical Selection of SVM Parameters and Noise Estimation for SVM Regression. *Neural Networks*, 17:113–126, 2004. 84

S.A. Christopher and J. Wang. Intercomparison between multi-angle imaging spectroradiometer (misr) and sunphotometer aerosol optical thickness in dust source regions over china: Implications for satellite aerosol retrievals and radiative forcing calculations. *Tellus.*, B56(5):451–456, 2004. 21

D. A. Chu, Y. J. Kaufman, C. Ichoku, L. A. Remer, D. Tanré, and B. N. Holben. Valication of MODIS aerosol optica depth retrieval over land. *Geophysical Research Letters*, 29(12):1617, 2002. 2

D. A. Chu, Y. J. Kaufman, G. Zibordi, J. D. Chern, J. Mao, C. Li, and B. N. Holben. Global monitoring of air pollution over land from EOS - Terra MODIS. *J. Geophys. Res*, 108(D21):4661, 2003. 2

CIMSS. International MODIS/AIRS processing Package, 2011. URL `http://cimss.ssec.wisc.edu/imapp/`. 57, 67

J. L. Deuzé, F. M. Bréon, C. Devaux, P. Goloub, M. Herman, B. Lafrance, F. Maignan, A. Marchand, F. Nadal, G. Perry, and D. Tanré. Remote sensing of aerosols over land surfaces from polder-adeos-1 polarized measurements. *J. Geophys. Res.*, 106(D5):4913–4926, 2001. 17

D. J. Diner, W. A. Abdou, T. P. Ackerman, K. Crean, H. R. Gordon, R. A. Kahn, J. V. Martonchik, S. McMuldroch, S. R. Paradise, B. Pinty, M. M. Verstraete, M. Wang, and R. A. West. Misr level 2 aerosol retrieval algorithm theoretical basis. *Jet Propul. Lab. Pasadena, CA, JPL-D 11400, Rev. G.*, 2008. 17, 27

D.J. Diner, W.A. Abdou, C.J. Bruegge, J.E. Conel, K.A. Crean, B.J. Gaitley, M.C. Helmlinger, R.A. Kahn, J.V. Martonchik, S.H. Pilorz, and B.N. Holben. MISR aerosol optical depth retrievals over southern Africa during the SAFARI-2000 dry season campaign. *Geophys. Res. Lett.*, 28:31273130, 2001. 21

Levy et. al. Evaluation of the modis retrievals of dust aerosol over the ocean during PRIDE. *J. Geophys. Res.*, 2003:8594, 108. 35

H. J. Fowler, S. Blenkinsop, and C. Tebaldi. Review linking climate change modelling to impacts studies: recent advances in downscaling techniques for hydrological modelling. *Int. J. Cli.*, 27:1547–1578, 2007. 4, 24

B. C. Gao, Y. J. Kaufman, D. Tanré, , and R. R. Li. Distinguishing tropospheric aerosols from thin cirrus clouds for improved aerosol retrievals using the ratio of 1.38-μm and 1.24μm channels. *Geophys. Res. Lett.*, 29:1890, 2002. 34

Y. Gao, Y. J. Kaufman, D. Tanré, and P. G. Falkowski. Seasonal distribution of aeolian iron fluxes to the global ocean. *Geophys. Res. Lett.*, 28:29–33, 2000. 2

P. Goloub, D. Tanré, J. L. Deuzé, M. Herman, A. Marchand, and F. M. Bréon. Validation of the first algorithm for deriving the aerosol properties over the ocean using the POLDER/ADEOS measurements. *IEEE Trans. Geosci.Remote Sens.*, 37(3):15861596, 1999. 20

H. R. Gordon and M. Wang. Retrieval of water-leaving radiance and aerosol optical thickness over the oceans with seawifs: A preliminary algorithm. *Appl. Opt.*, 33:443–452, 1994. 13, 20, 27

M.D. Graaf and P. Stammes. SCIAMACHY absorbing aerosol index calibration issues and global results from 20022004. *Atmos. Chem. Phys.*, 5:2385–2394, 2005. 27

M. Griggs. Measurements of atmospheric aerosol optical thickness over water using erts-1 data. *J. Air Pollut. Control. Assoc.*, 25:622626, 1975. 11

