Remote sensing of atmospheric aerosol distributions using supervised texture classification

Wiltshire, Ben

Award date:
2012

Awarding institution:
University of Bath

REMOTE SENSING OF ATMOSPHERIC AEROSOL DISTRIBUTIONS USING SUPERVISED TEXTURE CLASSIFICATION

Ben Wiltshire

A thesis submitted for the degree of Doctor of Philosophy
University of Bath
Department of Electronic and Electrical Engineering

2011

COPYRIGHT

Ben Wiltshire

Abstract

This thesis presents a new technique to identify a 2D mask showing the extent of particulate aerosol distributions in satellite imagery. This technique uses a supervised texture classification approach, and utilises data from two distinct satellite sources. The vertical feature mask (VFM) product from the CALIPSO lidar, provides an accurate description of the aerosol content of the atmosphere but has a limited footprint and coverage. The CALIPSO VFM is used to provide training data in order to form classifiers to be applied to other imagery, namely data from the spinning enhanced visible and infrared imager (SEVIRI) on the MSG satellite. The output from the classification is a 2D mask representing the locations of the particulate aerosol of interest within the SEVIRI image.

This approach has been demonstrated on test cases over land and ocean, and shows a good agreement with other techniques for the detection of particulate aerosol. However, the supervised texture approach provides outputs at a higher resolution than the existing methods and the same approach is applicable over land and ocean and therefore shows the advantages compared to the current techniques.

Furthermore, the coverage of the approach can be further extended using signature extension and chain classification. Signature extension was applied to one of the test cases to monitor the same geographical region with temporal extension away from the initial supervised classification. The experiments showed that it was possible to extend the coverage for ± 90 minutes from the original classification and indicates the possibility of greater extension over larger temporal windows.

Acknowledgements

I would like to thank my supervisors Dr. Adrian Evans and Dr. Ivan Astin for their advice and support during this project. In particular Dr. Evans for proof reading and suggestions for this thesis and finding the resources to allow me finish.

I would also like to thank Dr. Robert Watson for giving me the opportunity of employment at the university, his advice and support, and the moments of levity whilst I finished my thesis.

I am grateful for the help provided by my fellow lab mates; Dr Matt Foster for looking after me in my first year, and Dr Adrian Townsend and Dr Duncan Hodges for supplying the templates and numerous examples which provided me with a way of getting to grips with writing a thesis in LaTeX.

I would also like to acknowledge the financial support from EPSRC, without which I would not have been able to undertake the research.

I would like to acknowledge the BADC, the EUMETSAT Data Centre, the NASA Atmospheric Science Data Center and Goddard Space Flight Centre for providing the satellite data used, without which I would not have been able to complete the work.

Finally, I would like to thank my friends and family for supporting me throughout my Ph.D. Thank you to Sarah and Emma for all the lifts when I couldn't afford a car, to my parents for the free food, and to all the killer bees who made the second time round worthwhile.

Contents

List of Figures

List of Tables

Acronyms

ADRIEX	Aerosol Direct Radiative Impact Experiment
AERONET	Aerosol Robotic NETwork
AOD	Aerosol Optical Depth
AOT	Aerosol Optical Thickness
AVHRR	Advanced Very High Resolution Radiometer
BSFS	Backward Sequential Feature Selection
BT	Brightness Temperature
BTD	Brightness Temperature Difference
CAD	Cloud Aerosol Discrimination
CALIPSO	Cloud-Aerosol Lidar Infrared Pathfinder Satellite Observations
CCN	Cloud Condensing Nuclei
DABEX	Dust and Biomass Experiment
DAI	Deep blue Aerosol Index
DISORT	DIScrete Ordinates Radiative Transfer program
DODO	Dust Outflow and Deposition to Ocean
dpi	Dots per square inch
EARLINET	European Aerosol Research Lidar NETwork
EOS	Earth Observing Satellite
FAAM	Facility for Airborne Atmospheric Measurements
FOV	Field of View
FSFS	Forward Sequential Feature Selection
GFB	Gabor Filter Bank
GLCF	Grey Level Co-occurrence Features
GM	Geometric Mean
HDF	Hierarchical Data Format
HERA	Hybrid Extinction Retrieval Algorithm
IDDI	Infrared Difference Dust Index
IR	Infrared
LDA	Linear Discriminant Analysis
LUT	Look Up Table
MD	Mahalnobis Distance
MFG	Meteosat First Generation
MODIS	Moderate Resolution Imaging Spectroradiometer
MSG	Meteosat Second Generation
NDVI	Normalised Difference Vegetation Index
NERC	Natural Environment Research Council
NIR	Near Infrared
OMI	Ozone Monitoring Instrument

