

ATMOSPHERIC CORRECTION AND IMAGE CLASSIFICATION  
ON MODIS IMAGES BY NONPARAMETRIC REGRESSION SPLINES

A THESIS SUBMITTED TO  
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES  
OF  
MIDDLE EAST TECHNICAL UNIVERSITY

BY

SEMİH KUTER

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS  
FOR  
THE DEGREE OF DOCTOR OF PHILOSOPHY  
IN  
GEODETIC AND GEOGRAPHIC INFORMATION TECHNOLOGIES

FEBRUARY 2014



Approval of the thesis

**ATMOSPHERIC CORRECTION AND IMAGE CLASSIFICATION  
ON MODIS IMAGES BY NONPARAMETRIC REGRESSION SPLINES**

submitted by **SEMİH KUTER** in partial fulfillment of the requirements for the degree of **Doctor of Philosophy in Geodetic and Geographic Information Technologies Department, Middle East Technical University** by,

Prof. Dr. Canan ÖZGEN  
Dean, Graduate School of **Natural and Applied Sciences** \_\_\_\_\_

Prof. Dr. Ahmet COŞAR  
Head of Department, **Geodetic and Geographic Inf. Tech.** \_\_\_\_\_

Prof. Dr. Zuhal AKYÜREK  
Supervisor, **Civil Engineering Dept., METU** \_\_\_\_\_

Prof. Dr. Gerhard-Wilhelm WEBER  
Co-Supervisor, **Inst. of Applied Mathematics, METU** \_\_\_\_\_

**Examining Committee Members:**

Prof. Dr. Mehmet Lütfi SÜZEN  
Geological Engineering Dept., METU \_\_\_\_\_

Prof. Dr. Zuhal AKYÜREK  
Civil Engineering Dept., METU \_\_\_\_\_

Prof. Dr. Mahmut Onur KARSLIOĞLU  
Civil Engineering Dept., METU \_\_\_\_\_

Prof. Dr. İnci BATMAZ  
Statistics Dept., METU \_\_\_\_\_

Prof. Dr. Gül ERGÜN  
Statistics Dept., Hacettepe University \_\_\_\_\_

Date: 12.02.2014

**I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.**

Name, Last Name: SEMİH KUTER

Signature :

## ABSTRACT

### ATMOSPHERIC CORRECTION AND IMAGE CLASSIFICATION ON MODIS IMAGES BY NONPARAMETRIC REGRESSION SPLINES

KUTER, Semih

Ph.D., Department of Geodetic and Geographical Information Technologies

Supervisor : Prof. Dr. Zuhal AKYÜREK

Co-Supervisor : Prof. Dr. Gerhard-Wilhelm WEBER

February 2014, 173 pages

In this study, two novel applications of nonparametric regression splines are introduced within the frame of remote sensing: (i) For the first time ever, atmospheric correction models are generated for *moderate resolution imaging spectroradiometer* (MODIS) images by using *multivariate adaptive regression splines* (MARS), and its recently introduced version, *conic multivariate adaptive regression splines* (CMARS). The obtained models are applied on the predefined test areas of twenty four different MODIS scenes. *Simplified model for atmospheric correction* (SMAC) algorithm, a radiative transfer-based approach, is also employed on the same test data. The performance of the MARS, CMARS and SMAC models are assessed against MODIS surface reflectance products. (ii) Additionally, implementation of MARS algorithm in image classification for snow mapping on MODIS images is demonstrated within a well-elaborated framework. The relation between the variations in MARS model building parameters and their effect on the predictive performance are presented in various perspectives. Performance of MARS in classification is compared with the traditional maximum-likelihood method. For the atmospheric correction, results reveal that MARS and CMARS

approaches over perform SMAC method on all test areas. In classification, significant improvement in the classification accuracy of MARS models is observed as the number of basis functions and the degree of interaction increase. On three image sets out of four, the MARS approach gives better classification accuracies when compared to maximum-likelihood method.

Keywords: MARS, CMARS, Conic Quadratic Programming, Atmospheric Correction, Multispectral Image Classification

## ÖZ

### PARAMETRİK OLMAYAN REGRESYON EĞRİLERİ KULLANILARAK MODIS GÖRÜNTÜLERİ ÜZERİNDE ATMOSFERİK DÜZELTME VE GÖRÜNTÜ SINIFLANDIRMASI

KUTER, Semih

Doktora, Jeodezi ve Coğrafi Bilgi Teknolojileri Bölümü

Tez Yöneticisi : Prof. Dr. Zuhal AKYÜREK

Ortak Tez Yöneticisi : Prof. Dr. Gerhard-Wilhelm WEBER

Şubat 2014, 173 sayfa

Bu çalışmada, parametrik olmayan regresyon eğrilerinin uzaktan algılama alanındaki iki özgün uygulaması sunulmaktadır: (i) İlk kez *çok değişkenli uyarlanabilir regresyon eğrileri* (MARS), ve yeni geliştirilen versiyonu olan *konik çok değişkenli uyarlanabilir regresyon eğrileri* (CMARS) kullanılarak *orta çözünürlüklü görüntüleme spektrometresi* (MODIS) görüntüleri için atmosferik düzeltme modelleri üretilmiştir. Oluşturulan modeller, yirmi dört farklı MODIS görüntüsü üzerinde daha önceden belirlenmiş olan test alanlarında uygulanmıştır. Radyatif transfer tabanlı olan *basitleştirilmiş atmosferik düzeltme modeli* (SMAC) aynı test alanları üzerinde ayrıca kullanılmıştır. MARS, CMARS ve SMAC modellerinin performansları MODIS yüzey reflektans ürünleri baz alınarak değerlendirilmiştir. (ii) Ayrıca, MARS algoritmasının kar haritalaması amacına yönelik görüntü sınıflandırmasında MODIS görüntüleri üzerindeki uygulaması iyi kurgulanmış bir çerçeve içerisinde sunulmuştur. MARS model oluşturma parametrelerindeki değişikliklerin, oluşturulan modelin sınıflandırma performansı üzerindeki etkileri farklı açılardan değerlendirilmiştir. Sonuçlar, MARS ve CMARS yöntemlerinin atmosferik düzeltme için tüm test görüntüleri üzerinde SMAC

yönteminden daha iyi sonuçlar verdiđini ortaya koymaktadır. Görüntü sınıflandırması için, baz fonksiyonu sayısının ve etkileşim derecesinin artırılması ile MARS metodunun sınıflandırma doğruluğunda belirgin gelişme olduđu gözlemlenmiştir. Sonuçlar klasik maksimum-olasılık yöntemiyle karşılaştırıldığında, MARS metodu ile dört görüntü setinden üç'ü üzerinde daha iyi sınıflandırma doğruluđu elde edilmiştir.

Anahtar Kelimeler: MARS, CMARS, Konik Kuadratik Programlama, Atmosferik Düzeltme, Çok Bantlı Görüntü Sınıflandırması



*To my wife:*

*Nazan*

*And to my daughter:*

*Ela*

## ACKNOWLEDGEMENTS

First of all, I would like express my special appreciation and thanks to my supervisor Prof. Dr. Zuhal AKYÜREK and to my co-supervisor Prof. Dr. Gerhard-Wilhelm WEBER for their endless support and encouragement. Both of you have been tremendous mentors for me and your advice on both research as well as on my career have been priceless.

I would also like to thank Prof. Dr. Mehmet Lütfi SÜZEN, Prof. Dr. Mahmut Onur KARSLIOĞLU, Prof. Dr. İnci BATMAZ and Prof. Dr. Gül ERGÜN for serving as my thesis defense committee members. I also want to thank you for letting my defense be an enjoyable moment, and for your brilliant comments and suggestions, which have significantly improved the quality of the study, thanks to you.

I would especially like to express my gratitude to Ayşe ÖZMEN for her helps in data processing, and to Kenan BOLAT for his helps in code writing during this study. I would also like to thank Ayşenur ÖZCAN, Simge ÖRSÇELİK and Fatma YERLİKAYA-ÖZKURT for their valuable technical and grammatical comments on the manuscript of the thesis.

A special thanks to my family. Words cannot express how grateful I am to my beloved wife, Nazan, for all of the sacrifices that you've made on my behalf. And my daughter, Ela, without the joy and motivation that you have given me, it wouldn't have been possible to complete this work.

And last few words to the precious memory of my beloved friend Mithat ÇOLAK; it was really a great privilege for me to know you, May God rest your soul in peace...

# TABLE OF CONTENTS

ABSTRACT .....	v
ÖZ.....	vii
ACKNOWLEDGEMENTS .....	x
TABLE OF CONTENTS .....	xi
LIST OF TABLES .....	xv
LIST OF FIGURES .....	xix
LIST OF ABBREVIATIONS .....	xxiii
CHAPTERS	
1. INTRODUCTION.....	1
1.1 Motivation .....	1
1.2 Our Tasks .....	3
1.3 Contributions of the Thesis .....	4
1.4 Outline of the Thesis .....	5
2. OVERVIEW OF MARS AND CMARS METHODS .....	9
2.1 Nonparametric Regression Splines with Continuous Optimization and Conic Quadratic Programming.....	9
2.1.1 Advantage of using splines .....	10
2.2 MARS Method .....	12
2.3 CMARS Method .....	18

3. A BRIEF REVIEW ON RADIATIVE TRANSFER MODELS AND IMAGE CLASSIFICATION.....	23
3.1 Atmospheric Correction by RT Models .....	23
3.1.1 Definitions of reflectance and radiance .....	26
3.1.2 Basic RT formulation.....	27
3.1.3 Evolution of RT models.....	32
3.1.3.1 MODTRAN.....	32
3.1.3.2 5S.....	33
3.1.3.3 ATREM.....	34
3.1.3.4 SMAC .....	35
3.1.3.5 FLAASH .....	37
3.1.3.6 ACORN.....	38
3.1.3.7 ATCOR .....	39
3.1.3.8 6S.....	39
3.2 A Short Introductory to Image Classification .....	40
3.2.1 Supervised versus unsupervised .....	41
3.2.2 Parametric versus nonparametric .....	44
4. MODIS INSTRUMENT AND PRODUCTS.....	47
4.1 MODIS Instrument.....	47
4.2 MODIS Products .....	48
4.2.1 MODIS Level 1B product.....	51
4.2.2 MODIS Level 2G surface reflectance product .....	54
4.2.3 MODIS Level 3 daily atmospheric product.....	58

4.2.4 MODIS snow algorithm and Level 3 daily snow product .....	60
5. ATMOSPHERIC CORRECTION ON MODIS IMAGES BY MARS AND CMARS .....	65
5.1 Image Set and Model Training.....	65
5.2 Testing of MARS and CMARS Models .....	71
5.3 Results and Discussion.....	71
6. IMAGE CLASSIFICATION ON MODIS IMAGES BY MARS ....	87
6.1 Background of MARS in Multispectral Image Classification ....	87
6.2 Image Set and Model Training.....	89
6.3 Testing of MARS Models .....	94
6.3.1 Error matrix .....	95
6.4 Results and Discussion.....	97
7. CONCLUSIONS AND RECOMMENDATIONS .....	117
REFERENCES .....	123
APPENDICES	
A. TIKHONOV REGULARIZATION.....	141
B. CONIC QUADTRATIC PROGRAMMING .....	143
C. MARS AND CMARS MODELS .....	145
C.1 MARS and CMARS Models for Alps .....	145
C.2 MARS and CMARS Models for Turkey .....	147
D. ACCURACY MEASURES .....	151
E. MATLAB CODES .....	153
E.1 SMAC Code .....	153

E.2 MATLAB Code to Run the SMAC Algorithm on a MODIS Image.....	155
E.3 MATLAB Code to Prepare the Necessary Input Data for SMAC Processing .....	157
E.4 MATLAB Code to Prepare Response and Predictor Matrices for Multi-Response MARS Classification .....	160
E.5 MATLAB Code to Obtain Cloud Mask from MOD09GA State_1km QA Data .....	166
CURRICULUM VITAE .....	169

## LIST OF TABLES

### TABLES

Table 4.1: Description of MODIS spectral bands .....	48
Table 4.2: Level 1B product summary .....	52
Table 4.3: Units for reflective solar bands .....	54
Table 4.4: MOD/MYD09GA science data set .....	57
Table 4.5: Description of bits in a MOD/MYDD09GA State_1km QA data ..	58
Table 4.6: Several primary parameters in MOD/MYD08_D3 daily product...	59
Table 4.7: Seven MODIS snow data products .....	62
Table 4.8: MOD/MYD10A1 daily snow tile attributes.....	63
Table 5.1: TOA reflectance statistics for Alps (A) and Turkey (T) .....	67
Table 5.2: MARS model training settings.....	68
Table 5.3: Comparison of MARS models obtained with two different training sample sizes .....	69
Table 5.4: Shapiro-Wilk test of MAEs and R2s for two different MARS models obtained via two training samples with different sizes.....	70

Table 5.5: Paired-Samples t test of MAEs and R2s with respect to sample size .....	70
Table 5.6: Comparison of results for Alps (A) and Turkey (T) .....	73
Table 5.7: Shapiro-Wilk test of SMAC, MARS and CMARS results .....	73
Table 5.8: RANOVA test results .....	75
Table 5.9: Results of Huynh-Feldt tests .....	75
Table 5.10: Pairwise comparison of the results .....	76
Table 5.11: Processing times (in seconds) for Alps (A) and Turkey (T) .....	83
Table 6.1: MODIS data set used in image classification.....	90
Table 6.2: Details of training and test data .....	93
Table 6.3: Different MARS model building settings applied in image classification .....	93
Table 6.4: Merging of some classes in MOD10A1 reference image .....	94
Table 6.5: A sample error matrix.....	96
Table 6.6: Error matrix for AC1 data set.....	97
Table 6.7: Error matrix for AC2 data set.....	97
Table 6.8: Error matrix for TC1 data set .....	98



Table 6.9: Error matrix for TC2 data set .....	98
Table 6.10: MARS model training performance for AC1 and AC2 data sets (ST: number of selected terms after the backward pass).....	101
Table 6.11: MARS model training performance for TC1 and TC2 data sets (ST: number of selected terms after the backward pass).....	101
Table 6.12: Percentages of pixels for each class in training data sets.....	102
Table 6.13: Error matrices of MARS and ML methods for AC1 data set .....	107
Table 6.14: Error matrices of MARS and ML methods for AC2 data set .....	107
Table 6.15: Error matrices of MARS and ML methods for TC1 data set.....	108
Table 6.16: Error matrices of MARS and ML methods for TC2 data set.....	108



# LIST OF FIGURES

## FIGURES

Figure 3.1: Geometry of the problem, $\theta_s$ : solar zenith angle, $\theta_v$ : viewing zenith angle, $\Delta\phi$ : relative azimuth between the Sun and the sensor (adapted from [7]) .....	25
Figure 3.2: Reflectance from a) specular, and b) Lambertian surface (adapted from [2]).....	27
Figure 3.3: Steps in an RT-based atmospheric correction scheme (adapted from [2]) .....	28
Figure 3.4: Interaction of the solar radiation with atmosphere: a) direct, b) path, c) diffuse, d) multiple, and e) neighbor (adapted from [2])..	29
Figure 3.5: From a) image space, to b) feature space.....	41
Figure 3.6: A multispectral image classifier (adapted from [2]).....	42
Figure 4.1: Observation swath of MODIS (adapted from [106]).....	47
Figure 4.2: The MODIS sinusoidal grid scheme (adapted from [113]) .....	50
Figure 4.3: Scaled integer representation of $dn^b$ in Level 1B product (adapted from [106]).....	53
Figure 4.4: 16-bit binary representation of State_1km QA data .....	56

Figure 5.1: MOD09GA RGB color composite image of Alps (R: 1 <sup>st</sup> , G: 4 <sup>th</sup> , B: 3 <sup>rd</sup> band, A11: 06.11.2002) .....	66
Figure 5.2: MOD09GA RGB color composite image of Turkey (R: 1 <sup>st</sup> , G: 4 <sup>th</sup> , B: 3 <sup>rd</sup> band, T7: 16.07.2012).....	66
Figure 5.3: Images given by a) MARS, b) CMARS, c) SMAC, and d) MOD09GA for A10 data set.....	72
Figure 5.4: Images given by a) MARS, b) CMARS, c) SMAC, and d) MOD09GA for T11 data set .....	72
Figure 5.5: Box-Cox transformation plot for SMAC MAE values of Alps .....	74
Figure 5.6: Box-Cox transformation plot for SMAC R <sup>2</sup> values of Alps .....	74
Figure 5.7: Change in the mean relative difference between SMAC and MARS reflectance values with respect to atmospheric parameters for Alps.....	77
Figure 5.8: Change in the mean relative difference between SMAC and MARS reflectance values with respect to atmospheric parameters for Turkey.....	78
Figure 5.9: Image histograms for Alps data set: a) A1, b) A2, c) A3, d) A4, e) A5, and f) A6 .....	79
Figure 5.10: Image histograms for Alps data set: a) A7, b) A8, c) A9, d) A10, e) A11, and f) A12 .....	80
Figure 5.11: Image histograms for Turkey data set: a) T1, b) T2, c) T3, d) T4, e) T5, and f) T6 .....	81

Figure 5.12: Image histograms for Turkey data set: a) T7, b) T8, c) T9, d) T10, e) T11, and f) T12 .....	82
Figure 6.1: MOD09GA RGB color composite image of AC2 data set (R:1 <sup>st</sup> , G: 4 <sup>th</sup> , B: 3 <sup>rd</sup> band).....	91
Figure 6.2: Training samples for TC1 data set.....	92
Figure 6.3: Response-predictor matrix pair for MARS model training .....	93
Figure 6.4: Cloud mask generated from MOD09GA State_1km QA data for AC2 data set .....	95
Figure 6.5: Graph of overall accuracy for AC1 data set .....	99
Figure 6.6: Graph of overall accuracy for AC2 data set .....	99
Figure 6.7: Graph of overall accuracy for TC1 data set.....	100
Figure 6.8: Graph of overall accuracy for TC2 data set.....	100
Figure 6.9: Percentage overlapping of water, cloud and land with snow in MARS (OE: omission error, CE: commission error).....	103
Figure 6.10: Percentage overlapping of water, cloud and land with snow in ML (OE: omission error, CE: commission error).....	103
Figure 6.11: Percentage overlapping of snow, water and land with cloud in MARS (OE: omission error, CE: commission error) .....	104
Figure 6.12: Percentage overlapping of snow, water and land with cloud in ML (OE: omission error, CE: commission error).....	104

Figure 6.13: Percentage overlapping of snow, cloud and land with water in MARS (OE: omission error, CE: commission error) .....	105
Figure 6.14: Percentage overlapping of snow, cloud and land with water in ML (OE: omission error, CE: commission error).....	105
Figure 6.15: Percentage overlapping of snow, water and cloud with land in MARS (OE: omission error, CE: commission error) .....	106
Figure 6.16: Percentage overlapping of snow, water and cloud with land in ML (OE: omission error, CE: commission error).....	106
Figure 6.17: Classified images of AC1 data set: a) MOD09GA RGB color composite image (R:1st, G: 4th, B: 3rd band), b) MOD10A1 reference image, c) ML classification, and d) MARS classification .....	112
Figure 6.18: Classified images of AC2 data set: a) MOD09GA RGB color composite image (R:1st, G: 4th, B: 3rd band), b) MOD10A1 reference image, c) ML classification, and d) MARS classification .....	113
Figure 6.19: Classified images of TC1 data set: a) MOD09GA RGB color composite image (R:1st, G: 4th, B: 3rd band), b) MOD10A1 reference image, c) ML classification, and d) MARS classification .....	114
Figure 6.20: Classified images of TC2 data set: a) MOD09GA RGB color composite image (R:1 <sup>st</sup> , G: 4 <sup>th</sup> , B: 3 <sup>rd</sup> band), b) MOD10A1 reference image, c) ML classification, and d) MARS classification .....	115

## LIST OF ABBREVIATIONS

5S.....	Simulation of a Satellite Signal in the Solar Spectrum
6S.....	Second Simulation of a Satellite Signal in the Solar Spectrum
ACORN .....	Atmospheric Correction Now
ANN .....	Artificial Neural Networks
AOD .....	Atmospheric Optical Depth
ASTER .....	Advanced Spaceborne Thermal Emission and Reflection Radiometer
ATCOR .....	Atmospheric and Topographic Correction
ATREM.....	Atmosphere Removal Algorithm
AUC.....	Area Under the Curve
AVHRR .....	Advanced Very High Resolution Radiometer
AVIRIS.....	Airborne Visible Infra Red Imaging Spectrometer
BFs.....	Basis Functions
BRDF.....	Bidirectional Reflectance Distribution Function
CART .....	Classification and Regression Trees
CGPLM .....	Conic Generalized Partial Linear Models
CMARS .....	Conic Multivariate Adaptive Regression Splines
CQP .....	Conic Quadratic Programming
DEM.....	Digital Elevation Model

DN .....	Individual Pixel Values
DT.....	Decision Trees
EOS.....	Earth Observing System
EOSDIS .....	Earth Observing System Data and Information System
FC .....	Fuzzy Classification
FLAASH.....	Fast Line-of-sight Atmospheric Analysis of Spectral Hypercubes
GCV .....	Generalized Cross Validation
GIS.....	Geographic Information System
GPLM.....	Generalized Partial Linear Models
HDF .....	Hierarchical Data Format
HDF-EOS .....	Hierarchical Data Format – Earth Observing System
KM.....	K-Means
LFOV .....	Large Field Of View
LOF.....	Lack of Fit
LUT .....	Look up Table
MAE .....	Mean Absolute Error
MARS.....	Multivariate Adaptive Regression Splines
MD.....	Minimum-Distance
ML .....	Maximum-Likelihood
MOD.....	Prefix for MODIS instrument on board to Terra satellite
MODIS .....	Moderate Resolution Imaging Spectroradiometer



MODTRAN.....	Moderate Resolution Atmospheric Transmission
MSG-SEVIRI.....	Meteosat Second Generation - Spinning Enhanced Visible and Infrared Imager
MYD.....	Prefix for MODIS instrument on board to Aqua satellite
NASA.....	National Aeronautics and Space Administration
NDSI.....	Normalized Difference Snow Index
NDVI.....	Normalized Difference Vegetation Difference
NOAA.....	National Oceanic and Atmospheric Administration
PP.....	Parallelepiped
PRSS.....	Penalized Residual Sum of Squares
QA.....	Quality Assurance
$R^2$ .....	Multiple Coefficient of Determination
RANOVA.....	Related Samples Analysis of Variance
RCGPLM.....	Robust Conic Generalized Partial Linear Models
RCMARS.....	Robust Conic Multivariate Adaptive Regression Splines
ROI.....	Region of Interest
RS.....	Remote Sensing
RT.....	Radiative Transfer
SDS.....	Scientific Data Set
SI.....	Scaled Integer
SMAC.....	Simplified Model for Atmospheric Correction
SOS.....	Successive Orders of Scattering

SVM .....	Support Vector Machines
SWIR .....	Short Wave Infra Red
SZA.....	Solar Zenith Angle
TOA.....	Top of Atmospheric
TR.....	Tikhonov Regularization
uO <sub>3</sub> .....	Ozone Content
UV .....	Ultra-Violet
uWV .....	Water Vapor Content
VZA.....	Viewing Zenith Angle

# CHAPTER 1

## INTRODUCTION

### 1.1 Motivation

Remotely-sensed image data have been used as a primary information source in spatial technologies. However, a fundamental problem in *remote sensing* (RS) is the perturbations on surface reflectance data due to absorption and scattering by atmospheric gases and molecules [1-3]. So, in order to make correct analysis and interpretation of remotely-sensed images, the reflectance values of the image pixels must accurately represent the ground surface reflectance values.

There are two main approaches in atmospheric correction methodology: (i) relative and, (ii) absolute radiometric correction of atmospheric attenuation. According to the former, the image histograms of spectral bands are shifted to the left, based on the assumption that the offset of the histogram from zero brightness value is caused by the atmospheric scattering, which is easy to apply, but only provides an approximate correction [2, 4]. In the latter, detailed numerical *radiative transfer* (RT) codes are employed to model the whole process of attenuation of solar radiation in the atmosphere [5, 6].

However, it should be emphasized that the high accuracy available with detailed RT models may not actually be reached when working over large areas, due to unknown atmospheric parameters [7]. Additionally, for large areas and long time periods, researchers still need faster methods in order to derive surface reflectances.

To obtain a thematic map by image classification is another challenging task in RS because there exist many factors directly affecting the success of the process such as technical characteristic and quality of remotely sensed data, complexity of landscape, image processing techniques and classification scheme employed [8].

Two main approaches exist in image classification, namely, *supervised* and *unsupervised*. Until the mid-1990s, supervised classification methods originated from conventional statistical techniques such as *maximum-likelihood* (ML) or *minimum-distance* (MD) were the main classification methods used in RS [2]. Unsupervised classification is the contrary approach of supervised methods, where the user analyzes the whole image without reference to any training data to define separable clusters [9]. The most widely known unsupervised method is the *K-means* (KM) algorithm [10].

Other alternative image classification techniques based on *decision trees* (DT), *support vector machines* (SVM), *artificial neural networks* (ANN) and *fuzzy classification* (FC) have been proposed and assessed [11-14]. However, some of these approaches are known to have problems in explaining the connection between the predictor variables and the results of classification [8, 15].

In this thesis, we would like to propose new alternative approaches to the atmospheric correction and image classification by “smartly” and “innovatively” treating these issues within the frame of nonparametric regression splines. Indeed, nonparametric regression splines offer great advantages because in real life problems and natural phenomena, many effects often exhibit nonlinear behavior.

Firstly, for the first time ever in RS, we deal with the application of nonparametric regression splines in atmospheric correction by using two state-of-the-art tools, *multivariate adaptive regression splines* (MARS) [16], and its

recently introduced complementary alternative *conic multivariate adaptive regression splines* (CMARS) [17]. We introduce the novel application of these tools to develop atmospheric correction models for *moderate resolution imaging spectroradiometer* (MODIS) data.

Secondly, we treat the multispectral image classification issue by employing MARS approach. According to the results of our intensive literature survey, the study by Quirós et al. [15] is the only available one that used MARS as an alternative method in multispectral image classification. An image classification scheme for snow mapping by using multi-response MARS approach on MODIS images is introduced within a well-established framework and the results are presented.

## **1.2 Our Tasks**

In order to briefly summarize, our tasks in this thesis to achieve the objectives stated in the previous section are

- to generate atmospheric correction models for MODIS images through nonparametric MARS and CMARS methods,
- to assess the performance of the obtained models, and compare their results with an RT-based approach in terms of both accuracy and processing time,
- to employ MARS approach to perform multispectral image classification for snow mapping on MODIS images,
- to comprehensively analyze and reveal the effect of MARS model building parameters on the final classification results,

- to assess the accuracy obtained with MARS approach and to compare the results with the results obtained from classical ML classification method.

### **1.3 Contributions of the Thesis**

With this study, we recommend both MARS and our adaptive tool for regression and classification, CMARS, to be used in the area of RS, especially, as an alternative to traditional formula-based approaches which have their scientific origins in the natural sciences, especially, in physics. However, we very much respect and appreciate those formulas as they represent important relations between essential, natural dimensions, though on a certain level of abstraction and, sometimes, simplification. Our CMARS approach of learning and modeling is based on both the empirical data and the mathematics of applied probability and, especially, optimization theory, whereby accuracy becomes compromised with complexity.

This thesis has the following contributions related with the atmospheric correction and image classification subjects:

- Two nonparametric regression and classification tools, MARS and CMARS, are used for the first time to make atmospheric correction. In particular, the models obtained by this new approach do not require any atmospheric parameters as input such as atmospheric optical depth, water vapor and ozone content.
- Not only it illustrates the application of the obtained atmospheric correction models on MODIS images, but also assesses the performance of the two methods, in terms of both accuracy and processing time.
- Advantages of using CMARS method and further possible applications of its extensions are highlighted within the context of our problem in RS.

- A novel application of multi-response MARS approach in image classification for snow mapping is introduced within a well-elaborated frame.
- It clearly exposes the effects of different MARS model settings on the classification results.
- It demonstrates the application of the proposed approach on MODIS images, and reveals its applicability in image classification for snow mapping by comparing the results with traditional ML classification method.
- And in overall, it constitutes a base for those who further intend to employ these two state-of-the art nonparametric regression and classification tools in their RS-related applications.

#### 1.4 Outline of the Thesis

The organization of the thesis as follows

- **Overview of MARS and CMARS Methods:** In Chapter 2, we begin with the discussion on the advantages of using spline functions. Then, mathematical basis of MARS algorithm is given, together with the details of the parts of the algorithm, namely, *forward* and *backward step*. Then the alternative formulation of the backward step of MARS to obtain CMARS by using the modern methods of continuous optimization is introduced.
- **A Brief Review on Radiative Transfer Models and Image Classification:** In Chapter 3, we first present the basics of interaction between the solar radiation and the atmosphere. Derivation of RT formulation is explained briefly. Then, widely-recognized RT-based

atmospheric correction algorithms are revisited in a chronological order. Secondly, in Chapter 3, the methods and techniques used in multispectral image classification are covered, yet not in a detailed manner. Instead, well-known approaches are shortly introduced together with their pros and cons.

- **MODIS Instrument and Products:** Since our main data source in this study is MODIS, we reserve Chapter 4 to the details related with this instrument and its products. First, technical specifications of MODIS instrument are given. Then, the inner structure of the scientific data sets obtained by different data processing levels that are used in this thesis are introduced in details.
- **Atmospheric Correction on MODIS Images by MARS and CMARS:** Chapter 5 starts with the basic introductory of the data set used in this part of the study. Next, the details about the MARS and CMARS model training process are mentioned. The test results associated with the performance of the MARS and CMARS atmospheric correction models are given and compared with the performance of *simplified model for atmospheric correction* (SMAC), an RT-based approach, together with the results of statistical evaluations. And the results derived from the empirical findings are discussed.
- **Image Classification on MODIS Images by MARS:** A novel application of MARS algorithm for multispectral image classification on MODIS images are introduced in Chapter 6. We begin with the introduction of the data set. Then, the details of the approach used in multi-response MARS model training are introduced. By systematically varying the model building parameters, different MARS models are obtained, and then applied on the test images. The performance of the models is also compared against ML classification method, and the results are presented by using error matrices. The effect of employing



different model building parameters in the predictive performance of MARS is introduced based on the experimental findings.

- **Conclusions and Recommendations:** Finally, Chapter 7 wraps up the thesis by presenting the conclusions and overall findings of the study. Possible extensions of proposed methods for our future studies are given.



## CHAPTER 2

### OVERVIEW OF MARS AND CMARS METHODS

#### 2.1 Nonparametric Regression Splines with Continuous Optimization and Conic Quadratic Programming

To estimate unknown (predictor, response) dependency (or model) from training data (consisting of a finite number of observations) with good prediction (generalization) capabilities for future (test) data is the basic aim of the learning task. While regression is to learn a mapping from the input space,  $X$ , to the output space,  $Y$ , where this mapping,  $f$ , is called an *estimator*, which is used for predicting quantitative outputs (i.e.,  $X = \mathbb{R}^d, Y = \mathbb{R}$ ); the aim of classification is to learn a mapping from the feature space,  $X$ , to a label space,  $Y$ , where this mapping,  $f$ , is called a *classifier*, which is used for predicting qualitative outputs (i.e.,  $X = \mathbb{R}^d, Y = \{0,1\}$ ). Although the naming convention for the learning task depends on the output type, both have common characteristics and can be regarded as a task in function approximation [18].

As in almost all branches of science and engineering, regression and classification also play a crucial role in *geographical information systems* (GIS) and RS, where images taken by Earth-observing satellites are used to extract various kinds of information concerning Earth environment. Many kinds of regression and classification methods have been applied for information extraction from remotely sensed Earth data [19-25]. And at this stage, it is of value to emphasize that nonparametric regression and classification techniques are mostly the key data mining tools in explaining real

life problems and natural phenomena where many effects often exhibit nonlinear behavior.

Introduced by physicist and statistician Friedman [16], *multivariate adaptive regression splines* (MARS) is an innovative nonparametric regression procedure that does not make any specific assumption about the underlying functional relationship between response and predictor variables. The algorithm estimates the model function at two stages; forward step and backward step. In the first stage, similar to forward stepwise linear regression, the *basis functions* (BFs) and their products are used to generate a maximal model that overfits data. Then in backward step, the model is pruned stepwise by eliminating the terms which result in smallest increase in the residual squared error until a predefined threshold is reached.

*Conic multivariate adaptive regression splines* (CMARS) [17] has recently been developed as an alternative method to MARS. In CMARS, instead of applying backward step algorithm, a *penalized residual sum of squares* (PRSS) is introduced to MARS as a *Tikhonov regularization* (TR) problem, and this two-objective optimization problem is treated using the continuous optimization technique called *conic quadratic programming* (CQP). Within this context, CMARS is more model-based and employs continuous, well-structured convex optimization which uses *interior point methods* and their codes.

### **2.1.1 Advantage of using splines**

The use of spline functions offer great advantages, in general, and in the context of our modelling in RS, as it will be explained now compactly [26]:

- Splines, from the perspective of just one dimension (input variable), are “piecewise polynomials”. If polynomials are used only, then they would usually converge to plus or minus infinity when the absolute values of the (input) variables grow large. Since real-life processes stay in bounded margins (even if these bounds may be very large) generally, polynomials would need to have a high degree to oscillate sufficiently in order to remain within some reasonable margin. However, with high-degree polynomials it is hard and costly to work, in particular, as the real-life challenges are multivariate, which may need multiplications and, hence, a rapid increase in degree of the occurring multivariate splines. Using splines however permits us, in every dimension, to keep the degree of the polynomial “pieces” very low. Splines are indeed quite “flexible” and sufficiently “elastic” in order to approximate complex and high-dimensional data structures. So, they are often called “smoothing splines”, since they “smoothly” approximate discrete data.
- The splines of CMARS are more “smooth” even in the sense that their oscillation is kept under control via a penalization of their integrals of squared first- and, especially, second-order derivatives, i.e., their complexity. Then the integrals are discretized, and a program of TR is obtained and represented in the form of CQP.
- The multivariate splines of CMARS are multiplication products of “zig-zagging”, piecewise linear functions of degree 1 (or 0), and it can be carefully determined how many dimensions to include into the process of multiplications of those 1-dimensional splines in order to prevent the complexity of our model from becoming too high. Let us recall that reducing the complexity can also be expressed as an increase in stability.
- Those multiplications are an expression of the fact that there are input variables, which are dependent and, together in groups, contributing to an explanation of response variable by those explanatory input variables.

- Finally, differently from the use of “stiff” model formulas, which are motivated by the tradition in science, our approach by CMARS is very adaptive and is getting into the data set with its particular subsets and characteristics of shape.

## 2.2 MARS Method

Regression analysis in statistical learning is a method in which numerous independent variables are modeled and analyzed. As a form of regression analysis, MARS is a nonparametric regression technique. The algorithm is widely used in data mining and estimation theory in order to built flexible regression models for high dimensional nonlinear data. MARS is an evolved form of linear models that can automatically model nonlinear and interactive models, and it has great importance in both classification and regression.

MARS has many successful applications in various fields of science and engineering such as operational research, marketing and finance [27-29]; biology, ecology and forestry [30-36]; simulation and computation [37, 38]; geophysics [39]; engineering [40-43]; medical and biomedical sciences [44-48]; as well as in GIS and RS [15, 49-51].

From this point forward in this section, the basics of MARS method are given based on [16, 18, 26, 52].

Originated from a modified recursive partitioning methodology, MARS algorithm is an extension of *classification and regression trees* (CART), and both are similar in the partitioning of intervals, where two symmetric basis functions BFs are constructed at the knot location. However, continuous piecewise linear functions are used in MARS algorithm and a continuous model that provides a more effective way to model nonlinearities is produced,

whereas CART uses indicator functions that leads to lack of continuity, which severely affects the model accuracy, and the use of methods from calculus.

Selection of BFs is data-based and specific to the problem in MARS, which makes it an adaptive regression procedure suitable for solving high-dimensional problems. Additionally, interactive structure of MARS allows users to find relations (i.e., dependencies) between variables. In MARS model building, piecewise linear BFs are fitted in such a way that additive and interactive effects of the predictors are taken into account to determine the response variable.

MARS uses two stages when building up a regression model, namely, the *forward* and the *backward step* algorithms. In the forward step, BFs to be added to the model are determined by a fast searching algorithm, and at the end, a possibly large model which overfits the data set is obtained. The process ends as soon as the maximum number of BFs,  $M_{\max}$ , a user defined value, is reached. However, the model generated in this stage often contains BFs which together contribute most or least to the overall performance, and therefore, it is more complex and implies incorrect terms.

The backward step is applied in order to prevent the model obtained in the forward step from over-fitting by decreasing the complexity of the model without degrading the fit to the data. It removes the BFs that give the smallest increase in the residual sum of squares at each step, and eventually, an optimally estimated model is generated.

MARS uses expansions of the truncated piecewise linear basis functions of the form:

$$\begin{aligned}
(\tilde{x} - \tau)_+ &= \begin{cases} \tilde{x} - \tau, & \text{if } \tilde{x} > \tau, \\ 0, & \text{otherwise,} \end{cases} \\
(\tau - \tilde{x})_+ &= \begin{cases} \tau - \tilde{x}, & \text{if } \tilde{x} < \tau, \\ 0, & \text{otherwise,} \end{cases}
\end{aligned}$$

where  $\tau$  is a univariate knot ( $\tilde{x}, \tau \in \mathbb{R}$ ). These two functions are called as a *reflected pair* and the symbol ‘+’ indicates that only the positive parts are used, and otherwise zero. General model on the relation between the input variables and their response is defined in the following equation:

$$Y = f(\tilde{\mathbf{X}}) + \varepsilon, \quad (2.1)$$

where  $Y$  is the response variable,  $\tilde{\mathbf{X}} = (\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_p)^T$  is the vector of predictors and  $\varepsilon$  is an additive stochastic component with zero mean and finite variance. The logic in MARS is to generate reflected pairs for each input  $\tilde{X}_j$  ( $j = 1, 2, \dots, p$ ) with  $p$ -dimensional knots  $\boldsymbol{\tau}_i = (\tau_{i,1}, \tau_{i,2}, \dots, \tau_{i,p})^T$  at or just nearby each input data vectors  $\tilde{\mathbf{x}}_i = (\tilde{x}_{i,1}, \tilde{x}_{i,2}, \dots, \tilde{x}_{i,p})^T$  of that input ( $i = 1, 2, \dots, N$ ). Therefore, the collection of 1-dimensional BFs of MARS is as follows:

$$C := \left\{ \left[ \tilde{X}_j - \tau \right]_+, \left[ \tau - \tilde{X}_j \right]_+, \mid \tau \in \{x_{1,j}, x_{2,j}, \dots, x_{N,j}\}, j \in \{1, 2, \dots, p\} \right\}, \quad (2.2)$$

where  $N$  is the total number of observations,  $p$  is the dimension of the input space. Here,  $f(\tilde{\mathbf{X}})$  in Equation (2.1) can be represented as a linear combination, which is successively constructed by the set  $C$  and with the intercept  $\beta_0$ , and it has the following form:



$$Y = \beta_0 + \sum_{m=1}^M \beta_m B_m(\tilde{\mathbf{X}}^m) + \varepsilon, \quad (2.3)$$

where  $B_m$  is a BF or product of two or more BFs from the set  $C$ , and it is taken from a set of  $M$  linearly independent BFs. Here,  $\tilde{\mathbf{X}}^m$  is a subvector of  $\tilde{\mathbf{X}}$  contributing to the function  $B_m$ , and  $\beta_m$  denotes an unknown coefficient of the  $m$ th BF, or the constant 1 ( $m=0$ ). By multiplying an existing BF with another reflected pair including another variable, a new BF is generated that represents interaction between different variables, and both the existing BFs and the newly created BFs are included in the model. By this way, spline fitting in higher dimensions is achieved, leading to multivariate spline BFs with the following form:

$$B_m(\tilde{\mathbf{x}}_i^m) := \prod_{k=1}^{K_m} \left[ s_{\kappa_j^m} \cdot \left( \tilde{x}_{i\kappa_j^m} - \tau_{\kappa_j^m} \right) \right]_+ \quad (i=1, 2, \dots, N), \quad (2.4)$$

where the total number of truncated linear functions multiplied in the  $m$ th BF is denoted by  $K_m$ ,  $\tilde{x}_{i\kappa_j^m}$  indicates the input variable corresponding to the  $k$ th truncated linear function in the  $m$ th BF,  $\tau_{\kappa_j^m}$  is the knot location for  $\tilde{x}_{i\kappa_j^m}$ , and finally  $s_{\kappa_j^m} \in \{\pm 1\}$ .

In order to estimate  $\beta_0$ , forward step algorithm of MARS starts with the constant function  $B_0(X^0)=1$ . All functions from the set  $C$  are considered as candidate functions, and possible BFs have the following form:

- 1,
- $x_k$ ,
- $[x_k - \tau_i]_+$ ,

- $x_k x_l$ ,
- $[x_k - \tau_i]_+ x_l$ ,
- $[x_k - \tau_i]_+ [x_l - \tau_j]_+$ .

In MARS algorithm, self-interaction is not allowed (i.e., predictor variables cannot be the same for each BF); therefore,  $x_k$  and  $x_l$  in the above BFs represent distinct predictor variables, together with their corresponding knot locations  $\tau_i$  and  $\tau_j$ , respectively. At each step, with one of the reflected pair in the set  $C$ , all products of a function  $B_m(\mathbf{X}^m)$  in the model set are considered as a new function pair, which has the following form:

$$\beta_{M+1} B_k(\mathbf{X}^k) \cdot [X_j - \tau]_+ + \beta_{M+2} B_k(\mathbf{X}^k) \cdot [\tau - X_j]_+,$$

where  $\beta_{M+1}$  and  $\beta_{M+2}$  are coefficients estimated by least squares. For instance, the following BFs are potential candidates:

- 1,
- $x_k$ ,
- $[x_k - \tau_i]_+$ , if  $x_k$  is already in the model,
- $x_k x_l$ , if  $x_k$  and  $x_l$  are already in the model,
- $[x_k - \tau_i]_+ x_l$ , if  $x_k x_l$  and  $[x_k - \tau_i]_+$  are already in the model,
- $[x_k - \tau_i]_+ [x_l - \tau_j]_+$ , if  $[x_k - \tau_i]_+ x_l$  and  $[x_l - \tau_j]_+ x_k$  are already in the model.

At each step, the model chooses the knot and its corresponding pair of BFs that result in the largest decrease in residual error, and the products satisfying the

above mentioned condition are successively added to the model until a predefined value  $M_{max}$  is reached.

At the end, a large model that typically *overfits* the initial data is obtained. Then, the backward step is applied in order to prevent the model obtained in the forward step from over-fitting by decreasing the complexity of the model without degrading the fit to the data. It removes the BFs that give the smallest increase in the residual sum of squares at each step, which means that a predictor variable can be completely excluded from the model unless any of its BFs has a meaningful contribution to the predictive performance of the model, and this iterative procedure continues until an optimal number of effective terms are represented in the final model.

Among the sequence of models obtained from the above mentioned process, an estimated best model,  $\hat{f}_\alpha$ , with the optimum number of terms  $\alpha$  that gives the best predictive fit is chosen through a *lack-of-fit* (LOF) criteria defined by *generalized cross validation* (GCV), which is given in Equation (2.5):

$$\text{LOF}(\hat{f}_\alpha) = \text{GCV}(\alpha) := \frac{\sum_{i=1}^N (y_i - \hat{f}_\alpha(x_i))^2}{(1 - Q(\alpha)/N)^2}, \quad (2.5)$$

where  $N$  is the number of sample observations,  $Q(\alpha) = u + dK$  with  $K$  representing the number of knots which are selected in forward step and  $u$  is the number of linearly independent functions in the model,  $d$  denotes a cost for each BF optimization, and usually  $d = 3$  ( $d = 2$  is used when the model is additive). The numerator is the conventional residual sum of squares, which is penalized by the denominator that accounts for the increasing variance in case of increasing model complexity, i.e., while larger  $Q(\alpha)$  creates a smaller model with less number of BFs, smaller  $Q(\alpha)$  generates a larger model with

more BFs. Using the LOF criteria, the best model is chosen according to backward step that minimizes GCV.

### 2.3 CMARS Method

As briefly discussed so far, MARS uses two algorithms (i.e., forward and backward step), through which it tries to accomplish two tasks simultaneously: a good fit to the data, yet a simple model.

CMARS (“C” stands not only for *convex* but also *continuous* and *conic*) has been developed as an alternative method to MARS by utilizing statistical learning, inverse problems and multiobjective optimization theories, and its basics is briefly discussed in this section based on [17, 18, 26].