B. Han, S. Vucetic, A. Braverman, and Z. Obradovic. A statistical complement to deterministic algorithms for the retrieval of aerosol optical thickness from radiance data. *Engineering Applications of Artificial Intelligence*, 19:787–795, 2006. 3

H. Heiberg, S. Tsyro, A. Valdebenito, and H. Schyberg. Strategic review of satellite products and recommendations for future comparison with model results and data assimilation. *met.no note*, 2010. 18

J. R. Herman, P. K. Bhartia, O. Torres, C. Hsu, C. Seftor, and E. Celarier. Global distribution of uv-absorbing aerosols from nimbus 7/toms data. *J. Geophys. Res.*, 102:1691116922, 1997. 19, 20, 27

R. M. Hoff and S. A. Christopher. Remote sensing of particulate pollution from space: Have we reached the promised land? *J. Air Waste Manage. Assoc.*, 59: 645675, 2009. ix, 15

W. Von Hoyningen-Huene, M. Freitag, and J. P. Burrows. Retrieval of aerosol optical thickness over land surfaces from top-of-atmospher radiance. *J. Geophys. Res.*, 108:4260, 2003. 17, 21

C. Ichoku, D. A. Chu, S. Mattoo, Y. J. Kaufman, L. A. Remer, D. Tanré, I. Slutsker, and B. N. Holben. A spatio-temporal approach for global validation and analysis of MODIS aerosol products. *Geophysical Research Letters*, 29:8006–8010, 2002. 2, 4, 60, 61, 70, 79, 80, 81

C. Jamet, C. Moulin, and S. Thiria. Monitoring aerosol optical properties over the mediterranean from seawifs images using a neural network inversion. *Geophys. Res. Lett.*, 31(L13107):doi:10.1029/2004GL019951, 2004. 22

W. Karush. *Minima of functions of several variables with inequalities as side constraints*. Dep. of Mathematics, Univ. of Chicago, 1939. 47

Y. J. Kaufman and D. Tanré. Algorithm for Remote Sensing of Tropospheric Aerosol from MODIS. *MODIS ATBD*, 1997. x, 2, 3, 4, 16, 21, 52, 81, 82

Y. J. Kaufman, D. Tanré, and O. Boucher. A satellite view of aerosols in the climate system. *Nature.*, 419:215223, 2002. 10

K. R. Knapp, T. H. Vonder Harr, and Y. J. Kaufman. Aerosol optical depth retrieval from goes-8: Uncertainty study and retrieval validation over south america. *J. Geophys. Res*, 107(D7):4055, 2002. 18, 27

H. W. Kuhn and A. W. Tucker. Nonlinear programming. *Proc. 2nd Berkerley Symposium on Mathematical Statistics and Probabilistics*, pages 481–492, 1959. 47

D. J. Lary, L. A. Remer, D. MacNeill, B. Roscoe, and S. Paradise. Machine Learning Bias Correction of MODIS Aerosol Optical Depth. *IEEE Geo. Rem. Sen. Let.*, 6(4):694–698, 2009. 3, 23, 61, 63, 80

K. H. Lee, Z. Li, Y. J. Kim, and A. Kokhanovsky. Atmospheric Aerosol Monitoring from Satellite Observations: A History of Three Decades. *Atmospheric and Biological Environmental Monitoring*, pages 13–39, 2009. 18

R. C. Levy, L. A. Remer, J. V. Martins, Y. J. Kaufman, A. Plana-Fattori, J. Redemann, and B. Wenny. Evaluation of the MODIS Aerosol Retrievals over Ocean and Land during CLAMS. *Journal of the Atmospheric Sciences*, 62(4): 974–992, 2005. 2

R. C. Levy, L. A. Remer, S. Mattoo, E. F. Vermote, and Y. J. Kaufman. Second-generation operational algorithm: Retrieval of aerosol properties over land from inversion of Moderate Resolution Imaging Spectroradiometer spectral reflectance. *J. Geo. Res.*, 112, 2007. 21