In CMARS, backward step of MARS is not employed; instead, PRSS with  $M_{max}$  BFs is collected during forward step, and penalty terms are added to the least squares estimation in order to control the LOF, introducing an alternative point of view to the trade-off between the *complexity (accuracy)* and the *stability* of the estimation. PRSS summed up during the forward step of MARS has the following form:

$$\text{PRSS} := \sum_{i=1}^N (y_i - f(\tilde{\mathbf{x}}_i))^2 + \sum_{m=1}^{M_{max}} \varphi_m \sum_{|\delta|=1}^2 \sum_{r < s} \int_{Q^m} \lambda_m^2 [G_{r,s}^{\delta} B_m(\mathbf{h}^m)]^2 d\mathbf{h}^m. \quad (2.6)$$

$\delta = (\delta_1, \delta_2)^T \quad r, s \in V_m$

Here,  $V_m := \{\kappa_j^m | j = 1, 2, \dots, K_m\}$  denotes the variable set related with the  $m$ th BF,  $\mathbf{h}_m = (h_{m_1}, h_{m_2}, \dots, h_{m_{K_m}})^T$  is the vector of variables contributing to the  $m$ th BF. Furthermore,  $\varphi_m \geq 0$  are penalty parameters, and  $Q^m$  are suitable

integration domains  $(m=1, 2, \dots, M_{\max})$ . Finally,  $G_{r,s}^{\delta} B_m(\mathbf{h}^m) := (\partial^{\delta} B_m / \partial^{\delta_1} h_r^m \partial^{\delta_2} h_s^m)(\mathbf{h}^m)$  for  $\delta = (\delta_1, \delta_2)^T$ ,  $|\delta| := \delta_1 + \delta_2$ , where  $\delta_1, \delta_2 \in \{0, 1\}$ .

The trade-off between accuracy and complexity in this optimization problem is established through the penalties  $\varphi_m$ , and after approximating the multivariate integral  $\int_{\mathcal{Q}^m} \lambda_m^2 [G_{r,s}^{\delta} B_m(\mathbf{h}^m)]^2 d\mathbf{h}^m$  by discretizing, the PRSS in Equation (2.6) is rewritten as follows:

$$\text{PRSS} \approx \|\mathbf{y} - \mathbf{B}(\tilde{\mathbf{b}})\boldsymbol{\lambda}\|_2^2 + \sum_{m=1}^{M_{\max}} \varphi_m \sum_{i=1}^{(N+1)^{k_m}} L_{im}^2 \lambda_m^2, \quad (2.7)$$

where  $\mathbf{y} := (y_1, y_2, \dots, y_n)^T$  is the vector of response data,  $\mathbf{B}(\tilde{\mathbf{b}}) = (\mathbf{B}(\tilde{\mathbf{b}}_1), \mathbf{B}(\tilde{\mathbf{b}}_2), \dots, \mathbf{B}(\tilde{\mathbf{b}}_N))^T$  is an  $(N \times (M_{\max} + 1))$ -matrix,  $\sum_{i=1}^N (\mathbf{y}_i - \boldsymbol{\lambda}^T \mathbf{B}(\tilde{\mathbf{b}}_i))^2 = (\mathbf{y} - \mathbf{B}(\tilde{\mathbf{b}})\boldsymbol{\lambda})^T (\mathbf{y} - \mathbf{B}(\tilde{\mathbf{b}})\boldsymbol{\lambda}) = \|\mathbf{y} - \mathbf{B}(\tilde{\mathbf{b}})\boldsymbol{\lambda}\|_2^2$  with

$$L_{im} := \left[ \left[ \sum_{|\delta|=1}^2 \sum_{r < s} [G_{r,s}^{\delta} B_m(\hat{\mathbf{x}}_i^m)]^2 \right]_{\delta = (\delta_1, \delta_2)^T, r, s \in V_m} \Delta \hat{\mathbf{x}}_i^m \right]^{1/2}.$$

Then the new form of PRSS is expressed as follows:

$$\text{PRSS} \approx \|\mathbf{y} - \mathbf{B}(\tilde{\mathbf{b}})\boldsymbol{\lambda}\|_2^2 + \varphi \|\mathbf{L}\boldsymbol{\lambda}\|_2^2, \quad (2.8)$$

where  $\mathbf{L}$  is a diagonal  $((M_{\max} + 1) \times (M_{\max} + 1))$ -matrix and  $\boldsymbol{\lambda}$  denotes an  $((M_{\max} + 1) \times 1)$ -parameter vector estimated through the data points. The PRSS

problem turns into a classical TR problem with  $\varphi > 0$ ,  $\varphi = \phi^2$  for some  $\phi \in \mathbb{R}$ . A brief introductory to TR can be found in Appendix A.

The TR problem in Equation (2.8) can be treated by CQP, a technique from continuous, in fact, convex optimization, with a convenient choice of bound  $\tilde{Z} \in \mathbb{R}$  as given in the following equation:

$$\begin{aligned} & \underset{h, \boldsymbol{\lambda}}{\text{minimize}} && h, \\ & \text{subject to} && \left\| \mathbf{B}(\tilde{\mathbf{b}}) \boldsymbol{\lambda} - \mathbf{y} \right\|_2 \leq h, \quad \left\| \mathbf{L} \boldsymbol{\lambda} \right\|_2 \leq \sqrt{\tilde{Z}}. \end{aligned} \quad (2.9)$$

An introduction to CQP is available in Appendix B. At this point, one has to note that a careful learning process must be followed in order to obtain the values of bound  $\tilde{Z}$ . By applying the modern methods of continuous optimization, the CQP can be written in this basic notation:

$$\begin{aligned} & \underset{\mathbf{x}}{\text{minimize}} && \mathbf{c}^T \mathbf{x}, \\ & \text{subject to} && \left\| \mathbf{G}_i \mathbf{x} - \mathbf{g}_i \right\|_2 \leq \mathbf{p}_i^T \mathbf{x} - q_i \quad (1, 2, \dots, k), \end{aligned} \quad (2.10)$$

with  $\mathbf{c} = \left( 1, \mathbf{0}_{M_{\max}+1}^T \right)^T$ ,  $\mathbf{x} = \left( h, \boldsymbol{\lambda}^T \right)^T$ ,  $\mathbf{G}_1 = \left( \mathbf{0}_N, \mathbf{B}(\tilde{\mathbf{b}}) \right)$ ,  $\mathbf{g}_1 = \mathbf{y}$ ,  $\mathbf{p}_1 = \left( 1, 0, \dots, 0 \right)^T$ ,  $q_1 = 0$ ,  $\mathbf{G}_2 = \left( \mathbf{0}_{M_{\max}+1}, \mathbf{L} \right)$ ,  $\mathbf{g}_2 = \mathbf{0}_{M_{\max}+1}$ ,  $\mathbf{p}_2 = \mathbf{0}_{M_{\max}+2}$  and  $q_2 = -\sqrt{\tilde{Z}}$ .

After writing the optimality conditions for the problem (2.9):

$$\underset{h, \boldsymbol{\lambda}}{\text{minimize}} \quad h,$$

such that

$$\begin{aligned}\boldsymbol{\chi}^{i,j} &:= \begin{pmatrix} \mathbf{0}_N & \mathbf{B}(\tilde{\mathbf{b}}) \\ 1 & \mathbf{0}_{M_{\max}+1}^T \end{pmatrix} \begin{pmatrix} h \\ \boldsymbol{\lambda} \end{pmatrix} + \begin{pmatrix} -\mathbf{y} \\ \mathbf{0} \end{pmatrix}, \\ \boldsymbol{\eta} &:= \begin{pmatrix} \mathbf{0}_{M_{\max}+1} & \mathbf{L} \\ \mathbf{0} & \mathbf{0}_{M_{\max}+1}^T \end{pmatrix} \begin{pmatrix} h \\ \boldsymbol{\lambda} \end{pmatrix} + \begin{pmatrix} \mathbf{0}_{M_{\max}+1} \\ \sqrt{\tilde{Z}} \end{pmatrix},\end{aligned}$$

where  $\boldsymbol{\chi}^{i,j} \in L^{N+1}$ ,  $\boldsymbol{\eta} \in L^{M_{\max}+2}$ ,  $L^{N+1}$ , and  $L^{M_{\max}+2}$  are  $(N+1)$ - and  $(M_{\max}+2)$ -dimensional ice-cream (or *second-order*, or *Lorentz*) cones, respectively:

$$L^{N+1} := \left\{ \mathbf{x} = (x_1, \dots, x_{N+1})^T \in \mathbb{R}^{N+1} \mid x_{N+1} \geq \sqrt{x_1^2 + x_2^2 + \dots + x_N^2} \right\} \quad (N \geq 1).$$

The competitiveness of CMARS compared with MARS has been demonstrated in other applied contexts such as industrial engineering (e.g., quality management and control in manufacturing) [17, 53, 54], financial mathematics (e.g., prediction of credit default, and identification of stochastic differential equations in finance) [27, 55-59], or Earth sciences (e.g., prediction of ground motion, related to tectonics and earthquakes) [60]; in various cases of data sets, CMARS was superior even.





## CHAPTER 3

### A BRIEF REVIEW ON RADIATIVE TRANSFER MODELS AND IMAGE CLASSIFICATION

#### 3.1 Atmospheric Correction by RT Models

When examining Earth's surface from a space-based RS platform, the atmosphere is a large factor in the uncertainty associated with a surface reflectance measurement. In order to obtain high quality data, the correction of atmospheric perturbations acting upon land surface reflectance measurements recorded by a space-based sensor is an important topic within RS [61].

Dependent on the atmospheric conditions at any one time, the measured reflectance from a particular remotely sensed pixel may be increased or decreased by the atmosphere. A fundamental problem within RS of Earth is to correct land surface reflectance data gathered by space based sensors for the perturbations introduced by the passage of radiation through Earth's atmosphere [62, 63].

Failure to correct for the atmospheric effects can inevitably lead to different surface reflectance values, and will definitely have a significant effect on any conclusions drawn from such data. This issue is particularly important in the evaluation of images taken by *large field of view* (LFOV) sensors and spanning long time series, over which the atmospheric parameters do not remain constant [61, 63].

On its path from the Sun to the surface and onwards to the sensor, a photon can be disturbed from its course through the influence of atmospheric absorption and scattering. The former represents the amount of radiation that is attenuated by its passage from the Sun to surface and onwards to the satellite and decreases the radiance measured at satellite level. Similarly, the latter accounts for the component of radiation that is scattered within the atmosphere and directed toward the satellite, depending upon atmospheric conditions it may either increase or decrease the measured radiance. Both components can be calculated using knowledge of the atmospheric conditions at the time of measurement by examining the path length of the light through the atmosphere [61, 63].

Solar radiation is scattered by atmospheric gaseous molecules and aerosols, and two broadly known scattering mechanisms exist, namely *Rayleigh* and *Mie* (or aerosol) scattering [64]. The former is scattering by the air molecules themselves, and it is dominant within the short wavelength region between 0.4 and 0.7  $\mu\text{m}$ . It is an inverse fourth power function of the wavelength used, i.e., its effect diminishes rapidly with increasing wavelength ( $\lambda^{-4}$ ). The latter is a result of scattering of the radiation from larger particles such as those associated with smoke, haze and fumes. The effect of Mie scattering also decreases with increasing wavelength, but at a slower rate (typically,  $\lambda^{-2}$  to  $\lambda^{-1}$ ). The wavelength dependence disappears when the atmospheric particulates become much larger (i.e., fogs, clouds and dust) [65, 66].

On the other hand, absorption by atmospheric molecules is a selective process and the incoming energy is converted into heat during this process. Eight gases among the nearly thirty atmospheric gases, which are water vapor ( $\text{H}_2\text{O}$ ), carbon dioxide ( $\text{CO}_2$ ), ozone ( $\text{O}_3$ ), nitrous oxide ( $\text{N}_2\text{O}$ ), carbon monoxide ( $\text{CO}$ ), methane ( $\text{CH}_4$ ), oxygen ( $\text{O}_2$ ), and nitrogen dioxide ( $\text{NO}_2$ ), result in observable absorption features over the range 0.4 to 2.5  $\mu\text{m}$ . *Atmospheric water vapor* (uWV) absorption has a significant effect over the half of the spectral

region between 0.4 and 2.5  $\mu\text{m}$ , whereas the absorption effects due to the other seven gases usually take place within narrower wavelength intervals [65, 66].

As explained briefly, atmospheric components such as aerosol or water vapor content can substantially modify the *top of atmosphere* (TOA) radiance as seen by the satellite. In addition, the *view zenith angle* (VZA),  $\theta_v$ , and the *solar zenith angle* (SZA),  $\theta_s$ , also play a major role in determining the effects of the atmosphere (Figure 3.1). If the zenith angle is far from nadir then the photon must travel through a much larger portion of the atmosphere, and thus the chance of an absorption or scattering event greatly increases. Conversely, for angles close to nadir, the path length is greatly reduced, as is the uncertainty in radiance due to the atmosphere [61, 62, 66].

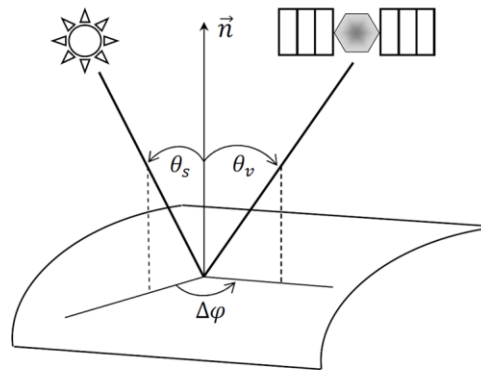


Figure 3.1: Geometry of the problem,  $\theta_s$ : solar zenith angle,  $\theta_v$ : viewing zenith angle,  $\Delta\varphi$ : relative azimuth between the Sun and the sensor (adapted from [7])

These atmospheric and geometric parameters are input values in RT models, and have to be a priori determined. The finally calculated value is the target reflectance, which is supposed to be equal to the in-situ spectroradiometrically measured reflectance value, as long as the RT model is accurate and reliable [4].

In images, acquired by instruments with high temporal resolution and LFOV (i.e., NOAA AVHRR, MSG-SEVIRI), each pixel has a different observational geometry and each atmospheric parameter has to be calculated for each pixel [7, 63, 65]. Hence, it is highly costly and time consuming to employ models based on a rigorous treatment of the RT problem, like *second simulation of a satellite signal in the solar spectrum* (6S) [67], for the atmospheric correction of bulk numbers of such satellite images in real-time.

The *simplified model for atmospheric correction* (SMAC) was first introduced nearly two decades ago as an alternative approach to tackle these problems. It is based on the parametrization of detailed RT models [7]. Despite its age and lower level of accuracy as compared to 6S, it is still a preferred atmospheric correction tool for many research groups due to its high speed [61].

Even though no atmospheric correction scheme can completely nullify the effects of the atmosphere, RT in clear atmosphere is now well understood and a significant amount of work has been done in improving the accuracy and quality of correction algorithms [61, 65].

### **3.1.1 Definitions of reflectance and radiance**

To help the readers, we think, it is necessary to briefly explain two common optical terms continuously used in this study: *reflectance* and *radiance*.

Basic definition of reflectance is the ratio of reflected energy to incident energy at a given wavelength, and it is unitless [68]. Three main factors affects the quantity of reflected energy [2]:

- magnitude of the incident energy,
- material roughness,

- and, material type.

In general, the first two factors are assumed as constants, and the third factor is taken into account. Surface roughness plays an important role on the behavior of reflectance [2]. If the incoming energy is reflected by a *specular* (i.e., perfectly smooth) surface, the angle of reflection is the same as the angle of incidence; whereas, equal reflection in all directions is observed in case of a *Lambertian* surface (i.e., perfectly rough surface) as illustrated in Figure 3.2. In real life, many materials exhibit reflectance behaviors between specular and Lambertian [10].

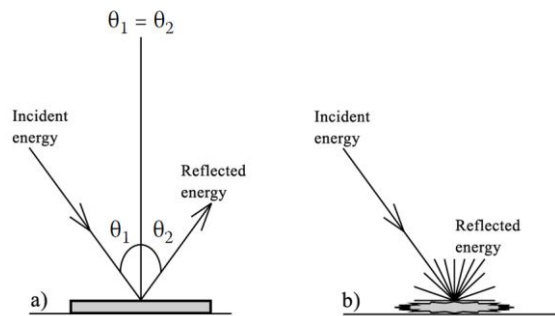


Figure 3.2: Reflectance from a) specular, and b) Lambertian surface (adapted from [2])

On the other hand, radiance is used to measure the amount of radiation, either passing through or being emitted from a surface, that falls within a given solid angle in a given direction [68]. Its unit is Watts per square meter for a given wavelength per steradian ( $\text{W}/\text{m}^2/\mu\text{m}/\text{sr}$ ).

### 3.1.2 Basic RT formulation

In RT models, the process for obtaining the ground target's reflectance value has three main stages as illustrated in Figure 3.3. In the first stage, the individual pixel values (DNs) are converted to apparent radiances ( $L_{app}$ ). Then,

the second stage includes the conversion from  $L_{app}$  to apparent reflectance values,  $\rho^*$ . The final step is to convert  $\rho^*$  to ground target reflectance values,  $\rho_s$ . In this section, in order to provide an introduction to atmospheric modeling, basic RT equation is described based on [2, 65, 69].

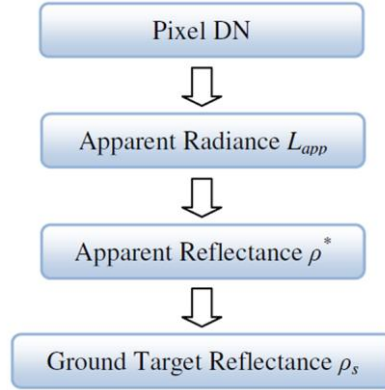


Figure 3.3: Steps in an RT-based atmospheric correction scheme (adapted from [2])

From the path between the Sun and the ground target, a certain fraction of the incident solar irradiance is scattered into the atmosphere. The remainder of the radiation that is incident on the ground target is direct solar radiation (Figure 3.4a), and this transmitted fraction,  $T(\theta_s)$ , is given as follows:

$$T(\theta_s) = \exp\left(\frac{-\tau}{\cos \theta_s}\right), \quad (3.1)$$

where  $\tau$  is the *atmospheric optical depth* (AOD), and  $\theta_s$  denotes the SZA.

Some portion of the solar radiation is first scattered by molecules or suspended particles in the atmosphere, and then reaches the surface. It is known as diffuse

skylight,  $t_d(\theta_s)$ , (Figure 3.4c). Once the effect of diffuse skylight is taken into account, Equation (3.1) becomes as follows:

$$T(\theta_s) = \exp\left(\frac{-\tau}{\cos\theta_s}\right) + t_d(\theta_s). \quad (3.2)$$

The contribution of the trapping mechanism, which is actually another scattering flux, should also be taken into account. The effect arises from the successive reflection and scattering of solar radiation between the neighboring ground target and the atmosphere (Figures 3.4d and 3.4e). Spherical albedo of the atmosphere,  $S$ , and the surface reflectance,  $\rho_s$ , determines the magnitude of this effect. Then, the illumination of the ground target is expressed as follows:

$$\frac{1}{1 - \rho_s S} \times T(\theta_s). \quad (3.3)$$

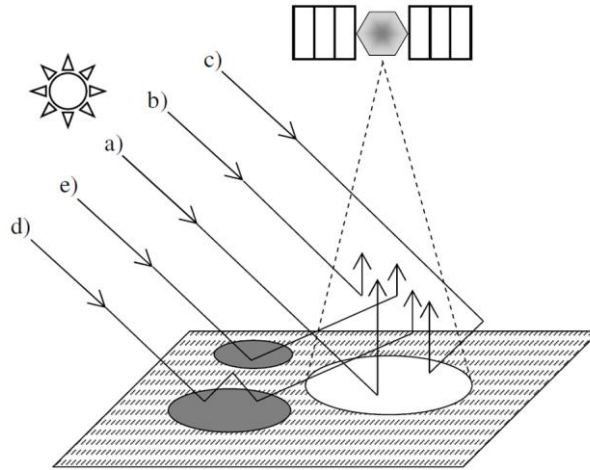


Figure 3.4: Interaction of the solar radiation with atmosphere: a) direct, b) path, c) diffuse, d) multiple, and e) neighbor (adapted from [2])

The proportion of the solar radiation reflected from the ground target is given by the following equation:

$$\frac{\rho_s}{1-\rho_s S} \times T(\theta_s). \quad (3.4)$$

The total solar radiation reflected by the ground target and directly transmitted from the surface to the sensor is not the only main source. The second source, which is also scattered into the field of view of the sensor, is the contribution from the neighboring target. The reflectance received by the sensor is given in Equation (3.5):

$$\frac{\rho_s}{1-\rho_s S} \times T(\theta_s) \times T(\theta_v). \quad (3.5)$$

Here,  $\theta_v$  is the VZA, and as in Equation (3.2),  $T(\theta_v)$  can be expressed as follows:

$$T(\theta_v) = \exp\left(\frac{-\tau}{\cos\theta_v}\right) + t_d(\theta_v). \quad (3.6)$$

A fraction of the solar radiation scattered out of the downward solar beam also enters into the sensor's field of view without interaction with the ground target (Figure 3.4b), and it is called as the atmospheric reflectance,  $\rho_a(\theta_s, \theta_v, \Delta\varphi)$ , where  $\Delta\varphi$  denotes the relative azimuth between the Sun and the sensor. Then, the apparent reflectance  $\rho^*$  at the sensor is given as follows:

$$\rho^* = \frac{\rho_s}{1-\rho_s S} \times T(\theta_s) \times T(\theta_v) + \rho_a(\theta_s, \theta_v, \Delta\varphi), \quad (3.7)$$



where  $\rho_s$  is the surface reflectance, and the term  $T(\theta_s) \times T(\theta_v)$  denotes the total atmospheric transmittance along the Sun–target–sensor path.

Equation (3.7) can be modified in order to consider the effect of atmospheric gaseous absorption, as mentioned in Section 3.1. Then, the apparent reflectance  $\rho^*$  is calculated by the following equation:

$$\rho^* = T_g(\theta_s, \theta_v) \times \left[ \frac{\rho_s}{1 - \rho_s S} \times T(\theta_s) \times T(\theta_v) + \rho_a(\theta_s, \theta_v, \Delta\varphi) \right], \quad (3.8)$$

where  $T_g(\theta_s, \theta_v)$  is the atmospheric gas transmittance after absorption. As obviously seen, Equation (3.8) deals with the atmospheric scattering and gaseous absorption as two independent processes, whereas they occur simultaneously in the real atmosphere.

One should further expand the Equation (3.8) in order to obtain the surface reflectance of the ground target as follows:

$$\begin{aligned} \frac{\rho^*}{T_g(\theta_s, \theta_v)} - \rho_a(\theta_s, \theta_v, \Delta\varphi) &= \frac{\rho_s}{1 - \rho_s} \times T(\theta_s) \times T(\theta_v) \\ \Rightarrow \frac{\rho^*}{T_g(\theta_s, \theta_v) \times T(\theta_s) \times T(\theta_v)} - \frac{\rho_a(\theta_s, \theta_v, \Delta\varphi)}{T(\theta_s) \times T(\theta_v)} &= \frac{\rho_s}{1 - \rho_s}. \end{aligned} \quad (3.9)$$

Then, Equation (3.9) is transformed into Equation (3.10)

$$\rho_s = \frac{A \times \rho^* + B}{\left[ 1 + S \times (A \times \rho^* + B) \right]}, \quad (3.10)$$

with  $A = \frac{1}{T_g(\theta_s, \theta_v) \times T(\theta_s) \times T(\theta_v)}$ , and  $B = -\frac{\rho_a(\theta_s, \theta_v, \Delta\varphi)}{T(\theta_s) \times T(\theta_v)}$ .

Hence, once the apparent reflectance  $\rho^*$ , the spherical albedo  $S$  and the coefficients  $A$  and  $B$  are known, one can obtain the surface reflectance  $\rho_s$  of the ground target.

### **3.1.3 Evolution of RT models**

Generally speaking, the goal of all correction schemes is to nullify atmospheric influence by modifying TOA radiance measurement recorded by a sensor in accordance with calculated values for the atmospheric absorption and scattering along the path travelled by incident light. This process results in a surface radiance, or after further computation, a surface reflectance that is free from influence by atmospheric scattering and absorption.

Numerous methods and algorithms, starting from the earlier empirical line method and the flat field method to more recent RT modeling approaches, have been proposed in order to remove atmospheric effects from images taken by sensors with different technical specs and operational objectives, installed either on aircraft or satellite platforms. The intend of the following subsections, however, is not to cover all these methods in a comprehensive manner. Instead, several of widely recognized correction schemes based on rigorous treatment of RT models are briefly introduced in a chronological fashion.

#### **3.1.3.1 MODTRAN**

*Moderate spectral resolution atmospheric transmittance* (MODTRAN) algorithm was developed by US Air Force Research Laboratory, Geophysics Directorate in collaboration with Spectral Sciences, Inc. in the late 1980s, and has been widely used to analyze AVIRIS data due to its computational speed

and its ability to model molecular and aerosol/cloud emissive and scattered radiance contributions as well as the atmospheric attenuation [70].

The MODTRAN code calculates atmospheric transmittance and radiance for frequencies from 0 to 50 000  $\text{cm}^{-1}$  (wavelength: 200 nm to  $+\infty$ ) at moderate spectral resolution, primarily 2  $\text{cm}^{-1}$  (20  $\text{cm}^{-1}$  in the UV). The basic approach in MODTRAN is to approximate the atmosphere and Earth surface as a sequence of quasi-homogeneous layers for which the individual layer radiance contributions from each of the source terms depicted is considered. The surface is treated as a layer of infinite opacity, an opaque boundary with variable emissivity/reflectivity. Spherical refraction geometry effects are incorporated into calculation of path sums and scattering angles, although multiple scattering radiances are based on plane-parallel models. Approximate corrections are made for the effects of inhomogeneous distributions of temperature and species concentrations within the atmospheric layers [71, 72].

### **3.1.3.2 5S**

*Simulation of a satellite signal in the solar spectrum (5S)* was developed for estimation of the solar radiation backscattered by Earth-surface-atmosphere system, as observed by a satellite sensor. Given the Lambertian ground reflectance, the apparent reflectance of the observed pixel is estimated by taking into account the effects of gaseous absorption, scattering by molecules and aerosols and, to some extent, inhomogeneity in the ground reflectance. Besides the pixel apparent reflectance, the code provides the gaseous transmittance, the irradiance at the surface and the different contributions to the satellite signal according to the origin of the measured radiance [69].

5S treats absorption and scattering separately since gaseous transmission is one factor in a series of terms describing scattering. For a given spectral band, 5S

calculates each term in Equation (3.8) at a step of 5 nm within the spectral interval, computes the TOA spectral reflectance, and then integrates TOA spectral reflectances over the whole spectral band, giving the TOA apparent reflectance,  $\rho^*$ . These integrations of each parameter of Equation (3.8) over the spectral band are highly demanding in computational time [7].

### 3.1.3.3 ATREM

The *atmosphere removal algorithm* (ATREM), developed by the University of Colorado in the early 1990s, allows the retrieval of surface reflectance spectra from imaging spectrometer data, such as those collected with AVIRIS, without the need for field measurements of reflectance spectra. Since water vapor has a spatially variable characteristic, the integrated water vapor amount is obtained on a pixel by pixel basis from the 0.94 and 1.14  $\mu\text{m}$  water vapor absorption bands using a channel ratio technique [73].

By using a narrow band spectral model and the derived water vapor values, the water vapor transmittance spectrum in the complete solar spectral region between 0.4 and 2.5  $\mu\text{m}$  is then simulated. Similarly, based on the solar and observational geometry, the transmission spectra of  $\text{CO}_2$ ,  $\text{O}_3$ ,  $\text{N}_2\text{O}$ ,  $\text{CO}$ ,  $\text{CH}_4$ , and  $\text{O}_2$  in the 0.4 - 2.5  $\mu\text{m}$  region are simulated. 5S computer code is used for the modeling of scattering effect due to atmospheric molecules and aerosols. An aerosol model and a surface visibility should be selected by the user to model the aerosol effect. The apparent reflectances are obtained by taking the ratio of measured radiances to the solar irradiances above the atmosphere. Simulated atmospheric gaseous transmittances and the simulated molecular and aerosol scattering data are used to derive the surface reflectances from the apparent reflectances. The final output of the ATREM consists of water vapor image and reflectance-corrected AVIRIS data without use of ground spectral

measurements. ATREM is no longer available to new users and ATREM v3.1 was the last publicly released version [65].

### 3.1.3.4 SMAC

*Simplified method for atmospheric correction* (SMAC) is a computationally fast and accurate technique for the atmospheric correction of satellite measurements in the solar spectrum and was developed by Rahman and Dedieu [7] in 1994. The main advantage of the method is that it is several hundred times faster than the more detailed RT models. SMAC is based on the parameterization of the RT equations. A separate equation for each of the atmospheric interaction processes is defined and the coefficients of these equations are adjusted to match an accurate RT model. For the development of the first version of SMAC, 5S was chosen as the reference model, whereas the current version is based on the 6S model.

Apart from a series of predefined coefficients, dependent upon the RS instrument, the algorithm requires 7 input variables:  $\rho^*$ ,  $\theta_s$ ,  $\theta_v$ ,  $\Delta\varphi$ , AOD, uWV and *ozone content* (uO<sub>3</sub>). Then, the surface reflectance of the ground target,  $\rho_s$ , is calculated on a pixel-by-pixel basis, computing the surface reflectance for each before moving on to the next.

Most of the equations used in the calculation of process variables are modified into a simpler form in SMAC, and by this way, calculation time is reduced by decreasing complexity. In order to calculate the gaseous transmission, an empirical band model is used in the following form [63]:

$$T_{g_i}(\theta_s, \theta_v) = \exp\left[a(mU)^n\right], \quad (3.11)$$

with  $m = \frac{1}{\mu_s} + \frac{1}{\mu_v}$ .

Here,  $U$  denotes the vertically integrated absorber amount,  $a$  and  $n$  are predefined coefficients by analysis of 6S data, which are read into the SMAC at run time, and finally,  $\mu_s$  and  $\mu_v$  are the cosine of the SZA and cosine of the VZA, respectively.

In the presence of several gases in the atmosphere, the total gaseous transmittance is retrieved in SMAC by multiplying the transmissions from each individual gas as given in Equation (3.12) [63]:

$$T_{g_i}(\theta_s, \theta_v) = \prod_{i=1}^x T_{g_i}(\theta_s, \theta_v). \quad (3.12)$$

An empirical approach for the total atmospheric transmission is used and expressed as a function of AOD at 550 nm and SZA (or VZA) in the following form [7]:

$$T(\theta_s) = a_0 + a_1 \frac{\tau_{550}}{\mu_s} + \frac{a_2}{\mu_v}, \quad (3.13)$$

where  $a_n$  are constants to be adjusted for a given spectral band,  $\tau_{550}$  is the AOD at 550 nm. In Equation (3.13), one has to note that (i) the (cosine) SZAs are replaced by the equivalent VZA values in the case of outgoing radiation; and (ii) this transmission accounts for both Rayleigh and aerosol scattering (i.e., when  $\tau_{550} = 0$ , the transmission corresponds to Rayleigh scattering).

Despite its age, the SMAC algorithm is still widely used by many groups and projects both with MSG-SEVIRI data [74-76] and that gathered from other satellites [77-79]. The primary reason why SMAC approach is still used by so

many groups is its simple nature, which means that it can perform an atmospheric correction in a short time.

In addition, when new sensors are introduced, only the band-specific coefficients need to be updated while the routines remain the same. Besides, SMAC is available as a free and open-source code. This means that it can easily be implemented in an operational data preprocessing computer and can be modified according to user requirements, which is another attractive feature of SMAC algorithm that makes it still popular. SMAC code is given in Appendix E.1.

### **3.1.3.5 FLAASH**

*Fast line-of-sight atmospheric analysis of spectral hypercubes* (FLAASH) is a MODTRAN4-based atmospheric correction software package developed by the Air Force Phillips Laboratory, Hanscom AFB and Spectral Sciences, Inc (SSI). It provides accurate, physics-based derivation of apparent surface reflectance through derivation of atmospheric properties such as surface albedo, surface altitude, water vapor column, aerosol and cloud optical depths, surface and atmospheric temperatures from hyperspectral imagery data. FLAASH operates in the 0.4 - 2.5  $\mu\text{m}$  spectral range. FLAASH uses the standard equation for spectral radiance at a sensor pixel in the solar wavelength range (neglecting thermal emission) from a flat Lambertian surface or its equivalent [80].

In several studies, it has been shown that at least under conditions of clear to moderate aerosol/haze, low to moderate water vapor, and nadir viewing from any altitude between the ground and the top of the atmosphere, the algorithm is capable of generating accurate surface reflectance spectra from hyperspectral imagery [81, 82].

### 3.1.3.6 ACORN

*Atmospheric correction now* (ACORN) is a commercially available, enhanced atmospheric model-based software that uses licensed MODTRAN4 technology to produce high quality surface reflectance without ground measurements [83]. The software provides an atmospheric correction of hyperspectral and multispectral data measured in the 0.4 - 2.5  $\mu\text{m}$  spectral range.

ACORN uses look-up tables (LUT) calculated with the MODTRAN4 RT code to model atmospheric gas absorption as well as molecular and aerosol scattering effects, converting the calibrated sensor radiance measurements to apparent surface reflectance. The well mixed gases are constrained by the elevation and the observation geometry. Water vapor is estimated from the data on a pixel-by-pixel basis using the water vapor absorption bands at 0.94 and/or 1.15  $\mu\text{m}$ . A lookup table for a range of water column vapor densities is generated using MODTRAN4 and then fitted in a least-squares sense against the imaging spectrometer data.

A key feature of ACORN is full spectral fitting to solve for the overlap of absorptions between water vapor and liquid water in surface vegetation. Visibility is estimated from the AVIRIS data using nonlinear least-squares spectral fitting between the AVIRIS radiance spectra and MODTRAN modeled radiance with the aerosol optical depth as the primary fitting parameter. The two-way transmitted radiance and atmospheric reflectance are calculated for each pixel using MODTRAN and the derived water vapor, pressure elevation, and aerosol optical depth estimations. Apparent surface reflectance is derived from the total upwelling spectral radiance for a given atmosphere using a variant of the RT equation [65, 81].



### 3.1.3.7 ATCOR

*Atmospheric and topographic correction* (ATCOR) code was created in the early 2000s, and is used for computing a ground reflectance image for the reflective spectral bands, and emissivity images for the thermal bands [84]. ATCOR has been evolved in the last decade in two different types as ATCOR2 and ATCOR3. The former is a spatially-adaptive fast atmospheric correction algorithm for nearly horizontal surface or flat terrain, whereas the latter is designed for rugged topographical surface. Hence, a *digital elevation model* (DEM) is used in the correction algorithm [85, 86].

Also the third type of the algorithm, ATCOR4, is available for the atmospheric correction of airborne imaging spectrometry data. ATCOR4 is an extension of the ATCOR3, and a LUT approach is used for the atmospheric correction employing the Lambertian assumption. It performs the combined atmospheric/topographic correction accounting for the angular and elevation dependence of the atmospheric correction functions and calculates surface reflectance (solar spectral wavelength region) and surface temperature (thermal wavelength region) based on the geocoded and orthorectified imagery [84].

### 3.1.3.8 6S

The *second simulation of a satellite signal in the solar spectrum* (6S) is a computer code that can accurately simulate the surface reflectance values measured by satellite or airborne platforms of land or sea surfaces in the visible and near infrared, which is strongly affected by the presence of the atmosphere along the Sun-ground target-sensor path, and its first version was introduced in the late 1990s as an improved version of 5S [87].

6S uses the surface/atmosphere *bidirectional reflectance distribution function* (BRDF) to compute the surface reflectance, as stated in Equation (3.8). It

accounts for the coupling between water vapor absorption and aerosol scattering, elevated targets, use of anisotropic and Lambertian surfaces, and for the calculation of gaseous absorption. The new vector version of 6S code (6SV) can work in both scalar and vector modes, and accounts for the polarization of solar radiation in the atmosphere [88]. It is based on the method of *successive orders of scattering* (SOS) approximations. Within this method, the atmosphere is divided into a number of layers and the RT equation is solved numerically for each layer with the help of iterations. The intensity is successively computed for photons scattered one, two, three times, etc., with the total intensity obtained as the sum of all orders. Numerical integration is performed using the decomposition in Fourier series for the azimuth angle and Gaussian quadratures for the zenith angle [89].

6SV is the basic code underlying the MODIS atmospheric correction algorithm, and several significant updates have been introduced into the code including [88, 89]:

- a more accurate calculation of highly asymmetric aerosol scattering phase functions,
- an arbitrary variation of a vertical aerosol profile,
- the ability to change the number of calculation angles and layers,
- and, the increase in the number of node wavelengths from 10 to 20.

### **3.2 A Short Introductory to Image Classification**

In RS, individual pixels each carrying the radiometric information (i.e., reflectance or radiance values recorded by the sensor within different a priori defined spectral band intervals) are used to form a multispectral image [2]. Therefore, in a multidimensional space, called *feature space*, each spectral

band is represented by an axis so that a pixel can be expressed as a point in that space [10] (Figure 3.5). If the user works with a 2-D feature space (i.e., number of spectral bands is two), lines or curves are used to separate clearly defined groups, *classes*, in which an individual pixel falls. These lines or curves are known as *decision boundaries*. Then the pixel is labeled, i.e., *classified*, according to its position with respect to the decision boundary. In case of higher-dimensional problems, these lines or curves are replaced with *hyperplanes* [10]. So, in a multidimensional feature space, classification can be regarded as positioning the hyperplanes, and the method that specifies the position of a pixel with respect to the hyperplane is called a *classifier*.

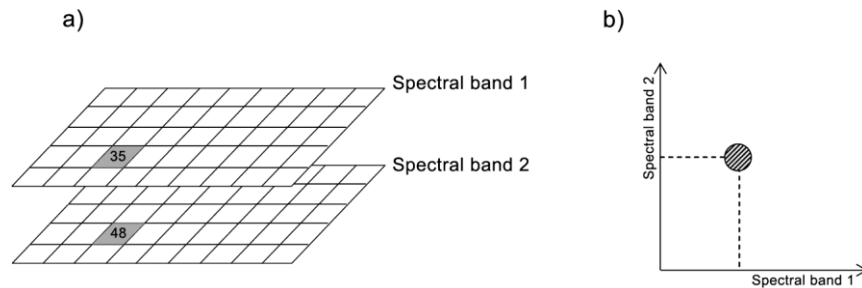


Figure 3.5: From a) image space, to b) feature space

### 3.2.1 Supervised versus unsupervised

Thus, the main objective in image classification is to establish a link between each pixel and one or more user-defined labels of classes by using the radiometric information contained in the image (Figure 3.6), and to obtain a thematic map of the study area [2]. There exist two main approaches:

- The class labels in the study area are known by the user beforehand,

- Or, the class labels are unknown, and the user may want to determine the number of separable classes. Then the labels of the separable classes are assigned by the user based on his/her expertise of the study area.

The former approach is known as *supervised* classification, whereas the latter as *unsupervised* classification or *clustering*.

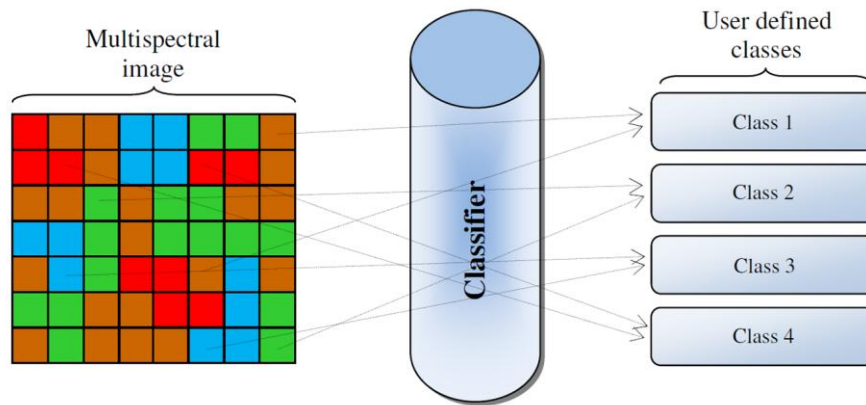


Figure 3.6: A multispectral image classifier (adapted from [2])

In supervised classification, the user divides the area of his/her interest in a specific number of groups in advance. Then he/she acquires samples from each group, known as *training data*, to train the classifier in order to determine the decision boundaries in feature space [64]. So, within this respect, supervised classification methods require rigorous and frequent user interaction. Moreover, the user needs to have sufficient background and experience regarding the type and number of groups existing in the study area [2].

On the other hand, unsupervised classification methods require less user interaction since they learn the characteristics of the groups (i.e., classes) directly from the input data. Unsupervised classification is usually preferred in cases where the user has less or insufficient knowledge about the data set, and wants to see if there exists any natural grouping in the data [2, 9]. Unsupervised classification methods are mostly iterative procedures. The

number of classes is a priori defined by the user, and the classifier continuously refines the class means at each iteration until they converge and remain stable in the feature space [2]. The final products of the unsupervised classification are subsets called clusters, each of which is linked to a specific group and labeled accordingly by the user.

KM [20] and FC [13] can be given as examples of unsupervised classification techniques. ML [11], *parallelepiped* (PP) [64] and MD [10] are the well-known techniques used in supervised classification. It is also of value to mention that in some image classification approaches, both supervised and unsupervised techniques can be used in combination [90, 91].

Even though unsupervised approach sounds more straightforward and automatic, the levels of accuracy achieved in this method are often lower than that obtained in supervised method. The user should pay special attention when classifying multispectral images acquired by RS devices since they are very complex entities including not only spectral attributes but also spatial attributes. The phenomenon known as the *mixed pixel problem* [2, 92] can often be encountered in real-life applications, especially, in complex classification problems, where classes are usually overlapping *spatially* and *spectrally*. The former means that a pixel can be contained in areas which are represented by more than one class, whereas the latter indicates that the radiometric characteristics of different classes may exhibit similar behaviors. Although the overall ability of supervised classification methods originated from conventional statistical techniques such as ML or MD approaches are limited in solving the mixed pixel problem, they were the main classification methods used in RS until the mid-1990s [2].

Recent improvements in applied mathematics, especially, in the statistical learning and data mining fields, together with the parallel developments in the computer technology, have allowed researchers to use many alternative image

classification techniques based on DT [93, 94], SVM [95, 96], ANN [97, 98] and FC [99, 100]. The reader can find a compact review and comparison of these techniques in [8].

### 3.2.2 Parametric versus nonparametric

As discussed in [8, 15], traditional methods in image classification can be grouped under two main categories: (i) parametric, and (ii) nonparametric methods, and each has its own drawbacks arising mainly from the inherent characteristic of multispectral image data. Parametric methods are mostly based on Bayesian approach, in which the training data are employed to estimate the parameters (i.e., mean and covariance matrices) of the probability density function of each class in order to generate the decision boundaries. These probability density functions are generally assumed to follow a normal multivariate distribution; however, this condition is hardly met in remotely sensed data [8]. Another problematic issue in parametric methods, especially, in case of complex landscapes with high-dimensional data, is the number of samples that the user has to collect in order to train the classifier should be increased. Additionally, certain level of uncertainty may further be introduced due to insufficient and non-representative training samples [8].

The most widely-used parametric classification method is perhaps the ML. The main reason behind this fact is that it is robust, and it is available in almost every image-processing software as a built-in feature [8]. In ML classification approach, a set of user-defined classes is a priori determined, and then, the probability  $P(\mathbf{x})$  that a pixel vector  $\mathbf{x}$  of  $p$  elements belonging to a class  $i$  is calculated based on the following Bayesian multivariate normal density function [2, 64]:

$$P(\mathbf{x}) = 2\pi^{-0.5p} |\mathbf{U}_i|^{-0.5} \exp\left[-0.5(\mathbf{y}^T \mathbf{U}_i^{-1} \mathbf{y})\right], \quad (3.14)$$

where  $|\cdot|$  is the determinant,  $\mathbf{U}_i$  denotes the variance-covariance matrix of the  $i$ th class,  $\mathbf{y} = (\mathbf{x} - \bar{\mathbf{x}}_i)$ , and finally,  $\bar{\mathbf{x}}_i$  is the multivariate mean of the  $i$ th class.

On the other hand, nonparametric methods are those that do not make any statistical a priori assumption about the underlying density function. Consequently, a nonparametric classifier does not need any statistical parameters, and the decision boundaries in a multidimensional feature space are obtained from training data of all classes [2, 8]. In the studies of [101] and [102], it has been shown that the performance of nonparametric methods, even with small training samples, are better than parametric ones. In case of mixed pixel problem (i.e., training samples overlap), a nonparametric approach should be preferred.

However, as indicated in [15] again, some of nonparametric methods used in image classification such as ANN and FC models suffer from certain problems related to their *black box* nature (i.e., no physical insight is available or used, but the chosen model is known to have good flexibility and has performed well in the past [103]). So, in such methods it is hardly possible to understand the relation between the predictor variables and the classification results, which hinders the generalizability of the classification. Let us note that in the sense of our studies, a “black box” nature may also be called an *inverse problem*.





## CHAPTER 4

### MODIS INSTRUMENT AND PRODUCTS

#### 4.1 MODIS Instrument

MODIS is the name of the two cross-track scanning radiometer instruments operated by NASA: one has been on board to Terra satellite since 1999 and the other to Aqua satellite since 2002 [104]. The former has a descending orbit (southward) over the equator about 10:30 local Sun time, whereas the latter has an ascending orbit (northward) over the equator about 13:30 local Sun time.

The swath width of each instrument is about 2330 km, and each scan is 10 km along track (Figure 4.1). They view Earth from an altitude of about 700 km above the surface with a  $\pm 55^\circ$  view scan, which means nearly the entire globe is observed on a daily basis [104]. Both instruments capture data in 36 spectral bands ranging in wavelength from  $0.4 \mu\text{m}$  to  $14.4 \mu\text{m}$  at varying spatial resolutions (Table 4.1) in order to investigate Earth's global dynamics such as radiation budget, cloud cover and atmosphere [105].

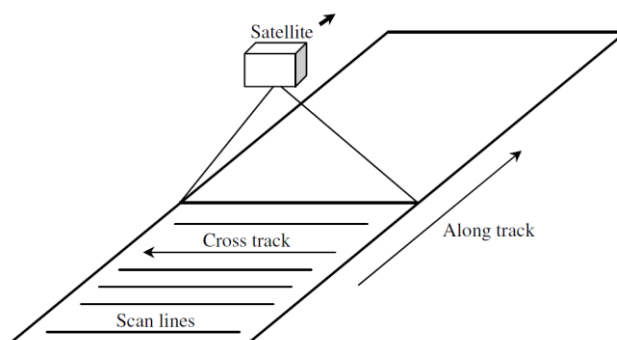


Figure 4.1: Observation swath of MODIS (adapted from [106])

Table 4.1: Description of MODIS spectral bands

Band	Range (μm)	Centered at (μm)	Use
<b>Reflective solar bands at 250 m spatial resolution</b>			
1	0.620-0.670	0.659	aerosol, cloud, land
2	0.841-0.876	0.865	
<b>Reflective solar bands at 500 m spatial resolution</b>			
3	0.459-0.479	0.470	aerosol and cloud optical thickness, cloud phase, cloud effective radius, cloud mask, snow, land
4	0.545-0.565	0.555	
5	1.230-1.250	1.240	
6	1.628-1.652	1.640	
7	2.105-2.155	2.130	
<b>Reflective solar bands at 1000 m spatial resolution</b>			
8	0.405-0.420	0.412	ocean color, chlorophyll, phytoplankton, biogeochemistry
9	0.438-0.448	0.443	
10	0.483-0.493	0.448	
11	0.526-0.536	0.531	
12	0.546-0.556	0.551	sediments, atmosphere
13	0.662-0.672	0.667	
14	0.673-0.683	0.678	fluorescence
15	0.743-0.753	0.748	atmosphere, aerosol
16	0.862-0.877	0.869	
17	0.890-0.920	0.905	atmospheric total precipitable water vapor, clouds
18	0.931-0.941	0.936	
19	0.915-0.965	0.940	
26	1.360-1.390	1.375	cirrus cloud
<b>Thermal emissive bands at 1000 m spatial resolution</b>			
20	3.660-3.840	3.750	cloud & surface, temperature, fire & volcano, sea surface temperature
21	3.929-3.989	3.960	
22	3.929-3.989	3.960	
23	4.020-4.080	4.050	
24	4.433-4.498	4.470	atmospheric temperature profile
25	4.482-4.549	4.520	
27	6.535-6.895	6.720	tropospheric water vapor
28	7.175-7.475	7.330	
29	8.400-8.700	8.550	cloud particle radius
30	9.580-9.880	9.730	total column ozone
31	10.780-11.280	11.030	cloud, surface temperature, fire
32	11.770-12.270	12.020	
33	13.185-13.485	13.340	cloud top height, temperature, pressure, temperature profile
34	13.485-13.785	13.640	
35	13.785-14.085	13.940	
36	14.085-14.385	14.240	

## 4.2 MODIS Products

Since MODIS data may be available at different collection and processing levels, users have to identify the type and the coverage of data that they wish to

use. MODIS data can be grouped in two broad categories; daily scenes and derived products. Products are further separated into Calibration, Atmosphere, Land, Cryosphere, and Ocean groups.

- Calibration product information is available at [107]:  
<http://mcst.gsfc.nasa.gov/>,
- Atmosphere product information is available at [108]:  
<http://modis-atmos.gsfc.nasa.gov/>,
- Land product information is available at [109, 110]:  
<http://modis-land.gsfc.nasa.gov/> and <https://lpdaac.usgs.gov/>,
- Cryosphere product information is available at [111]:  
<http://nsidc.org/data/modis/index.html>,
- Ocean product information is available at [112]:  
<http://oceancolor.gsfc.nasa.gov/>.

As in most satellite data processing systems, MODIS has five different processing levels [113]. Raw satellite feeds are given by level 0 data. Level 1 data is composed of radiometrically calibrated data. Level 2 data is obtained by atmospherical correction of level 1 data and it gives the surface reflectance values. When level 2 data is gridded into a map projection, and temporally composited or averaged as well, level 3 data is obtained. Level 4 data are products that are put through additional processing.

MODIS data up to level 2 are available as an ungridded orbital swath format, where each swath typically divided into small segments or granules for ease of processing. Higher level data are geolocated into a specific map projection, with the geolocated products typically in a set of non-overlapping tiles as illustrated in Figure 4.2.

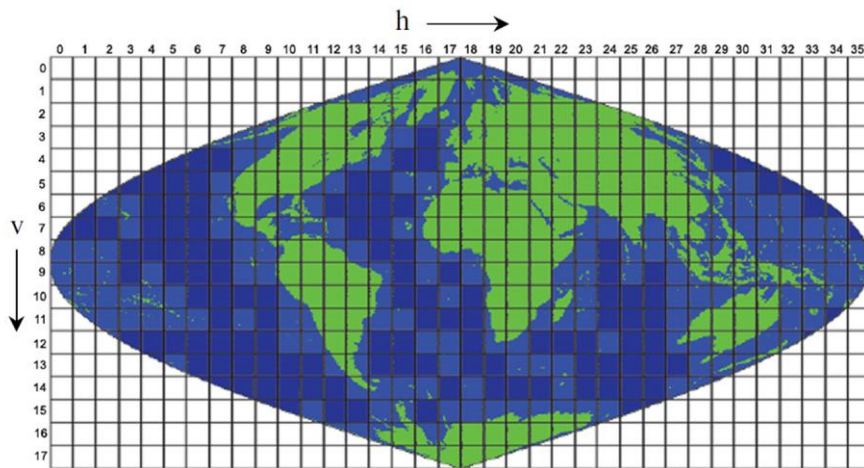


Figure 4.2: The MODIS sinusoidal grid scheme (adapted from [113])

The MODIS data files are in *hierarchical data format* (HDF) [114], which is the standard data storage format selected by *Earth observing system data and information system* (EOSDIS). These data files use the extension of HDF-EOS (i.e., *hierarchical data format – Earth observing system*), which has been defined by EOSDIS in order to establish standards for storing and applying search services to *Earth observing system* (EOS) data [115].

MODIS filenames uses a simple, yet very introductive convention that gives brief and useful information associated with the specific product. In order to give a specific example,

“MOD09GA.A2006013.h18v04.005.2008059145954.hdf” file name indicates:

- MOD09GA – Short name of the product (MOD: Terra / MYD: Aqua),
- .A2006013 – Julian date of data retrieval (A-YYYYDDD),
- .h18v04 – Tile (horizontalXXverticalYY),
- .005 – collection version,

- .2008059145954 – Julian date of production (YYYYDDHMMSS),
- .hdf – Data format (HDF-EOS).

In the following subsections, types of MODIS products used in this study are briefly introduced.

#### **4.2.1 MODIS Level 1B product**

Short introductory of MODIS Level 1B product is given in this subsection based on [106, 116, 117].

Level 1 processing consists of two phases, namely, Level 1A and Level 1B. The former is used to unpack and verify Level 0 MODIS data, whereas the latter involves radiometric calibration of the raw detector output values contained in the Level 1A data product. Level 1B calibrated data is used as input by other applications to generate the images. The Level 1A geolocation and Level 1B software systems are used to process the data collected by Terra and Aqua, and then to generate the associated product files.

A 2-hour Level 0 dataset is organized into a set of granules, each of which contains approximately 5 minutes of MODIS data, by the Level 1A code. Level 1A geolocation algorithm calculates the geodetic position (latitude, longitude, and height), ground to satellite direction and range, and the Sun direction for each MODIS spatial element. It takes the MODIS mirror 1.4771 seconds for one scan; therefore, a 5-minute Level 1B product file typically contains 203 full scans, and occasionally 204 full scans. Level 1B products are summarized in Table 4.2.

Table 4.2: Level 1B product summary

Product name	Product content	Bands	Spatial resolution
MOD/MYD02QKM	Calibrated Earth view data	1,2	250 m
MOD/MYD02HKM	Calibrated Earth view data (also includes bands 1 and 2, which are aggregated to 500 m resolution)	3-7	500 m
MOD/MYD02HKM	Calibrated Earth view data (also includes bands 1-7, which are aggregated to 1000 m resolution)	8-36	1000 m
MOD/MYD02OBC	On board calibrator and engineering data	N/A	N/A

Two calibrated data products, namely the reflectance and Earth-exiting radiance, are supplied in Level 1B for the reflective solar bands. In order to calculate the reflectances and radiances, user should know how to convert the data written in the associated *scientific data set* (SDS) in the Level 1B product.

The raw digital signals measured by the reflective band detectors are indicated as DN and their values are corrected in Level 1B for the following instrumental effects:

- electronic offsets,
- nonlinearities in the analog-to-digital converters,
- angular variations of the scan mirror reflectance,
- variations in gain caused by variations in the instrument and focal plane temperatures, and finally
- out-of-band spectral response in the *short wave infrared* (SWIR) bands 5, 6, 7 and 26.

Then the corrected digital signals,  $dn^a$ , are further adjusted for the effects of certain variations in calibration parameters from detector to detector within each band. The new values of adjusted  $dn^a$  are now  $dn^b$  (each  $dn^b$  is a 32-bit floating point number), and they are scaled to the 16-bit representation in the SDS. Each  $dn^b$  in the whole range of  $[dn_{\min}, dn_{\max}]$  is scaled into a range of  $[0, 32767]$ , which is called *scaled integer* (SI) representation and illustrated in Figure 4.3.

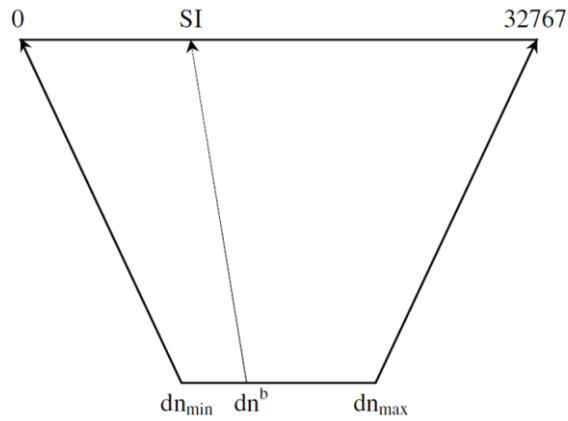


Figure 4.3: Scaled integer representation of  $dn^b$  in Level 1B product (adapted from [106])

In order to calculate the reflectance values,  $\Theta$ , for the reflective solar bands, the following equation is used:

$$\Theta = \delta \cdot (SI - \gamma), \quad (4.1)$$

where  $\delta$  and  $\gamma$  are the corresponding reflectance scale and reflectance offset values of the corresponding reflective solar band, respectively, and they are available as attribute values in the associated band's SDS. One can also calculate the radiances similarly by the help of the following equation:

$$\Psi = \alpha \cdot (SI - \omega). \quad (4.2)$$

Here,  $\alpha$  and  $\omega$  are the related radiance scale and radiance offset values, respectively. The units of the data in the reflective solar bands are given in Table 4.3.

The radiances for thermal emissive bands are represented by 32-bit floating-point format in Level 1B. The DN values measured by the emissive band detectors are corrected for the following effects in Level 1B:

- nonlinearities in the analog-to-digital converters,
- self-emission of the instrument and optics, and
- optical crosstalk in the long-wave bands.

Scaling of radiance values into 16-bit integers is carried out in the same manner as explained for the reflective solar bands. The calculation of the radiance values for the thermal emissive bands is again similar as indicated in Equation (4.2).

Table 4.3: Units for reflective solar bands

Quantity	Unit
dn <sup>b</sup>	unitless
Reflectance	unitless
Reflectance scale	unitless
Reflectance offset	unitless
Radiance	W/m <sup>2</sup> /μm/sr
Radiance scale	W/m <sup>2</sup> /μm/sr
Radiance offset	unitless

#### 4.2.2 MODIS Level 2G surface reflectance product

A basic outline about MODIS Level 2G MOD/MYD09GA surface reflectance product is given based on [113, 118] under this subsection.



MODIS surface reflectance product (i.e., MOD09/MYD09) comprises the seven MODIS Level 1B reflective solar bands 1-7 (Table 4.1). As briefly explained in the Subsection 4.2.1, Level 2 data is obtained via the atmospheric correction of Level 1 data and it gives estimated surface reflectance values for each band in case of no atmosphere between the ground target pixel and the sensor (i.e., no atmospheric scattering or absorption). The Level 2G format consists of gridded Level 2 data and its purpose is to separate geolocating from compositing and averaging. The entire data mapping to a given pixel is conserved as observations at that pixel in the Level 2G format.

MOD/MYD09 data product series includes several subgroups of products:

- Level 2G daily products (MOD/MYD09GHK, MOD/MYD09GQK, and MOD/MYD09GST),
- Level 2G-lite daily products (MOD/MYD09GA and MOD/MYD09GQ),
- Level 3, 8-day composited products (MOD/MYD09A1 and MOD/MYD09Q1),
- and daily Level 3 climate modeling grid products (MOD/MYD09CMG and MOD/MYD09CMA).

Since this subsection aims to introduce only MOD/MYD09GA product, those who need further details regarding the other types of MOD/MYD09 products should refer to [113].

MOD/MYD09GA surface reflectance product has two data groups. The first group provides daily surface reflectance values for bands 1-7 at 500 m spatial resolution. The second group gives observation and geolocation statistics at 1 km spatial resolution. The details of the MOD/MYD09GA product can be found in Table 4.4.

*Quality assurance* (QA) information for all MODIS land products is also available for users in order to understand and make best use of the data. For MOD/MYD09GA product, the associated QA information can be obtained by reading the State\_1km data, which is in 16-bit unsigned integer format (Figure 4.4). In order to interpret the QA values written in the State\_1km product, one should first convert the pixel's decimal value into the 16-bit binary representation. Users should also note that MODIS data QA layers use big-endian reference, which means the "first" bit is the farthest to the right and the "last" bit is the farthest to the left. Required number of zeros should be added to the most significant bit (i.e., to the left) in order to complete the 16-bit string.

To give a specific example, suppose that a pixel's value in the State\_1km QA data of a MOD/MYD09GA product reads 1025. When this value is converted into binary representation, it gives "10000000001". To complete the 16-bit string, five "0"s should be added to the left, which results in "0000010000000001". If the user needs to read the cloud state info for this particular pixel, first two bits (bit 0 and bit 1; i.e., first two bits from the right) should be read, and here those bits are now represented in bold "000001000000000**1**". Even though bits are numbered right to left, bit words should be still read left to right, so "01" would not be reversed to "10". According to the State\_1km QA info, which is given in Table 4.5, "01" means the cloud state of this pixel is defined as "cloudy".

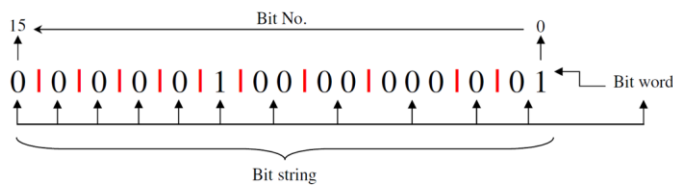


Figure 4.4: 16-bit binary representation of State\_1km QA data

Table 4.4: MOD/MYD09GA science data set

Data Group	SDS	Unit	Data Type	Valid Range	Scale Factor
Spatial resolution: 500 m	num_observations_500m	none	8-bit signed integer	0-127	N/A
	sur_refl_b01: Surface reflectance band 01	reflectance	16-bit signed integer	-100 - 16000	0.0001
	sur_refl_b01: Surface reflectance band 02	reflectance	16-bit signed integer	-100 - 16000	0.0001
	sur_refl_b01: Surface reflectance band 03	reflectance	16-bit signed integer	-100 - 16000	0.0001
	sur_refl_b01: Surface reflectance band 04	reflectance	16-bit signed integer	-100 - 16000	0.0001
	sur_refl_b01: Surface reflectance band 05	reflectance	16-bit signed integer	-100 - 16000	0.0001
	sur_refl_b01: Surface reflectance band 06	reflectance	16-bit signed integer	-100 - 16000	0.0001
	sur_refl_b01: Surface reflectance band 07	reflectance	16-bit signed integer	-100 - 16000	0.0001
	QC_500m: 500m reflectance band quality	bit field	32-bit unsigned integer	N/A	N/A
	Obs_cov_500m: Observation coverage	percent	8-bit signed integer	0-100	0.0001
	iobs_res: Observation number	none	8-bit unsigned integer	0-254	N/A
	q_scan: 250m scan value info	none	8-bit unsigned integer	0-255	N/A
	Spatial resolution: 1 km	num_observations_1km: Number of observations	none	8-bit signed integer	0-127
State_1km: Reflectance quality assurance data		bit field	16-bit unsigned integer	N/A	N/A
Sensor zenith angle		degree	16-bit signed integer	0 - 18000	0.01
Sensor azimuth angle		degree	16-bit signed integer	-18000 - 18000	0.01
Range: pixel to sensor		meter	16-bit unsigned integer	27000 - 65535	0.04
Solar zenith angle		degree	16-bit signed integer	0 - 18000	0.01
Solar azimuth angle		degree	16-bit signed integer	-18000 - 18000	0.01
gflags: Geolocation flags		bit field	8-bit unsigned integer	0 - 248	N/A
orbit_pnt: Orbit pointer		none	8-bit signed integer	0 - 15	N/A

Table 4.5: Description of bits in a MOD/MYDD09GA State\_1km QA data

Bit No.	Quality data parameter	Bit combination	Data state
0-1	cloud state	00	clear
		01	cloudy
		10	mixed
		11	not set, assumed clear
2	cloud shadow	1	yes
		0	no
3-5	land/water flag	000	shallow ocean
		001	land
		010	ocean coastlines and lake shorelines
		011	shallow inland water
		100	ephemeral water
		101	deep inland water
		110	continental/moderate ocean
6-7	aerosol quantity	00	climatology
		01	low
		10	average
		11	high
8-9	cirrus detected	00	none
		01	small
		10	average
		11	high
10	internal cloud algorithm flag	1	cloud
		0	no cloud
11	internal fire algorithm flag	1	fire
		0	no fire
12	MOD35 snow/ice flag	1	yes
		0	no
13	pixel is adjacent to cloud	1	yes
		0	no
14	BRDF correction performed	1	yes
		0	no
15	internal snow mask	1	snow
		0	no snow

### 4.2.3 MODIS Level 3 daily atmospheric product

MODIS Level 3 MOD/MYD08\_D3 daily atmospheric product is briefly introduced in this subsection based on [118-121].

MODIS Level 3 atmospheric products are available on a global  $1^{\circ} \times 1^{\circ}$  latitude-longitude grid (180×360 cells). Level 3 atmospheric products are produced at

3 different temporal resolutions: daily (MOD/MYD08\_D3), eight-day (MOD/MYD08\_E3), and monthly (MOD/MYD08\_M3), and they are all stored in HDF-EOS files. As well as simple statistics (i.e., mean, standard deviation, etc.) for each parameter, marginal density and probability density functions between selected parameters are also available in Level 3 data. Aerosol properties, cloud radiative properties, atmospheric water vapor and temperature are the key geophysical parameters provided in the Level 3 products and they are mainly used in applications related to climate and ecosystem monitoring and modeling, cloud radiative properties, atmospheric properties, and atmospheric correction.

Parameters such as AOD, uWV and uO<sub>3</sub>, which are the atmospheric input parameters for SMAC, or any other RT-based correction algorithm, are stored as a SDS in the associated MOD/MYD08\_D3 data. Level 3 daily product spans nearly 600 statistical datasets, which is difficult to enlist; however, several primary parameters are given in Table 4.6.

Table 4.6: Several primary parameters in MOD/MYD08\_D3 daily product

<b>Name of SDS</b>	<b>Unit</b>	<b>Valid Range</b>	<b>Scale Factor</b>
Optical_Depth_Land_And_Ocean_Mean: Aerosol optical thickness at 0.55 $\mu\text{m}$	none	-100 - 5000	0.001
Corrected_Optical_Depth_Land_Mean: Corrected aerosol optical depth at 0.47, 0.55, 0.66 $\mu\text{m}$	none	-100 - 5000	0.001
Water_Vapor_Near_Infrared_Clear_Mean: Water vapor near infrared –clear column (bright land and ocean sunlight only): Mean	cm	0 - 20000	0.001
Water_Vapor_Near_Infrared_Cloud_Mean: Water vapor near infrared –cloudy column: Mean	cm	0 - 20000	0.001
Atmospheric_Water_Vapor_Mean: Precipitable Water Vapor (IR retrieval) Total Column: Mean	cm	0 - 20000	0.001
Total_Ozone_Mean: Total Ozone Burden	Dobson Units	0 - 5000	0.1

#### 4.2.4 MODIS snow algorithm and Level 3 daily snow product

Brief description of MODIS snow algorithm and Level 3 MOD/MYD10A1 daily snow product is given based on [122-124].

Since its launch in 1999, data collected by MODIS on the Terra satellite have been extensively used for mapping global snow cover through the SNOWMAP algorithm, where each MODIS 500-m pixel is classified as snow or non-snow [125]. The snow mapping algorithm is mainly based on the *normalized difference snow index* (NDSI), in which the MODIS bands 4 (centered at 0.555  $\mu\text{m}$ ) and 6 (centered at 1.640  $\mu\text{m}$ ) are used, along with a series of threshold tests and the MODIS cloud mask [122, 125].

Due to the fact that band 6 of MODIS instrument on board to Aqua has 15 nonfunctional detectors (out of 20), NDSI calculations based on bands 4 and 6 measurements cannot be used for this instrument [125].

Main logic behind snow detection is the fact that the reflectance of snow is high in the visible and low in the near infrared region. MODIS uses a fully automated snow mapping algorithm, in which bands 4 (0.545-0.565  $\mu\text{m}$ ) and 6 (1.628-1.652  $\mu\text{m}$ ) are used to calculate the NDSI value as given in Equation (4.3):

$$\text{NDSI} = \frac{\text{band}_4 - \text{band}_6}{\text{band}_4 + \text{band}_6}. \quad (4.3)$$

If the NDSI of a pixel in a non-densely forested area is  $\geq 0.4$  and its reflectance in band 2 (0.841-0.876  $\mu\text{m}$ ) is  $> 11\%$ , it is labeled as snow. However, very low reflectance values make the denominator of the NDSI considerably small, and pixels with very dark targets, like black spruce forests, can be erroneously classified as snow. Therefore, when the reflectance in band 4 (0.545-0.565  $\mu\text{m}$ )

is  $< 10\%$ , then the pixel is not labeled as snow even if the other conditions are met.

When a forest stand is covered by snow, its spectral response changes in such a way that reflectance in the visible generally increases with respect to near infra-red reflectance, which results in a decrease in the *normalized difference vegetation index* (NDVI). In order to improve the snow mapping in dense forests, NDSI is used together with NDVI. MODIS bands 1 (0.620-0.670  $\mu\text{m}$ ) and 2 (0.841-0.876  $\mu\text{m}$ ) are employed in the calculation of NDVI, and if its value  $\approx 0.1$ , the pixel can be labeled as snow even if the value of NDSI is  $< 0.4$ .

Additionally, MODIS infra-red bands 31 (10.780-11.280  $\mu\text{m}$ ) and 32 (11.770-12-270  $\mu\text{m}$ ) are employed via a split-window technique to estimate the ground temperature in order to eliminate the spurious snow and to increase the accuracy of snow mapping. By using this thermal mask, pixels with temperature  $> 283 \text{ K}$  are not labeled as snow.

Since the functionality of MODIS band 6 on Aqua is roughly 30%, snow mapping in the swath level algorithm had to be shifted to band 7 (2.105-2.155  $\mu\text{m}$ ) for this sensor, and the details about use of different bands and its effect on snow mapping are not in the scope of this subsection, but they can be found in [125].

There are seven MODIS snow products produced at Level 2 or Level 3, and they are at different temporal and spatial resolutions (Table 4.7). The file format for snow products is HDF-EOS. Each daily snow product, MOD/MYD10A1, is a  $10^\circ \times 10^\circ$  tile (1200 $\times$ 1200km) with a sinusoidal projection. Four SDSs are available in a MOD/MYD10A1: snow cover map, fractional snow cover, snow albedo, and, finally QA data. In order to select an observation for the day, the scoring algorithm given in Equation (4.4) is employed:

$$\text{score} = 0.5SE + 0.3ND + 0.2OC, \quad (4.4)$$

where  $SE$  indicates the solar elevation,  $ND$  is the distance from nadir, and  $OC$  is the observation coverage. In order to generate the snow cover map, observation closest to local noon time (i.e., highest solar elevation angle) nearest to nadir with the greatest coverage is selected by the scoring algorithm, and then classified as snow, snow-covered water bodies, land, water, cloud or other condition.

Table 4.7: Seven MODIS snow data products

Name of the product	Level	Data dimension	Spatial resolution	Temporal resolution
MOD/MYD10_L2	L2	1354 × 2000 km	500 m	swath
MOD/MYD10L2G	L2G	1200 × 1200 km	500 m	day of multiple coincident swaths
MOD/MYD10A1	L3	1200 × 1200 km	500 m	day
MOD/MYD10A2	L3	1200 × 1200 km	500 m	8 days
MOD/MYD10C1	L3	360° × 180° (global)	0.05° × 0.05°	day
MOD/MYD10C2	L3	360° × 180° (global)	0.05° × 0.05°	8 days
MOD/MYD10C	L3	360° × 180° (global)	0.05° × 0.05°	month

Observations from the fractional snow cover of L2G product are used to determine the daily fractional snow cover, again by using the same scoring algorithm in Equation (4.4). Fractional snow is given within 0-100% range by the fractional snow cover map. This map includes inland water bodies, and non-snow pixels are classified as water, cloud or other condition.



By using the MODIS surface reflectance product, the snow albedo is determined for the visible and near infra-red bands. The resultant map shows the snow albedo within the range of 0-100, and non-snow features are labeled with different values. A summary of attributes in a MOD/MYD10A1 daily snow tile is given in Table 4.8.

Table 4.8: MOD/MYD10A1 daily snow tile attributes

<b>Coordinate system</b>	Cartesian
<b>Valid range</b>	0-254
<b>Fill value</b>	255 (used to fill gaps in the swath)
<b>Key to data values</b>	0: missing data
	1: no decision
	11: night
	25: no snow
	37: lake
	39: ocean
	50: cloud
	100: lake ice
	200: snow
	254: detector saturated
	255:fill



## CHAPTER 5

# ATMOSPHERIC CORRECTION ON MODIS IMAGES BY MARS AND CMARS

### 5.1 Image Set and Model Training

As discussed in Subsection 4.2.2, MOD09GA surface reflectance product comprises the atmospherically corrected surface reflectance values for the first seven bands, and the vector version of 6S (6SV) is the basic correction algorithm behind this product. The algorithm takes the TOA reflectance values from MODIS Level 1B (Subsection 4.2.1) product, and makes a series of corrections for the absorption and scattering by atmospheric gases and molecules. It also takes into account the coupling between atmospheric and surface bi-directional reflectance function and adjacency effect.

Another product family of MODIS is the Level 3 MOD08\_D3 daily atmospheric product (Subsection 4.2.3), from which AOD, uWV and uO<sub>3</sub> can be retrieved in order to be employed as input in SMAC. All these features make MODIS an attractive and convenient platform for this study.

24 MODIS images taken by Terra satellite are selected as data set. Half of the images are over Alps region (Figure 5.1) and the other half is over Turkey (Figure 5.2). Detailed information about the image data set can be found in Table 5.1. At this point, we emphasize that all the calculations and operations are performed on the 4<sup>th</sup> reflective solar band (0.545 – 0.565  $\mu\text{m}$ ).

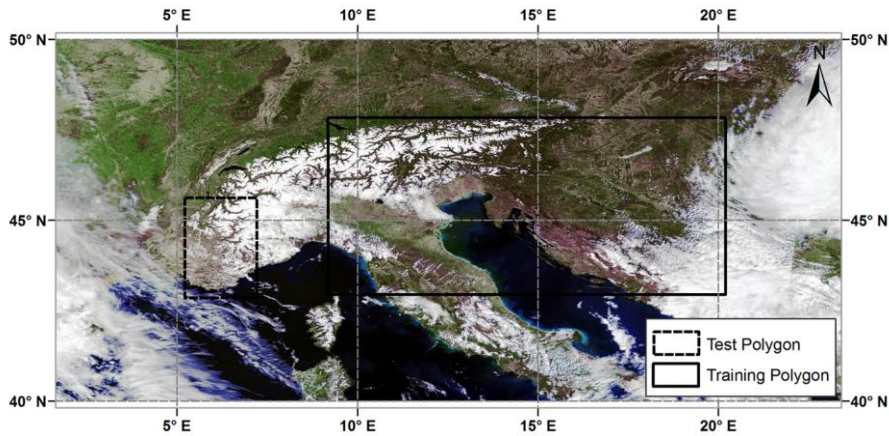


Figure 5.1: MOD09GA RGB color composite image of Alps (R: 1<sup>st</sup>, G: 4<sup>th</sup>, B: 3<sup>rd</sup> band, A11: 06.11.2002)

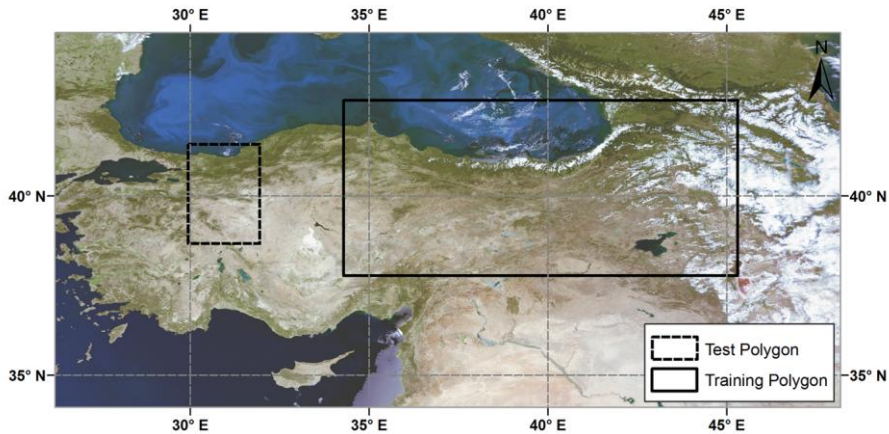


Figure 5.2: MOD09GA RGB color composite image of Turkey (R: 1<sup>st</sup>, G: 4<sup>th</sup>, B: 3<sup>rd</sup> band, T7: 16.07.2012)

In the model training phase, for each region (i.e., Alps and Turkey), a training polygon with fixed dimension (i.e., rectangular in shape with an area of 483 000 km<sup>2</sup>), which includes all basic land-cover types, is drawn on each image. In order to reveal the effect of the training sample size on the final models, two different random point distributions, one with size of 5000, and the other with 10 000, are generated in each training polygon of both regions. Next, for each point, TOA reflectance values and atmospherically corrected surface reflectance values are extracted from MOD02HKM and MOD09GA

products, respectively, together with the corresponding geographic longitudes and latitudes.

Atmospheric parameters (i.e., AOD, uWV and uO<sub>3</sub>) required to perform atmospheric correction through SMAC algorithm on the test areas are obtained from MOD08\_D3 daily atmospheric product. In order to prepare the input data for SMAC, a code written in MATLAB<sup>®</sup> software [126] is used, and it is given in Appendix E.2.

Table 5.1: TOA reflectance statistics for Alps (A) and Turkey (T)

Data Set	Date	Training Set (Area: 483 000 km <sup>2</sup> )				Testing Set (Area: 53 700 km <sup>2</sup> )			
		Min	Max	Mean	Std	Min	Max	Mean	Std
A1	13.01.2006	0.023006	0.515043	0.111236	0.066780	0.021848	0.540773	0.065938	0.062607
A2	03.02.2005	0.021381	0.612208	0.187324	0.084896	0.027219	0.570827	0.052124	0.038957
A3	10.03.2002	0.036336	0.634253	0.123546	0.098884	0.052459	0.591130	0.093768	0.061479
A4	22.04.2005	0.037909	0.815975	0.117174	0.118288	0.045157	0.045157	0.132936	0.154975
A5	03.05.2003	0.043902	0.831254	0.120384	0.080038	0.042749	0.893523	0.157483	0.158453
A6	04.06.2007	0.051815	0.984628	0.409352	0.252784	0.039096	1.039224	0.231790	0.193401
A7	01.07.2007	0.041622	0.617155	0.103806	0.064886	0.051844	1.010623	0.176865	0.142061
A8	06.08.2008	0.052038	1.018599	0.134000	0.123403	0.042518	0.904316	0.135226	0.100786
A9	03.09.2001	0.044417	0.730243	0.122082	0.094248	0.031365	0.680813	0.107640	0.108413
A10	11.10.2001	0.024596	0.706770	0.082712	0.059985	0.020154	0.623411	0.059328	0.039499
A11	06.11.2002	0.023207	0.499682	0.115683	0.112962	0.019956	0.730805	0.087927	0.099035
A12	06.12.2003	0.019304	0.497581	0.112805	0.082715	0.020898	0.559915	0.074968	0.08527
T1	22.01.2007	0.022460	0.438156	0.072725	0.072429	0.018987	0.531954	0.063973	0.061618
T2	08.02.2004	0.030369	0.494759	0.072528	0.057265	0.025595	0.571422	0.076824	0.080169
T3	22.03.2009	0.041930	0.767917	0.174026	0.149042	0.038070	0.834106	0.171948	0.146781
T4	13.04.2008	0.074651	0.735924	0.161762	0.105837	0.046226	0.833497	0.152259	0.119837
T5	16.05.2011	0.046915	0.763972	0.100172	0.076170	0.038706	0.747831	0.107228	0.077950
T6	17.06.2003	0.056582	0.981983	0.147624	0.110149	0.077487	0.923808	0.211551	0.138608
T7	16.07.2012	0.044871	0.627486	0.097143	0.029799	0.055703	0.398507	0.100335	0.030883
T8	01.08.2005	0.036527	0.909753	0.229990	0.175922	0.036824	0.891506	0.135249	0.094953
T9	26.09.2009	0.040853	0.783761	0.104974	0.093539	0.033658	0.743776	0.098932	0.056643
T10	18.10.2002	0.035435	0.173049	0.068862	0.021872	0.032394	0.123435	0.061462	0.015683
T11	02.11.2010	0.033515	0.279079	0.057913	0.017563	0.024014	0.507661	0.056879	0.033940
T12	18.12.2006	0.024390	0.443912	0.060216	0.038547	0.025587	0.446532	0.073266	0.034281

As seen in Table 5.1, each month appears once in the dataset. So, month value of the image, from which the corresponding model training parameters are retrieved, is also written in that point data file. In the next step, the training data from each of the training polygons for two different sample sizes are separately combined in the associated text files, which results in training data with 60 000 and 120 000 points on aggregate for each region.

In MARS model building,  $\mathbf{x} = (\lambda, \phi, \kappa, \zeta)^T$  is the vector of predictor variables, where  $\lambda$  and  $\phi$  represents the geographic latitude and longitude, respectively,  $\kappa$  is the TOA reflectance value, and  $\zeta$  is the month indicator ( $\zeta = 1, 2, \dots, 12$ ). The observation equation for MARS is given by  $y_i = \text{SRef}(\mathbf{x}_i) + \varepsilon_i$  ( $i = 1, 2, \dots, N$ ), where  $y_i$  is the atmospherically corrected surface reflectance obtained from MODIS Level 2G product at  $\mathbf{x}_i = (\lambda_i, \phi_i, \kappa_i, \zeta_i)^T$  with a measurement error  $\varepsilon_i$ . SRef is the function that gives the surface reflectance predicted by MARS, and  $N$  is the total number of observations.

Since MARS algorithm allows users to define model-building dimensions such as maximal number of terms, the degree of interaction among predictor variables, different settings are applied for two different training sample sizes, and the models that give the optimal GCV and *multiple coefficient of determination* ( $R^2$ ) [127] values are chosen. All settings in the model training are given in Table 5.2. Salford Systems MARS<sup>®</sup> Ver. 3 [128] is used for MARS model building.

Table 5.2: MARS model training settings

	Sample size: 60 000		Sample size: 120 000	
	Alps	Turkey	Alps	Turkey
Number of basis functions	40	40	40	40
Degree of interaction	3	2	3	2
GCV	0.00575	0.00274	0.00562	0.0271
$R^2$	0.9175	0.8938	0.9189	0.8937
Adjusted $R^2$	0.9174	0.8937	0.9188	0.8936

The obtained MARS models for two different training sample sizes are applied on the related test areas in both regions. Then, the results are compared against the MOD09GA product in terms of *mean absolute error* (MAE) and  $R^2$  values (Table 5.3). Details on the calculations of MAE and  $R^2$  are given in Appendix D. The atmospheric correction algorithm used in MOD09GA is based on 6SV

method as mentioned earlier. 6SV has been seriously cross-checked with other complex correction algorithms and has given highly accurate results under various conditions; therefore, it is preferred as a reference in the comparison of the results.

Table 5.3: Comparison of MARS models obtained with two different training sample sizes

Data Set	Sample size: 60 000		Sample size: 120 000	
	MAE	R <sup>2</sup>	MAE	R <sup>2</sup>
A1	0.0524	0.7619	0.0680	0.7300
A2	0.0221	0.8247	0.0181	0.8311
A3	0.0204	0.9138	0.0305	0.8888
A4	0.0454	0.8341	0.0397	0.8394
A5	0.0381	0.8721	0.0419	0.8723
A6	0.0449	0.8953	0.0476	0.8935
A7	0.0461	0.7990	0.0513	0.7809
A8	0.0327	0.7878	0.0351	0.7700
A9	0.0358	0.8628	0.0317	0.8713
A10	0.0189	0.7132	0.0194	0.7051
A11	0.0462	0.8324	0.0464	0.8333
A12	0.0603	0.7988	0.0622	0.8007
T1	0.0337	0.8133	0.0318	0.8169
T2	0.0332	0.8779	0.0298	0.8895
T3	0.0323	0.8235	0.0381	0.7350
T4	0.0318	0.8810	0.0439	0.8539
T5	0.0271	0.6758	0.0251	0.6997
T6	0.0463	0.8839	0.0415	0.8987
T7	0.0085	0.9184	0.0099	0.9051
T8	0.0354	0.6912	0.0358	0.6964
T9	0.0240	0.6861	0.0246	0.6836
T10	0.0114	0.7879	0.0131	0.7506
T11	0.0169	0.7913	0.0178	0.7784
T12	0.0237	0.8995	0.0171	0.9274

In order to see if the MAEs and R<sup>2</sup>s differ with respect to the sample size, an appropriate statistical test should be employed. However, prior to any statistical test, it is necessary to check whether the accuracy measures follow the normal distribution assumption. For this purpose, Shapiro-Wilk test is used to test the normality of the MAEs and R<sup>2</sup>s for both sample sizes. The results of the test are given in Table 5.4. In all tables that represent the result of statistical tests, \* indicates statistically significant result at significance level  $\alpha=0.05$ .

Table 5.4: Shapiro-Wilk test of MAEs and R<sup>2</sup>s for two different MARS models obtained via two training samples with different sizes

Shapiro-Wilk ( $\alpha=0.05$ )				
Sample size	60 000		120 000	
	Statistic	Sig.	Statistic	Sig.
<b>Alps</b>				
MAE	0.939	0.479	0.971	0.918
R <sup>2</sup>	0.982	0.990	0.937	0.455
<b>Turkey</b>				
MAE	0.995	0.713	0.957	0.743
R <sup>2</sup>	0.888	0.111	0.914	0.239

Results of the test implies that the accuracy measures of each region calculated for two different sample sizes are normally distributed (Sig.>0.05). Therefore, parametric paired-samples t test is employed for both regions to see if there is a statistically significant difference in MAEs and R<sup>2</sup>s with respect to the training sample size. The results of the test are presented in Table 5.5.

Table 5.5: Paired-Samples t test of MAEs and R<sup>2</sup>s with respect to sample size

Paired-Samples t test ( $\alpha=0.05$ )		
Compared Pair (60 000 vs. 120 000)	t score	Sig. (2-tailed)
Alps MAE	-1.368	0.198
Alps R <sup>2</sup>	1.713	0.115
Turkey MAE	-0.244	0.811
Turkey R <sup>2</sup>	-0.118	0.908

As the results of the paired-samples t test indicate, statistically, there is no significant difference in the MAE and R<sup>2</sup> values for the two sample sizes (Sig.>0.05). Thus, sample size of 5000 in each training polygon (i.e., 60 000 on aggregate for each region) is used with the associated model building parameters given in Table 5.2.

All statistical analysis mentioned above are carried out by using SPSS statistical software [129], and the details of the analysis can be found in [130].



In CMARS, basis functions, which are considered for the formulation of PRSS in the form of a CQP problem, are created by using Salford MARS<sup>®</sup>. Next, CMARS model is obtained by running a MATLAB code. Then, the CQP problem is solved by using MOSEK<sup>™</sup> software [131]. The details of obtained MARS and CMARS models are given in Appendix C.

## **5.2 Testing of MARS and CMARS Models**

After generating MARS and CMARS models for each region through the training phase, the obtained models are applied on the predefined test areas on each image in the related data set. In addition to MARS and CMARS models, SMAC algorithm is also employed on the test areas by using a MATLAB code given in Appendix E.2. Then, the surface reflectance values produced by MARS, CMARS and SMAC algorithms for each test area are compared with the ones given by the associated MOD09GA products in terms of MAE and  $R^2$ . The results are presented in Table 5.6.

The images of the test areas obtained from MARS, CMARS, SMAC algorithms and from the MOD09GA data for data sets A10 and T11 are given in Figures 5.3 and 5.4, respectively.

## **5.3 Results and Discussion**

When the surface reflectance values obtained by applying MARS, CMARS and SMAC algorithms are compared against the reflectance values of MOD09GA product in terms of MAE and  $R^2$  (Table 5.6), it seems that MARS and CMARS outperforms SMAC method. However, it is necessary to statistically check the significance of the results.