C. Li, A. K. H. Lau, J. Mao, and D. A. Chu. Retrieval, Validation, and Application of the 1-km Aerosol Optical Depth from MODIS Measurements over Hong Kong. *IEEE Transactions on Geoscience and Remote Sensing*, 43:2650–2658, 2005. 4, 5, 25, 77

R. R. Li, Y. J. Kaufman, B. C. Gao, and C. O. Davis. Remote sensing of suspended sediments and shallow coastal waters. *IEEE Trans. Geosci. Remote Sens.*, 41:559–566, 2003. 34

Z. Li, A. Khananian, R. H. Fraser, and J. Cihlar. Automatic detection of fire smoke using artificial neural networks and threshold approaches applied to avhrr imagery. *IEEE Trans. Geo.Remote. Sens.*, 39(9), 2001. 22

G. P. Limone. AOT - SVR application, a software tool to predict aerosol optical thickness maps from MODIS data. Undergraduation Thesis, 2011. 85

Y. Liu, J.A. Sarnat, B.A. Coull, P. Koutrakis, and D.J. Jacob. Validation of Multiangle Imaging Spectroradiometer (MISR) aerosol optical thickness mea-

surements using Aerosol Robotic Network (AERONET) observations over the contiguous United States. *J. Geophys. Res.*, 109(D06205, 2004. 22

W. Lu, W. Wang, A. Y. T. Leung, S. M. Lo, R. K. K. Yuen, Z. Xu, and H. Fan. Air Pollutant Parameter Forecasting Using Support Vector Machine. *Proc. The 2002 International Joint Conference on Neural Network (IJCNN02)*, pages 630–635, 2002. 3

Y. Ma, W. Gong, Z. Zhu, L. Zhang, and P. Li. Cloud amount and aerosol characteristic research in the atmosphere over hubei province, china. *IEEE/IGARSS*, III:631–634, 2009. 23

J.V. Martins, D. Tanré, L. Remer, Y. Kaufman, S. Matto, and R. Levy. MODIS cloud screening for remote sensing of aerosols over oceans using spatial variability. *Geophysical Research Letters*, 29:8009, 2002. 34, 56, 86

J.V. Martonchik, D.J. Diner, R.A. Kahn, T.P. Ackerman, M.M. Verstraete, B. Pinty, and H.R. Gordon. Techniques for the retrieval of aerosol properties over land and ocean using multiangle imaging. *IEEE Trans. Geosci. Remote Sens.*, 36:1212–1227, 1998. 21, 27

MEEO. Pm mapper. Internal Documents, MEEO S.r.l, Ferrara, Italy, 2010a. 6, 54, 67, 103

MEEO. Sensor Project. Internal Documents, MEEO S.r.l, Ferrara, Italy, 2010b. 7, 104

MEEO. Aqua Project. Internal Documents, MEEO S.r.l, Ferrara, Italy, 2012. 7, 104

Meteorological Environmental Earth Observation MEEO. SOIL MAPPER®, 2011. URL `http://www.meeo.it/`. 71

M. I. Mishchenko, I. V. Geogdzhayev, B. Cairns, W. B. Rossow, and A. A. Lacis. Aerosol retrievals over the ocean by use of channels 1 and 2 AVHRR data: Sensitivity analysis and Preliminary results. *Appl. Opt.*, 38:7325–7341, 1999. 19, 27

T. N. T. Nguyen, S. Mantovani, and M. Bottoni. Estimation of Aerosol and Air Quality Fields with PM MAPPER An Optical Multispectral Data Processing Package. *ISPRS TC VII Symposium 100 year ISPRS*, XXXVIII(7A):257–261, 2010a. 6, 7, 54, 56, 66, 74, 77, 103

T. N. T. Nguyen, S. Mantovani, and M. Bottoni. PM MAPPER: An air quality monitoring system with fine spatial resolution product and integrated surface information. *Proc. of Hyperspectral Workshop 2010, European Space Agency (ESA)*, XXXVIII(7A):257–261, 2010b. 6, 7, 54, 56, 66, 77, 103

T. N. T. Nguyen, S. Mantovani, P. Campalani, M. Cavicchi, and M. Bottoni. Aerosol Optical Thickness Retrieval from Satellite Observation Using Support Vector Regression. *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications - 15th Iberoamerican Congress on Pattern Recognition (CIARP2010)*, pages 492–499, 2010c. 6, 22, 60, 69, 80, 104