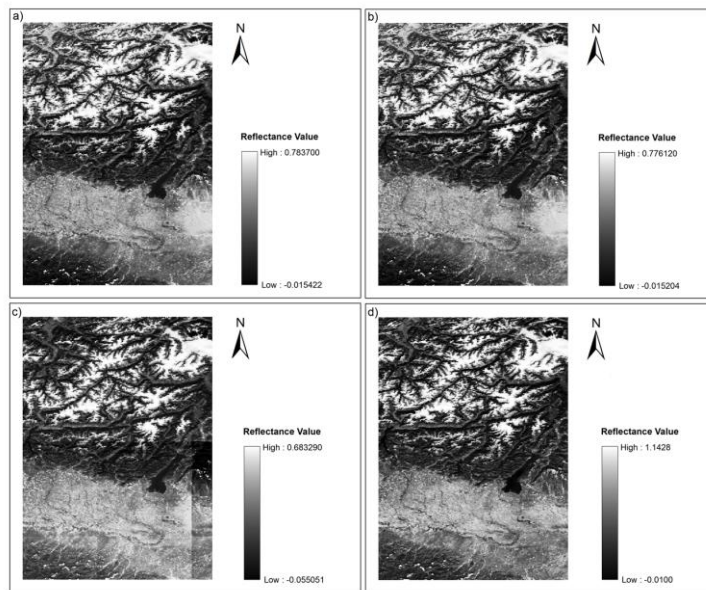


Figure 5.3: Images given by a) MARS, b) CMARS, c) SMAC, and d) MOD09GA for A10 data set

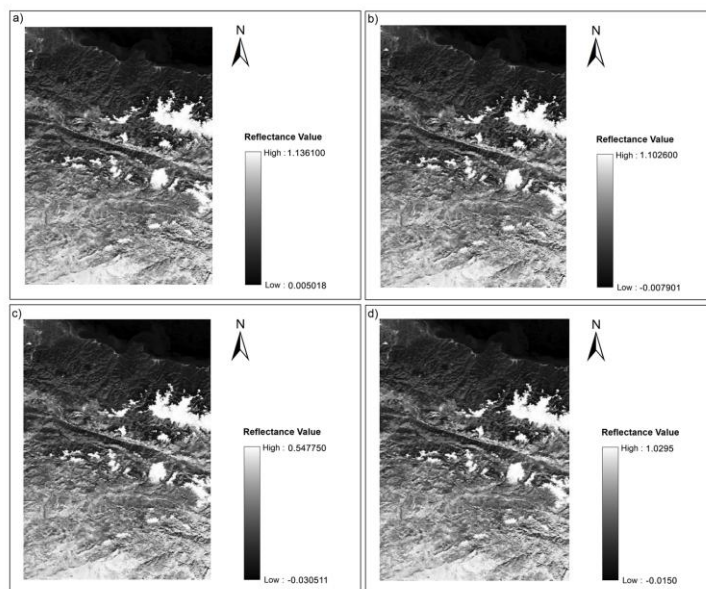


Figure 5.4: Images given by a) MARS, b) CMARS, c) SMAC, and d) MOD09GA for T11 data set

Table 5.6: Comparison of results for Alps (A) and Turkey (T)

Data Set	MAE			R <sup>2</sup>		
	SMAC	MARS	CMARS	SMAC	MARS	CMARS
A1	0.1291	0.0524	0.0523	0.1672	0.7619	0.7622
A2	0.0898	0.0221	0.0219	0.0019	0.8247	0.8253
A3	0.1018	0.0204	0.0201	0.2863	0.9138	0.9141
A4	0.0606	0.0454	0.0454	0.7934	0.8341	0.8345
A5	0.0469	0.0381	0.0381	0.8465	0.8721	0.8722
A6	0.0439	0.0449	0.0449	0.8932	0.8953	0.8953
A7	0.0484	0.0461	0.0461	0.7688	0.7990	0.7990
A8	0.0399	0.0327	0.0327	0.7533	0.7878	0.7878
A9	0.0416	0.0358	0.0358	0.8255	0.8628	0.8628
A10	0.0564	0.0189	0.0188	0.1717	0.7132	0.7134
A11	0.1267	0.0462	0.0462	0.4123	0.8324	0.8324
A12	0.1871	0.0603	0.0603	0.1745	0.7988	0.7988
T1	0.1017	0.0337	0.0336	0.2680	0.8133	0.8134
T2	0.0906	0.0332	0.0330	0.5171	0.8779	0.8786
T3	0.0686	0.0323	0.0320	0.7253	0.8235	0.8256
T4	0.0497	0.0318	0.0316	0.8262	0.8810	0.8812
T5	0.0294	0.0271	0.0266	0.6667	0.6758	0.6791
T6	0.0474	0.0463	0.0460	0.8801	0.8839	0.8871
T7	0.0205	0.0085	0.0075	0.6974	0.9184	0.9294
T8	0.0387	0.0354	0.0353	0.6754	0.6912	0.6949
T9	0.0544	0.0240	0.0234	0.4248	0.6861	0.6901
T10	0.0518	0.0114	0.0100	0.3420	0.7879	0.8048
T11	0.0675	0.0169	0.0151	0.2760	0.7913	0.8132
T12	0.1407	0.0237	0.0228	0.2240	0.8995	0.9046

Thus, *repeated measures analysis of variance* (RANOVA) [130] can be employed to make an analysis for repeated measures on the same individuals (i.e., if there exists a difference between SMAC, MARS and CMARS methods on the same image in terms of accuracy measures used). But prior to RANOVA, the normality of MAEs and R<sup>2</sup>s must be checked for each group. Therefore, Shapiro-Wilk test is employed once again, and the results are given in Table 5.7.

Table 5.7: Shapiro-Wilk test of SMAC, MARS and CMARS results

Shapiro-Wilk Test ( $\alpha=0.05$ )					
Accuracy measure	Method	Alps		Turkey	
		Statistic	Sig.	Statistic	Sig.
MAE	SMAC	0.836	0.025*	0.918	0.273
	MARS	0.939	0.479	0.955	0.713
	CMARS	0.937	0.457	0.951	0.651
R <sup>2</sup>	SMAC	0.848	0.034*	0.918	0.270
	MARS	0.982	0.990	0.888	0.111
	CMARS	0.982	0.991	0.887	0.107

According to the test results, MAE and  $R^2$  values of SMAC for Alps do not obey normal distribution assumption (Sig.<0.05). Therefore, Box-Cox transformation [132] is applied to these values by using Minitab® [133]. The graphs of the Box-Cox transformations for MAE and  $R^2$  values of SMAC are given in Figures 5.5 and 5.6, respectively.

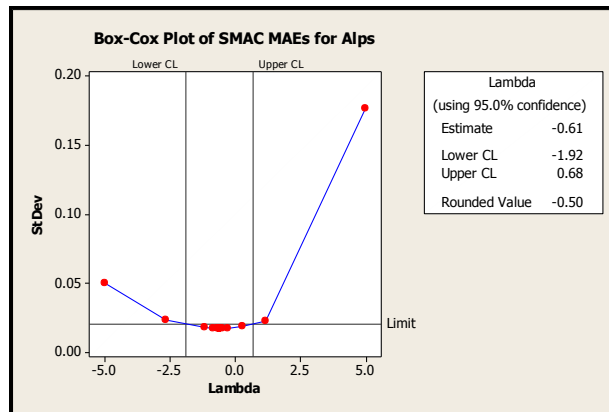


Figure 5.5: Box-Cox transformation plot for SMAC MAE values of Alps

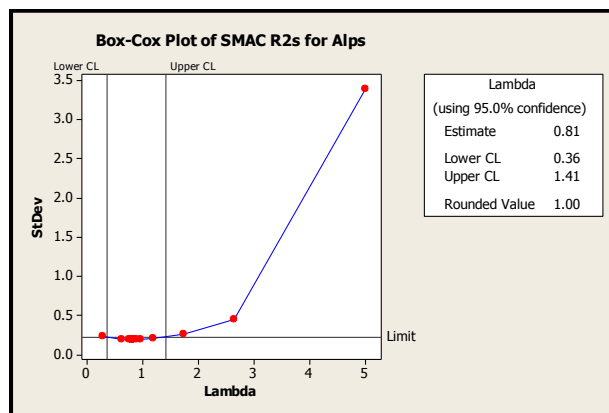


Figure 5.6: Box-Cox transformation plot for SMAC  $R^2$  values of Alps

As indicated in the Box-Cox plot of  $R^2$  values for SMAC in Alps (Figure 5.6), the associated Lambda value is equal to 1.00; therefore, it is not necessary to apply a transformation on  $R^2$  values. For MAE values of SMAC in Alps, lambda is equal to -0.50, so MAE values are transformed accordingly. Then,

RANOVA test is carried out. The results of the RANOVA test are presented in Table 5.8.

Table 5.8: RANOVA test results

	<b>Mauchly's Test of Sphericity (<math>\alpha=0.05</math>)</b>		<b>RANOVA (<math>\alpha=0.05</math>)</b>	
	<b>Chi-Square</b>	<b>Sig.</b>	<b>F</b>	<b>Sig.</b>
<b>MAE Alps</b>	90.596	0.000*	14.087	0.000*
<b>R<sup>2</sup> Alps</b>	139.638	0.000*	12.204	0.000*
<b>MAE Turkey</b>	72.058	0.000*	14.020	0.000*
<b>R<sup>2</sup> Turkey</b>	61.493	0.000*	15.333	0.000*

Even though the results of RANOVA indicate significant difference between group means (Sig.<0.05), sphericity condition is not satisfied (Sig.<0.05), which is a necessary condition for RANOVA. Therefore, the results of Huynh-Feldt tests are also checked (Table 5.9).

Table 5.9: Results of Huynh-Feldt tests

	<b>Huynh-Feldt Test (<math>\alpha=0.05</math>)</b>	
	<b>F</b>	<b>Sig.</b>
<b>MAE Alps</b>	14.087	0.003*
<b>R<sup>2</sup> Alps</b>	12.204	0.005*
<b>MAE Turkey</b>	14.020	0.003*
<b>R<sup>2</sup> Turkey</b>	15.333	0.002*

The results of the Huynh-Feldt tests indicate statistically significant difference between the group means (Sig.<0.05). However, these tests only tell there is at least one difference between the group means.

Thus, pairwise comparisons in SPSS are carried out, and the results are presented in Table 5.10. The reader can find the details of the above mentioned statistical tests in [130].

As observed from the pairwise comparisons, there is a statistically significant difference between SMAC and MARS (Sig.<0.05), and also between SMAC

and CMARS (Sig.<0.05) for all data sets. Therefore, we can conclude that MARS and CMARS methods have better performances as compared to the traditional RT-based method SMAC.

Table 5.10: Pairwise comparison of the results

Pairwise Comparison ( $\alpha=0.05$ )				
Dependent Variable	(I) METHOD	(J) METHOD	Mean Difference (I-J)	Sig.
MAE Alps	SMAC	MARS	-1.468	0.003*
		CMARS	-1.476	0.003*
	MARS	SMAC	1.468	0.003*
		CMARS	-0.009	0.095
	CMARS	SMAC	-1.476	0.003*
		MARS	0.009	0.095
R <sup>2</sup> Alps	SMAC	MARS	-0.317	0.005*
		CMARS	-0.317	0.005*
		MARS	0.317	0.005*
	MARS	SMAC	0.317	0.005*
		CMARS	0.000	0.020*
	CMARS	SMAC	0.317	0.005*
MARS		0.000	0.020*	
MAE Turkey	SMAC	MARS	0.036	0.003*
		CMARS	0.037	0.003*
	MARS	SMAC	-0.036	0.003*
		CMARS	0.001	0.003*
	CMARS	SMAC	-0.037	0.003*
		MARS	-0.001	0.003*
R <sup>2</sup> Turkey	SMAC	MARS	-0.267	0.002*
		CMARS	-0.273	0.002*
		MARS	0.267	0.002*
	MARS	SMAC	0.267	0.002*
		CMARS	-0.006	0.012*
	CMARS	SMAC	0.273	0.002*
MARS		0.006	0.012*	

However, an interesting finding of our study should be further discussed here. As seen from Table 5.6, MAE and R<sup>2</sup> values of both regions converge starting roughly from the mid-spring until the beginning of fall. Therefore, monthly variations in the atmospheric parameters (i.e., AOD, uWV and uO<sub>3</sub>) should be checked in order to see if they contribute to this effect. The mean relative reflectance difference between SMAC and MARS for each month is calculated and plotted against the mean value of each atmospheric parameter of that month for Alps (Figure 5.7) and Turkey (Figure 5.8). Difference between SMAC and CMARS is not considered at this stage due to the close similarity

between MARS and CMARS results. The readers should note that in Figures 5.7 and 5.8,  $uO_3$  and  $uWV$  values on the secondary y-axis are multiplied by 0.001 and 0.1, respectively.

In both graphs  $uO_3$  follows a constant trend around 310 Dobson Unit for all months. So, it is hardly possible to make any comment about the impact of  $uO_3$  on the monthly variations in the performances of SMAC and MARS. However, significant inverse relationship is observed between the relative differences in reflectance values and both AOD and  $uWV$ . Differences in reflectance values increases with decreasing AOD and  $uWV$ , which mainly corresponds to the fall and winter seasons. For AOD around 0.20-0.50 and  $uWV$  around 2.0-4.0 cm, drastic decrease in relative differences between the reflectance values of SMAC and MARS are observed. At first glance, it may be inferred that both AOD and  $uWV$  have large effect on the performance of SMAC for the 4<sup>th</sup> reflective solar band of MODIS (0.545-0.565  $\mu\text{m}$ ).

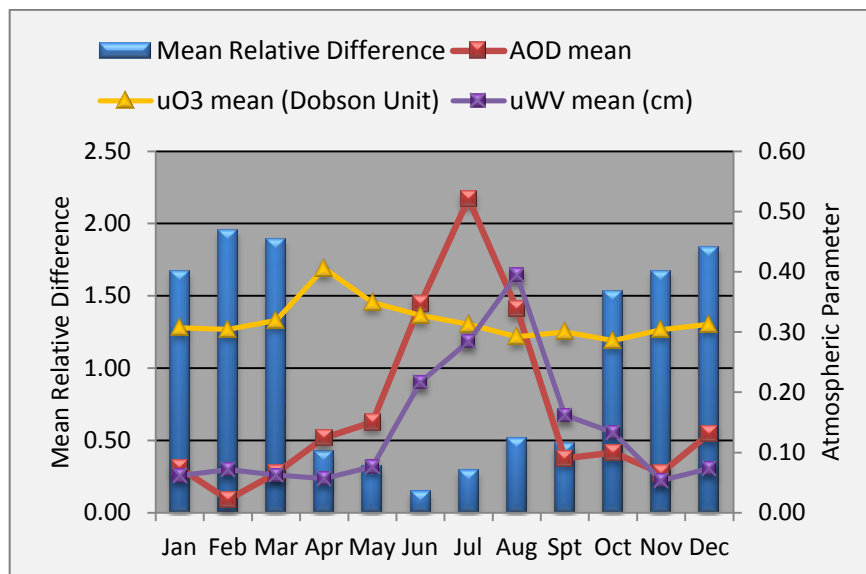


Figure 5.7: Change in the mean relative difference between SMAC and MARS reflectance values with respect to atmospheric parameters for Alps

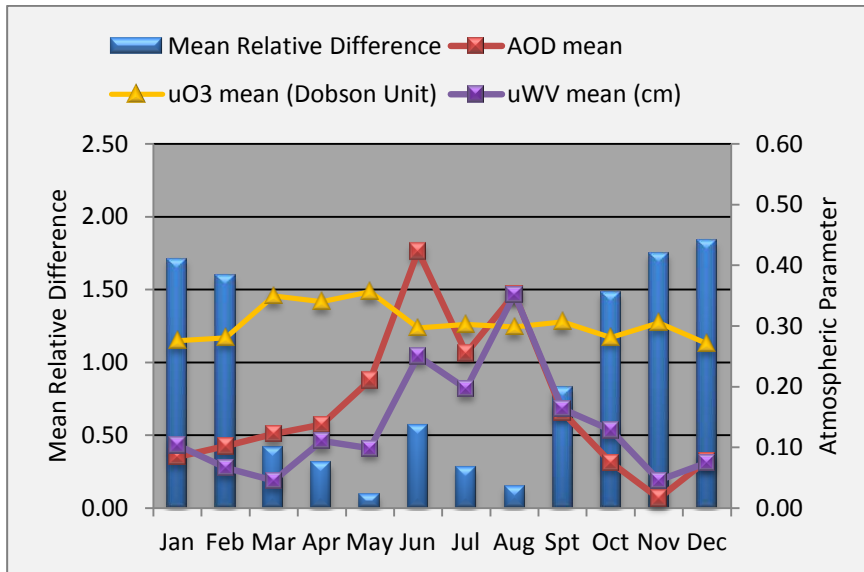


Figure 5.8: Change in the mean relative difference between SMAC and MARS reflectance values with respect to atmospheric parameters for Turkey

However, absorption effect by water vapor at  $0.55 \mu\text{m}$  for uWV values ranging from 0.01 to 5.0 cm is negligible according to [134]. Therefore, we can conclude that in case of high AOD, SMAC and MARS seem to merge more closely.

In [63], performances of SMAC and 6S were compared, and a similar relation between AOD and the two models were observed. Our findings on SMAC and MARS also match up with the results found in [63]. Since the MARS model is trained with MOD09GA data, which is based on 6S, it eventually “mimics” the behavior of 6S.

The above mentioned characteristics of SMAC and MARS (and CMARS) are more obvious when the image histograms given in Figures 5.9 - 5.12 are analyzed.



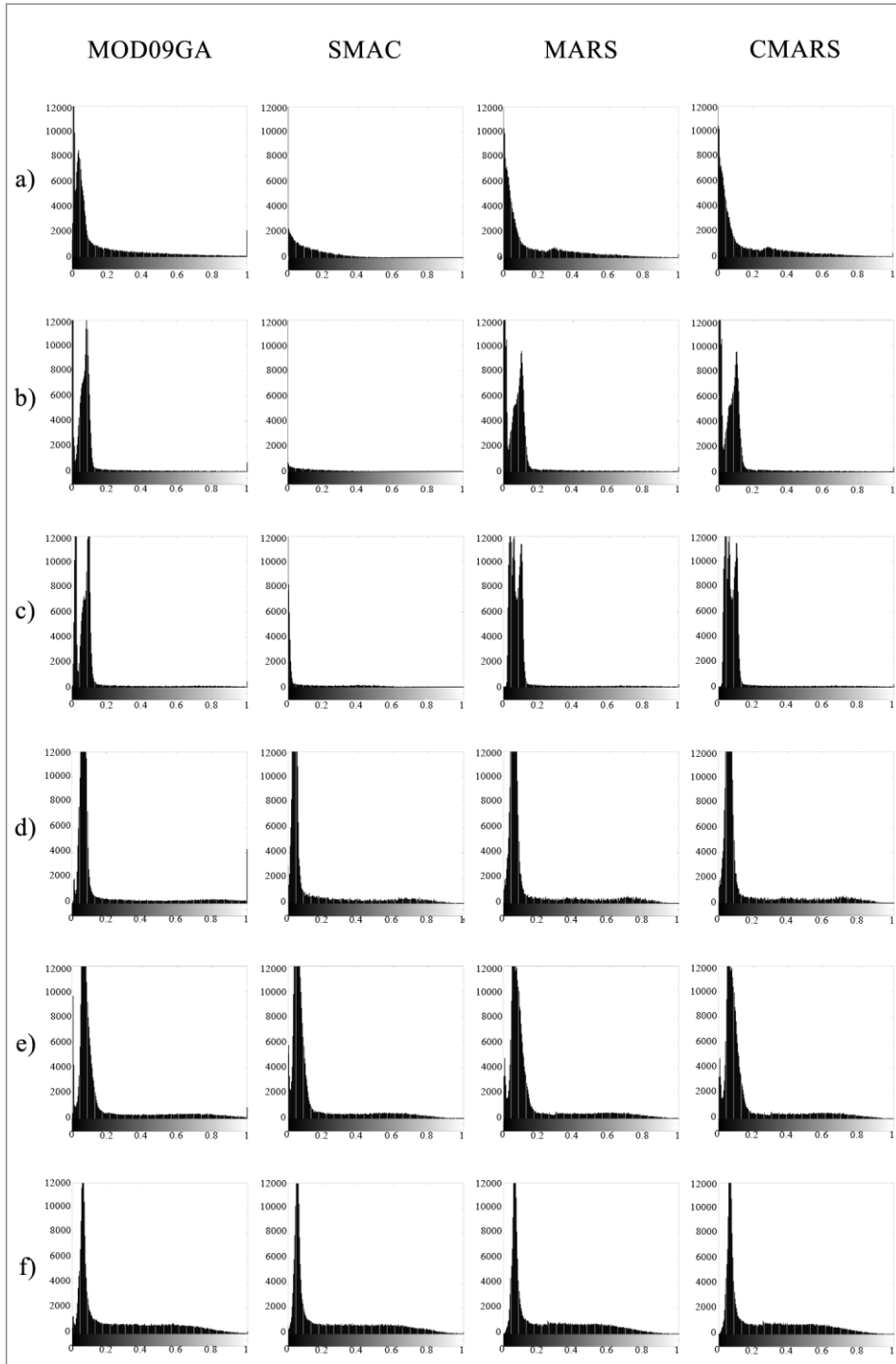


Figure 5.9: Image histograms for Alps data set: a) A1, b) A2, c) A3, d) A4, e) A5, and f) A6

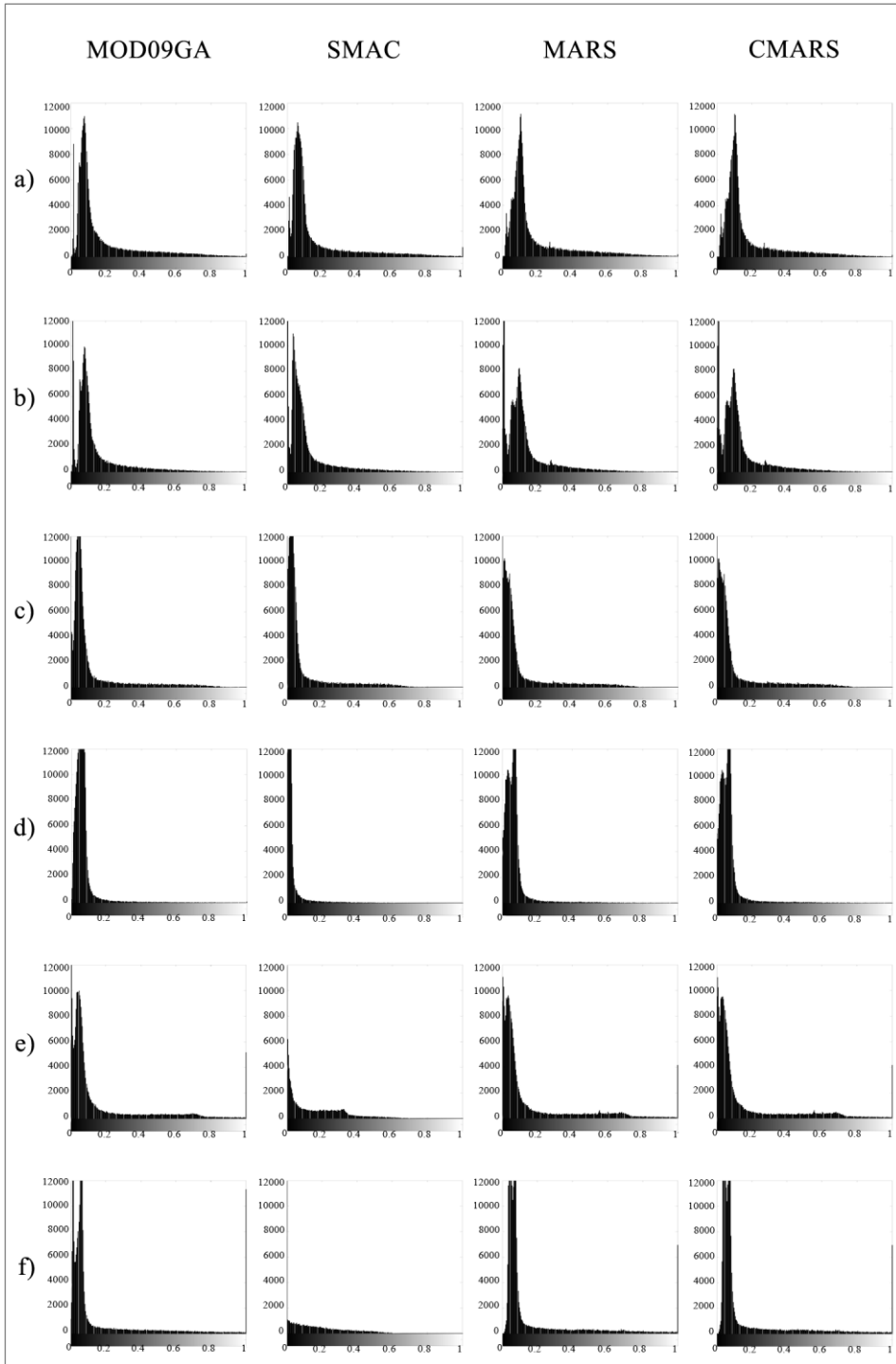


Figure 5.10: Image histograms for Alps data set: a) A7, b) A8, c) A9, d) A10, e) A11, and f) A12

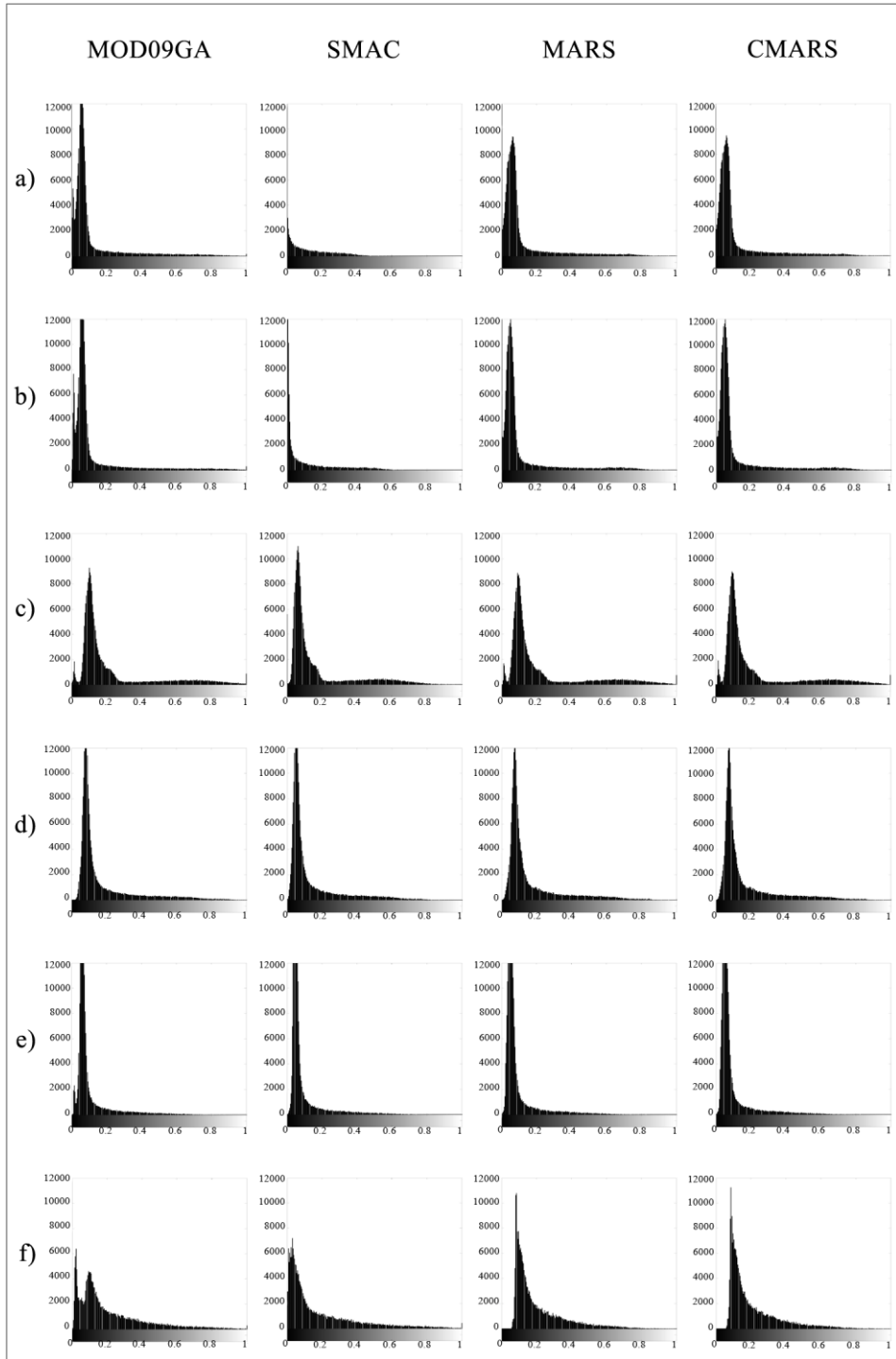


Figure 5.11: Image histograms for Turkey data set: a) T1, b) T2, c) T3, d) T4, e) T5, and f) T6

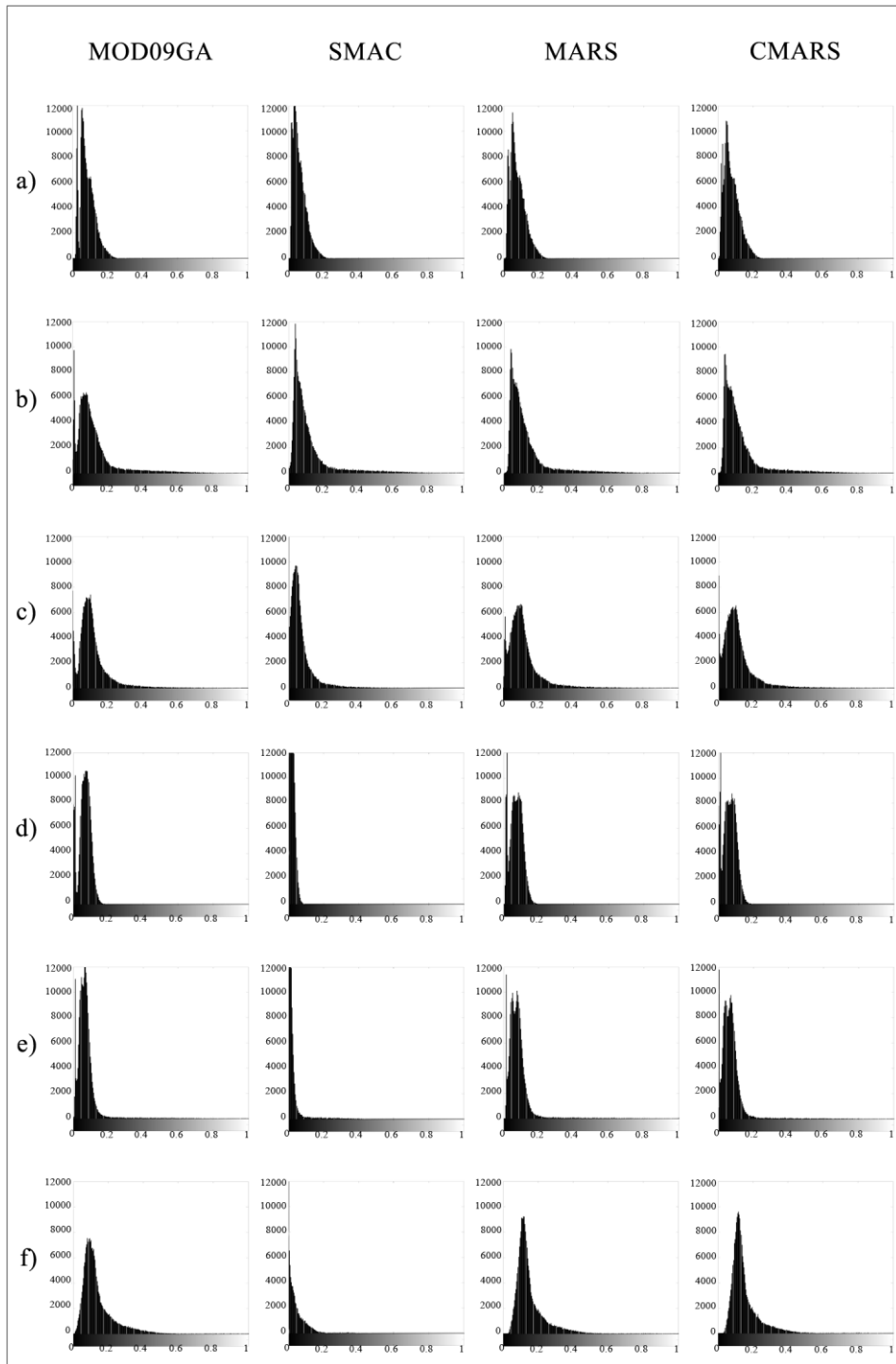


Figure 5.12: Image histograms for Turkey data set: a) T7, b) T8, c) T9, d) T10, e) T11, and f) T12

The processing time for the CMARS in the training phase is evidently much longer than the MARS (Table 5.11); however, it must be noted that MARS has already been in the statistical learning and data mining community for more than two decades, and it is available as a standalone and sophisticated software package, whereas different software packages (i.e., MARS<sup>®</sup>, MATLAB<sup>®</sup> and MOSEK<sup>™</sup>) have to be used for the time being in order to obtain the CMARS model, which naturally increases the processing time.

Table 5.11: Processing times (in seconds) for Alps (A) and Turkey (T)

Data Set	SMAC		MARS		CMARS	
	Training	Testing	Training	Testing	Training	Testing
A1	N/A	24	57	32	1148	233
A2		27		34		241
A3		23		32		239
A4		25		33		243
A5		24		35		241
A6		26		37		237
A7		28		32		246
A8		24		38		244
A9		24		33		238
A10		25		35		240
A11		23		33		242
A12		24		36		239
T1	N/A	25	59	33	1183	243
T2		23		36		229
T3		27		36		233
T4		24		35		231
T5		25		32		239
T6		24		37		242
T7		24		34		238
T8		26		34		234
T9		23		31		241
T10		24		34		232
T11		25		33		238
T12		26		35		240

The average processing time of MARS (34 seconds) in the testing phase is almost comparable with that of SMAC (25 seconds). Although the average testing time for CMARS (238 seconds) is longer than the other two models, it still seems acceptable and competitive for this kind of huge and complex real-life data. As new refinements are made on CMARS, improvement on the processing time will be achieved.

Here, we should further mention that CMARS proposes a “model-based” approach on controlling the complexity, in the form of TR which is very well-established in engineering and realized by us in the form of CQP together with its efficient Interior Point Methods. And by this, a remarkable support is gained from the more model-based methods of calculus and of modern mathematics in general, in particular, through state-of-the-art continuous optimization theory.

Moreover, via the tradeoff between precision and complexity, CMARS lets us open our investigations to real-world challenges which include noise in the response (output) variable, namely, to the *Theory of Inverse Problems*, to *Data Mining* and to *Statistical Learning*.

Since we are interested in the multi-criteria antagonism (tradeoff) between accuracy and stability (stability may be called as small complexity), this tradeoff, keeping the complexity under “control”, is realized by us through CMARS in different ways:

- restraining the discretized integral of first- and second-order derivatives of the basis functions within some margin or “tolerance” (under a chosen and prescribed upper bound),
- and using modern optimization theory (model-based approach) for a more integrated approach of both the forward and the backward step, which we have in MARS.

By this, we are closer also to exploiting the “power” of differential and integral calculus and to future developments in this calculus and in optimization theory.

Another specific feature of CMARS when viewed within the RS perspective regarding our particular problem, CMARS stays “longer” model-based than MARS since it offers a more direct and intriguing comparison with the traditional “formulas” of RS, which are originated from physics and other

natural sciences, where functional relationships between variables are assumed. CMARS permits this by its interactive and adaptive “functional model”, which is supported by optimization and applied mathematics and can learn under different forms of uncertainty.





## CHAPTER 6

### IMAGE CLASSIFICATION ON MODIS IMAGES BY MARS

#### 6.1 Background of MARS in Multispectral Image Classification

MARS has many successful applications in many fields of science and engineering, as mentioned in Chapter 2. Works such as [15, 35, 135-137] indicate that MARS not only performs better in cases where there exist a large number of variables, nonlinearity, multicollinearity, and even a high degree of interaction between predictor variables. But MARS also has a clear advantage over black-box approaches due to its well-elaborated statistical basis.

However, there is only one example dealing with the application of MARS in multispectral image classification [15]. This study used an *advanced spaceborne thermal emission and reflection radiometer* (ASTER) image of a part of Spanish province of Badajoz. It has an area of  $60 \times 60$  km, taken on 4 August 2000. The forest stand map in vector format that covers some of study area was used for both training and testing. Total 17 land classes were a priori determined. Training polygons (they are often called *region of interest*, ROI), areas in which the training data were collected, were defined in the forest stand map by buffer analysis. Necessary training data were obtained by superimposing the ROIs for each class onto the satellite image, and then extracting the necessary pixel reflectance values for each spectral band.

The MARS classification on the image was carried out in binary fashion, i.e., each time, one of the classes was fixed and labeled as “1”, and the rest was considered belonging to the other class, so labeled as “0”. In this way, class probability maps were generated for each class, and then they were combined

to obtain the final classification result. The result of the MARS classification was then compared with that of ML and PP methods in terms of *area under the curve* (AUC) statistics [138, 139], and as the results revealed, 14 out of 17 classes, MARS gave better classification accuracies.

Even though this study revealed the potential of MARS algorithm for the classification of multispectral satellite images, it lacks certain important aspects, in our opinion:

- Total area of the ROIs (976 km<sup>2</sup>), where the training samples were taken, was nearly 64% of the area used for accuracy assessment (1523 km<sup>2</sup>), which can easily be assumed as *over safe*,
- MARS algorithm allows the user to set certain model building parameters that directly affects the model's predictive performance such as the maximum number of BFs and the degree of interaction between predictor variables. However, this important issue and its impact on the final classification were not addressed,
- The binary classification approach used in the study is a bit *over complex* and *time consuming* in operational perspective, especially, when the number of classes is high,
- All the analysis covered a single geographic area, and were carried out on a single image taken by a medium spatial resolution sensor (i.e., ASTER). So, it is really difficult to reach a fair conclusion about the classification performance of MARS. We strongly believe that it should also be tested on images of different geographic regions taken by coarse resolution sensors such as MODIS, where the effect of mixed pixel can easily prevail due to the heterogeneous structure of landscapes over large areas.

As a result, in order to clarify these shortcomings, and to evaluate the classification performance of MARS approach on multispectral images in a more comprehensible manner: four MODIS images, two of them taken over Alps and the other two over Turkey, are classified by using MARS algorithm and traditional ML approach. The results are compared with the corresponding reference images by using error matrices (i.e., confusion matrix). In the following subsections, details of the analysis are given.

## 6.2 Image Set and Model Training

The data set used in this part of the study consists of four MODIS Terra images taken over two different geographic regions: Alps and Turkey. This set comprises three different products for each date, MOD02HKM calibrated Earth view image, MOD09GA surface reflectance image, and MOD10A1 daily snow cover image. For Alps, a single MOD09GA and a single MOD10A1 tile are used, whereas four tiles of the same product types are used for Turkey. All images have 500 m spatial resolution. The details of the images are given in Table 6.1.

The first step is to decide the number of classes in the images. MODIS is a coarse resolution instrument with LFOV sensors, and its tiles cover large areas. Therefore, the basic land cover types already available in MOD10A1 are chosen, namely, *snow*, *water*, *cloud*, and *land* (i.e., no snow). We have two specific reasons for making this choice:

- The test regions have surface areas changing roughly from 62 500 to 67 500 km<sup>2</sup>, and it is hardly possible to find a reference data to assess the accuracy of classifiers on such large areas. So, we prefer to use MODIS' own daily snow product,

- Different land features such as forest, bare soil and pasture; or different cloud types such as low and high level clouds, they all have different spectral signatures. Confining each of these spectrally complex features in a single class enables us to make the mixed pixel problem more apparent in our study, and to evaluate the performances of the classifiers.

Table 6.1: MODIS data set used in image classification

Data Set	Date	Image product type	Image size (row × col.)	Image tile info (.hdf)
AC1	10.03.2002	MOD02HKM	3423 × 5713	MOD02HKM.A2002069.0950.005.2010080042428
		MOD09GA	2176 × 3384	MOD09GA.A2002069.h18v04.005.2008233075920
		MOD10A1	2176 × 3384	MOD10A1.A2002069.h18v04.005.2008234025715
AC2	13.01.2006	MOD02HKM	3409 × 5822	MOD02HKM.A2006013.1055.005.2010203044449
		MOD09GA	2176 × 3384	MOD09GA.A2006013.h18v04.005.2008059145954
		MOD10A1	2176 × 3384	MOD10A1.A2006013.h18v04.005.2008060044128
TC1	18.12.2006	MOD02HKM	3550 × 5703	MOD02HKM.A2006352.0810.005.2010186235551
				MOD09GA.A2006352.h20v04.005.2008245162115
				MOD09GA.A2006352.h20v05.005.2008245153323
				MOD09GA.A2006352.h21v04.005.2008245161752
		MOD09GA	4790 × 9372	MOD09GA.A2006352.h21v05.005.2008245162301
				MOD10A1.A2006352.h20v04.005.2008245163543
				MOD10A1.A2006352.h20v05.005.2008245154937
				MOD10A1.A2006352.h21v04.005.2008245163033
MOD10A1	4790 × 9372	MOD10A1.A2006352.h21v05.005.2008245163046		
		MOD02HKM.A2009081.0805.005.2010238222139		
		MOD09GA.A2009081.h20v04.005.2009084013618		
		MOD09GA.A2009081.h20v05.005.2009084025517		
TC2	22.03.2009	MOD09GA	4790 × 9372	MOD09GA.A2009081.h21v04.005.2009084002955
				MOD09GA.A2009081.h21v05.005.2009084010416
				MOD10A1.A2009081.h20v04.005.2009084053142
				MOD10A1.A2009081.h20v05.005.2009084070048
		MOD10A1	4790 × 9372	MOD10A1.A2009081.h21v04.005.2009084052449
				MOD10A1.A2009081.h21v05.005.2009084052839

It is a characteristic of MODIS instrument that in most of the scenes, it is highly probable to observe cloud banks, whether dense or sparse. Another difficulty is to differentiate snow from low level clouds, because both exhibit large reflectance in visible bands and have similar thermal properties in infrared bands [140]. So, this also provides an excellent opportunity to test and compare the effectiveness of both classifiers under such challenging situation.

In model training phase, reflectance values from the solar reflective bands (bands 1-7) in MOD02HKM images are used as predictor variables. For each

solar reflective band, the SI values are converted to corresponding reflectance values. By using the 1<sup>st</sup>, 3<sup>rd</sup> and 4<sup>th</sup> bands from MOD09GA images, an RGB color composite image of each area is obtained (R=1<sup>st</sup>, G=4<sup>th</sup>, and B=3<sup>rd</sup> band). On each image, a test area is defined (Figure 6.1) in ArcMap™ software [141].

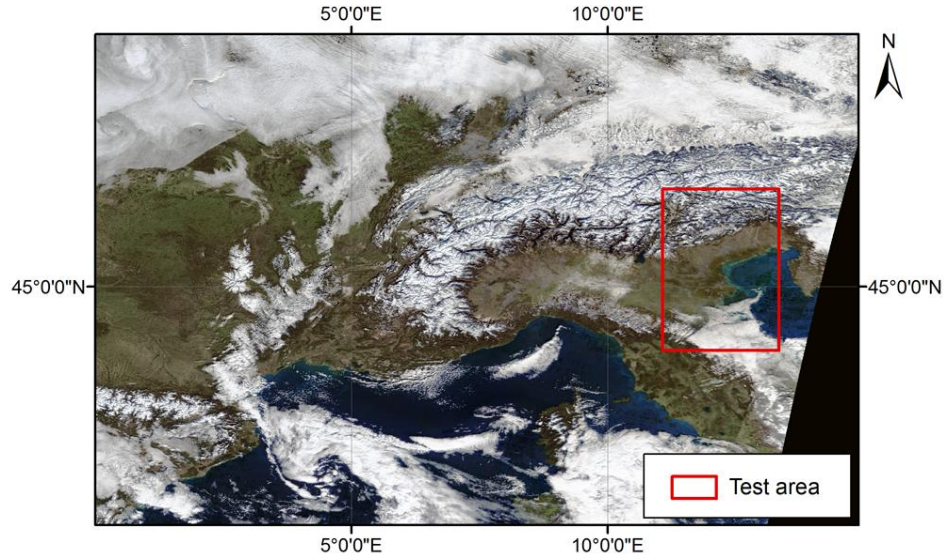


Figure 6.1: MOD09GA RGB color composite image of AC2 data set (R:1<sup>st</sup>, G: 4<sup>th</sup>, B: 3<sup>rd</sup> band)

In MARS model training, the “earth” module [142] under R statistical software [143] is used. The earth module requires two matrices to build MARS models, the matrix of predictor variables (i.e., reflectance values of bands 1-7), and the matrix of responses (i.e., pixel’s corresponding class). Since we are dealing with a multi-response MARS classification, instead of a binary classification as applied in [15], a special design for the response matrix is necessary. The response matrix should have four columns, each of which represents one of the classes.

At each row, the corresponding class is assigned “1” under the associated column and the other cells are labeled as “0”. By this way, earth produces four simultaneous models, each have the same set of BFs, but different coefficients.

For each class, ROIs for training samples are marked on the associated RGB image in such a way that they never fall inside or overlap with the test area on that image (i.e., ROIs are excluded from the test areas), and they are saved as shape files in ArcMap (Figure 6.2).

In order to prepare the necessary response-predictor matrix pairs for each data set, reflectance values from bands 1-7, together with the associated class labels, are extracted onto the pixels delineated by ROIs by using a code written in MATLAB, which is given in Appendix E.4.

The results are saved in text files. The average percentage of training data to the associated test area is nearly 25%. The details on the test areas and the training data can be found in Table 6.2.

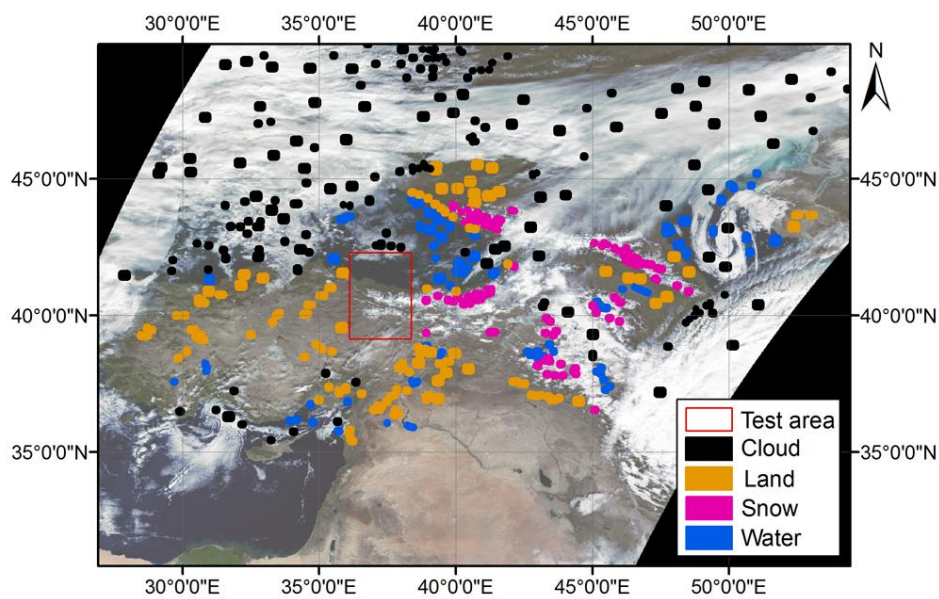


Figure 6.2: Training samples for TC1 data set

Table 6.2: Details of training and test data

Data Set	Test image size (row × col.)	Training data size (pixels)					% to test area
		Snow	Water	Cloud	Land	Total	
AC1	689 × 494 =340366 pixels	7796	17700	12154	40216	77866	22.9
AC2	688 × 494 =339872 pixels	6214	15800	26678	41300	89992	26.5
TC1	758 × 544 =412352 pixels	8065	16261	47555	32452	104333	25.3
TC2	758 × 544 =412352 pixels	9821	18031	36286	48732	112870	27.4

An example of a predictor-response matrix pair is given in Figure 6.3.

Response variables					Predictor variables						
SNOW	WATER	CLOUD	LAND		B01	B02	B03	B04	B05	B06	B07
1	0	0	0	→	0.11438	0.13688	0.13066	0.11435	0.07668	0.02652	0.01336
0	1	0	0		0.21561	0.24288	0.21334	0.21016	0.12669	0.02909	0.01476
0	0	1	0		0.22261	0.24897	0.21835	0.21314	0.12032	0.03194	0.01381
1	0	0	0		0.09884	0.12634	0.12178	0.1022	0.08829	0.02542	0.01534
0	0	0	1		0.18175	0.20506	0.19202	0.18143	0.11692	0.02769	0.01426

Figure 6.3: Response-predictor matrix pair for MARS model training

Table 6.3: Different MARS model building settings applied in image classification

Setting	Degree of interaction	Max. number of BFs
S1	1	20
S2		40
S3		60
S4		80
S5		100
S6	2	20
S7		40
S8		60
S9		80
S10		100
S11	3	20
S12		40
S13		60
S14		80
S15		100

The predictor-response matrix pair of each data set is introduced into R. Then the MARS classification models are generated for different settings of model building parameters (i.e., the maximum numbers of BFs and the degree of interaction between predictor variables). For all settings, threshold for stopping criteria is  $10^{-6}$ . These settings are summarized in Table 6.3.

### 6.3 Testing of MARS Models

For each data set, the reference image of the related test area is generated from the associated MOD10A1 daily snow product. During this process, some data values in MOD10A1 are merged, and then the whole image is reclassified to represent four classes: snow, water, cloud, and land Table 6.4.

Table 6.4: Merging of some classes in MOD10A1 reference image

MOD10A1 data value	Data value name	Reclassified as
37	lake	water
39	ocean	
25	no snow	land
50	cloud	cloud
100	lake ice	snow
200	snow	

In MOD10A1 daily snow images, water mask is superimposed on all other classes, including cloud. Consequently, the cloud cover over water bodies, if exists any, is suppressed by the water mask. To deal with the issue, a cloud mask is generated from the State\_1km QA data available in MOD09GA image by using a MATLAB code (Appendix E.5). The cloud mask generated for AC2 is given in Figure 6.4. The cloud mask is resampled to 500 m spatial resolution, and then applied onto the test area in order to obtain the final reference image.

Each MARS model, generated for each data set with different model building settings, given in Table 6.3, is applied on the test areas. ML approach is also



employed for each test area. The performance of MARS and ML classifiers on the test areas is presented in the associated error matrices.

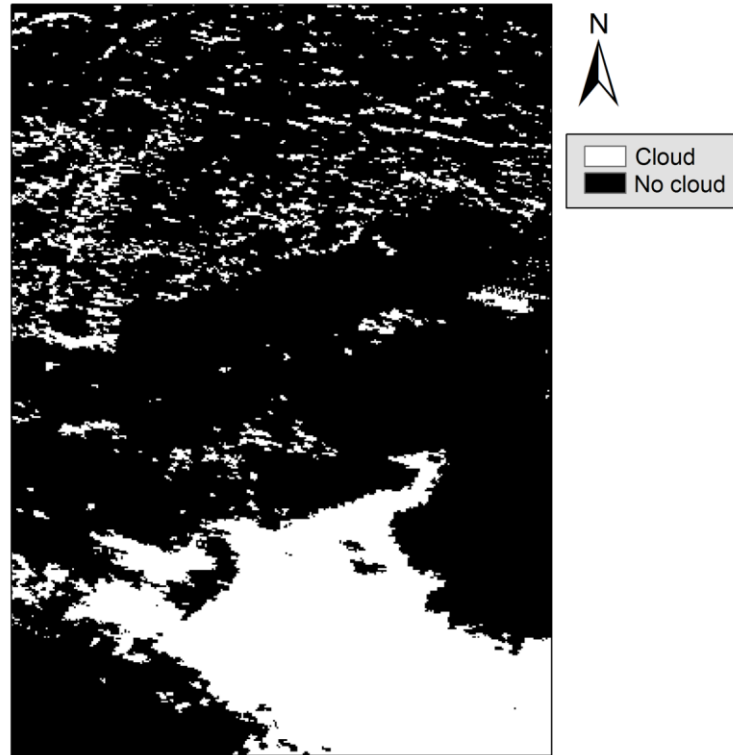


Figure 6.4: Cloud mask generated from MOD09GA State\_1km QA data for AC2 data set

### 6.3.1 Error matrix

The readers can find a brief introductory to the concept of *error matrix* in this subsection.

The classification accuracy of remotely sensed data is often expressed by an error matrix [64]. It is composed of square arrays of numbers arranged in the form of rows and columns, as illustrated in Table 6.5. The cells in the matrix represent the number of pixels assigned to a particular class with respect to the actual class as verified in the reference data [144].

Table 6.5: A sample error matrix

		Classified Data				Row Total
		Class 1	Class 2	Class 3	Class 4	
Reference Data	Class 1	130	8	44	48	230
	Class 2	12	162	10	16	200
	Class 3	0	22	170	38	230
	Class 4	8	14	6	180	208
Column Total		150	206	230	282	<b>868</b>

When the total number of correctly classified pixels in a class is divided by the total number of pixels of that class as derived from the reference data (i.e., the row total), *omission error* (i.e., producer's accuracy) is obtained. This measure gives the probability of a reference pixel being correctly classified.

The *commission error* (i.e., user's accuracy) is calculated by taking the ratio of the total number of correct pixels in a class to the total number of pixels that are classified in that class (i.e., the column total). Commission error refers to the probability that a pixel labeled as a certain class in the map is really this class.

*Overall accuracy*, the simplest descriptive statistic, is calculated by dividing the total correctly classified pixels (i.e., main diagonal) by the total number of pixels. Example calculations are given below.

Omission Error			Commission Error			Overall Accuracy	
Class 1=	130/230	=57%	Class 1=	130/150	=87%	642/868	=74%
Class 2=	162/200	=81%	Class 2=	162/206	=79%		
Class 3=	170/230	=74%	Class 3=	170/230	=74%		
Class 4=	180/208	=87%	Class 4=	180/282	=64%		

## 6.4 Results and Discussion

The related error matrices for AC1, AC2, TC1 and TC2 data sets are given in Tables 6.6 - 6.9, respectively (\* indicates the highest overall accuracy value, and ML means classification by maximum-likelihood method).

Table 6.6: Error matrix for AC1 data set

Setting	Omission Error (%)				Commission Error (%)				Overall (%)
	Snow	Water	Cloud	Land	Snow	Water	Cloud	Land	
S1	69.989	98.798	25.797	97.382	93.510	97.555	8.192	95.783	94.316
S2	69.507	98.677	27.908	97.171	93.745	97.566	7.619	95.890	94.113
S3	69.564	98.675	27.720	97.190	93.719	97.566	7.622	95.897	94.129
S4	69.566	98.675	27.580	97.190	93.705	97.566	7.588	95.897	94.128
S5	69.555	98.675	27.720	97.190	93.718	97.566	7.623	95.895	94.128
S6	71.044	98.196	42.777	96.345	94.569	97.892	8.544	96.180	93.726
S7	73.852	97.907	38.555	96.619	93.231	98.410	8.682	96.280	94.107
S8	74.898	97.575	34.428	96.683	92.137	98.742	8.350	96.133	94.153
S9	75.156	97.559	34.146	96.608	91.942	98.754	8.196	96.162	94.127
S10	76.021	97.599	34.475	96.546	91.923	98.717	8.368	96.242	94.194
S11	69.940	98.727	39.353	96.790	93.784	97.691	8.931	96.248	94.000
S12	72.751	97.989	41.839	96.710	94.152	98.285	8.628	96.517	94.086
S13	73.822	97.658	41.041	96.674	93.742	98.576	8.548	96.442	94.093
S14	75.766	97.637	34.897	96.636	91.986	98.583	8.257	96.442	94.236
S15	77.368	97.634	33.255	96.619	91.378	98.591	8.623	96.448	94.387*
ML	82.692	96.247	22.889	94.050	84.567	99.444	3.432	97.912	92.918

Table 6.7: Error matrix for AC2 data set

Setting	Omission Error (%)				Commission Error (%)				Overall (%)
	Snow	Water	Cloud	Land	Snow	Water	Cloud	Land	
S1	61.248	96.177	63.436	89.562	88.007	57.941	97.190	79.772	78.154
S2	61.368	96.118	63.465	89.573	88.134	57.977	97.345	79.698	78.179
S3	61.321	96.332	63.458	89.586	88.130	58.032	97.361	79.745	78.207
S4	61.321	96.327	63.463	89.586	88.132	58.035	97.356	79.743	78.208
S5	61.321	96.327	63.463	89.586	88.132	58.035	97.356	79.743	78.208
S6	64.193	95.033	66.499	90.882	85.738	67.059	96.501	77.564	79.862
S7	61.905	94.509	67.284	92.284	87.973	71.835	95.626	75.213	80.076
S8	61.031	93.518	68.290	92.409	86.613	73.092	92.773	75.816	80.064
S9	61.690	93.629	67.742	92.634	86.205	73.436	93.800	75.629	80.149
S10	63.509	93.668	67.813	92.653	85.783	73.865	94.896	75.921	80.535
S11	74.821	95.281	68.210	90.839	84.869	73.610	95.659	79.804	82.407
S12	74.831	92.878	67.876	92.954	83.171	81.929	94.816	77.838	82.743
S13	74.747	93.809	67.733	92.735	82.626	80.638	95.850	78.210	82.753*
S14	73.995	94.168	66.541	91.850	80.629	80.479	95.719	77.558	82.010
S15	77.244	94.390	66.803	91.693	80.082	80.747	95.317	79.263	82.687
ML	93.714	81.712	66.266	89.007	62.635	98.465	93.391	87.210	82.711

Table 6.8: Error matrix for TC1 data set

Setting	Omission Error (%) TC1				Commission Error (%)				Overall (%)
	Snow	Water	Cloud	Land	Snow	Water	Cloud	Land	
S1	61.945	99.755	35.258	97.615	76.077	91.748	88.211	71.036	79.240
S2	57.144	99.757	37.092	97.570	79.293	91.762	86.127	71.163	79.351
S3	56.671	99.755	37.403	97.555	79.524	91.788	85.775	71.249	79.392
S4	56.585	99.755	37.383	97.554	79.528	91.787	85.730	71.237	79.380
S5	56.571	99.755	37.385	97.554	79.532	91.787	85.723	71.237	79.379
S6	50.443	99.846	43.871	96.267	84.004	89.133	79.222	74.699	80.148
S7	53.458	99.824	47.140	95.928	80.671	90.041	80.471	75.906	81.086
S8	50.950	99.824	47.215	96.107	81.303	89.621	80.241	75.965	80.992
S9	50.381	99.813	53.381	96.115	80.200	90.731	80.527	78.477	82.566
S10	49.785	99.815	53.284	96.148	79.947	90.448	80.955	78.391	82.511
S11	37.324	99.803	46.989	96.186	85.351	90.974	73.182	75.366	79.960
S12	46.016	99.850	55.201	94.583	82.923	89.438	75.789	79.913	82.131
S13	53.282	99.851	51.587	94.761	78.616	89.276	77.607	78.859	81.787
S14	55.053	99.819	54.535	94.707	77.915	90.031	79.345	79.722	82.659
S15	59.546	99.818	53.103	94.939	71.102	89.931	81.280	80.092	82.704
ML	75.230	98.320	53.577	92.741	60.615	96.855	81.754	78.984	82.706*

Table 6.9: Error matrix for TC2 data set

Setting	Omission Error (%) TC2				Commission Error (%)				Overall (%)
	Snow	Water	Cloud	Land	Snow	Water	Cloud	Land	
S1	51.853	98.096	44.732	95.836	96.560	94.937	51.347	83.677	83.087
S2	51.397	98.116	44.208	95.812	96.601	94.347	50.800	83.640	82.941
S3	51.866	98.118	44.216	95.809	96.566	94.373	50.861	83.724	83.018
S4	51.872	98.118	44.219	95.810	96.563	94.373	50.868	83.725	83.020
S5	51.872	98.118	44.219	95.810	96.563	94.373	50.868	83.725	83.020
S6	38.178	97.960	57.360	93.793	96.682	95.068	38.512	87.591	80.970
S7	45.477	97.696	56.552	94.398	96.405	98.240	41.475	87.406	82.428
S8	46.074	97.479	56.580	94.148	96.494	98.852	43.381	86.293	82.355
S9	53.528	97.382	57.153	93.854	96.552	98.946	46.319	86.648	83.471
S10	57.538	97.370	56.959	93.919	96.586	98.979	48.802	86.628	84.155
S11	41.932	97.793	56.191	93.970	96.849	95.288	41.339	86.595	81.551
S12	78.479	97.762	58.498	92.753	95.077	94.874	58.877	88.892	87.160*
S13	78.031	97.550	64.003	91.600	95.202	97.994	55.734	89.510	86.979
S14	78.942	97.439	66.704	89.899	94.480	98.145	52.046	90.367	86.394
S15	82.138	97.326	66.267	89.762	93.730	98.087	52.020	91.121	86.782
ML	65.835	96.054	67.146	90.778	93.124	99.099	43.258	92.160	84.624

At first glance, MARS gives better overall accuracy for AC1, AC2 and TC2 data sets. Even though the performance of ML method on TC1 seems better than MARS, their overall accuracies are very close to each other.

On all data sets, when the degree of interaction is set to one (i.e., additive modeling) and the number of BFs is increased, no remarkable change in the overall accuracy is achieved. This behavior can be seen more evidently in Figures 6.5 - 6.8, where the overall accuracies of MARS models with different degree of interactions (DG) are plotted against the number of BFs for AC1, AC2, TC1 and TC2, respectively. We have to emphasize that the vertical axes of these graphs are not in common scale for better illustrative purpose.

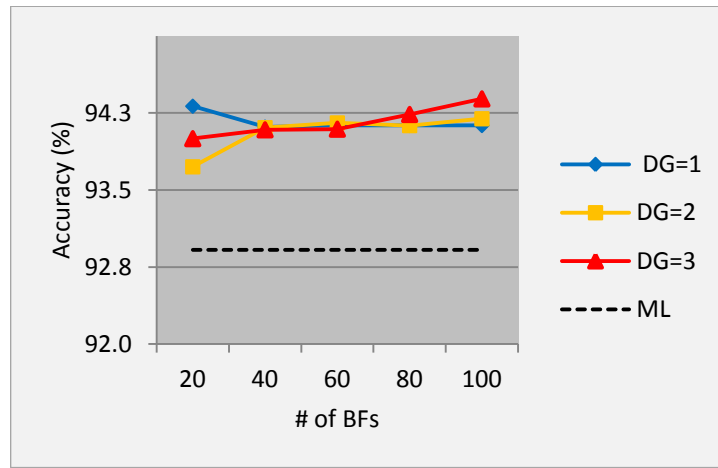


Figure 6.5: Graph of overall accuracy for AC1 data set

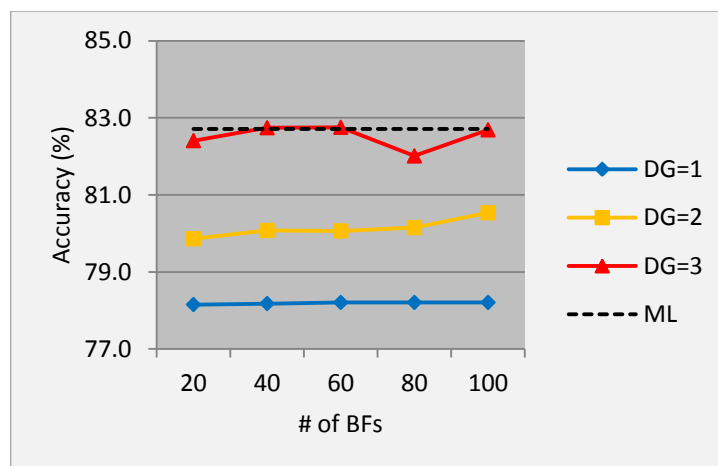


Figure 6.6: Graph of overall accuracy for AC2 data set

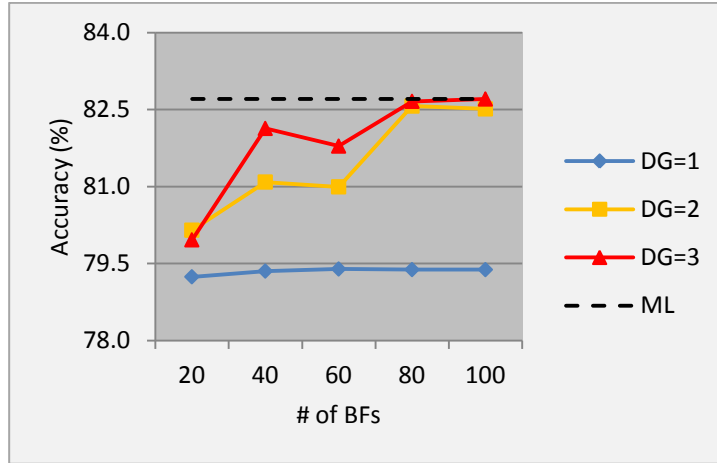


Figure 6.7: Graph of overall accuracy for TC1 data set

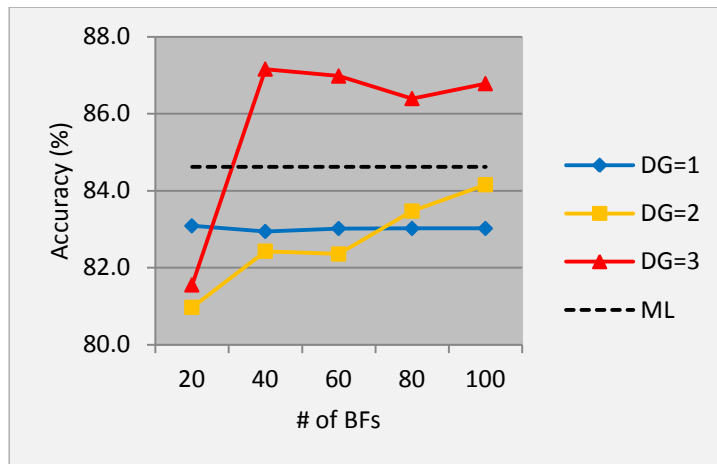


Figure 6.8: Graph of overall accuracy for TC2 data set

A similar behavior is also observed on AC1 and AC2 for the second and the third degree of interactions. Therefore, it can be concluded that when the degree of interaction is fixed, increasing the number of BFs may not always contribute significantly to the model's predictive performance. On the other hand, increase in both the number of BFs and the degree of interaction results in better classification performance. This is a typical, and expected, characteristic of MARS method because higher settings for the number of BFs and the degree of interaction allow MARS to enlarge its search space in order to add more terms into the model, which increase its predictive ability in turn.

Table 6.10: MARS model training performance for AC1 and AC2 data sets  
(ST: number of selected terms after the backward pass)

Setting	AC1			AC2		
	GCV	R <sup>2</sup>	ST	GCV	R <sup>2</sup>	ST
S1	0.04576	0.92940	17	0.03658	0.94510	15
S2	0.04371	0.93250	27	0.03571	0.94640	26
S3	0.04358	0.93280	37	0.03565	0.94660	33
S4	0.04356	0.93280	45	0.03564	0.94660	35
S5	0.04356	0.93280	45	0.03564	0.94660	35
S6	0.02945	0.95450	19	0.01614	0.97580	17
S7	0.02176	0.96650	34	0.01239	0.98140	29
S8	0.01944	0.97010	49	0.01085	0.98380	39
S9	0.01855	0.97150	62	0.01026	0.98460	53
S10	0.01784	0.97260	74	0.00987	0.98520	63
S11	0.03076	0.95250	17	0.01382	0.97930	16
S12	0.02089	0.96780	32	0.00972	0.98540	28
S13	0.01743	0.97310	43	0.00827	0.98760	39
S14	0.01567	0.97590	55	0.00733	0.98900	49
S15	0.01437	0.97790	65	0.00659	0.99010	59

Table 6.11: MARS model training performance for TC1 and TC2 data sets  
(ST: number of selected terms after the backward pass)

Setting	TC1			TC2		
	GCV	R <sup>2</sup>	ST	GCV	R <sup>2</sup>	ST
S1	0.05561	0.91650	16	0.04207	0.93790	16
S2	0.05409	0.91880	26	0.04075	0.93990	26
S3	0.05374	0.91930	36	0.04063	0.94010	37
S4	0.05364	0.91950	45	0.04061	0.94010	45
S5	0.05364	0.91950	47	0.04061	0.94010	45
S6	0.03279	0.95070	17	0.02651	0.96090	19
S7	0.02279	0.96580	33	0.02180	0.96780	36
S8	0.01923	0.97110	44	0.01960	0.97110	50
S9	0.01727	0.97410	55	0.01876	0.97240	65
S10	0.01663	0.97510	68	0.01847	0.97280	75
S11	0.03218	0.95170	16	0.02563	0.96220	19
S12	0.02147	0.96780	30	0.01852	0.97270	33
S13	0.01736	0.97390	44	0.01541	0.97730	45
S14	0.01450	0.97830	54	0.01381	0.97970	56
S15	0.01292	0.98060	64	0.01308	0.98070	66

In Tables 6.10 and 6.11, it can easily be observed that GCV and R<sup>2</sup> values for S1, S2, S3, S4 and S5 (DG=1) remain stable, supporting our conclusion. In these settings, although the number of terms included in the final model after the backward pass increase, no remarkable change in the GCV and R<sup>2</sup> values is

seen. For example,  $GCV=0.05561$  and  $R^2=0.91650$  for S1 in TC1 data set (number of BFs=20, DG=1). Increasing the number of BFs from 20 to 100 (i.e., S5), only gives 3.5% and 0.33% improvement in GCV (0.05364) and  $R^2$  (0.91950) values, respectively.

When the number of BFs is fixed and the degree of interaction is increased, as for S5, S10 and S15 in AC2, significant change is achieved in GCV and  $R^2$  values. If one compares S5 with S15, improvement in GCV and  $R^2$  are 83.8% and 4.6%, respectively.

As inferred from the results shown in Tables 6.6 - 6.9, there is no single setting of model building parameters that gives the best classification accuracy for each class in multi-response MARS classification. Before making further comments on the results, it would be helpful to check the amount of overlapping between individual classes, and also to see the percentage of pixels for each class in its own training data, which is given in Table 6.12.

Table 6.12: Percentages of pixels for each class in training data sets

<b>Data Set</b>	<b>Snow</b>	<b>Water</b>	<b>Cloud</b>	<b>Land</b>
<b>AC1</b>	10.01	22.73	15.61	51.65
<b>AC2</b>	6.91	17.56	29.64	45.89
<b>TC1</b>	7.73	15.59	45.58	31.10
<b>TC2</b>	8.70	15.98	32.15	43.18

Bar graphs that show the percentage of overlapping pixels between classes in MARS and ML methods for each data set are drawn. For MARS method, bar graphs are obtained only for the model setting that gives the highest overall accuracy for that data set (i.e., for AC1: S15, AC2: S13, TC1: S15, and TC2: S12). The graphs are given in Figures 6.9 - 6.16. The associated error matrices for MARS and ML methods are also given in Tables 6.13 - 6.16.



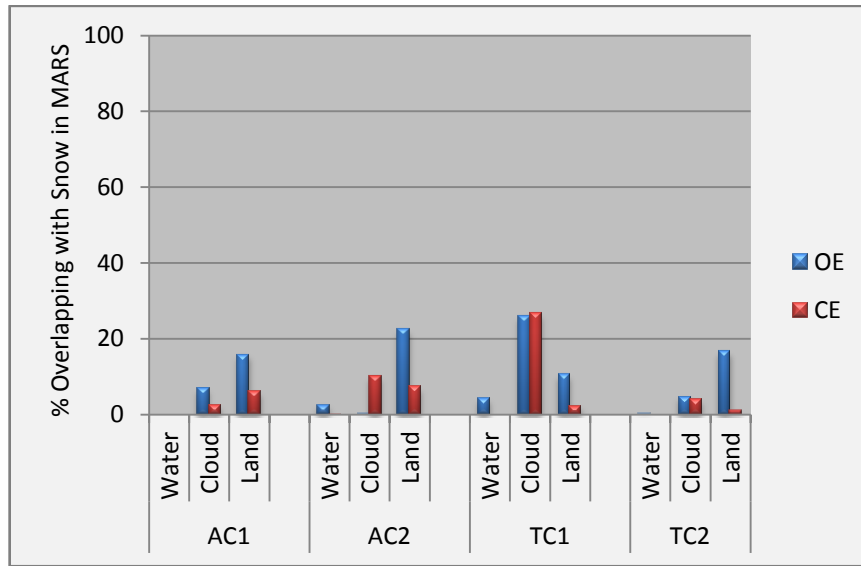


Figure 6.9: Percentage overlapping of water, cloud and land with snow in MARS (OE: omission error, CE: commission error)

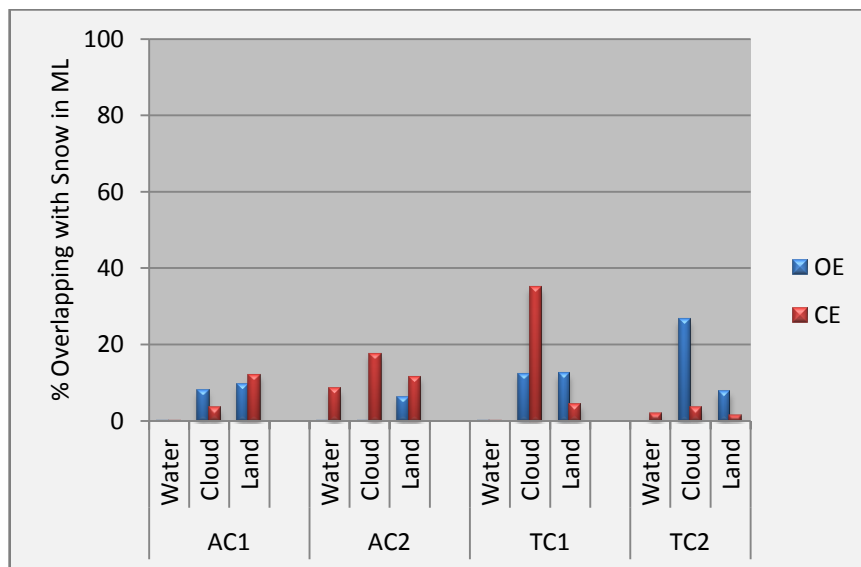


Figure 6.10: Percentage overlapping of water, cloud and land with snow in ML (OE: omission error, CE: commission error)

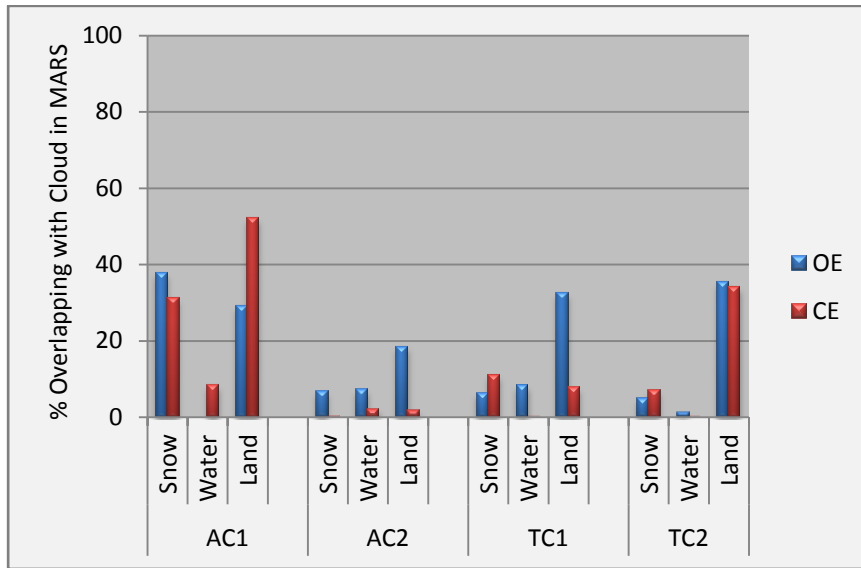


Figure 6.11: Percentage overlapping of snow, water and land with cloud in MARS (OE: omission error, CE: commission error)

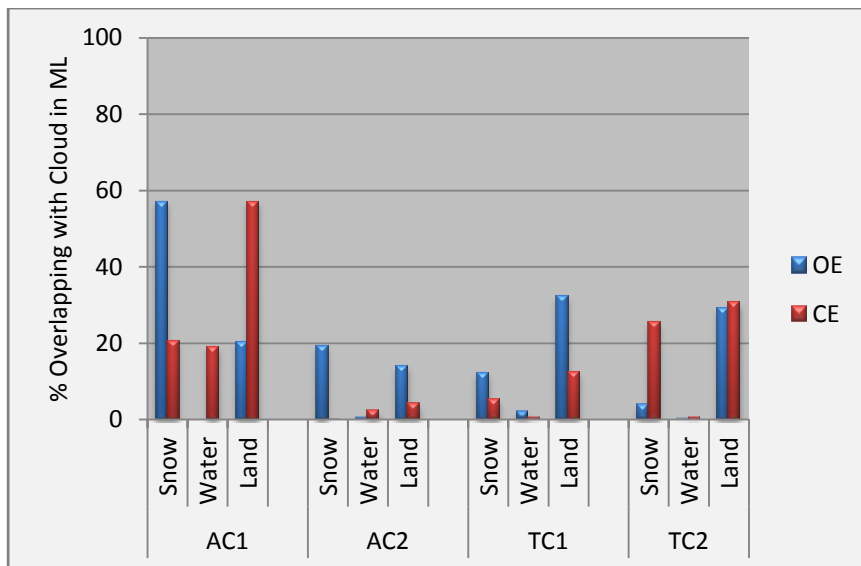


Figure 6.12: Percentage overlapping of snow, water and land with cloud in ML (OE: omission error, CE: commission error)

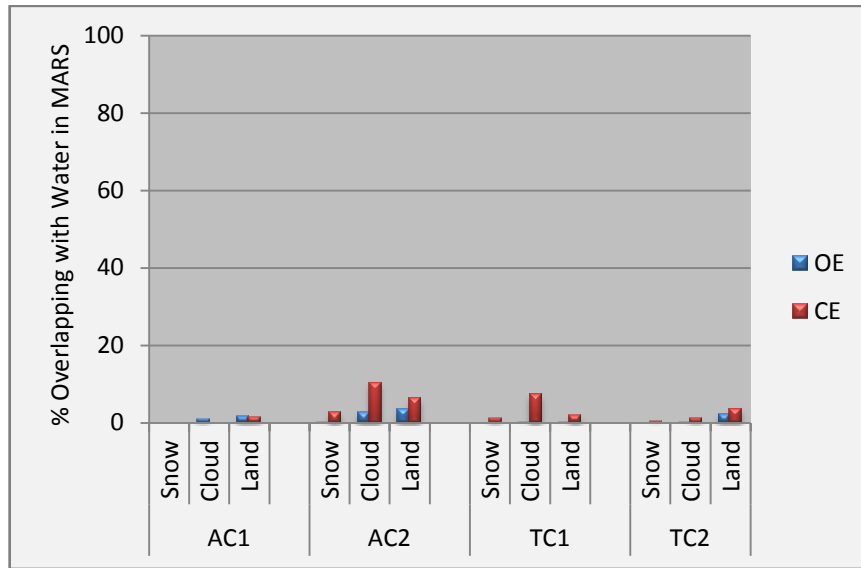


Figure 6.13: Percentage overlapping of snow, cloud and land with water in MARS (OE: omission error, CE: commission error)

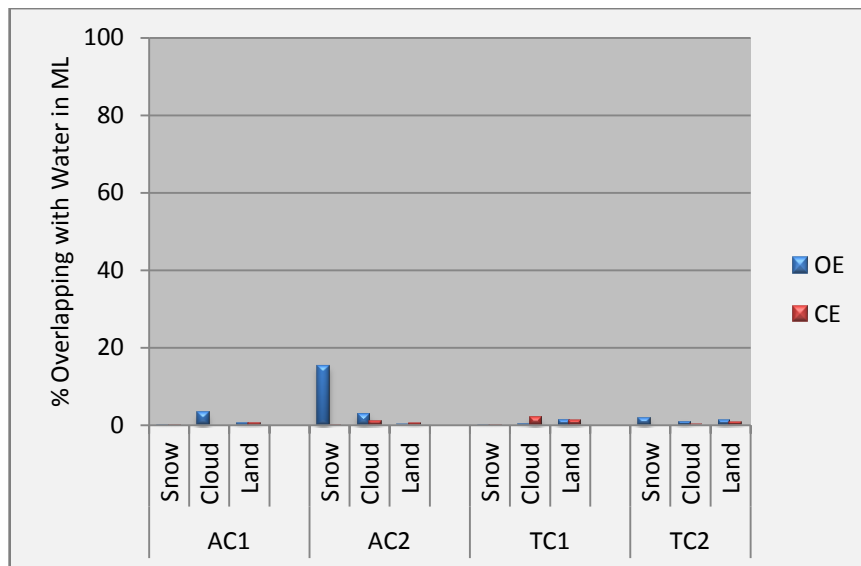


Figure 6.14: Percentage overlapping of snow, cloud and land with water in ML (OE: omission error, CE: commission error)

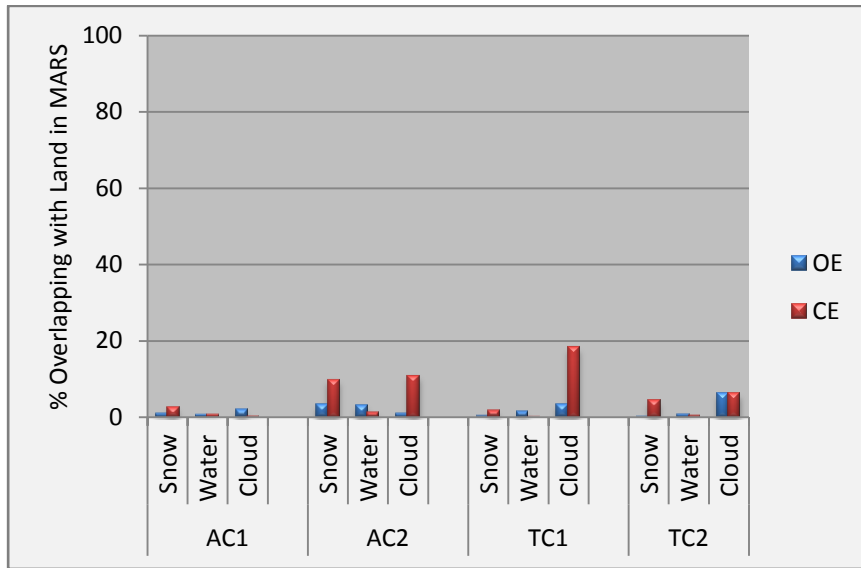


Figure 6.15: Percentage overlapping of snow, water and cloud with land in MARS (OE: omission error, CE: commission error)

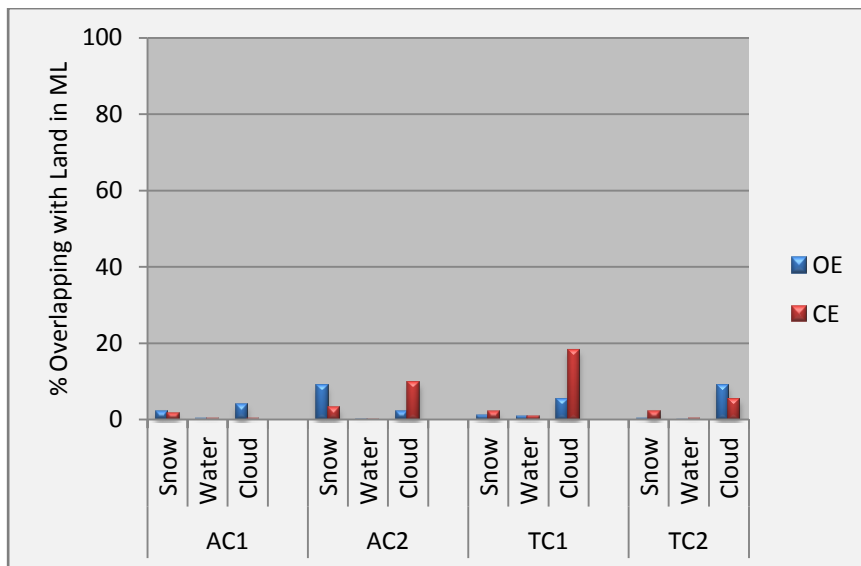


Figure 6.16: Percentage overlapping of snow, water and cloud with land in ML (OE: omission error, CE: commission error)

Table 6.13: Error matrices of MARS and ML methods for AC1 data set

		Classified by MARS (S15)				
		Snow	Water	Cloud	Land	Total
Reference (MOD10A1)	Snow	28 541	0	2550	5799	<b>36 890</b>
	Water	0	82 382	692	1304	<b>84 378</b>
	Cloud	805	0	709	618	<b>2132</b>
	Land	1888	1177	4271	209 630	<b>216 966</b>
	Total	<b>31 234</b>	<b>83 559</b>	<b>8222</b>	<b>217 351</b>	<b>340 366</b>
		Classified by ML				
		Snow	Water	Cloud	Land	Total
Reference (MOD10A1)	Snow	30 505	1	2914	3470	<b>36 890</b>
	Water	6	81 211	2707	454	<b>84 378</b>
	Cloud	1216	0	488	428	<b>2132</b>
	Land	4345	453	8112	204 056	<b>216 966</b>
	Total	<b>36 072</b>	<b>81 665</b>	<b>14 221</b>	<b>208 408</b>	<b>340 366</b>

Table 6.14: Error matrices of MARS and ML methods for AC2 data set

		Classified by MARS (S13)				
		Snow	Water	Cloud	Land	Total
Reference (MOD10A1)	Snow	49 406	1681	136	14 875	<b>66 098</b>
	Water	46	50 669	1387	1911	<b>54 013</b>
	Cloud	6049	6515	61 269	16 623	<b>90 456</b>
	Land	4294	3970	1130	119 911	<b>129 305</b>
	Total	<b>59 795</b>	<b>62 835</b>	<b>63 922</b>	<b>153 320</b>	<b>339 872</b>
		Classified by ML				
		Snow	Water	Cloud	Land	Total
Reference (MOD10A1)	Snow	61 943	18	24	4113	<b>66 098</b>
	Water	8226	44 135	1545	107	<b>54 013</b>
	Cloud	17 401	454	59 942	12 659	<b>90 456</b>
	Land	11 325	216	2673	115 091	<b>129 305</b>
	Total	<b>98 895</b>	<b>44 823</b>	<b>64 184</b>	<b>131 970</b>	<b>339 872</b>

Table 6.15: Error matrices of MARS and ML methods for TC1 data set

		Classified by MARS (S15)				
		Snow	Water	Cloud	Land	Total
Reference (MOD10A1)	Snow	17 996	1268	7780	3178	<b>30 222</b>
	Water	0	111 569	52	151	<b>111 772</b>
	Cloud	6745	8962	57 383	34 970	<b>108 060</b>
	Land	569	2261	5384	154 084	<b>162 298</b>
	Total	<b>25 310</b>	<b>124 060</b>	<b>70 599</b>	<b>192 383</b>	<b>412 352</b>
		Classified by ML				
		Snow	Water	Cloud	Land	Total
Reference (MOD10A1)	Snow	22 736	6	3710	3770	<b>30 222</b>
	Water	9	109 894	352	1517	<b>111 772</b>
	Cloud	13 142	2261	57 895	34 762	<b>108 060</b>
	Land	1622	1301	8859	150 516	<b>162 298</b>
	Total	<b>37 509</b>	<b>113 462</b>	<b>70 816</b>	<b>190 565</b>	<b>412 352</b>

Table 6.16: Error matrices of MARS and ML methods for TC2 data set

		Classified by MARS (S12)				
		Snow	Water	Cloud	Land	Total
Reference (MOD10A1)	Snow	53 905	154	3188	11 440	<b>68 687</b>
	Water	0	48 362	29	1078	<b>49 469</b>
	Cloud	2236	575	26 874	16 255	<b>45 940</b>
	Land	555	1884	15 553	230 264	<b>248 256</b>
	Total	<b>56 696</b>	<b>50 975</b>	<b>45 644</b>	<b>259 037</b>	<b>412 352</b>
		Classified by ML				
		Snow	Water	Cloud	Land	Total
Reference (MOD10A1)	Snow	45 220	0	18 182	5285	<b>68 687</b>
	Water	918	47 517	429	605	<b>49 469</b>
	Cloud	1717	94	30 847	13 282	<b>45 940</b>
	Land	704	338	21 851	225 363	<b>248 256</b>
	Total	<b>48 559</b>	<b>47 949</b>	<b>71 309</b>	<b>244 535</b>	<b>412 352</b>

As observed in Figures 6.13 - 6.16, and also in Tables 6.13 - 6.16, water and land have no significant overlapping with other classes in overall for both MARS and ML classification. For water, there is a slight overlapping with cloud for MARS and with snow for ML method in AC2 data set, 10.37%-CE and 15.23%-OE, respectively. For land, the highest overlapping is with cloud for both MARS and ML methods in TC1 data set, both around 18%-CE. So, high and comparable accuracy levels are obtained for land and water by MARS and ML methods in AC1, TC1 and TC2 data sets. Due to the above mentioned overlapping problem, performances of MARS and ML methods differ in AC2 data set, in favor of MARS for omission error, and in favor of ML for commission error. Additionally, land has the greatest percentage in terms of pixels in the training data sets, 43% in average (Table 6.12). Although the average percentage of water pixels in the training data is 18%, water exhibits the lowest overlapping with other classes.