T. N. T. Nguyen, S. Mantovani, and P. Campalani. Validation of support vector regression in deriving aerosol optical thickness maps at 1 x 1 km^2 spatial resolution from satellite observations. *Proc. The IEEE International Symposium on Signal Processing and Information Technology*, pages 551–556, 2011. 6, 7, 78, 104

T. N. T. Nguyen, S. Mantovani, P. Campalani, and G. P. Limone. Downscaling Aerosol Optical Thickness from Satellite Observations using Support Vector Regression replied on Domain Knowledge. *Proc. The 1st International Conference on Pattern Recognition Applications and Methods*, pages 230–239, 2012. 6, 7, 78, 104

J. E. Nichol and M. S. Wong. High Resolution Remote Sensing of Densely Urbanised Regions: a Case Study of Hong Kong. *Sensor*, 9:4695–4780, 2009. 5, 25, 77

Z. Obradovic, D. Das, V. Radosavljevic, K. Ristovski, and S. Vucetic. Spatio-Temporal Characterization of Aerosols through Active Use of Data from Multiple Sensors. *Proc. ISPRS TC VII Symposium 100 year ISPRS*, XXVIII(7B): 424–429, 2010. 3, 22, 23, 80

Y. Okada, S. Mukai, and I. Sano. Neural Network Approach for Aerosol Retrieval. *Proc. IEEE 2001 International Geoscience And Remote Sensing Symposium (IGARSS01)*, 4:1716–1718, 2001. 3, 22

M. M. Oo, E. Hernandez, M. Jerg, B. M. G. Moshary, and S. A. Ahmed. Improved MODIS Aerosol Retrieval Using Modified VIS/MIR Surface Albedo Ratio over Urban Scenes. *IEEE 2008 International Geoscience and Remote Sensing Symposium (IGARSS08)*, 3:977–979, 2008. 5, 25, 77

S. Osowski and K. Garanty. Wavelets and Support Vector Machine for Forecasting the Meteorological Pollution. *Proc. The 7th Nordic Signal Processing Symposium (NORSIG)*, pages 158–161, 2006. 3

E. Osuna, R. Freund, and F. Girosi. An improved training algorithm for support vector machines. *Principe J., Gile L., Morgan N., and Wilson E. (Eds), Neural Networks for Signal Processing VII - Proceedings of the 1997 IEEE Workshop*, pages 276–285, 1997. 50

J. Platt. *Fast training of support vector machines using sequential minimal optimization.* MIT Press, Scholkopf B., Burges C.J.C., and Smola A.J. (Eds) Advances in Kernel Methods - Support Vector Learning edition, 1999. 50

R. Ramakrishnan, J. J. Schauer, L. Chen, Z. Huang, M. M. Shafer, D. S. Gross, and D. R. Musicant. The EDAM project: Mining atmospheric aerosol datasets. *Int. J. Int. Sys.*, 20:759–787, 2005. 3

L. Remer, R. Levy, S. Mattoo, R. Kleidman, I. Koren, and V. Martins. The MODIS aerosol algorithm: Critical evaluation and plans for Collection 6. *Geo. Res. Abs.*, 12, 2010. 5, 25

L. A. Remer, D. Tanré, and Y. J. Kaufman. Algorithm for Remote Sensing of Troposheric Aerosol from MODIS: Collection 5. *MODIS ATBD*, 2004. ix, 2, 3, 32, 38, 52, 86

L. A. Remer, Y. J. Kaufman, D. Tanré, S. Matto, D. A. Chu, J. V. Martins, R. R. Li, C. Ichoku, R. C. Levy, R. G. Kleidman, T. F. Eck, E. Vermote, and

B. N. Holben. The modis aerosol algorithm, products, and validation. *J. Atm. Sci.*, 62:947–973, 2005. ix, 2, 17, 21, 27, 33, 90

V. V. Salomonson, w. L. Barnes, P. W. Maymon, H. E. Montgomery, and H. Ostrow. MODIS Advanced Facility Instrument for Study of the Earth as a System. *IEEE Transcations on Geoscience and Remote Sensing*, 27:145–153, 1989. 2, 51