For snow (Figures 6.9 - 6.10, and Tables 6.13 - 6.16), no considerable overlapping exists with water (the highest value: 8.32%-CE in AC2 for ML classification). The most significant overlapping with land is in AC1, AC2 and TC2 data sets for MARS method with values of 15.72%-OE, 22.50%-OE and 16.65%-OE, respectively. Overlapping with land in other data sets for both MARS and ML methods is roughly around 10%. However, in TC1 data set, overlapping with cloud is about 25%-OE/CE for MARS method, and 35%-CE for ML method. A mid-level overlapping with cloud is also observed for ML method in AC2 and TC data sets. In general, overlapping with cloud in MARS method is relatively lower than in ML method.

For cloud (Figures 6.11 - 6.12, and Tables 6.13 - 6.16), serious overlapping exists with snow in AC1 data set for MARS (38%-OE, 31%-CE) and ML (57%-OE, 20%-CE) methods, and in AC2 (19%-CE) and TC2 (26%-CE) data sets for ML method. However, MARS shows lower overlapping with snow than ML method. Additionally, MARS and ML methods exhibit almost similar

behaviors in all data sets. Considerable amount of overlapping with land is observed (for MARS: AC1: 29%-OE, 52%-CE, AC2: 18%-OE, TC1: 32%-OE, TC2: 35%-OE/CE; for ML: AC1: 20%-OE, 57%-CE, AC2: 14%-OE, TC1: 32%-OE, 13%-CE, TC2: 30%-OE/CE). Only moderate overlapping with water is seen in AC1 data set (19%-CE) for ML method. In AC1 training data set, percentage of cloud pixels is relatively lower than that of land pixels, 16% and 52%, respectively. This difference may cause overlapping with land pixels in AC1 data set. Cloud has ~30% and ~32% of the training pixels in AC2 and TC2 data sets, respectively; whereas ~46% and ~43% of the training pixels belong to land for the same two data sets (Table 6.12). In both AC2 and TC2, relatively lower overlapping with land than in AC1 is observed. However, overlapping with land in TC2 is higher than in AC2 for both MARS and ML methods, although these two data sets have the same percentage of pixels in training data sets. Therefore, it is hard to relate their overlapping with the percentage of training pixels.

It is also of value to mention that the reference data used in the accuracy assessment is MODIS daily snow product (i.e., MOD10A1). This product is composed of different layers of data obtained by various algorithms such as snow algorithm, land/water mask and cloud mask algorithms [104]. Therefore, errors inherent with each layer also pass to MOD10A1 product, and can eventually have a negative contribution to our accuracy assessment. On the other hand, it is the best available product when the type of the classification scheme and the sizes of the test areas are considered.

According to the results, larger number of BFs and higher degree of interaction should be preferred for multispectral image classification by multi-response MARS model. However, it should also be noted that there is no unique setting for the best model since the performance of the MARS is highly dependent on the training data. Therefore, the user should certainly “play” with the primary MARS model building parameters, i.e., the maximum number of BFs and the



degree of interaction, in order to observe their effects on the model's behavior during the model training phase.

One drawback of building multi-response MARS model by using R is that it does not allow users to adjust classification thresholds (i.e., cut-off values) for individual classes as studied by [15] in a binary fashion. On the other hand, binary classification approach may not seem operationally practical if the number of classes is large. In Salford MARS, only binary classification scheme can be implemented.

Additionally, we have to emphasize that neither our MARS and CMARS approaches nor a traditional RT-based method are employed to atmospherically correct the TOA reflectance values of the four image sets prior to the classification process. Because our classification scheme is not based on an *end-member extraction* approach, in which the mixed pixel problem is treated by decomposing those pixels into a set of pure signatures for a certain spectral class, and their corresponding proportions [10, 145]. As indicated by [146], calibration and atmospheric correction on image data are required in most end-member approaches. A review of end-member approaches is available in [147].

In our study; however, we follow a *hard classification* approach, in which a pixel is assigned to a specific class by selecting the greatest likelihood of belonging to that class (i.e., one pixel-one class) [10, 148]. In many hard classification studies where homogenous atmospheric conditions prevail or a single image is classified, atmospheric correction is not considered as a necessary pre-processing stage [149, 150].

Besides, our correction scheme is only valid for the 4<sup>th</sup> MODIS band, so an assessment on the variation in the results of classification based on an atmospheric correction on a single band (i.e., MODIS 4<sup>th</sup> reflective solar band) would probably lead no significant outcome. For those who are interested in

the effect of the atmosphere on spectral signatures measured by satellites, detailed information can be found in [151-153].

The classified images of AC1, AC2, TC1 and TC2 data sets are given in Figures 6.17 - 6.20, respectively.

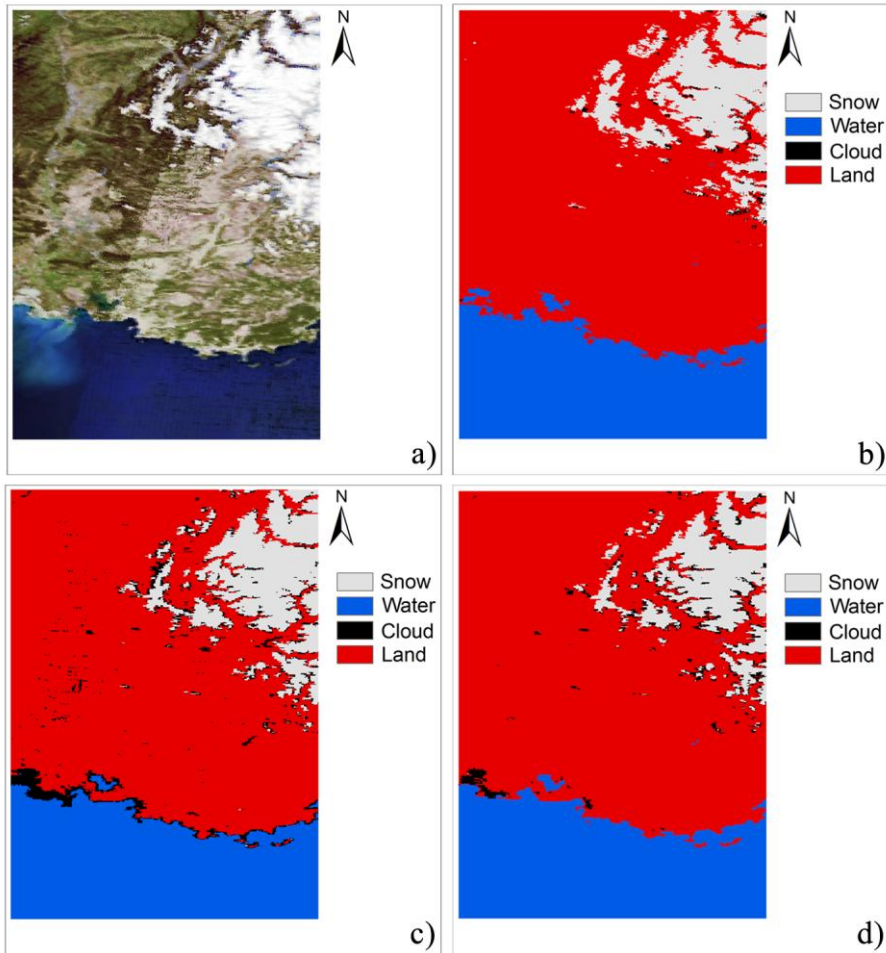


Figure 6.17: Classified images of AC1 data set: a) MOD09GA RGB color composite image (R:1<sup>st</sup>, G: 4<sup>th</sup>, B: 3<sup>rd</sup> band), b) MOD10A1 reference image, c) ML classification, and d) MARS classification

Overlapping of cloud with land and water is apparent in the above figure for ML method.

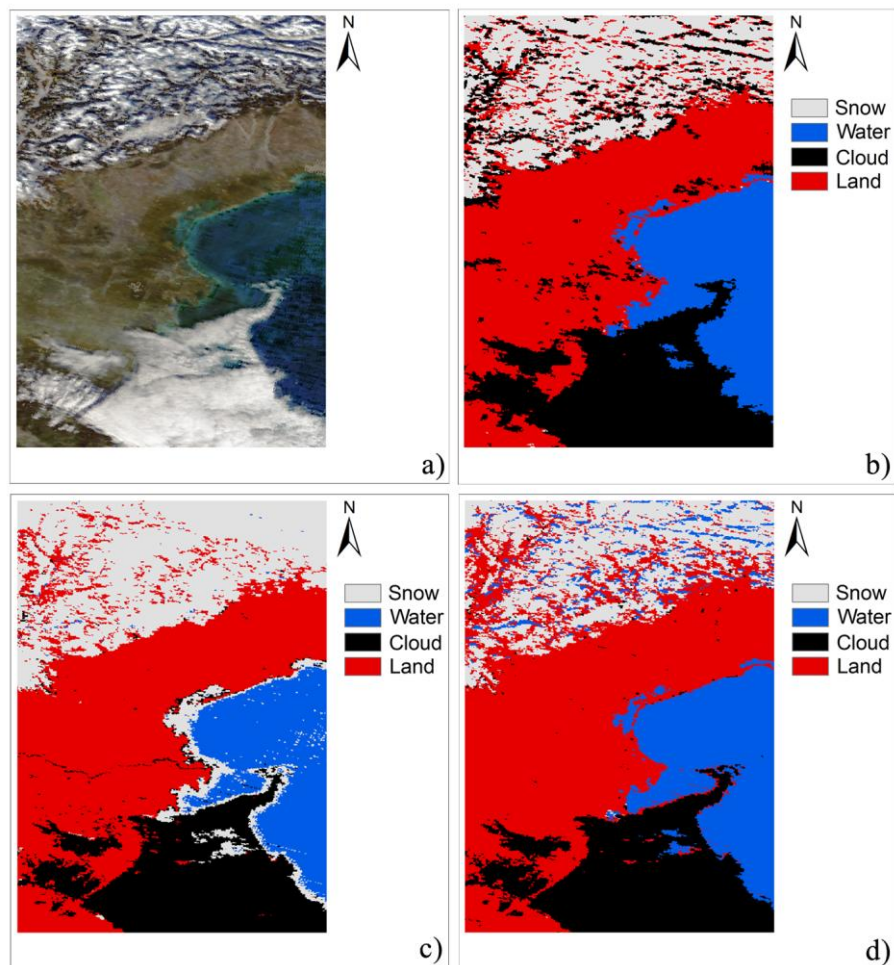


Figure 6.18: Classified images of AC2 data set: a) MOD09GA RGB color composite image (R:1<sup>st</sup>, G: 4<sup>th</sup>, B: 3<sup>rd</sup> band), b) MOD10A1 reference image, c) ML classification, and d) MARS classification

In Figure 6.18c, overlapping of water with snow for ML method is clearly seen as also indicated in Figure 6.14. For MARS method (Figure 6.18d), relatively slighter overlapping between cloud, water and land is observed.

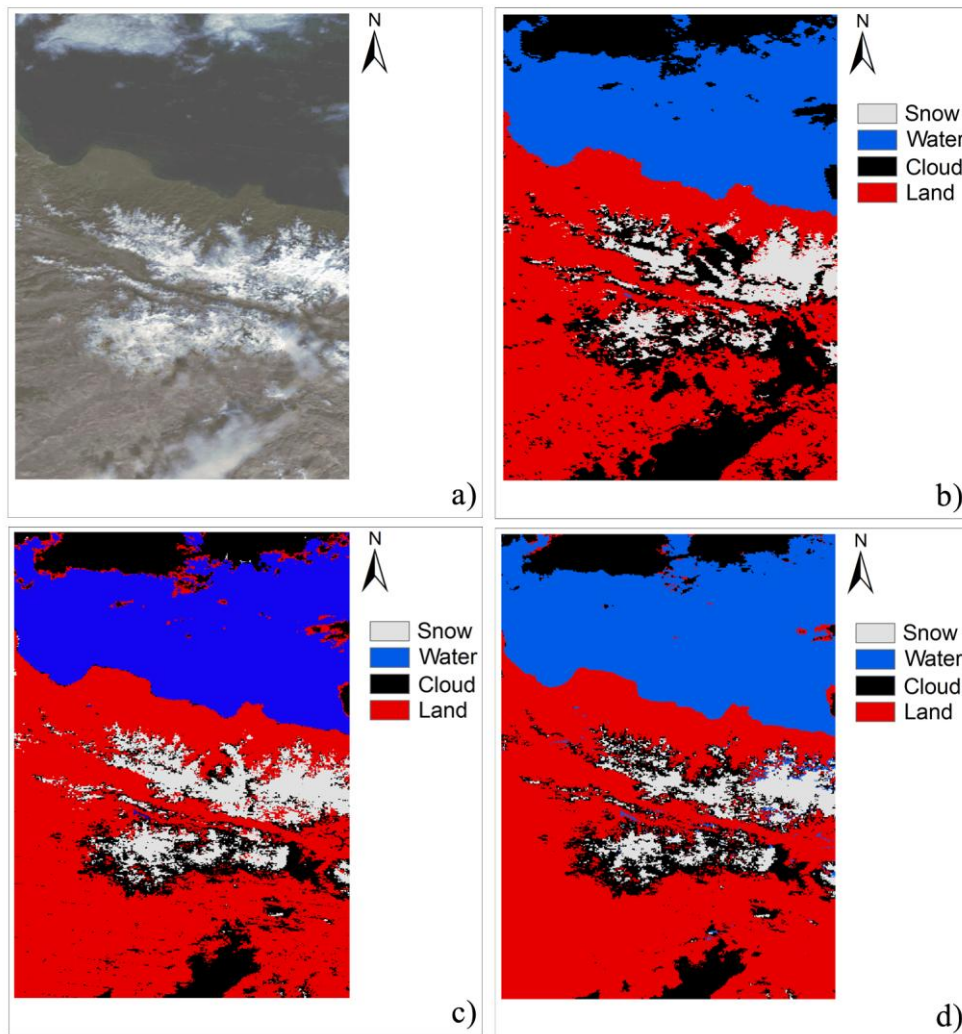


Figure 6.19: Classified images of TC1 data set: a) MOD09GA RGB color composite image (R:1<sup>st</sup>, G: 4<sup>th</sup>, B: 3<sup>rd</sup> band), b) MOD10A1 reference image, c) ML classification, and d) MARS classification

For ML method in Figure 6.19c, relatively high overlapping between land and cloud exists, as compared to MARS method shown in Figure 6.19d.

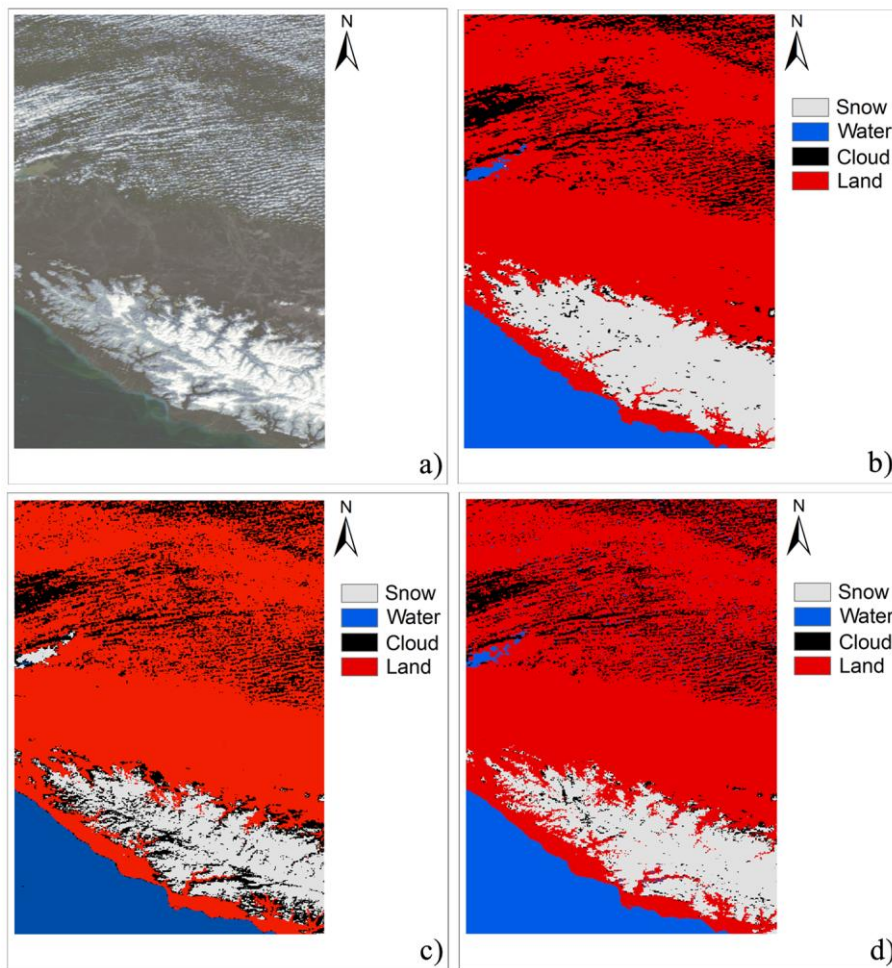


Figure 6.20: Classified images of TC2 data set: a) MOD09GA RGB color composite image (R:1<sup>st</sup>, G: 4<sup>th</sup>, B: 3<sup>rd</sup> band), b) MOD10A1 reference image, c) ML classification, and d) MARS classification

In Figure 6.20, overall classification performance of MARS seems better than ML method. Small water body in the western part of Figure 6.20b is not recognized by ML method, but successfully detected by MARS.

According to our visual inspections on the classified images with respect to RGB and MOD10A1 images, MARS has relatively better performance than ML classifier for all data sets.



## CHAPTER 7

### CONCLUSIONS AND RECOMMENDATIONS

In this study, we have approached the two main active research topics in RS within a different and progressive perspective. First, atmospheric correction issue has been treated by using two powerful nonparametric regression tools: MARS and CMARS. Then, we have dealt with image classification for snow mapping within the frame of multi-response MARS approach.

Our MARS- and CMARS-based modeling schemes in atmospheric correction allow users to perform atmospheric correction on images without the need for retrieval of atmospheric parameters, which is required in all RT-based approaches.

Nonparametric atmospheric correction models (nonphysical-based models) have been generated for MODIS data (for the 4<sup>th</sup> reflective solar band) by employing MARS and CMARS methods. Performances of MARS and CMARS models have been tested on MODIS images, and compared with an RT-based correction algorithm, SMAC. According to results of the statistical tests on the accuracy measures, MARS and CMARS have demonstrated their superiority.

In the model training phase, atmospherically corrected surface reflectance values, which are used as predictor variable for MARS and CMARS, are extracted from MOD09GA products. This issue may naturally lead to a criticism since these atmospherically corrected surface reflectance values are not observed values in real sense, yet predicted values computed from a physical-based model, 6S. However, when the spatial resolution and the surface area of the images are considered, it is practically impossible to acquire

the each pixel's ground surface reflectance value by in-situ measurements. Instead, surface reflectance values predicted by a model with high-accuracy have been preferred in model training.

According to our findings in this study, performance of SMAC method is highly dependent on the variations in AOD. For AOD values between 0.20 and 0.50, our MARS and CMARS methods perform in a similar way with SMAC, whereas, MARS and CMARS are more accurate and reliable for lower AOD levels.

Although the comparison reported in Table 5.6 reveals that CMARS has accuracy quite similar to MARS, the results of pairwise comparisons given in Table 5.10 indicate statistically significant difference in favor of CMARS method.

On the other hand, the computational time spent by CMARS for model building is much higher (Table 5.11). This may seem like a shortcoming at first glance, but should not be considered as a permanent one. It essentially arises from the need for use of several software packages in CMARS processing. However, once the model is built (i.e., trained), it is applied on the image data consecutively. Therefore, we strongly believe that the testing time, 238 seconds in average for our data set, will certainly be confined within more reasonable margins as further improvements are made on CMARS algorithm.

In image classification for snow mapping, use of MARS method has been demonstrated in a well-elaborated framework. A multi-response MARS classification scheme has been implemented on MODIS images. The performance of the approach has been tested under various settings of primary model building parameters (i.e., maximum number of BFs and degree of interaction between predictor variables). The ML classification has also been applied on the same test data for comparison.



The multi-response classification has been carried out by using earth module in R software. It produces  $k$  simultaneous models, where  $k$  is the number of classes, and each model has the same set of BFs but different coefficients. One certain limitation of this approach is that the user is not allowed to define certain cut-off values to increase the separability of individual classes as implemented in [15] by following a binary fashion classification scheme. Binary classification approach would naturally be considered as an alternative. However, it may not be practical in cases where the number of classes is high. Because for each binary pair, a different MARS model has to be trained.

The results have revealed that the model building parameters have substantial impact on the predictive performance of MARS approach in image classification. As the maximum number of BFs and the degree of interaction increase, significant improvements are observed in MARS classification, and better overall classification accuracies are obtained as compared to ML method.

Spatial and spectral overlapping of classes is a challenging issue in image classification, especially when working with coarse-spatial resolution instruments. As the results have indicated, significant overlapping between snow and cloud is observed, and in fact, it has already been expected. Because as mentioned in Section 6.2, snow and cloud have high reflectance in visible spectrum, and they exhibit similar thermal characteristics in infrared region. Additionally, when classifying for snow, it has also a high tendency to overlap with land. MARS approach exhibit lower overlapping in CE values for snow as compared to ML method, whereas ML has the same behavior for OE values.

As compared to the overlapping of land when classifying for snow, cloud has a much higher tendency to overlap with land. Again, overall overlapping in CE values is less in MARS for cloud.

More “realistic” and better performance of MARS approach is also apparent when the classified images have been visually inspected and compared with the associated RGB and MOD10A1 images. Thus, we hope that the demonstrated MARS classification scheme and the presented results would serve as a guide and be helpful for those who will consider MARS for their image classification applications.

As an overall conclusion, natural phenomena and real-life problems often exhibit nonparametric and multivariate behavior, in which the curse of dimensionality may easily prevail. This means, controlled successive multiplications may sometimes be inevitable and useful to represent the multi-dimensional interaction of the model. We have demonstrated that high-dimensional and complex RS data sets can be wisely handled with the inherent smoothing characteristic of splines.

As our study has showed us, CMARS (and MARS) can be employed as an alternative tool for atmospheric correction. For image classification, MARS approach has proven its competitiveness and applicability as compared to traditional Bayesian-based ML approach. Thus, MARS and CMARS can also be used for other problems associated with Earth sciences. By integrating the dynamical progress of scientific advances in modern continuous optimization with the spatial technologies, we can enhance our understanding of the value of spatial data and its inherent structure to reach better modeling capabilities in GIS and RS.

The following points would be considered as further extensions of our study:

- In this study, the atmospheric correction models produced for MODIS data by MARS and CMARS were only applicable on the 4<sup>th</sup> solar band of the instrument. Implementation of MARS and CMARS methods to generate atmospheric correction models that can be used for all solar

reflective bands of MODIS (bands 1-7) will be an interesting direction to pursue. While generating these models, it is also worth experimenting with *Robust* CMARS (RCMARS) [154] through the involvement of robust optimization. By this, the regularization in CMARS becomes rigorously extended towards a robustification which especially addresses uncertainty in the input variables, too, and goes beyond noise in the response variable by uncertainty.

- In this thesis, we presented the application of multi-response MARS method in image classification for snow mapping on MODIS images and compared its performance with the classical ML approach. However, multi-response MARS modeling scheme imposes that the user has no freedom to choose specific cut-off values to increase correct classification probabilities of individual classes. Implementation of both MARS and CMARS methods in binary fashion on larger MODIS image sets by making extensive testing with different model building settings will be worth examining. Evaluating the performances of binary MARS and CMARS classification schemes on other conventional parametric and nonparametric methods such as PP, MD, SVM and ANN will have great potential.
- Finally, an interesting future direction would be to see if MARS and CMARS can be incorporated in studying the time dependent variety of the problem, i.e. processing of large image data sets, both time-discretely and time-continuously.



## REFERENCES

1. Trotter, C.M., *Remotely-sensed data as an information source for geographical information systems in natural resource management a review*. International Journal of Geographical Information Systems, 1991. **5**: p. 225-239.
2. Tso, B. and Mather, P.M., *Classification Methods for Remotely Sensed Data*. 2nd ed. 2009: CRC Press.
3. Zagolski, F. and Gastellu-Etchegorry, J.P., *Atmospheric corrections of AVIRIS images with a procedure based on the inversion of 5S model*, in *Geoscience and Remote Sensing Symposium, 1994. IGARSS '94. Surface and Atmospheric Remote Sensing: Technologies, Data Analysis and Interpretation., International*, 8-12 Aug 1994, vol. 1, p. 44–46.
4. Nikolakopoulos, K.G., Vaiopoulos, D.A. and Skianis, G.A., *A comparative study of different atmospheric correction algorithms over an Area with complex geomorphology in western peloponnese, Greece*, in *Geoscience and Remote Sensing Symposium, 2002. IGARSS '02. 2002 IEEE International*, 24-28 June 2002, vol. 4, p. 2492–2494.
5. Chavez, P.S., *Image-based atmospheric corrections-revisited and improved*. Photogrammetric engineering and remote sensing, 1996. **62**: p. 1025-1035.
6. Moran, M.S., Jackson, R.D., Slater, P.N. and Teillet, P.M., *Evaluation of simplified procedures for retrieval of land surface reflectance factors from satellite sensor output*. Remote Sensing of Environment, 1992. **41**: p. 169-184.
7. Rahman, H. and Dedeiu, G., *SMAC: A simplified method for the atmospheric correction of satellite measurements in the solar spectrum*. International Journal of Remote Sensing, 1994. **15**: p. 123–143.
8. Lu, D. and Weng, Q., *A survey of image classification methods and techniques for improving classification performance*. International Journal of Remote Sensing, 2007. **28**: p. 823-870.

9. Rees, W.G., *Physical Principles of Remote Sensing*. 2nd ed. 2001, United Kingdom: Cambridge University Press.
10. Schowengerdt, R.A., *Remote sensing: models and methods for image processing*. 3rd ed. 2006: Academic press.
11. Otukey, J.R. and Blaschke, T., *Land cover change assessment using decision trees, support vector machines and maximum likelihood classification algorithms*. International Journal of Applied Earth Observation and Geoinformation, 2010. **12, Supplement 1**: p. S27-S31.
12. Paola, J.D. and Schowengerdt, R.A., *A detailed comparison of backpropagation neural network and maximum-likelihood classifiers for urban land use classification*. Geoscience and Remote Sensing, IEEE Transactions on, 1995. **33**: p. 981-996.
13. Wang, F., *Fuzzy supervised classification of remote sensing images*. Geoscience and Remote Sensing, IEEE Transactions on, 1990. **28**: p. 194-201.
14. Yuan, H., Van Der Wiele, C.F. and Khorram, S., *An Automated Artificial Neural Network System for Land Use/Land Cover Classification from Landsat TM Imagery*. Remote Sensing, 2009. **1**: p. 243-265.
15. Quirós, E., Felicísimo, Á.M. and Cuartero, A., *Testing Multivariate Adaptive Regression Splines (MARS) as a Method of Land Cover Classification of TERRA-ASTER Satellite Images*. Sensors, 2009. **9**: p. 9011-9028.
16. Friedman, J.H., *Multivariate adaptive regression splines*. The Annals of Statistics, 1991. **19**: p. 1-67.
17. Weber, G.-W., Batmaz, İ., Köksal, G., Taylan, P. and Yerlikaya-Özkurt, F., *CMARS: a new contribution to nonparametric regression with multivariate adaptive regression splines supported by continuous optimization*. Inverse Problems in Science and Engineering, 2011. **20**: p. 371-400.

18. Kuter, S., Weber, G.-W., Özmen, A. and Akyürek, Z., *Modern applied mathematics for alternative modelling of the atmospheric effects on satellite images*, in *Modeling, Optimization, Dynamics and Bioeconomy, Springer Proceedings in Mathematics*, D. Zilberman and A. Pinto, Editors. 2013, Springer.
19. Lhermitte, S., Verbesselt, J. and Verstraeten, W.W.C., P., *A comparison of time series similarity measures for classification and change detection of ecosystem dynamics*. *Remote Sensing of Environment*, 2011. **115**: p. 3129-3152.
20. Permuter, H., Francos, J. and Jermyn, I., *A study of Gaussian mixture models of color and texture features for image classification and segmentation*. *Pattern Recognition*, 2006. **39**: p. 695-706.
21. Pouteau, R., Rambal, S., Ratte, J.P., Gogé, F., Joffre, R. and Winkel, T., *Downscaling MODIS-derived maps using GIS and boosted regression trees: The case of frost occurrence over the arid Andean highlands of Bolivia*. *Remote Sensing of Environment*, 2011. **115**: p. 117-129.
22. Steele, B.M., *Combining Multiple Classifiers: An Application Using Spatial and Remotely Sensed Information for Land Cover Type Mapping*. *Remote Sensing of Environment*, 2000. **74**: p. 545-556.
23. Xu, M., Watanachaturaporn, P., Varshney, P.K. and Arora, M.K., *Decision tree regression for soft classification of remote sensing data*. *Remote Sensing of Environment*, 2005. **97**: p. 322-336.
24. Yu, J. and Ekström, M., *Multispectral image classification using wavelets: a simulation study*. *Pattern Recognition*, 2003. **36**: p. 889-898.
25. Zhang, W., Wang, W. and Wu, F., *The Application of Multi-variable Optimum Regression Analysis to Remote Sensing Imageries in Monitoring Landslide Disaster*. *Energy Procedia*, 2012. **16**: p. 190-196.
26. Özmen, A., Kropat, E. and Weber, G.-W., *Spline regression models for complex multi-modal regulatory networks*. *Optimization Methods and Software*, 2012. DOI: 10.1080/10556788.2013.821611.

27. Alp, Ö.S., Büyükbebeci, E., Çekiç, A.İ., Özkurt, F.Y., Taylan, P. and Weber, G.-W., *CMARS and GAM & CQP-Modern optimization methods applied to international credit default prediction*. Journal of Computational and Applied Mathematics, 2011. **235**: p. 4639-4651.
28. Deichmann, J., Eshghi, A., Haughton, D., Sayek, S. and Teebagy, N., *Application of Multiple Adaptive Regression Splines (MARS) in direct response modeling*. Journal of Interactive Marketing, 2002. **16**: p. 15-27.
29. Lee, T., Chiu, C., Chou, Y. and Lu, C., *Mining the customer credit using classification and regression tree and multivariate adaptive regression splines*. Computational Statistics and Data Analysis, 2006. **50**: p. 1113-1130.
30. Balshi, M.S., McGuire, A.D., Duffis, P., Flannigan, M., Walsh, J. and Melillo, J., *Assessing the response of area burned to changing climate in western boreal North America using a multivariate adaptive regression splines (MARS) approach*. Global Change Biology, 2009. **15**: p. 578-600.
31. Felicísimo, Á.M., Gómez, A. and Muñoz, J., *Potential distribution of forest species in dehesas of Extremadura (Spain)*, in *Sustainability of Agrosilvopastoral Systems-Dehesas, Montados, Advances in GeoEcology 37*, S. Schnabel and A. Ferreira, Editors. 2005, Catena Verlag: Reiskirchen. p. 231-246.
32. Guitérrez, A.G., Schnabel, S. and Contador, J.F.L., *Using and comparing two nonparametric methods (CART and MARS) to model the potential distribution of gullies*. Ecological Modelling, 2009. **220**: p. 3630-3637.
33. Heikkinen, R.K., Luoto, M., Kuussaari, M. and Toivonen, T., *Modelling the spatial distribution of a threatened butterfly: Impacts of scale and statistical technique*. Landscape and Urban Planning, 2007. **79**: p. 347-357.
34. Henne, P.D., Hu, F.S. and Cleland, D.T., *Lake-effect snow as the dominant control of mesic-forest distribution in Michigan, USA*. Journal of Ecology, 2007. **95**: p. 517-529.



35. Leathwick, J.R., Elith, J. and Hastie, T., *Comparative performance of generalized additive models and multivariate adaptive regression splines for statistical modelling of species distributions*. Ecological Modeling, 2006. **199**: p. 188-196.
36. Viscarra Rossel, R.A. and Behrens, T., *Using data mining to model and interpret soil diffuse reflectance spectra*. Geoderma, 2010. **158**: p. 46-54.
37. Banks, D.L., Olszewski, R.T. and Maxion, R.A., *Comparing Methods for Multivariate Nonparametric Regression*. Communications in Statistics - Simulation and Computation, 2003. **32**: p. 541-571.
38. Ben-Ari, E.N. and Steinberg, D.M., *Modeling Data from Computer Experiments: An empirical comparison of kriging with MARS and projection pursuit regression*. Quality Engineering, 2007. **19**: p. 327-338.
39. Krzyścin, J.W., Eerme, K. and Janouch, M., *Long-term variations of the UV-B radiation over Central Europe as derived from the reconstructed UV time series*. Annales Geophysicae, 2004. **22**: p. 1473-1485.
40. Briand, L.C., Freimut, B. and Vollei, F., *Using multiple adaptive regression splines to support decision making in code inspections*. The Journal of Systems and Software, 2004. **73**: p. 205-217.
41. Mukhopadhyay, A. and Iqbal, A., *Prediction of mechanical property of steel strips using multivariate adaptive regression splines*. Journal of Applied Statistics, 2009. **36**: p. 1-9.
42. Samui, P., Das, S. and Kim, D., *Uplift capacity of suction caisson in clay using multivariate adaptive regression spline*. Ocean Engineering, 2011. **38**: p. 2123-2127.
43. Zhou, Y. and Leung, H., *Predicting object-oriented software maintainability using multivariate adaptive regression splines*. The Journal of Systems and Software, 2007. **80**: p. 1349-1361.

44. Chou, S.M., Lee, T.S., Shao, Y.E. and Chen, I.F., *Mining the breast cancer pattern using artificial neural networks and multivariate adaptive regression splines*. Expert Systems with Applications, 2004. **27**: p. 133-142.
45. Cook, N.R., Zee, R.Y.L. and Ridker, P.M., *Tree and spline based association analysis of gene–gene interaction models for ischemic stroke*. Statistics in Medicine, 2004. **23**: p. 1439-1453.
46. Put, R., Xua, Q.S., Massart, D.L. and Vander Heyden, Y., *Multivariate adaptive regression splines (MARS) in chromatographic quantitative structure–retention relationship studies*. Journal of Chromatography, 2004. **A 1055**: p. 11-19.
47. York, T.P., Eaves, L.J. and van den Oord, E.J.C.G., *Multivariate adaptive regression splines: a powerful method for detecting disease-risk relationship differences among subgroups*. Statistics in Medicine, 2006. **25**: p. 1355-1367.
48. Zakeri, I.F., Adolph, A.L., Puyau, M.R., Vohra, F.A. and Butte, N.F., *Multivariate adaptive regression splines models for the prediction of energy expenditure in children and adolescents*. Journal of Applied Physiology, 2010. **108**: p. 128-136.
49. Durmaz, M. and Karslıoğlu, M.O., *Non-parametric regional VTEC modeling with Multivariate Adaptive Regression B-Splines*. Advances in Space Research, 2011. **48**: p. 1523-1530.
50. Durmaz, M., Karslıoğlu, M.O. and Nohutcu, M., *Regional VTEC modeling with multivariate adaptive regression splines*. Advances in Space Research, 2010. **46**: p. 180-189.
51. Miao, Z., Shi, W., Zhang, H. and Wang, X., *Road Centerline Extraction From High-Resolution Imagery Based on Shape Features and Multivariate Adaptive Regression Splines*. IEEE Geoscience and Remote Sensing Letters, 2013. **10**: p. 583-587.
52. Hastie, T., Tibshirani, R. and Friedman, J., *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2nd ed. 2009: Springer.

53. Batmaz, İ., Yerlikaya- Özkurt, F., Kartal-Koç, E., Köksal, G. and Weber, G.-W., *Evaluating the CMARS performance for modeling non-linearities*, in *AIP Conference Proceedings*, 2–4 February, 2010, vol. 1239, p. 351-357.
54. Özkurt, F.Y., Batmaz, İ. and Weber, G.W., *A review of conic multivariate adaptive regression splines (CMARS): a powerful tool for predictive data mining*, in *Modeling, Optimization, Dynamics and Bioeconomy*, D. Zilberman and A. Pinto, Editors. to appear as chapter in book, Springer Proceedings in Mathematics.
55. Taylan, P. and Weber, G.W., *Organization in finance prepared by stochastic differential equations with additive and nonlinear models and continuous optimization*. Organizacija (Organization - Journal of Management, Information Systems and Human Resources), 2008. **41**: p. 185-193.
56. Taylan, P., Weber, G.W. and Kropat, E., *Approximation of stochastic differential equations by additive models using splines and conic programming*. International Journal of Computing Anticipatory Systems, 2008. **21**: p. 341-352.
57. Taylan, P., Weber, G.-W. and Beck, A., *New approaches to regression by generalized additive models and continuous optimization for modern applications in finance, science and technology*. Optimization: A Journal of Mathematical Programming and Operations Research, 2007. **56**: p. 675-698.
58. Weber, G.W., Taylan, P., Görgülü, Z.K., Rahman, H.A. and Bahar, A., *Parameter Estimation in Stochastic Differential Equations*, in *Dynamics, Games and Science II*, M.M. Peixoto, A.A. Pinto, and D.A. Rand, Editors. 2011, Springer Berlin Heidelberg. p. 703-733.
59. Weber, G.-W., Çavuşoğlu, Z. and Özmen, A., *Predicting default probabilities in emerging markets by new conic generalized partial linear models and their optimization*. Optimization: A Journal of Mathematical Programming and Operations Research, 2012. **61**: p. 443-457.

60. Yerlikaya-Özkurt, F., Askan, A. and Weber, G.-W., *An alternative approach to ground motion prediction problem by a non-parametric adaptive regression method.* to appear in Engineering Optimization.
61. Proud, S.R., Rasmussen, M.O., Fensholt, R., Sandholt, I., Shisanya, C., Mutero, W., Mbow, C. and Anyamba, A., *Improving the SMAC atmospheric correction code by analysis of Meteosat Second Generation NDVI and surface reflectance data.* Remote Sensing of Environment, 2010. **114**: p. 1687–1698.
62. Herman, B. and Browning, S., *The effect of aerosols on the earth-atmosphere albedo.* Journal of the Atmospheric Sciences, 1975. **32**: p. 1430–1445.
63. Proud, S.R., Fensholt, R., Rasmussen, M.O. and Sandholt, I., *A comparison of the effectiveness of 6S and SMAC in correcting for atmospheric interference in Meteosat Second Generation images.* Journal of Geophysical Research – Atmospheres, 2010. **115**: p. (D17209)1-14.
64. Mather, P.M., *Computer Processing of Remotely-Sensed Images An Introduction.* 3rd ed. 2004, England: John Wiley & Sons Ltd.
65. Gao, B.-C., Montes, M.J., Davis, C.O. and Goetz, A.F.H., *Atmospheric correction algorithms for hyperspectral remote sensing data of land and ocean.* Remote Sensing of Environment, 2009. **113**: p. S17–S24.
66. Richards, J.A. and Jia, X., *Remote sensing digital image analysis: An introduction.* 4th ed. 2006, Germany: Springer.
67. Vermote, E.F., El Saleous, N.Z. and Justice, C.O., *Atmospheric correction of MODIS data in the visible to middle infrared: first results.* Remote Sensing of Environment, 2002. **83**: p. 97–111.
68. Bass, M., Van Stryland, E.W., Williams, D.R. and Wolfe, W.L., *Handbook of optics.* Vol. 2. 2001: McGraw-Hill.
69. Tanre, D., Deroo, C., Duhaut, P., Herman, M., Morcrette, J.J., Perbos, J. and Deschamps, P.Y., *Description of a computer code to simulate the*

*satellite signal in the solar spectrum - The 5S code*. International Journal of Remote Sensing, 1990. **11**: p. 659-668.

70. Berk, A., Bernstein, L.S. and Robertson, D.C., *MODTRAN: A moderate resolution model for LOWTRAN7*. Final report, GL-TR-89-0122, AFGL, Hanscom AFB, MA 42 pp, 1989.
71. Berk, A., Bernstein, L.S., Anderson, G.P., Acharya, P.K., Robertson, D.C., Chetwynd, J.H. and Adler-Golden, S.M., *MODTRAN Cloud and Multiple Scattering Upgrades with Application to AVIRIS*. Remote Sensing of Environment, 1998. **65**: p. 367-375.
72. Xu, Y., Wang, R., Liu, S., Yang, S. and Yan, B., *Atmospheric correction of hyperspectral data using MODTRAN model*, in *Remote Sensing of the Environment: 16th National Symposium on Remote Sensing of China. Proceedings of the SPIE*, November 24, 2008, vol. 7123, p. 712306-1-712306-7.
73. Gao, B.-C., Heidebrecht, K.B. and Goetz, A.F.H., *Derivation of scaled surface reflectances from AVIRIS data*. Remote Sensing of Environment, 1993. **44**: p. 165-178.
74. Fensholt, R., Sandholt, I., Stisen, S. and Tucker, C., *Analysing NDVI for the African continent using the geostationary meteosat second generation SEVIRI sensor*. Remote Sensing of Environment, 2006. **101**: p. 212-229.
75. Geiger, B., Franchistéguy, L. and Roujean, J.-L., *Land surface albedo from Meteosat Second Generation (MSG) observations*, in *Geoscience and Remote Sensing Symposium, 2003. IGARSS 2003. Proceedings*, 2003, vol. 4, p. 2617-2619.
76. Popp, C., Foppa, N., Hauser, A. and Wunderle, S., *Aerosol optical depth retrieval over land using Meteosat-8 SEVIRI data*, in *ESA Atmospheric Science Conference*, 08-12 May 2006.
77. Beal, D., Baret, F., Weiss, M., Gu, X. and Verbrugge, M., *A method for MERIS atmospheric correction based on the spectral and spatial observation*, in *Geoscience and Remote Sensing Symposium, 2003. IGARSS 2003. Proceedings*, 21-25 July 2003 vol. 6, p. 3601-3603.

78. Hagolle, O., Dedieu, G., Mougenot, B., Debaecker, V., Duchemin, B. and Meygret, A., *Correction of aerosol effects on multi-temporal images acquired with constant viewing angles: Application to Formosat-2 images*. Remote Sensing of Environment, 2008. **112**: p. 1689-1701.
  
79. Maisongrande, P., Duchemin, B. and Dedieu, G., *VEGETATION/SPOT: An operational mission for the Earth monitoring; presentation of new standard products*. International Journal of Remote Sensing, 2004. **25**: p. 9-14.
  
80. Adler-Golden, S.M., Matthew, M.W., Bernstein, L.S., Levine, R.Y., Berk, A., Richtsmeier, S.C., Acharya, P.K., Anderson, G.P., Felde, G., Gardner, J., Hoke, M., Jeong, L.S., Pukall, B., Mello, J., Ratkowski, A. and Burke, H.-H., *Atmospheric correction for shortwave spectral imagery based on MODTRAN4*. Imaging Spectrometry V, 1999. **3753**.
  
81. Kruse, F.A., *Comparison of ATREM, ACORN, and FLAASH Atmospheric Corrections using Low-Altitude AVIRIS Data of Boulder, Colorado*, in *Proceedings 13th JPL Airborne Geoscience Workshop, Pasadena, CA, USA*, March 2004.
  
82. Matthew, M., Adler-Golden, S., Berk, A., Felde, G., Anderson, G., Gorodetzky, D., Paswaters, S. and Shippert, M., *Atmospheric correction of spectral imagery - Evaluation of the FLAASH algorithm with AVIRIS data*, in *Applied Imagery Pattern Recognition Workshop, 2002. Proceedings. 31st*, 16-17 Oct. 2002, p. 157-163.
  
83. ACORN 4.0 User's Guide Stand-alone Version, Version 4.0, January 2002. Analytical Imaging and Geophysics LLC.
  
84. Richter, R. and Schlaepfer, D., *Geo-atmospheric processing of airborne imaging spectrometry data. Part 2: atmospheric/topographic correction*. International Journal of Remote Sensing, 2002. **23**: p. 2631-2649.
  
85. ATCOR-2/3 User Guide (Ver. 7.1), 2010. DLR-German Aerospace Center, Remote Sensing Data Center.

86. San, B.T. and Süzen, M.L., *Evaluation of different atmospheric correction algorithms for EO-1 HYPERION imagery*, in *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Science*, 2010, vol. XXXVIII, p. 392-397.
87. Vermote, E., Tanre, D., Deuze, J., Herman, M. and Morcette, J.-J., *Second Simulation of the Satellite Signal in the Solar Spectrum, 6S: An overview*. IEEE Transactions on Geoscience and Remote Sensing, 1997. **35**: p. 675–686.
88. Kotchenova, S.Y. and Vermote, E.F., *Validation of a vector version of the 6S radiative transfer code for atmospheric correction of satellite data. Part II: Homogeneous Lambertian and anisotropic surfaces*. Applied Optics, 2007. **46**: p. 4455-4464.
89. Kotchenova, S.Y., Vermote, E.F., Matarrese, R. and Klemm, F.J., *Validation of a vector version of the 6S radiative transfer code for atmospheric correction of satellite data. Part I: Path radiance*. Applied Optics, 2006. **45**: p. 6762-6774.
90. Castellana, L., D'Addabbo, A. and Pasquariello, G., *A composed supervised/unsupervised approach to improve change detection from remote sensing*. Pattern Recognition Letters, 2007. **28**: p. 405-413.
91. Thomas, N., Hendrix, C. and Congalton, R.G., *A comparison of urban mapping methods using high-resolution digital imagery*. Photogrammetric engineering and remote sensing, 2003. **69**: p. 963-972.
92. Ines, A.V. and Honda, K., *On quantifying agricultural and water management practices from low spatial resolution RS data using genetic algorithms: A numerical study for mixed-pixel environment*. Advances in Water Resources, 2005. **28**: p. 856-870.
93. Friedl, M.A. and Brodley, C.E., *Decision tree classification of land cover from remotely sensed data*. Remote sensing of environment, 1997. **61**: p. 399-409.
94. Friedl, M.A., Brodley, C.E. and Strahler, A.H., *Maximizing land cover classification accuracies produced by decision trees at continental to*

- global scales*. Geoscience and Remote Sensing, IEEE Transactions on, 1999. **37**: p. 969-977.
95. Mercier, G. and Lennon, M., *Support vector machines for hyperspectral image classification with spectral-based kernels*, in *Geoscience and Remote Sensing Symposium, 2003. IGARSS'03. Proceedings. 2003 IEEE International*, vol. 1, p. 288-290.
  96. Pal, M. and Mather, P., *Support vector machines for classification in remote sensing*. International Journal of Remote Sensing, 2005. **26**: p. 1007-1011.
  97. Bischof, H., Schneider, W. and Pinz, A.J., *Multispectral classification of Landsat-images using neural networks*. Geoscience and Remote Sensing, IEEE Transactions on, 1992. **30**: p. 482-490.
  98. Civco, D.L., *Artificial neural networks for land-cover classification and mapping*. International Journal of Geographical Information Science, 1993. **7**: p. 173-186.
  99. Foody, G.M., *Approaches for the production and evaluation of fuzzy land cover classifications from remotely-sensed data*. International Journal of Remote Sensing, 1996. **17**: p. 1317-1340.
  100. Chanussot, J., Benediktsson, J.A. and Fauvel, M., *Classification of remote sensing images from urban areas using a fuzzy possibilistic model*. Geoscience and Remote Sensing Letters, IEEE, 2006. **3**: p. 40-44.
  101. Cortijo, F. and De La Blanca, N.P., *The performance of regularized discriminant analysis versus non-parametric classifiers applied to high-dimensional image classification*. International journal of remote sensing, 1999. **20**: p. 3345-3365.
  102. Raudys, S., *On dimensionality, sample size, and classification error of nonparametric linear classification algorithms*. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 1997. **19**: p. 667-671.



103. Sjöberg, J., Zhang, Q., Ljung, L., Benveniste, A., Delyon, B., Glorennec, P.-Y., Hjalmarsson, H. and Juditsky, A., *Nonlinear black-box modeling in system identification: a unified overview*. Automatica, 1995. **31**: p. 1691-1724.
104. Qu, J.J., Gao, W., Kafatos, M., Murphy, R.E. and Salomonson, V.V., *Earth science satellite remote sensing. Volume 1: Science and Instruments*. 2006, Beijing: Springer.
105. Salomonson, V.V., Barnes, W., Maymon, P.W., Montgomery, H.E. and Ostrow, H., *MODIS: advanced facility instrument for studies of the Earth as a system*. Geoscience and Remote Sensing, IEEE Transactions on, 1989. **27**: p. 145-153.
106. MODIS level 1B product user's guide, 2009. NASA/Goddard Space Flight Center. [cited 2013, 29 October]; Available from: <http://ccplot.org/pub/resources/Aqua/MODIS%20Level%201B%20Product%20User%20Guide.pdf>.
107. MODIS calibration product information. [cited 2014, 5 March]; Available from: <http://mcst.gsfc.nasa.gov/>.
108. MODIS atmospheric products information. [cited 2014, 5 March]; Available from: <http://modis-atmos.gsfc.nasa.gov/>.
109. MODIS land products information - 1. [cited 2014, 5 March]; Available from: <http://modis-land.gsfc.nasa.gov/>.
110. MODIS land products information - 2. [cited 2014, 5 March]; Available from: <https://lpdaac.usgs.gov/>.
111. MODIS cryosphere product information. [cited 2014, 5 March]; Available from: <http://nsidc.org/data/modis/index.html>.
112. MODIS ocean product information. [cited 2014, 5 March]; Available from: <http://oceancolor.gsfc.nasa.gov/>.

113. Vermote, E.F., Kotchenova, S.Y. and Ray, J.P. *MODIS Surface Reflectance User's Guide (Ver. 1.3)*. 2011 [cited 2012, 10 November]; Available from: [http://dratmos.geog.umd.edu/products/MOD09\\_UserGuide\\_v1\\_3.pdf](http://dratmos.geog.umd.edu/products/MOD09_UserGuide_v1_3.pdf).
114. HDF4 User's Guide, February 2010. The HDF Group. [cited 2013, 29 October]; Available from: [http://www.hdfgroup.org/ftp/HDF/Documentation/HDF4.2.5/HDF425\\_UserGd.pdf](http://www.hdfgroup.org/ftp/HDF/Documentation/HDF4.2.5/HDF425_UserGd.pdf).
115. HDF-EOS Library User's Guide for the ECS Project, Volume 1: Overview and Examples, June 1999. Raytheon Systems Company, Upper Marlboro, Maryland, USA. [cited 2013, 23 April]; Available from: <http://edhs1.gsfc.nasa.gov/waisdata/sdp/pdf/tp17050001.pdf>.
116. MODIS Level 1A Earth Location: Algorithm Theoretical Basis Document Version 3.0, 1997. [cited 2012, 19 May]; Available from: [http://modis.gsfc.nasa.gov/data/atbd/atbd\\_mod28\\_v3.pdf](http://modis.gsfc.nasa.gov/data/atbd/atbd_mod28_v3.pdf).
117. Xiong, X., Isaacman, A. and Barnes, W., *MODIS Level-1B Products*, in *Earth Science Satellite Remote Sensing*, J. Qu, et al., Editors. 2006, Springer Berlin Heidelberg. p. 33-49.
118. Parkinson, C.L. and Greenstone, R., eds. *EOS Data Products Handbook Volume 2*. October 2000: NASA Goddard Space Flight Center Greenbelt, Maryland.
119. Ahmad, S.P., King, M.D., Koziara, C.V., Leptoukh, G.G., Serafino, G.N. and Shams, S., *MODIS Cloud, Aerosol, and Water Vapor Products for Climate and Global Change Studies*, in *13th Symposium on Global Change and Climate Variations*, 13-17 January 2002.
120. Hubanks, P.A., King, M.D., Platnick, S. and Pincus, R. *MODIS atmosphere L3 gridded product algorithm theoretical basis document (Collection 005 Version 1.1)*. 2008 [cited 2012, 29 October]; Available from: [http://modis-atmos.gsfc.nasa.gov/docs/L3\\_ATBD\\_2008\\_12\\_04.pdf](http://modis-atmos.gsfc.nasa.gov/docs/L3_ATBD_2008_12_04.pdf).
121. King, M.D., Menzel, W.P., Kaufman, Y.J., Tanre, D., Gao, B.-C., Platnick, S., Ackerman, S.A., Remer, L.A., Pincus, R. and Hubanks,

P.A., *Cloud and aerosol properties, precipitable water, and profiles of temperature and water vapor from MODIS*. Geoscience and Remote Sensing, IEEE Transactions on, 2003. **41**: p. 442-458.

122. Hall, D.K., Riggs, G.A., Salomonson, V.V., DiGirolamo, N.E. and Bayr, K.J., *MODIS snow-cover products*. Remote Sensing of Environment, 2002. **83**: p. 181-194.
123. Riggs, G.A., Hall, D.K. and Salomonson, V.V. *MODIS Snow Products User Guide to Collection 5*. 2006 [cited 2012, 29 October]; Available from: [http://modis-snow-ice.gsfc.nasa.gov/uploads/sug\\_c5.pdf](http://modis-snow-ice.gsfc.nasa.gov/uploads/sug_c5.pdf).
124. Tekeli, A.E., Akyürek, Z., Şorman, A.A., Şensoy, A. and Şorman, Ü., *Using MODIS snow cover maps in modeling snowmelt runoff process in the eastern part of Turkey*. Remote Sensing of Environment, 2005. **97**: p. 216-230.
125. Salomonson, V.V. and Appel, I., *Development of the Aqua MODIS NDSI Fractional Snow Cover Algorithm and Validation Results*. IEEE Transactions on Geoscience and Remote Sensing, 2006. **44**: p. 1747-1756.
126. MATLAB<sup>®</sup>, R2012b (8.0.0.783). The MathWorks, Inc.
127. Cameron, A.C. and Windmeijer, F.A.G., *An R-squared measure of goodness of fit for some common nonlinear regression models*. Journal of Econometrics, 1997. **77**: p. 329-342.
128. MARS<sup>®</sup>, Salford Systems (software available at <http://www.salfordsystems.com>).
129. SPSS, Released 2006. SPSS for Windows, Version 15.0. Chicago, SPSS Inc. .
130. Ho, R., *Handbook of univariate and multivariate data analysis and interpretation with SPSS*. 2006: CRC Press.
131. MOSEK<sup>™</sup>, (software available at <http://www.mosek.com>).

132. Osborne, J.W., *Improving your data transformations: Applying the Box-Cox transformation*. Practical Assessment, Research & Evaluation, 2010. **15**: p. 1-9.
133. Minitab®, Version 16.2.4. 2013 Minitab Inc.
134. Eldridge, R.G., *Water vapor absorption of visible and near infrared radiation*. Applied optics, 1967. **6**: p. 709-713.
135. Muñoz, J. and Felicísimo, Á.M., *Comparison of statistical methods commonly used in predictive modelling*. Journal of Vegetation Science, 2004. **15**: p. 285-292.
136. Dwinnell, W., *Exploring MARS: an alternative to neural networks*. PC AI, 2000. **14**: p. 21-24.
137. Francis, L., *Martian Chronicles: Is MARS Better than Neural Networks?*, in *Casualty Actuarial Society Forum*, p. 75-102.
138. Bradley, A.P., *The use of the area under the ROC curve in the evaluation of machine learning algorithms*. Pattern Recognition, 1997. **30**: p. 1145-1159.
139. Kürüm, E., Yildirak, K. and Weber, G.-W., *A classification problem of credit risk rating investigated and solved by optimisation of the ROC curve*. Central European Journal of Operations Research, 2012. **20**: p. 529-557.
140. Li, Z., Li, J., Menzel, W.P., Schmit, T.J. and Ackerman, S.A., *Comparison between current and future environmental satellite imagers on cloud classification using MODIS*. Remote Sensing of Environment, 2007. **108**: p. 311-326.
141. ArcMap, ESRI ArcMap Version 9.3.1. 1999-2009 ESRI Inc.
142. Milborrow, S., *earth: Multivariate Adaptive Regression Spline Models - Derived from mda:mars by Trevor Hastie and Rob Tibshirani*. R

package version 3.2-2. Available from: <http://CRAN.R-project.org/package=earth>, 2012.

143. R Development Core Team, *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, 2012. Available from: <http://www.R-project.org/>.
144. Congalton, R.G., *A Review of Assessing the Accuracy of Classifications of Remotely Sensed Data*. Remote Sensing of Environment, 1991. **37**: p. 35-46.
145. Miao, L. and Qi, H., *Endmember extraction from highly mixed data using minimum volume constrained nonnegative matrix factorization*. Geoscience and Remote Sensing, IEEE Transactions on, 2007. **45**: p. 765-777.
146. Theseira, M., Thomas, G., Taylor, J., Gemmell, F. and Varjo, J., *Sensitivity of mixture modelling to end-member selection*. International Journal of Remote Sensing, 2003. **24**: p. 1559-1575.
147. Plaza, A., Martínez, P., Pérez, R. and Plaza, J., *A quantitative and comparative analysis of endmember extraction algorithms from hyperspectral data*. Geoscience and Remote Sensing, IEEE Transactions on, 2004. **42**: p. 650-663.
148. Melgani, F., Al Hashemy, B.A. and Taha, S.M., *An explicit fuzzy supervised classification method for multispectral remote sensing images*. Geoscience and Remote Sensing, IEEE Transactions on, 2000. **38**: p. 287-295.
149. Foody, G.M. and Mathur, A., *The use of small training sets containing mixed pixels for accurate hard image classification: Training on mixed spectral responses for classification by a SVM*. Remote Sensing of Environment, 2006. **103**: p. 179-189.
150. Liang, S., Fang, H., Morissette, J.T., Chen, M., Shuey, C.J., Walthall, C. and Daughtry, C.S., *Atmospheric correction of Landsat ETM+ land surface imagery. II. Validation and applications*. Geoscience and Remote Sensing, IEEE Transactions on, 2002. **40**: p. 2736-2746.

151. Kaufman, Y.J., *The atmospheric effect on the separability of field classes measured from satellites*. Remote Sensing of Environment, 1985. **18**: p. 21-34.
152. Kaufman, Y.J., *Atmospheric effect on spectral signature—measurements*. Advances in Space Research, 1987. **7**: p. 203-206.
153. Kaufman, Y.J., *Atmospheric effect on spectral signature-measurements and corrections*. Geoscience and Remote Sensing, IEEE Transactions on, 1988. **26**: p. 441-450.
154. Özmen, A., Weber, G.-W., Batmaz, İ. and Kropat, E., *RCMARS: Robustification of CMARS with different scenarios under polyhedral uncertainty set*. Communications in Nonlinear Science and Numerical Simulation, 2011. **16**: p. 4780-4787.
155. Hansen, P.C., *Rank-Deficient and Discrete Ill-Posed Problems: Numerical Aspects of Linear Inversion*. 1998, Philadelphia: SIAM.
156. Andersen, E.D., Roos, C. and Terlaky, T., *On implementing a primal-dual interior-point method for conic quadratic optimization*. Mathematical Programming, 2003. **95**: p. 249-277.
157. Ben-Tal, A. and Nemirovski, A., *Lectures on modern convex optimization: analysis, algorithms, and engineering applications*. Vol. 2. 2001: Siam.
158. Lobo, M.S., Vandenberghe, L., Boyd, S. and Lebret, H., *Applications of second-order cone programming*. Linear algebra and its applications, 1998. **284**: p. 193-228.
159. Neter, J., Kutner, M.H., Nachtsheim, C.J. and Wasserman, W., *Applied linear statistical models*. Vol. 4. 1996: Irwin Chicago.
160. SMAC Code. [cited 2014, 5 March]; Available from: <http://www.cesbio.ups-tlse.fr/us/serveurs4.htm>.

## APPENDIX A

### TIKHONOV REGULARIZATION

*Tikhonov regularization* (TR) is the most common and well-known form to make ill-posed problems regular and stable [155]. Tikhonov solution can be expressed easily in terms of the *singular value decomposition* (SVD) of the coefficient matrix  $\mathbf{A}$  of regarded linear systems of equations  $\mathbf{Ax} - \mathbf{b}$ , where  $\mathbf{A}$  is an ill-conditioned  $(N \times m)$ -matrix. The standard approach to approximately solve this system of equations is known as *least squares estimation* (LSE). It seeks to minimize the residual  $\|\mathbf{Ax} - \mathbf{b}\|_2^2$ . To solve different kinds of TR problem, we use SVD to have a solution that minimizes the objective function including  $\|\mathbf{x}\|_2$ . However, in many cases, it is preferred to achieve a solution that minimizes some other measure of  $\mathbf{x}$ , such as the norm of first- or second-order derivatives. They are, in approximate sense, given by first- or second-order difference quotients of  $\mathbf{x}$ , considered as a function that is evaluated at the points  $k$  and  $k+1$ . These difference quotients approximate those derivatives; altogether, they are comprised by products  $\mathbf{Lx}$  of  $\mathbf{x}$  with some suitably chosen Tikhonov matrices  $\mathbf{L}$ . These matrices represent the discrete differential operators of first- and second-order, respectively. Hereby, the optimization problem is in the following form:

$$\underset{\mathbf{x}}{\text{minimize}} \quad \|\mathbf{Ax} - \mathbf{b}\|_2^2 + \lambda^2 \|\mathbf{Lx}\|_2^2.$$

Generally, the above equation consists of high order TR problems, and generalized SVD is used to solve them. In many situations, to obtain a solution which minimizes some other measure  $\mathbf{x}$ , the norm of the first- or second-derivative is preferred.





## APPENDIX B

### CONIC QUADRATIC PROGRAMMING

Conic quadratic optimization is defined as the problem of minimizing a linear objective function subject to the intersection of an affine set and the direct product of quadratic cones of the form

$$\left\{ \mathbf{x} \in \mathbb{R}^{m+1} \mid x_{m+1}^2 \geq \sum_{j=1}^m x_j^2 \right\}.$$

Many optimization problems may be created in the form of quadratic cone, which is also known as the second-order (i.e., Lorentz or ice-cream) cone. Linear, convex quadratic and convex quadratically constrained quadratic optimization are some examples. The problem of minimizing the sum of norms and robust linear optimization can also be given as examples [156]. In [157, 158], different implementations of conic quadratic optimization can be found.

A CQP problem, a conic problem, can be given in the following form [157]:

$$\begin{aligned} & \underset{\mathbf{x}}{\text{minimize}} && \mathbf{c}^T \mathbf{x} \\ & \text{subject to} && \mathbf{Ax} - \mathbf{b} \succeq_K \mathbf{0}, \end{aligned}$$

for which the cone  $K$  is the direct product of several Lorentz cones. Therefore, the following expression can be used to represent  $K$ :

$$K = L^m \times \dots \times L^{m_r} \subseteq E.$$

In general, a CQP problem is an optimization problem with linear objective function and finitely many “ice-cream constraints”,  $A_i \mathbf{x} - \mathbf{b}_i \geq_{L^m} \mathbf{0}$  ( $i = 1, 2, \dots, r$ ).

Therefore, a CQP problem can be expressed in the following form [157]:

$$\begin{aligned} & \underset{\mathbf{x}}{\text{minimize}} && \mathbf{c}^T \mathbf{x} \\ & \text{subject to} && A_i \mathbf{x} - \mathbf{b}_i \geq_{L^m} \mathbf{0} \quad (i = 1, 2, \dots, r). \end{aligned}$$

If the data matrix,  $[A_i, \mathbf{b}_i]$ , is subdivided as follows:

$$[A_i, \mathbf{b}_i] = \begin{bmatrix} \mathbf{D}_i & \mathbf{d}_i \\ \mathbf{p}_i^T & q_i \end{bmatrix},$$

where  $\mathbf{D}_i$  is of the size  $(m_i - 1) \times \dim x$ , the problem can be written as follows:

$$\begin{aligned} & \underset{\mathbf{x}}{\text{minimize}} && \mathbf{c}^T \mathbf{x}, \\ & \text{subject to} && \|\mathbf{D}_i \mathbf{x} - \mathbf{d}_i\|_2 \leq \mathbf{p}_i^T \mathbf{x} - q_i \quad (i = 1, 2, \dots, r). \end{aligned}$$

This form is the most explicit one, in which  $\mathbf{D}_i$  indicate matrices of the same row dimensions as  $\mathbf{x}$ ,  $\mathbf{d}_i$  represent vectors of the same dimensions as the column dimensions of the matrices  $\mathbf{D}_i$ ,  $\mathbf{p}_i$  are vectors of the same dimensions as  $\mathbf{x}$ , and  $q_i$  indicate real numbers [157].

## APPENDIX C

### MARS AND CMARS MODELS

The notation used here is as follows:  $X_1$  and  $X_2$  are geographic latitude and longitude (i.e.,  $\lambda$  and  $\phi$ ), respectively,  $X_3$  is the TOA reflectance value (i.e.,  $\kappa$ ), and finally,  $X_4$  denotes the month indicator (i.e.,  $\zeta$ ).

#### C.1 MARS and CMARS Models for Alps

- Basis functions generated in the forward step of MARS algorithm:

$$\begin{aligned} B_1(\mathbf{x}) &= \max\{0, X_3 - 0.2331\}, & B_2(\mathbf{x}) &= \max\{0, 0.2331 - X_3\}, \\ B_3(\mathbf{x}) &= \max\{0, X_4 - 6.0001\}, & B_4(\mathbf{x}) &= \max\{0, 6.0001 - X_4\}, \\ B_5(\mathbf{x}) &= \max\{0, X_4 - 6.0001\} \cdot \max\{0, 0.2331 - X_3\}, \\ B_6(\mathbf{x}) &= \max\{0, 6.0001 - X_4\} \cdot \max\{0, 0.2331 - X_3\}, \\ B_7(\mathbf{x}) &= \max\{0, X_4 - 9.0001\}, & B_8(\mathbf{x}) &= \max\{0, 9.0001 - X_4\}, \\ B_9(\mathbf{x}) &= \max\{0, X_4 - 3.0001\}, & B_{10}(\mathbf{x}) &= \max\{0, 3.0001 - X_4\}, \\ B_{11}(\mathbf{x}) &= \max\{0, X_4 - 2.0001\}, & B_{12}(\mathbf{x}) &= \max\{0, 2.0001 - X_4\}, \\ B_{13}(\mathbf{x}) &= \max\{0, X_3 - 0.14383\} \cdot \max\{0, X_4 - 9.0001\}, \\ B_{14}(\mathbf{x}) &= \max\{0, 0.14383 - X_3\} \cdot \max\{0, X_4 - 9.0001\}, \\ B_{15}(\mathbf{x}) &= \max\{0, X_2 - 49.8934\}, & B_{16}(\mathbf{x}) &= \max\{0, 49.8934 - X_2\}, \\ B_{17}(\mathbf{x}) &= \max\{0, X_4 - 3.0001\} \cdot \max\{0, 49.8934 - X_2\}, \\ B_{18}(\mathbf{x}) &= \max\{0, 3.0001 - X_4\} \cdot \max\{0, 49.8934 - X_2\}, \\ B_{19}(\mathbf{x}) &= \max\{0, X_1 - 2.8508\} \cdot \max\{0, X_4 - 6.0001\}, \\ B_{20}(\mathbf{x}) &= \max\{0, 2.8508 - X_1\} \cdot \max\{0, X_4 - 6.0001\}, \\ B_{21}(\mathbf{x}) &= \max\{0, X_3 - 0.1783\} \cdot \max\{0, 2.8508 - X_1\} \cdot \max\{0, X_4 - 6.0001\}, \\ B_{22}(\mathbf{x}) &= \max\{0, 0.1783 - X_3\} \cdot \max\{0, 2.8508 - X_1\} \cdot \max\{0, X_4 - 6.0001\}, \\ B_{23}(\mathbf{x}) &= \max\{0, X_4 - 4.0001\} \cdot \max\{0, X_3 - 0.2331\}, \\ B_{24}(\mathbf{x}) &= \max\{0, 4.0001 - X_4\} \cdot \max\{0, X_3 - 0.2331\}, \\ B_{25}(\mathbf{x}) &= \max\{0, X_1 - 11.8273\} \cdot \max\{0, X_4 - 4.0001\} \cdot \max\{0, X_3 - 0.2331\}, \end{aligned}$$

$$\begin{aligned}
B_{26}(\mathbf{x}) &= \max\{0, 11.8273 - X_1\} \cdot \max\{0, X_4 - 4.0001\} \cdot \max\{0, X_3 - 0.2331\}, \\
B_{27}(\mathbf{x}) &= \max\{0, X_1 - 3.0547\} \cdot \max\{0, 6.0001 - X_4\}, \\
B_{28}(\mathbf{x}) &= \max\{0, 3.0547 - X_1\} \cdot \max\{0, 6.0001 - X_4\}, \\
B_{29}(\mathbf{x}) &= \max\{0, X_3 - 0.2325\} \cdot \max\{0, X_4 - 3.0001\}, \\
B_{30}(\mathbf{x}) &= \max\{0, 0.2325 - X_3\} \cdot \max\{0, X_4 - 3.0001\}, \\
B_{31}(\mathbf{x}) &= \max\{0, X_1 - 13.1440\} \cdot \max\{0, 3.0001 - X_4\} \cdot \max\{0, 49.8934 - X_2\}, \\
B_{32}(\mathbf{x}) &= \max\{0, 13.1440 - X_1\} \cdot \max\{0, 3.0001 - X_4\} \cdot \max\{0, 49.8934 - X_2\}, \\
B_{33}(\mathbf{x}) &= \max\{0, X_2 - 41.2841\} \cdot \max\{0, X_1 - 2.8508\} \cdot \max\{0, X_4 - 6.0001\}, \\
B_{34}(\mathbf{x}) &= \max\{0, 41.2841 - X_2\} \cdot \max\{0, X_1 - 2.8508\} \cdot \max\{0, X_4 - 6.0001\}, \\
B_{35}(\mathbf{x}) &= \max\{0, X_1 - 3.1696\} \cdot \max\{0, 6.0001 - X_4\} \cdot \max\{0, 0.2331 - X_3\}, \\
B_{36}(\mathbf{x}) &= \max\{0, 3.1696 - X_1\} \cdot \max\{0, 6.0001 - X_4\} \cdot \max\{0, 0.2331 - X_3\}, \\
B_{37}(\mathbf{x}) &= \max\{0, X_3 - 0.0906\} \cdot \max\{0, 3.0001 - X_4\} \cdot \max\{0, 49.8934 - X_2\}, \\
B_{38}(\mathbf{x}) &= \max\{0, 0.0906 - X_3\} \cdot \max\{0, 3.0001 - X_4\} \cdot \max\{0, 49.8934 - X_2\}, \\
B_{39}(\mathbf{x}) &= \max\{0, X_4 - 2.0001\} \cdot \max\{0, 0.2331 - X_3\}, \\
B_{40}(\mathbf{x}) &= \max\{0, 2.0001 - X_4\} \cdot \max\{0, 0.2331 - X_3\}.
\end{aligned}$$

- MARS model obtained after the backward step:

$$\begin{aligned}
\hat{Y} &= \beta_0 + \sum_{m=1}^M \beta_m B_m(\mathbf{x}) \\
&= 0.2053 + 1.538B_1(\mathbf{x}) - 1.1465B_2(\mathbf{x}) - 0.0792B_3(\mathbf{x}) + 0.1342B_4(\mathbf{x}) + 0.2270B_5(\mathbf{x}) \\
&\quad - 0.5909B_6(\mathbf{x}) + 0.0902B_7(\mathbf{x}) + 0.3106B_9(\mathbf{x}) - 0.2191B_{11}(\mathbf{x}) \\
&\quad + 0.3591B_{13}(\mathbf{x}) - 0.5781B_{14}(\mathbf{x}) - 0.6562B_{15}(\mathbf{x}) \\
&\quad - 0.0003B_{17}(\mathbf{x}) - 0.0164B_{18}(\mathbf{x}) - 0.0008B_{19}(\mathbf{x}) \\
&\quad - 0.0044B_{20}(\mathbf{x}) - 0.1813B_{21}(\mathbf{x}) + 0.022B_{22}(\mathbf{x}) \\
&\quad + 0.5208B_{23}(\mathbf{x}) + 0.5506B_{24}(\mathbf{x}) + 0.0075B_{25}(\mathbf{x}) \\
&\quad + 0.0078B_{26}(\mathbf{x}) - 0.0213B_{28}(\mathbf{x}) - 0.5072B_{29}(\mathbf{x}) \\
&\quad - 1.0255B_{30}(\mathbf{x}) + 0.0083B_{31}(\mathbf{x}) + 0.0019B_{32}(\mathbf{x}) \\
&\quad + 0.0002B_{33}(\mathbf{x}) + 0.0072B_{35}(\mathbf{x}) + 0.1310B_{36}(\mathbf{x}) \\
&\quad - 0.1419B_{37}(\mathbf{x}) - 0.0524B_{38}(\mathbf{x}) + 0.7232B_{39}(\mathbf{x}).
\end{aligned}$$

- CMARS model:

$$\begin{aligned}
\hat{Y} &= \beta_0 + \sum_{m=1}^{M_{\max}} \beta_m B_m(\mathbf{x}) \\
&= 0.1660 + 1.5154B_1(\mathbf{x}) - 1.1652B_2(\mathbf{x}) + 0.0193B_3(\mathbf{x}) \\
&\quad + 0.0375B_4(\mathbf{x}) + 0.4728B_5(\mathbf{x}) - 0.8445B_6(\mathbf{x}) \\
&\quad + 0.0623B_7(\mathbf{x}) + 0.0279B_8(\mathbf{x}) + 0.0740B_9(\mathbf{x}) \\
&\quad + 0.2362B_{10}(\mathbf{x}) - 0.0528B_{11}(\mathbf{x}) - 0.1648B_{12}(\mathbf{x}) \\
&\quad + 0.3591B_{13}(\mathbf{x}) - 0.5786B_{14}(\mathbf{x}) - 0.6539B_{15}(\mathbf{x}) \\
&\quad + 0.0002B_{16}(\mathbf{x}) - 0.0004B_{17}(\mathbf{x}) - 0.0163B_{18}(\mathbf{x}) \\
&\quad - 0.0008B_{19}(\mathbf{x}) - 0.0044B_{20}(\mathbf{x}) - 0.1816B_{21}(\mathbf{x}) \\
&\quad + 0.0223B_{22}(\mathbf{x}) + 0.4976B_{23}(\mathbf{x}) + 0.5700B_{24}(\mathbf{x}) \\
&\quad + 0.0075B_{25}(\mathbf{x}) + 0.0078B_{26}(\mathbf{x}) - 0.0002B_{27}(\mathbf{x}) \\
&\quad - 0.0217B_{28}(\mathbf{x}) - 0.4842B_{29}(\mathbf{x}) - 1.0259B_{30}(\mathbf{x}) \\
&\quad + 0.0084B_{31}(\mathbf{x}) + 0.0019B_{32}(\mathbf{x}) + 0.0002B_{33}(\mathbf{x}) \\
&\quad + 0.0001B_{34}(\mathbf{x}) + 0.0082B_{35}(\mathbf{x}) + 0.1331B_{36}(\mathbf{x}) \\
&\quad - 0.1448B_{37}(\mathbf{x}) - 0.0485B_{38}(\mathbf{x}) + 0.4781B_{39}(\mathbf{x}) + 0.2443B_{40}(\mathbf{x}).
\end{aligned}$$

## C.2 MARS and CMARS Models for Turkey

- Basis functions generated in the forward step of MARS algorithm:

$$\begin{aligned}
B_1(\mathbf{x}) &= \max\{0, X_3 - 0.7039\}, & B_2(\mathbf{x}) &= \max\{0, 0.7039 - X_3\}, \\
B_3(\mathbf{x}) &= \max\{0, X_4 - 6.0001\}, & B_4(\mathbf{x}) &= \max\{0, 6.0001 - X_4\}, \\
B_5(\mathbf{x}) &= \max\{0, X_4 - 6.0001\} \cdot \max\{0, 0.7039 - X_3\}, \\
B_6(\mathbf{x}) &= \max\{0, 6.0001 - X_4\} \cdot \max\{0, 0.7039 - X_3\}, \\
B_7(\mathbf{x}) &= \max\{0, X_3 - 0.1591\}, & B_8(\mathbf{x}) &= \max\{0, 0.1591 - X_3\}, \\
B_9(\mathbf{x}) &= \max\{0, X_1 - 33.0836\} \cdot \max\{0, 6.0001 - X_4\}, \\
B_{10}(\mathbf{x}) &= \max\{0, 33.0836 - X_1\} \cdot \max\{0, 6.0001 - X_4\}, \\
B_{11}(\mathbf{x}) &= \max\{0, X_4 - 8.0001\}, & B_{12}(\mathbf{x}) &= \max\{0, 8.0001 - X_4\}, \\
B_{13}(\mathbf{x}) &= \max\{0, X_2 - 39.5459\} \cdot \max\{0, X_3 - 0.7039\}, \\
B_{14}(\mathbf{x}) &= \max\{0, 39.5459 - X_2\} \cdot \max\{0, X_3 - 0.7039\}, \\
B_{15}(\mathbf{x}) &= \max\{0, X_3 - 0.3128\} \cdot \max\{0, 8.0001 - X_4\}, \\
B_{16}(\mathbf{x}) &= \max\{0, 0.3128 - X_3\} \cdot \max\{0, 8.0001 - X_4\},
\end{aligned}$$

$$\begin{aligned}
B_{17}(\mathbf{x}) &= \max\{0, X_2 - 40.8146\} \cdot \max\{0, 0.1591 - X_3\}, \\
B_{18}(\mathbf{x}) &= \max\{0, 40.8146 - X_2\} \cdot \max\{0, 0.1591 - X_3\}, \\
B_{19}(\mathbf{x}) &= \max\{0, X_2 - 40.0364\} \cdot \max\{0, X_3 - 0.7039\}, \\
B_{20}(\mathbf{x}) &= \max\{0, 40.0364 - X_2\} \cdot \max\{0, X_3 - 0.7039\}, \\
B_{21}(\mathbf{x}) &= \max\{0, X_1 - 33.3572\} \cdot \max\{0, X_4 - 6.0001\}, \\
B_{22}(\mathbf{x}) &= \max\{0, 33.3572 - X_1\} \cdot \max\{0, X_4 - 6.0001\}, \\
B_{23}(\mathbf{x}) &= \max\{0, X_3 - 0.8951\}, \\
B_{24}(\mathbf{x}) &= \max\{0, 0.8951 - X_3\}, \\
B_{25}(\mathbf{x}) &= \max\{0, X_4 - 4.0001\}, \\
B_{26}(\mathbf{x}) &= \max\{0, 4.0001 - X_4\}, \\
B_{27}(\mathbf{x}) &= \max\{0, X_3 - 0.3159\} \cdot \max\{0, 4.0001 - X_4\}, \\
B_{28}(\mathbf{x}) &= \max\{0, 0.3159 - X_3\} \cdot \max\{0, 4.0001 - X_4\}, \\
B_{29}(\mathbf{x}) &= \max\{0, X_2 - 39.5027\} \cdot \max\{0, X_3 - 0.7039\}, \\
B_{30}(\mathbf{x}) &= \max\{0, 39.5027 - X_2\} \cdot \max\{0, X_3 - 0.7039\}, \\
B_{31}(\mathbf{x}) &= \max\{0, X_1 - 34.9980\} \cdot \max\{0, X_3 - 0.7039\}, \\
B_{32}(\mathbf{x}) &= \max\{0, 34.9980 - X_1\} \cdot \max\{0, X_3 - 0.7039\}, \\
B_{33}(\mathbf{x}) &= \max\{0, X_1 - 35.9806\} \cdot \max\{0, X_3 - 0.7039\}, \\
B_{34}(\mathbf{x}) &= \max\{0, 35.9806 - X_1\} \cdot \max\{0, X_3 - 0.7039\}, \\
B_{35}(\mathbf{x}) &= \max\{0, X_1 - 35.8098\} \cdot \max\{0, X_3 - 0.7039\}, \\
B_{36}(\mathbf{x}) &= \max\{0, 35.8098 - X_1\} \cdot \max\{0, X_3 - 0.7039\}, \\
B_{37}(\mathbf{x}) &= \max\{0, X_1 - 37.4083\} \cdot \max\{0, 8.0001 - X_4\}, \\
B_{38}(\mathbf{x}) &= \max\{0, 37.4083 - X_1\} \cdot \max\{0, 8.0001 - X_4\}, \\
B_{39}(\mathbf{x}) &= \max\{0, X_2 - 40.6531\} \cdot \max\{0, 4.0001 - X_4\}, \\
B_{40}(\mathbf{x}) &= \max\{0, 40.6531 - X_2\} \cdot \max\{0, 4.0001 - X_4\}.
\end{aligned}$$

- MARS model obtained after the backward step:

$$\begin{aligned}
\hat{Y} &= \beta_0 + \sum_{m=1}^M \beta_m B_m(\mathbf{x}) \\
&= 0.4528 + 3.4737B_1(\mathbf{x}) - 0.6470B_2(\mathbf{x}) + 0.09075B_3(\mathbf{x}) + 0.0447B_4(\mathbf{x}) - 0.3631B_5(\mathbf{x}) \\
&\quad + 0.1670B_6(\mathbf{x}) - 0.1715B_7(\mathbf{x}) - 0.0011B_9(\mathbf{x}) - 0.0036B_{10}(\mathbf{x}) + 0.0867B_{11}(\mathbf{x}) \\
&\quad + 53.3818B_{13}(\mathbf{x}) + 0.2635B_{15}(\mathbf{x}) - 0.3062B_{16}(\mathbf{x}) \\
&\quad - 0.1083B_{17}(\mathbf{x}) + 0.0535B_{18}(\mathbf{x}) - 2.0498B_{19}(\mathbf{x}) \\
&\quad + 0.0006B_{21}(\mathbf{x}) - 0.0019B_{22}(\mathbf{x}) - 2.6863B_{23}(\mathbf{x}) \\
&\quad + 0.07088B_{25}(\mathbf{x}) - 0.1896B_{28}(\mathbf{x}) - 51.0555B_{29}(\mathbf{x}) \\
&\quad - 0.5592B_{32}(\mathbf{x}) + 12.8766B_{33}(\mathbf{x}) - 12.702B_{35}(\mathbf{x}) + 0.0010B_{37}(\mathbf{x}) \\
&\quad + 0.0007B_{38}(\mathbf{x}) - 0.0043B_{39}(\mathbf{x}) - 0.0092B_{40}(\mathbf{x})
\end{aligned}$$

- CMARS model:

$$\begin{aligned}
\hat{Y} &= \beta_0 + \sum_{m=1}^{M_{\max}} \beta_m B_m(\mathbf{x}) \\
&= 0,7883 - 0,0107B_1(\mathbf{x}) + 0,5609B_2(\mathbf{x}) + 0,1036B_3(\mathbf{x}) \\
&\quad + 0,0441B_4(\mathbf{x}) + -0,3407B_5(\mathbf{x}) + 0,1152B_6(\mathbf{x}) \\
&\quad - 0,2084B_7(\mathbf{x}) + 0,0023B_8(\mathbf{x}) - 0,0012B_9(\mathbf{x}) \\
&\quad - 0,0035B_{10}(\mathbf{x}) + 0,0844B_{11}(\mathbf{x}) - 0,0001B_{12}(\mathbf{x}) \\
&\quad - 0,0062B_{13}(\mathbf{x}) + 55,0503B_{14}(\mathbf{x}) + 0,2311B_{15}(\mathbf{x}) \\
&\quad - 0,2592B_{16}(\mathbf{x}) - 0,1112B_{17}(\mathbf{x}) + 0,0716B_{18}(\mathbf{x}) \\
&\quad + 0,2624B_{19}(\mathbf{x}) - 2,2228B_{20}(\mathbf{x}) - 0,4299B_{21}(\mathbf{x}) \\
&\quad + 0,7220B_{22}(\mathbf{x}) - 1,3133B_{23}(\mathbf{x}) - 1,3056B_{24}(\mathbf{x}) \\
&\quad + 0,0495B_{25}(\mathbf{x}) + 0,0232B_{26}(\mathbf{x}) - 0,0506B_{27}(\mathbf{x}) \\
&\quad - 0,1882B_{28}(\mathbf{x}) - 0,0197B_{29}(\mathbf{x}) - 51,3736B_{30}(\mathbf{x}) \\
&\quad - 0,0576B_{31}(\mathbf{x}) - 0,7970B_{32}(\mathbf{x}) + 0,4116B_{33}(\mathbf{x}) \\
&\quad + 12,2530B_{34}(\mathbf{x}) + 0,3086B_{35}(\mathbf{x}) - 12,5382B_{36}(\mathbf{x}) \\
&\quad + 0,0011B_{37}(\mathbf{x}) + 0,0007B_{38}(\mathbf{x}) - 0,0041B_{39}(\mathbf{x}) - 0,0100B_{40}(\mathbf{x})
\end{aligned}$$





## APPENDIX D

### ACCURACY MEASURES

The accuracy measures to evaluate the performances of SMAC, MARS and CMARS methods are presented as follows:

- **General notation:**

$y_i$  :  $i$ th observed response value,

$\hat{y}_i$  :  $i$ th fitted response,

$\bar{y}$  : mean of observed values,

$N$  : number of observations.

- **Multiple coefficient of determination ( $R^2$ ):** It is a value that indicates how much variation in response is explained by the model. Higher  $R^2$  values means better fit to the data. It is expressed as:

$$R^2 := 1 - \left( \frac{\sum_{i=1}^N (y_i - \hat{y}_i)}{\sum_{i=1}^N (y_i - \bar{y})} \right)^2.$$

- **Mean absolute error (MAE):** It gives the average magnitude of error. The smaller the MAE, the better it is. The formula is:

$$MAE := \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i|.$$

These measures of accuracy can be found in statistic text books such as [159].