B. Scholkopf and A. J. Smola. *Learning with Kernels*. MIT Press, 2002. ix, 46

K. Siwek, S. Osowski, K. Garanty, and M. Sowinski. Ensemble of Neural Predictors for Forecasting the Atmospheric Pollution. *Proc. IEEE International Joint Conference on Neural Network*, pages 643–648, 2008. 3

J. D. Spinhirne, S. P. Palm, D. L. Hlavka, W. D. Hart, and E. J. Welton. Global aerosol distribution from the glas polar orbiting lidar instrument. remote sensing of atmospheric aerosols. *IEEE Workshop on aerosols*, pages 2–8, 2005. 13

L. L. Stowe. Cloud and aerosol products at noaa/nesdis. *Paleogeogr. Paleoclimatol. Paleoecol.*, 90:2532, 1991. 19, 27

L. L. Stowe, A. M. Ignatov, and R. R. Singh. Development, validation, and potential enhancements to the second-generation operational aerosol product at the national environmental satellite, data, and information service of the national oceanic and atmospheric administration. *J. Geophys. Res.*, 102:1692316934, 1997. 19, 27

O. Torres, P. K. Bhartia, J. R. Herman, A. Sinyuk, and B. Holben. A long term record of aerosol optical thickness from toms observations and comparison to aeronet measurements. *J. Atm. Sci.*, 59:398413, 2002. 20, 27

O. Torres, P. K. Bhartia, A. Sinyuk, E. J. Welton, and B. Holben. Total ozone mapping spectrometer measurements of aerosol absorption from space: Comparison to safari 2000 ground-based observations. *J. Geophys. Res.*, 110(D10S18), 2005. 20

V. N. Vapnik. *Estimation of Dependences based on Empirical Data*. Springer, Berlin, 1982. 50

V.N. Vapnik. *The nature of statistical learning theory*. Springer-Verlag, Berlin, 1995. 44

W. von Hoyningen-Huene, A. A. Kokhanovsky, J. P. Burrows, V. Bruniquel-Pinel, P. Regner, and F. Baret. Simultaneous determination of aerosol and surface chracteristics from meris top-of-atmosphere reflectance. *Adv. Space Res.*, 37: 2172–2177, 2006. 27

S. Vucetic, B. Han, W. Mi, Z. Li, and Z. Obradovic. A Data-Mining Approach for the Validation of Aerosol Retrievals. *IEEE Geo. Rem. Sen. Let.*, 5:113–117, 2008. 3, 22, 23, 61, 80

M. Wang, K. D. Knobelspiesse, and C. R. McClain. Study of the sea-viewing wide field-of-view sensor (seawifs) aerosol optical property data over ocean in combination with the ocean color products. *J. Geophys. Res.*, 110(D10S06): doi:10.1029/2004JD004950, 2005. 20

D. M. Winker, R. H. Couch, and M. P. McCormick. An overview of lite: Nasas lidar in-space technology experiment. *Proc. IEEE*, 84:164180, 1996. 13

M. S. Wong, J. E. Nichol, K. H. Lee, and Z. Q. Li. High resolution aerosol optical thickness retrieval over the Pearl River Delta region with improved aerosol modelling. *Sci. China Ser. D-Earth Sci.*, 52(10):1641–1649, 2009. 5, 25, 77

Q. Xu, Z. Obradovic, B. Han, Y. Li, A. Braverman, and S. Vucetic. Improving Aerosol Retrieval Accuracy by Integrating AERONET, MISR and MODIS Data. *Proc. The 8th Intenational Conference on Information Fusion*, 1, 2005. 3, 22, 23, 80

H. Yu, Y. J. Kaufman, M. Chin, G. Feingold, L. Remer, T. Anderson, Y. Balkanski, N. Bellouin, O. Boucher, S. Christopher, P. DeCola, R. Kahn, D. Koch, N. Loeb, M. S. Reddy, M. Schulz, T. Takemura, and M. Zhou. A review of measurement-based assessments of aerosol direct radiative effect and forcing. *Atmos. Chem . Phys.*, 6:613–666, 2006. 2