## APPENDIX E

### MATLAB CODES

#### E.1 SMAC Code

The SMAC code is available at <http://www.cesbio.ups-tlse.fr/us/serveurs4.htm> [160]. The code given below calculates the atmospherically corrected surface reflectance of a target pixel.

```
function y = smac(tetas, tetav, phis, phiv, uh2o, uo3, taup550, pression,
    r_toa, ah2o, nh2o, ao3, no3, ao2, no2, po2, aco2, nco2, pco2, ach4,
    nch4, pch4, ano2, nno2, pno2, aco, nco, pco, a0s, a1s, a2s, a3s, a0T,
    a1T, a2T, a3T, taur, a0taup, altaup, wo, gc, a0P, a1P, a2P, a3P, a4P,
    Rest1, Rest2, Rest3, Rest4, Resr1, Resr2, Resr3, Resa1, Resa2,
    Resa3, Resa4)

%Adapted to MATLAB by: Semih Kuter
%Date : June 2008

%SMAC Calculations

crd=180./pi;
cdr=pi/180.;

us = cos(tetas*cdr);
uv = cos(tetav*cdr);
dphi=(phis-phiv)*cdr;
Peq=pression/1013.0;

%----- 1) air mass
m = 1./us + 1./uv;

%----- 2) aerosol optical depth in the spectral band, taup
taup = (a0taup) + (altaup) * taup550 ;

%----- 3) gaseous transmissions (downward and upward paths)
to3 = 1. ;
th2o= 1. ;
to2 = 1. ;
tco2= 1. ;
tch4= 1. ;

uo2= Peq^po2;
uco2= Peq^pco2;
uch4= Peq^pch4;
uno2= Peq^pno2;
uco = Peq^pco;
```

```

%----- 4) if uh2o <= 0 and uo3 <= 0 no gaseous absorpion is computed
if( (uh2o > 0.) || (uo3 > 0.) )

    to3 = exp ( (ao3) * ((uo3 *m)^(no3)) ) ;
    th2o = exp ( (ah2o) * ((uh2o*m)^(nh2o)) ) ;
    to2 = exp ( (ao2) * ((uo2 *m)^(no2)) ) ;
    tco2 = exp ( (aco2) * ((uco2*m)^(nco2)) ) ;
    tch4 = exp ( (ach4) * ((uch4*m)^(nch4)) ) ;
    tno2 = exp ( (ano2) * ((uno2*m)^(nno2)) ) ;
    tco = exp ( (aco) * ((uco*m)^(nco)) ) ;

end

%----- 5) Total scattering transmission
ttetas = a0T + (a1T*taup550/us) + ((a2T*Peq + a3T)/(1.+us)) ; %downward
ttetav = a0T + (a1T*taup550/uv) + ((a2T*Peq + a3T)/(1.+uv)) ; %upward

%----- 6) spherical albedo of the atmosphere
s = (a0s * Peq) + a3s + (als*taup550) + (a2s *taup550^2) ;

%----- 7) scattering angle cosine
cksi = - ( (us*uv) + (sqrt(1. - us*us) * sqrt(1. - uv*uv)*cos(dphi) ) ) ;
if (cksi < -1 )
    cksi=-1.0 ;
end

%----- 8) scattering angle in degree
ksid = crd*acos(cksi) ;

%----- 9) rayleigh atmospheric reflectance
% pour 6s on a delta = 0.0279
ray_phase = 0.7190443 * (1. + (cksi*cksi)) + 0.0412742 ;

taurz = taur*Peq;

ray_ref = ( taurz*ray_phase ) / (4.*us*uv) ;

%-----Residu Rayleigh -----
Res_ray= Resr1 + ((Resr2 * taurz*ray_phase) / (us*uv))
          + (Resr3 * ((taurz*ray_phase/(us*uv))^2));

%----- 10) aerosol atmospheric reflectance
aer_phase = (a0P) + (a1P)*ksid + (a2P)*ksid*ksid + (a3P)*(ksid^3)
            + (a4P) * (ksid^4);

ak2 = (1. - (wo))*(3. - (wo)*3*(gc)) ;
ak = sqrt(ak2) ;
e = -3.*us*us*(wo) / (4.*(1. - ak2*us*us) ) ;
f = -(1. - (wo))*3.*(gc)*us*us*(wo) / (4.*(1. - ak2*us*us) ) ;
dp = e / (3.*us) + us*f ;
d = e + f ;
b = 2.*ak / (3. - (wo)*3*(gc));
del = exp( ak*taup )*(1. + b)*(1. + b) - exp(-ak*taup)*(1. - b)*(1. - b) ;
ww = (wo)/4.;
ss = us / (1. - ak2*us*us) ;
q1 = 2. + 3.*us + (1. - (wo))*3.*(gc)*us*(1. + 2.*us) ;
q2 = 2. - 3.*us - (1. - (wo))*3.*(gc)*us*(1. - 2.*us) ;
q3 = q2*exp( -taup/us ) ;
c1 = ((ww*ss) / del) * ( q1*exp(ak*taup)*(1. + b) + q3*(1. - b) ) ;
c2 = -((ww*ss) / del) * ( q1*exp(-ak*taup)*(1. - b) + q3*(1. + b) ) ;
cp1 = c1*ak / ( 3. - (wo)*3.*(gc) ) ;
cp2 = -c2*ak / ( 3. - (wo)*3.*(gc) ) ;

```

```

z = d - (wo)*3.*(gc)*uv*dp + (wo)*aer_phase/4. ;
x = c1 - (wo)*3.*(gc)*uv*cp1 ;
y = c2 - (wo)*3.*(gc)*uv*cp2 ;
aa1 = uv / (1. + ak*uv) ;
aa2 = uv / (1. - ak*uv) ;
aa3 = us*uv / (us + uv) ;

aer_ref = x*aa1* (1. - exp( -taup/aa1 ) ) ;
aer_ref = aer_ref + y*aa2*( 1. - exp( -taup / aa2 ) ) ;
aer_ref = aer_ref + z*aa3*( 1. - exp( -taup / aa3 ) ) ;
aer_ref = aer_ref / ( us*uv ) ;

%-----Residu Aerosol -----
Res_aer= ( (Resa1) + (Resa2) * ( taup * m *cksi )
          + (Resa3) * ((taup*m*cksi)^2) ) + (Resa4) * ((taup*m*cksi)^3);

%-----Residu 6s-----
tautot=taup+taurz;
Res_6s= ( (Rest1) + (Rest2) * ( tautot * m *cksi )
          + (Rest3) * ((tautot*m*cksi)^2) ) + (Rest4) * ((tautot*m*cksi)^3);

%----- 11) total atmospheric reflectance
atm_ref = ray_ref - Res_ray + aer_ref - Res_aer + Res_6s;

%----- reflectance at toa
tg      = th2o * to3 * to2 * tco2 * tch4* tco * tno2 ;

%reflectance at surface

r_surf = r_toa - (atm_ref * tg) ;
y = r_surf / ( (tg * ttetas * ttetav) + (r_surf * s) );

```

## E.2 MATLAB Code to Run the SMAC Algorithm on a MODIS Image

```

%function runsmacnew runs the smac code on selected sets of MODIS images
function y = runsmacnew(pressure)

%pressure=1013

%Author: Semih Kuter
%Date : June 2008

%Get the Text File containing SMAC Inputs
CurrentFolder = pwd;
[FileName1,PathName1] = uigetfile('*.csv', 'SELECT THE SMAC INPUT TEXT FILE');
cd(PathName1);
fid = fopen(FileName1);
smacinput = textscan(fid, '%f %f %f %f %f %f %f %f %f',
'Delimiter',' ','HeaderLines',1);
fclose(fid);

%SOLAR ZENITH ANGLE
tetas = double(smacinput[106])/100;

%SENSOR ZENITH ANGLE
tetav = double((smacinput[85]))/100;

```

```

%SOLAR AZIMUTH ANGLE
phis = double((smacinput[85]))/100;

%SENSOR AZIMUTH ANGLE
phiv = double((smacinput{4}))/100;

%WATER VAPOR CONTENT
uh2o = double((smacinput{5}))/1000;

%OZONE CONTENT
uo3 = double((smacinput{6}))/10000;

%ATMOSPHERIC OPTICAL DEPTH
taup550 = double((smacinput{7}))/1000;

%MOD02 TOP OF ATMOSPHERIC REFLECTANCE
r_toa = double(smacinput{8});

%MOD09 ATMOSPHERIC REFLECTANCE
mod09 = double(smacinput{9});

%Get the sensor calibration parameters
CurrentFolder = pwd;
[FileName9,PathName9] = uigetfile('*.dat', 'SELECT THE SENSOR CALIBRATION
FILE');
cd(PathName9);
fid = fopen(FileName9);
C = fscanf(fid,'%f',49);
fclose(fid);
F=C';
Fcell=num2cell(F);

[ah2o,nh2o,ao3,no3,ao2,no2,po2,aco2,nco2,pco2,ach4,nch4,pch4,ano2,nno2,pno2,aco
,nco,pco,a0s,als,a2s,a3s,a0T,a1T,a2T,a3T,taur,sr,a0taup,altaup,wo,gc,a0P,a1P,a2
P,a3P,a4P,Rest1,Rest2,Rest3,Rest4,Resr1,Resr2,Resr3,Resa1,Resa2,Resa3,Resa4]=Fc
ell{:};

%Run SMAC Calculations
pression = pressure;

ImgSize = size(tetas);
y = zeros(ImgSize(1),ImgSize(2));
tic;
tstart=tic;
for i=1:ImgSize(1)
    for j=1:ImgSize(2)
        y(i,j) = smacnew(tetas(i,j), tetav(i,j), phis(i,j), phiv(i,j),
uh2o(i,j), uo3(i,j), taup550(i,j), pression, r_toa(i,j),
ah2o,nh2o,ao3,no3,ao2,no2,po2,aco2,nco2,pco2,ach4,nch4,pch4,ano2,nno2,pno2,aco
,nco,pco,a0s,als,a2s,a3s,a0T,a1T,a2T,a3T,taur,a0taup,altaup,wo,gc,a0P,a1P,a2P,a3
P,a4P,Rest1,Rest2,Rest3,Rest4,Resr1,Resr2,Resr3,Resa1,Resa2,Resa3,Resa4);
    end
end
telapsed=toc(tstart);

%Reference ASCII File for the SMAC output in ASCII format
[FileName10,PathName10] = uigetfile('*.asc', 'SELECT THE REFERENCE ASCII
FILE');
cd(PathName10);
fid = fopen(FileName10);
out = textscan(fid, '%s %s');
C = cellstr(out{1,2});
fclose(fid);

```

```

ncols = ['ncols',' ',char(C(1,1))]; nrows = ['nrows',' ',char(C(2,1))];
xllcorner = ['xllcorner',' ',char(C(3,1))];
yllcorner = ['yllcorner',' ',char(C(4,1))]; cellsize = ['cellsize',' ',
',char(C(5,1))]; NODATA_value = ['NODATA_value',' ',char(C(6,1))];
img_dat=strvcat(ncols,nrows,xllcorner,yllcorner,cellsize,NODATA_value);

%Write output to a space-delimited ASCII file
[FileName1,PathName1] = uiputfile('*.asc', 'SAVE SMAC RESULT AS');
cd(PathName1);
dlmwrite('del01.asc',img_dat,'delimiter',' ','newline','pc');
dlmwrite('del02.asc',y,' ');
system(['for %f in ("del01.asc", "del02.asc") do type "%f" >> "' FileName1
'"]');
delete('del01.asc', 'del02.asc');

%CALCULATE RMSE ERROR BY USING SMAC & MOD09 DATA
r=sqrt(sum((y(:)-mod09(:)).^2)/numel(y));
fprintf('RMSE in SMAC = %f\n',r)
fprintf('Process time in sec. = %f\n',telapsed)
cd(CurrentFolder);

%MAX & MIN VALUES FOR SMAC IMAGE
fprintf('Maximum Reflectance Value in SMAC = %f\n',max(max(y)))
fprintf('Minimum Reflectance Value in SMAC = %f\n',min(min(y)))

```

### E.3 MATLAB Code to Prepare the Necessary Input Data for SMAC Processing

```

%function valextract2points extract raster values to points stored in a
%shape file and prepares the data necessary for SMAC processing
function y = valextract2points(refscales)

%"refscales" is the reflectance scale value of the corresponding MODIS
%band and must be entered without its exponent (i.e., -e5)

%Author: Semih Kuter
%Date : March 2012

%GET THE ARCGIS SHAPE FILE IN WHICH POINTS ARE STORED
CurrentFolder = pwd;
[FileName1,PathName1] = uigetfile('*.shp', 'SELECT THE SHAPE FILE');
cd(PathName1);
S = shaperead(FileName1);
dimensions = size(S);
row=dimensions(1);
column=dimensions(2);

%GET THE SZA FILE
[FileName2,PathName2] = uigetfile('*.tif', 'SELECT THE SZA FILE');
cd(PathName2);
sza = geotiffread(FileName2);
info_sza=geotiffinfo(FileName2);
refmatrix_sza=info_sza.RefMatrix;

%GET THE VZA FILE
[FileName3,PathName3] = uigetfile('*.tif', 'SELECT THE VZA FILE');
cd(PathName3);
vza = geotiffread(FileName3);
info_vza=geotiffinfo(FileName3);

```

```

refmatrix_vza=info_vza.RefMatrix;

%GET THE SAA FILE
[FileName4,PathName4] = uigetfile('*.tif', 'SELECT THE SAA FILE');
cd(PathName4);
saa = geotiffread(FileName4);
info_saa=geotiffinfo(FileName4);
refmatrix_saa=info_saa.RefMatrix;

%GET THE VAA FILE
[FileName5,PathName5] = uigetfile('*.tif', 'SELECT THE VAA FILE');
cd(PathName5);
vaa = geotiffread(FileName5);
info_vaa=geotiffinfo(FileName5);
refmatrix_vaa=info_vaa.RefMatrix;

%GET THE WV FILE
[FileName6,PathName6] = uigetfile('*.tif', 'SELECT THE WV FILE');
cd(PathName6);
wv = geotiffread(FileName6);
info_wv=geotiffinfo(FileName6);
refmatrix_wv=info_wv.RefMatrix;

%GET THE O3 FILE
[FileName7,PathName7] = uigetfile('*.tif', 'SELECT THE O3 FILE');
cd(PathName7);
o3 = geotiffread(FileName7);
info_o3=geotiffinfo(FileName7);
refmatrix_o3=info_o3.RefMatrix;

%GET THE AOD FILE
[FileName8,PathName8] = uigetfile('*.tif', 'SELECT THE AOD FILE');
cd(PathName8);
aod = geotiffread(FileName8);
info_aod=geotiffinfo(FileName8);
refmatrix_aod=info_aod.RefMatrix;

%GET THE MOD02 FILE
[FileName9,PathName9] = uigetfile('*.tif', 'SELECT THE MOD02 FILE');
cd(PathName9);
mod02 = double(geotiffread(FileName9))*refscases*0.00001;
info_mod02=geotiffinfo(FileName9);
refmatrix_mod02=info_mod02.RefMatrix;

%GET THE MOD09 FILE
CurrentFolder = pwd;
[FileName10,PathName10] = uigetfile('*.tif', 'SELECT THE MOD09 FILE');
cd(PathName10);
mod09 = double(geotiffread(FileName10))/10000;
info_mod09=geotiffinfo(FileName10);
refmatrix_mod09=info_mod09.RefMatrix;

y=zeros(row,9);
y2=zeros(row,3);
y3=zeros(row,1);
r1=zeros(1,row);c1=zeros(1,row);
r2=zeros(1,row);c2=zeros(1,row);
r3=zeros(1,row);c3=zeros(1,row);
r4=zeros(1,row);c4=zeros(1,row);
r5=zeros(1,row);c5=zeros(1,row);
r6=zeros(1,row);c6=zeros(1,row);
r7=zeros(1,row);c7=zeros(1,row);
r8=zeros(1,row);c8=zeros(1,row);

```



```

r9=zeros(1,row);c9=zeros(1,row);
a1=zeros(1,row);a2=zeros(1,row);
a3=zeros(1,row);a4=zeros(1,row);
a5=zeros(1,row);a6=zeros(1,row);
a7=zeros(1,row);a8=zeros(1,row);
a9=zeros(1,row);a10=zeros(1,row);
a11=zeros(1,row);

tstart=tic;
for i=column:row

    [r1(i),c1(i)] = map2pix(refmatrix_sza, S(i,1).X,S(i,1).Y);
    a1(i)= sza(ceil(r1(i)),ceil(c1(i)));

    [r2(i),c2(i)] = map2pix(refmatrix_vza, S(i,1).X,S(i,1).Y);
    a2(i)= vza(ceil(r2(i)),ceil(c2(i)));

    [r3(i),c3(i)] = map2pix(refmatrix_saa, S(i,1).X,S(i,1).Y);
    a3(i)= saa(ceil(r3(i)),ceil(c3(i)));

    [r4(i),c4(i)] = map2pix(refmatrix_vaa, S(i,1).X,S(i,1).Y);
    a4(i)= vaa(ceil(r4(i)),ceil(c4(i)));

    [r5(i),c5(i)] = map2pix(refmatrix_wv, S(i,1).X,S(i,1).Y);
    a5(i)= wv(ceil(r5(i)),ceil(c5(i)));

    [r6(i),c6(i)] = map2pix(refmatrix_o3, S(i,1).X,S(i,1).Y);
    a6(i)= o3(ceil(r6(i)),ceil(c6(i)));

    [r7(i),c7(i)] = map2pix(refmatrix_aod, S(i,1).X,S(i,1).Y);
    a7(i)= aod(ceil(r7(i)),ceil(c7(i)));

    [r8(i),c8(i)] = map2pix(refmatrix_mod02, S(i,1).X,S(i,1).Y);
    a8(i)= mod02(ceil(r8(i)),ceil(c8(i)));

    [r9(i),c9(i)] = map2pix(refmatrix_mod09, S(i,1).X,S(i,1).Y);
    a9(i)= mod09(ceil(r9(i)),ceil(c9(i)));

    a10(i)=S(i,1).X;

    a11(i)=S(i,1).Y;
end

y(:,1)=a1;
y(:,2)=a2;
y(:,3)=a3;
y(:,4)=a4;
y(:,5)=a5;
y(:,6)=a6;
y(:,7)=a7;
y(:,8)=a8;
y(:,9)=a9;
y2(:,1)=a10;
y2(:,2)=a11;
y2(:,3)=a8;
y3(:,1)=a9;

%COLUMN HEADERS & WRITING THE OUTPUT TO A COMMA SEPARATED CSV FILE
headers1=['SZA',' ','VZA',' ','SAA',' ','VAA',' ','WV',' ','O3',' ','AOD',' ',' '
MOD02',' ','MOD09'];
headers2=['X',' ','Y',' ','MOD02'];
headers3=('MOD09');

%SMAC input csv file
[FileName11,PathName11] = uiputfile('*.csv', 'SAVE THE SMAC INPUT FILE AS');
cd(PathName11);
fid = fopen(FileName11,'w');
fprintf(fid,'%s\r\n',headers1);
fclose(fid);

```

```

dlmwrite(FileName11,y,'-append','delimiter',' ','precision',9);

%MARS xtest csv file
b1='xtest.csv';
fid = fopen(b1,'w');
fprintf(fid,'%s\r\n',headers2);
fclose(fid);
dlmwrite(b1,y2,'-append','delimiter',' ','precision',9);

%MARS ytest csv file
b2='ytest.csv';
fid = fopen(b2,'w');
fprintf(fid,'%s\r\n',headers3);
fclose(fid);
dlmwrite(b2,y3,'-append','delimiter',' ','precision',9);

telapsed=toc(tstart);
fprintf('Process time in sec. = %f\n',telapsed)

```

## E.4 MATLAB Code to Prepare Response and Predictor Matrices for Multi-Response MARS Classification

```

%function trngpointextract_snow extract raster values onto points stored in a
%shape file to prepare training file for MARS snow classification
function y = trngpointextract_snow

%class indicates the landuse type of the shape file, and valid input ranges:
%snow =1
%water =2
%cloud =3
%land =4

%Author: Semih Kuter
%Date : January 2013

%GET THE ARCGIS SHAPE FILES IN WHICH THE SAMPLING POINTS ARE STORED
%CurrentFolder = pwd;
[FileName1,PathName1] = uigetfile('*.shp', 'SELECT THE SNOW SHAPE FILE');
cd(PathName1);
S = shaperead(FileName1);
Sdimensions = size(S);
Srow=Sdimensions(1);
Scolumn=Sdimensions(2);

[FileName12,PathName12] = uigetfile('*.shp', 'SELECT THE WATER SHAPE FILE');
cd(PathName12);
W = shaperead(FileName12);
Wdimensions = size(W);
Wrow=Wdimensions(1);
Wcolumn=Wdimensions(2);

[FileName13,PathName13] = uigetfile('*.shp', 'SELECT THE CLOUD SHAPE FILE');
cd(PathName13);
C = shaperead(FileName13);
Cdimensions = size(C);
Crow=Cdimensions(1);
Ccolumn=Cdimensions(2);

```

```

[FileName14,PathName14] = uigetfile('*.shp', 'SELECT THE LAND SHAPE FILE');
cd(PathName14);
L = shaperead(FileName14);
Ldimensions = size(L);
Lrow=Ldimensions(1);
Lcolumn=Ldimensions(2);

%GET THE B01 FILE
[FileName2,PathName2] = uigetfile('*.tif', 'SELECT THE BAND 01 IMAGE');
cd(PathName2);
b01 = geotiffread(FileName2);
info_b01=geotiffinfo(FileName2);
refmatrix_b01=info_b01.RefMatrix;

%GET THE B02 FILE
[FileName3,PathName3] = uigetfile('*.tif', 'SELECT THE BAND 02 IMAGE');
cd(PathName3);
b02 = geotiffread(FileName3);
info_b02=geotiffinfo(FileName3);
refmatrix_b02=info_b02.RefMatrix;

%GET THE B03 FILE
[FileName4,PathName4] = uigetfile('*.tif', 'SELECT THE BAND 03 IMAGE');
cd(PathName4);
b03 = geotiffread(FileName4);
info_b03=geotiffinfo(FileName4);
refmatrix_b03=info_b03.RefMatrix;

%GET THE B04 FILE
[FileName5,PathName5] = uigetfile('*.tif', 'SELECT THE BAND 04 IMAGE');
cd(PathName5);
b04 = geotiffread(FileName5);
info_b04=geotiffinfo(FileName5);
refmatrix_b04=info_b04.RefMatrix;

%GET THE B05 FILE
[FileName6,PathName6] = uigetfile('*.tif', 'SELECT THE BAND 05 IMAGE');
cd(PathName6);
b05 = geotiffread(FileName6);
info_b05=geotiffinfo(FileName6);
refmatrix_b05=info_b05.RefMatrix;

%GET THE B06 FILE
[FileName7,PathName7] = uigetfile('*.tif', 'SELECT THE BAND 06 IMAGE');
cd(PathName7);
b06 = geotiffread(FileName7);
info_b06=geotiffinfo(FileName7);
refmatrix_b06=info_b06.RefMatrix;

%GET THE B07 FILE
[FileName8,PathName8] = uigetfile('*.tif', 'SELECT THE BAND 07 IMAGE');
cd(PathName8);
b07 = geotiffread(FileName8);
info_b07=geotiffinfo(FileName8);
refmatrix_b07=info_b07.RefMatrix;

```

```

%GET THE B31 FILE
[FileName9,PathName9] = uigetfile('*.tif', 'SELECT THE BAND 31 IMAGE');
cd(PathName9);
b31 = geotiffread(FileName9);
info_b31=geotiffinfo(FileName9);
refmatrix_b31=info_b31.RefMatrix;

%GET THE B32 FILE
[FileName10,PathName10] = uigetfile('*.tif', 'SELECT THE BAND 32 IMAGE');
cd(PathName10);
b32 = geotiffread(FileName10);
info_b32=geotiffinfo(FileName10);
refmatrix_b32=info_b32.RefMatrix;

%MATRIX PREALLOCATIONS
y1=zeros(Srow,12);
y2=zeros(Wrow,12);
y3=zeros(Crow,12);
y4=zeros(Lrow,12);
y=zeros(Srow+Wrow+Crow+Lrow,12);

r1=zeros(1,Srow);c1=zeros(1,Srow);
r2=zeros(1,Srow);c2=zeros(1,Srow);
r3=zeros(1,Srow);c3=zeros(1,Srow);
r4=zeros(1,Srow);c4=zeros(1,Srow);
r5=zeros(1,Srow);c5=zeros(1,Srow);
r6=zeros(1,Srow);c6=zeros(1,Srow);
r7=zeros(1,Srow);c7=zeros(1,Srow);
r8=zeros(1,Srow);c8=zeros(1,Srow);
r9=zeros(1,Srow);c9=zeros(1,Srow);

r10=zeros(1,Wrow);c10=zeros(1,Wrow);
r11=zeros(1,Wrow);c11=zeros(1,Wrow);
r12=zeros(1,Wrow);c12=zeros(1,Wrow);
r13=zeros(1,Wrow);c13=zeros(1,Wrow);
r14=zeros(1,Wrow);c14=zeros(1,Wrow);
r15=zeros(1,Wrow);c15=zeros(1,Wrow);
r16=zeros(1,Wrow);c16=zeros(1,Wrow);
r17=zeros(1,Wrow);c17=zeros(1,Wrow);
r18=zeros(1,Wrow);c18=zeros(1,Wrow);

r19=zeros(1,Crow);c19=zeros(1,Crow);
r20=zeros(1,Crow);c20=zeros(1,Crow);
r21=zeros(1,Crow);c21=zeros(1,Crow);
r22=zeros(1,Crow);c22=zeros(1,Crow);
r23=zeros(1,Crow);c23=zeros(1,Crow);
r24=zeros(1,Crow);c24=zeros(1,Crow);
r25=zeros(1,Crow);c25=zeros(1,Crow);
r26=zeros(1,Crow);c26=zeros(1,Crow);
r27=zeros(1,Crow);c27=zeros(1,Crow);

r28=zeros(1,Lrow);c28=zeros(1,Lrow);
r29=zeros(1,Lrow);c29=zeros(1,Lrow);
r30=zeros(1,Lrow);c30=zeros(1,Lrow);
r31=zeros(1,Lrow);c31=zeros(1,Lrow);
r32=zeros(1,Lrow);c32=zeros(1,Lrow);
r33=zeros(1,Lrow);c33=zeros(1,Lrow);
r34=zeros(1,Lrow);c34=zeros(1,Lrow);
r35=zeros(1,Lrow);c35=zeros(1,Lrow);
r36=zeros(1,Lrow);c36=zeros(1,Lrow);

a1=zeros(1,Srow);a2=zeros(1,Srow);
a3=zeros(1,Srow);a4=zeros(1,Srow);
a5=zeros(1,Srow);a6=zeros(1,Srow);
a7=zeros(1,Srow);a8=zeros(1,Srow);

```

```

a9=zeros(1,Srow);a10=zeros(1,Srow);
a11=zeros(1,Srow);a12=zeros(1,Srow);

a13=zeros(1,Wrow);a14=zeros(1,Wrow);
a15=zeros(1,Wrow);a16=zeros(1,Wrow);
a17=zeros(1,Wrow);a18=zeros(1,Wrow);
a19=zeros(1,Wrow);a20=zeros(1,Wrow);
a21=zeros(1,Wrow);a22=zeros(1,Wrow);
a23=zeros(1,Wrow);a24=zeros(1,Wrow);

a25=zeros(1,Crow);a26=zeros(1,Crow);
a27=zeros(1,Crow);a28=zeros(1,Crow);
a29=zeros(1,Crow);a30=zeros(1,Crow);
a31=zeros(1,Crow);a32=zeros(1,Crow);
a33=zeros(1,Crow);a34=zeros(1,Crow);
a35=zeros(1,Crow);a36=zeros(1,Crow);

a37=zeros(1,Lrow);a38=zeros(1,Lrow);
a39=zeros(1,Lrow);a40=zeros(1,Lrow);
a41=zeros(1,Lrow);a42=zeros(1,Lrow);
a43=zeros(1,Lrow);a44=zeros(1,Lrow);
a45=zeros(1,Lrow);a46=zeros(1,Lrow);
a47=zeros(1,Lrow);a48=zeros(1,Lrow);

tstart=tic;
%SNOW
for i=Scolumn:Srow

    a1(i)= 1;

    [r1(i),c1(i)] = map2pix(refmatrix_b01, S(i,1).X,S(i,1).Y);
    a2(i)= b01(round(r1(i)),round(c1(i)));

    [r2(i),c2(i)] = map2pix(refmatrix_b02, S(i,1).X,S(i,1).Y);
    a3(i)= b02(round(r2(i)),round(c2(i)));

    [r3(i),c3(i)] = map2pix(refmatrix_b03, S(i,1).X,S(i,1).Y);
    a4(i)= b03(round(r3(i)),round(c3(i)));

    [r4(i),c4(i)] = map2pix(refmatrix_b04, S(i,1).X,S(i,1).Y);
    a5(i)= b04(round(r4(i)),round(c4(i)));

    [r5(i),c5(i)] = map2pix(refmatrix_b05, S(i,1).X,S(i,1).Y);
    a6(i)= b05(round(r5(i)),round(c5(i)));

    [r6(i),c6(i)] = map2pix(refmatrix_b06, S(i,1).X,S(i,1).Y);
    a7(i)= b06(round(r6(i)),round(c6(i)));

    [r7(i),c7(i)] = map2pix(refmatrix_b07, S(i,1).X,S(i,1).Y);
    a8(i)= b07(round(r7(i)),round(c7(i)));

    [r8(i),c8(i)] = map2pix(refmatrix_b31, S(i,1).X,S(i,1).Y);
    a9(i)= b31(round(r8(i)),round(c8(i)));

    [r9(i),c9(i)] = map2pix(refmatrix_b32, S(i,1).X,S(i,1).Y);
    a10(i)= b32(round(r9(i)),round(c9(i)));

    a11(i)=S(i,1).X;

    a12(i)=S(i,1).Y;

end

y1(:,1)=a1;      %class snow
y1(:,2)=a11;     %X
y1(:,3)=a12;     %Y
y1(:,4)=a2;      %b01

```

```

y1(:,5)=a3;      %b02
y1(:,6)=a4;      %b03
y1(:,7)=a5;      %b04
y1(:,8)=a6;      %b05
y1(:,9)=a7;      %b06
y1(:,10)=a8;     %b07
y1(:,11)=a9;     %b31
y1(:,12)=a10;    %b32

```

```
%WATER
```

```
for i=Wcolumn:Wrow
```

```
    a13(i) = 2;
```

```
    [r10(i),c10(i)] = map2pix(refmatrix_b01, W(i,1).X,W(i,1).Y);
    a14(i)= b01(round(r10(i)),round(c10(i)));
```

```
    [r11(i),c11(i)] = map2pix(refmatrix_b02, W(i,1).X,W(i,1).Y);
    a15(i)= b02(round(r11(i)),round(c11(i)));
```

```
    [r12(i),c12(i)] = map2pix(refmatrix_b03, W(i,1).X,W(i,1).Y);
    a16(i)= b03(round(r12(i)),round(c12(i)));
```

```
    [r13(i),c13(i)] = map2pix(refmatrix_b04, W(i,1).X,W(i,1).Y);
    a17(i)= b04(round(r13(i)),round(c13(i)));
```

```
    [r14(i),c14(i)] = map2pix(refmatrix_b05, W(i,1).X,W(i,1).Y);
    a18(i)= b05(round(r14(i)),round(c14(i)));
```

```
    [r15(i),c15(i)] = map2pix(refmatrix_b06, W(i,1).X,W(i,1).Y);
    a19(i)= b06(round(r15(i)),round(c15(i)));
```

```
    [r16(i),c16(i)] = map2pix(refmatrix_b07, W(i,1).X,W(i,1).Y);
    a20(i)= b07(round(r16(i)),round(c16(i)));
```

```
    [r17(i),c17(i)] = map2pix(refmatrix_b31, W(i,1).X,W(i,1).Y);
    a21(i)= b31(round(r17(i)),round(c17(i)));
```

```
    [r18(i),c18(i)] = map2pix(refmatrix_b32, W(i,1).X,W(i,1).Y);
    a22(i)= b32(round(r18(i)),round(c18(i)));
```

```
    a23(i)=W(i,1).X;
```

```
    a24(i)=W(i,1).Y;
```

```
end
```

```

y2(:,1)=a13;     %class water
y2(:,2)=a23;     %X
y2(:,3)=a24;     %Y
y2(:,4)=a14;     %b01
y2(:,5)=a15;     %b02
y2(:,6)=a16;     %b03
y2(:,7)=a17;     %b04
y2(:,8)=a18;     %b05
y2(:,9)=a19;     %b06
y2(:,10)=a20;    %b07
y2(:,11)=a21;    %b31
y2(:,12)=a22;    %b32

```

```
%CLOUD
```

```
for i=Ccolumn:Crow
```

```
    a25(i) = 3;
```

```

[r19(i),c19(i)] = map2pix(refmatrix_b01, C(i,1).X,C(i,1).Y);
a26(i)= b01(round(r19(i)),round(c19(i)));

[r20(i),c20(i)] = map2pix(refmatrix_b02, C(i,1).X,C(i,1).Y);
a27(i)= b02(round(r20(i)),round(c20(i)));

[r21(i),c21(i)] = map2pix(refmatrix_b03, C(i,1).X,C(i,1).Y);
a28(i)= b03(round(r21(i)),round(c21(i)));

[r22(i),c22(i)] = map2pix(refmatrix_b04, C(i,1).X,C(i,1).Y);
a29(i)= b04(round(r22(i)),round(c22(i)));

[r23(i),c23(i)] = map2pix(refmatrix_b05, C(i,1).X,C(i,1).Y);
a30(i)= b05(round(r23(i)),round(c23(i)));

[r24(i),c24(i)] = map2pix(refmatrix_b06, C(i,1).X,C(i,1).Y);
a31(i)= b06(round(r24(i)),round(c24(i)));

[r25(i),c25(i)] = map2pix(refmatrix_b07, C(i,1).X,C(i,1).Y);
a32(i)= b07(round(r25(i)),round(c25(i)));

[r26(i),c26(i)] = map2pix(refmatrix_b31, C(i,1).X,C(i,1).Y);
a33(i)= b31(round(r26(i)),round(c26(i)));

[r27(i),c27(i)] = map2pix(refmatrix_b32, C(i,1).X,C(i,1).Y);
a34(i)= b32(round(r27(i)),round(c27(i)));

a35(i)=C(i,1).X;

a36(i)=C(i,1).Y;
end

y3(:,1)=a25;      %class cloud
y3(:,2)=a35;      %X
y3(:,3)=a36;      %Y
y3(:,4)=a26;      %b01
y3(:,5)=a27;      %b02
y3(:,6)=a28;      %b03
y3(:,7)=a29;      %b04
y3(:,8)=a30;      %b05
y3(:,9)=a31;      %b06
y3(:,10)=a32;     %b07
y3(:,11)=a33;     %b31
y3(:,12)=a34;     %b32

%LAND
for i=Lcolumn:Lrow

a37(i)= 4;

[r28(i),c28(i)] = map2pix(refmatrix_b01, L(i,1).X,L(i,1).Y);
a38(i)= b01(round(r28(i)),round(c28(i)));

[r29(i),c29(i)] = map2pix(refmatrix_b02, L(i,1).X,L(i,1).Y);
a39(i)= b02(round(r29(i)),round(c29(i)));

[r30(i),c30(i)] = map2pix(refmatrix_b03, L(i,1).X,L(i,1).Y);
a40(i)= b03(round(r30(i)),round(c30(i)));

[r31(i),c31(i)] = map2pix(refmatrix_b04, L(i,1).X,L(i,1).Y);
a41(i)= b04(round(r31(i)),round(c31(i)));

[r32(i),c32(i)] = map2pix(refmatrix_b05, L(i,1).X,L(i,1).Y);
a42(i)= b05(round(r32(i)),round(c32(i)));

[r33(i),c33(i)] = map2pix(refmatrix_b06, L(i,1).X,L(i,1).Y);

```

```

a43(i)= b06(round(r33(i)),round(c33(i)));

[r34(i),c34(i)] = map2pix(refmatrix_b07, L(i,1).X,L(i,1).Y);
a44(i)= b07(round(r34(i)),round(c34(i)));

[r35(i),c35(i)] = map2pix(refmatrix_b31, L(i,1).X,L(i,1).Y);
a45(i)= b31(round(r35(i)),round(c35(i)));

[r36(i),c36(i)] = map2pix(refmatrix_b32, L(i,1).X,L(i,1).Y);
a46(i)= b32(round(r36(i)),round(c36(i)));

a47(i)=L(i,1).X;

a48(i)=L(i,1).Y;
end

y4(:,1)=a37;      %class land
y4(:,2)=a47;      %X
y4(:,3)=a48;      %Y
y4(:,4)=a38;      %b01
y4(:,5)=a39;      %b02
y4(:,6)=a40;      %b03
y4(:,7)=a41;      %b04
y4(:,8)=a42;      %b05
y4(:,9)=a43;      %b06
y4(:,10)=a44;     %b07
y4(:,11)=a45;     %b31
y4(:,12)=a46;     %b32

y=vertcat(y1,y2,y3,y4);

%Scolumn HEADERS & WRITING THE OUTPUT TO A COMMA SEPARATED CSV FILE
headers1=['CLASS',' ','X',' ','Y',' ','B01',' ','B02',' ','B03',' ','B04',' ','B05',' ','B06',' ','B07',' ','B31',' ','B32'];

telapsed=toc(tstart);

%SNOW TEST DATA csv file
[FileName11,PathName11] = uinputfile('*.csv', 'SAVE THE TEST DATA AS');
cd(PathName11);
fid = fopen(FileName11,'w');
fprintf(fid,'%s\r\n',headers1);
fclose(fid);
dlmwrite(FileName11,y,'-append','delimiter',' ','precision',9);

fprintf('Process time in sec. = %f\n',telapsed)

```

## E.5 MATLAB Code to Obtain Cloud Mask from MOD09GA State\_1km

### QA Data

```

%function cloudmask makes a cloud mask from MOD09GA State_1km QA data
function y = cloudmask

```

```

%Author: Semih Kuter

```



```

%Date : March 2013

%GET THE MOD09GA STATE1KM QA FILE
CurrentFolder = pwd;
[FileName1,PathName1] = uigetfile('*.tif', 'SELECT THE MOD09GA STATE_1KM_QA
IMAGE');
cd(PathName1);
[mod09,data] = geotiffread(FileName1);
row = size(mod09,1);
column = size(mod09,2);

%CALCULATIONS
y = zeros(row,column);
tstart=tic;
for i=1:row;
    for j=1:column;
        S=dec2bin(mod09(i,j),16);
        a=str2double(S(15));
        b=str2double(S(16));
        if a==0 && b==1;
            y(i,j)=0;
        elseif a==1 && b==0;
            y(i,j)=0;
        else
            y(i,j)=1;
        end;
    end;
end;
telapsed=toc(tstart);

%WRITE CLOUD MASK INTO A GEOTIFF IMAGE
[FileName2,PathName2] = uiputfile('*.tif', 'SAVE THE CLOUD MASK IMAGE AS');
cd(PathName2);
geotiffwrite(FileName2,y,data);

fprintf('Process time in sec. = %f\n',telapsed)

cd(CurrentFolder);

```



## CURRICULUM VITAE

### PERSONEL INFORMATION

**Surname, Name:** KUTER, Semih

**Nationality:** Turkish (TC)

**Date and Place of Birth:** 13/02/1973, Ankara

**Marital Status:** Married

### EDUCATION

Degree	Institution	Graduation
M.Sc.	METU Dept. of Physics	2000
B.Sc.	METU Dept. of Physics	1996
High School	Bahçelievler Deneme High School	1990

### PROFESSIONAL EXPERIENCE

Year	Place	Enrollment
1996-1998	METU Dept. of Physics	Research Assistant
1999-2002	TİTAŞ A.Ş.	Technical Service Engineer
2003-2005	DRAEGER Türkiye	Technical Service Supervisor
2006-	ÇKÜ Dept. of Forest Eng.	Instructor

### PUBLICATIONS

#### International Reviewed Papers (SCI & SCI-E)

- **Kuter, S.**, Weber, G.-W., Akyürek, Z. and Özmen, A. “Inversion of top of atmospheric reflectance values by conic multivariate adaptive regression splines”. to appear in IPSE - Inverse Problems in Science and Engineering.

- **Kuter, S.**, Usul, N., Kuter, N., 2011. “Bandwidth Determination for Kernel Density Analysis of Wildfire Events at Forest Sub-District Scale”. *Ecological Modelling*. Vol. 222, Issue: 17, p. 3033-3040.
- Kuter, N., Yenilmez, F., **Kuter, S.**, 2011. “Forest Fire Risk Mapping by Kernel Density Estimation”. *Croatian Journal of Forest Engineering*. Vol. 32, Issue: 2, p. 599-610.
- Kuter, N., **Kuter, S.**, 2010. “Accuracy Comparison between GPS and DGPS: A Field Study at METU Campus”. *Italian Journal of Remote Sensing*. Vol. 42 Issue: 3, p. 3-14.

### **International Conference Publications**

- **Kuter, S.**, Weber, G.-W., Akyürek, Z. and Özmen, A., 2013. “Incorporating the modern techniques of continuous optimization within spatial analysis”, EURO XXVI EURO-INFORMS Joint International Conference, Rome, Italy, July 1-4, 2013 (oral presentation).
- Weber, G.-W., **Kuter, S.**, Yerlikaya- Özkurt, F., Kılıç, E. and Karimov, A., “Investigating stochastic hybrid systems within the frame of portfolio optimization by modern applied mathematics”, in the proceedings (CD) of International Istanbul Finance Congress - 2013 (IIFC 2013), Kadir Has University, Istanbul, May 30-31, 2013 (May 31, 2013, Session 2-2) 94-116. (oral presentation).
- **Kuter, S.**, Akyürek, Z., Weber, G.-W. and Özmen, A., 2012. “Testing of MARS on MODIS Images for Local Atmospheric Correction”, 25<sup>th</sup> European Conference on Operational Research (EURO XXV 2012), Vilnius, Lithuania, July 8-11, 2012 (oral presentation).
- Weber, G.-W., **Kuter, S.**, Kerimov, A., Temoçin, B.Z. and Kürüm, E., 2012. “Prediction of Network Dynamics under Various Time and Uncertainty Assumptions. Investigations in Education, Finance,

Economics, Biology and Medicine Recent Contributions Supported by Optimization”, Constructive Nonsmooth Analysis and Related Topics, Euler International Mathematical Institute, St. Petersburg, Russia, June 18-23, 2012 (oral presentation).

- **Kuter, S.**, Usul, N., Kuter, N., 2010. “Evaluation of Forest Fires with GIS”. ESRI International User Conference, 12-16 July 2010, San Diego Convention Center, USA (oral presentation).
- Kuter, N, **Kuter, S.**, 2010. “Wildfire Risk Mapping by GIS: A Case Study in Bodrum Forest Sub-District”. The 2nd International Geography Symposium Mediterranean Environment, GEOMED 2010, 2-5 June 2010, Kemer, Antalya, Page: 196 (oral presentation).
- Kuter, N., Yenilmez, F. and **Kuter, S.**, 2009. “Evaluation of Forest Fires with GIS at METU Campus”. Ecology and Forests for Public Health International Conference, 17-20 September 2009, Oslo, Norway, Book of Abstracts, Page: 65 (oral presentation).

#### **Presentations at International Scientific Meetings**

- **Kuter, S.**, Weber, G.-W., Akyürek, Z. and Özmen, A., 2013. “Nonparametric Regression Splines for Regional Atmospheric Correction”, Seminar at Higher School of Economics, National Research University, School of Applied Mathematics and Information Science, Moscow, Russia, October 16, 2013. (oral presentation).
- **Kuter, S.**, Weber, G.-W., Akyürek, Z. and Özmen, A., 2013. “Incorporating the Modern Methods of Continuous Optimization within Spatial Analysis”, AACIMP-2013|8th Summer School on Achievements and Applications of Contemporary Informatics, Mathematics and Physics, National University of Technology of the Ukraine, Kyiv, Ukraine, August 1-17, 2013. (oral presentation).

- **Kuter, S.**, Weber, G.-W., Akyürek, Z. and Özmen, A., 2013. “Benefitting from Modern Methods of Continuous Optimization in Spatial Analysis and Remote Sensing”, VII Moscow International Conference on Operations Research (ORM2013), Moscow, Russia, October 15-19, 2013. (oral presentation).
- **Kuter, S.**, Akyürek, Z., Weber, G.-W. and Özmen, A., 2012. “Testing of MARS on MODIS Images for Local Atmospheric Correction”, AACIMP-2012|7th Summer School on Achievements and Applications of Contemporary Informatics, Mathematics and Physics, National University of Technology of the Ukraine, Kyiv, Ukraine, August 3-16, 2012. (oral presentation).

### **Book Chapters**

- **Kuter, S.**, Weber, G.-W., Özmen, A. and Akyürek, Z., 2012. “Modern applied mathematics for alternative modelling of the atmospheric effects on satellite images”, to appear as chapter in book, *Springer volume Modeling, Optimization, Dynamics and Bioeconomy, series Springer Proceedings in Mathematics*, D. Zilberman and A. Pinto (Eds.).
- Kuter, N, **Kuter, S.**, 2011. “Wildfire Risk Mapping by GIS: A Case Study in Bodrum Forest Sub-District”. In R. Efe, M. Öztürk & İ. Atalay (Eds.), *Natural Environment and Culture in the Mediterranean Region II* (pp. 325-337). UK: Cambridge Scholars Publishing.

### **INTERNATIONAL VISITS AS INVITED LECTURER**

- Estonian University of Life Sciences – The Institute of Forestry and Rural Engineering, 7-11 October 2013, Tartu, Estonia.

## REFEERING IN SCI & SCI-E JOURNALS

- Rohani, V.A., Kasirun, Z.M., Kumar., S and Shamshirband, S., 2014. “An Effective Recommender Algorithm for Cold-start Problem in Academic Social Networks”. Mathematical Problems in Engineering.
- Bastos, A.S. and Hasegawa, H., 2013. “Behavior of GPS Signal Interruption Probability under Tree Canopies in Different Forest Conditions”. European Journal of Remote Sensing.
- Choi, Y.R., 2013. “Analytic Approach for the Study of Air and/or Liquid-Filled Geomembrane Tube Sections on a Horizontal Foundation”. Mathematical Problems in Engineering.
- Azizi-Sultan, A.S., 2013. “Pseudo-Automatic Beam's Eye View for Multi-Criteria Intensity-Modulated Radiation Therapy”. Optimization: A Journal of Mathematical Programming and Operations Research.
- Zhou, G., Lin, J., Ma, X. and Zhang, W., 2012. “A Web-based GIS for crime mapping and decision support”. Earth Science Informatics.

## VOLUNTARY CONTRIBUTIONS IN SCIENTIFIC ASSOCIATIONS

- Web Page Administrator of EUROPT - The Continuous Optimization Working Group of EURO (The Association of European Operational Research Societies) (<http://europt.iam.metu.edu.tr/>)
- Contributor at IFORS DC Resources Web Page – The International Federation of Operational Research Societies Developing Countries Operational Research Resources ([http://ifors.org/developing\\_countries/index.php?title=Main\\_Page](http://ifors.org/developing_countries/index.php?title=Main_Page))