PCA	Principal Component Analysis
PCASP	Passive Cavity Aerosol Spectrometer
PSAP	Particle Soot Absorption Photometer
RGB	Red Green Blue colour model
RTM	Radiative Transfer Model
SCA	Scene Classification Algorithm
SeaWiFS	Sea-viewing Wide Field of view Sensor
SEVIRI	Spinning Enhanced Visible and Infrared Imager
SIBYL	Selective Iterated BoundarY Locator
SNR	Signal to Noise Ratio
SSA	Single Scattering Albedo
SVM	Support Vector Machine
SWIR	Short Wave Infrared
TIR	Thermal Infrared
TNR	True Negative Rate
TOMS	Total Ozone Mapping Spectrometer
TPR	True Positive Rate
UCL	Upper Control Limit
UV	Ultraviolet
VFM	Vertical Feature Mask
VIS	Visible
XOR	Exclusive OR

Chapter 1

Introduction

This chapter introduces the effect particulate aerosols have on multiple scientific disciplines, and explains why an increasing number of sensors networks and platforms are being implemented to specifically study particulate aerosols. There are three main platforms currently employed for monitoring tropospheric aerosols; ground-based networks, aircraft-based instruments, and satellites. Satellite based sensors offer the only viable approach for global monitoring of atmospheric aerosols and a survey of the current satellites and the techniques used are discussed in greater depth in Chapter 2.

Despite there being multiple dedicated sensors and variety of approaches, no one technique provides coverage over all terrestrial surfaces or at all times. Therefore, multiple approaches are currently required to monitor particulate aerosols and these outputs are still limited by the location of the sensors for ground- and aircraft-based sensors, or the repeat cycle of the satellite platforms.

1.1 The Importance of Particulate Aerosols

The apparent importance of Greenhouse Gases, such as carbon dioxide CO_2, methane CH_4 and Nitrous Oxide NO_x, and the effect their build up may have on global warming is well known. The global appreciation is reflected by 187 states ratifying the Kyoto Protocol, as of Nov 2009, which aims to stabilise and reduce the amount of greenhouse gas in the atmosphere to a level that would prevent dangerous anthropogenic interference with the climate [UNFCCC, 1997].

An emphasis has been placed on monitoring of anthropogenic aerosols, in particular

those released from combustion of fossil fuels, as they have a significant affect on radiation propagation and cloud nucleation, and due to a relatively recent massive influx of these aerosols into the atmosphere [Prospero, 1999].

One of the reason particulate aerosols have been overlooked is due to their presence in the atmosphere throughout human history as evidenced in ice and snow core records. In fact, human civilization is closely linked to the transportation of particulate minerals, which can create loess deposits in the soil which are highly fertile and are found where many early European and Asian Civilizations developed.

Mineral dust is a major constituent of particulate aerosols, and it has been estimated that up to 50% of the mineral dust present in the atmosphere is due to wind-blown mineral dust [Tegen et al., 1996]. Therefore, the ability to identify these deflated mineral distributions would contribute significantly to quantifying the affect of particulate aerosols has on the climate.

Other contributions to tropospheric particulate aerosols include smoke from biomass burning, urban haze from urban and industrial pollution, and ash from volcanic eruptions.

1.2 Affect of Aerosols on Climatic Processes

The recent increased awareness of the impact of particulate aerosols on climatic processes is partly due to strong affects with low volumetric ratios [Buseck and Psfai, 1999]. The affect on the climate is difficult to assess, which is in part due to an atmospheric lifetime of weeks compared to $10^2 - 10^3$ years of anthropogenic greenhouse gases, limiting the time available to detect and assess any affects on the climate and atmosphere [Schwartz and Andreae, 1996].

1.2.1 Direct Radiative Forcing

Radiative forcing measures the influence a factor has on the incoming and outgoing energy in the Earth's atmosphere, and is used to measure the importance of the factor as a potential climate change mechanism [IPCC, 2007], and is measured in watts per square metre (Wm^{-2}). Positive forcing means a net warming of the Earth's surface, whilst negative forcing represents a net cooling.

Particulate aerosols in the troposphere can have both a positive and negative forcing affect. Particles will more efficiently scatter light with a wavelength close to the particle size, and hence submicron particles will scatter incoming sunlight and have a cooling effect in the Earth's surface. Outgoing radiation is emitted as Infra red (IR), which can be absorbed by silicate materials and so mineral dust can also have a greenhouse effect [Buseck and Psfai, 1999, Kaufman et al., 2002a].

It has been suggested that mineral aerosols have a direct forcing of $-0.75 W m^{-2}$ [Duce et al., 1995], which is similar to that of sulphates released through the used of fossil fuels. If this is the case, atmospheric dust distributions could be masking positive forcing effects of anthropogenic greenhouse gases and explain why global temperatures have not increased by as much as climate models for greenhouse gases have predicted.

The overall forcing effect will vary depending on the exact makeup of the particulate aerosols and their size distribution, the geographical location and environmental factors such as humidity and the sunlight intensity.

1.2.2 Cloud Formation

Cloud droplets form around small particles, without which droplets would not be able to form under normal atmospheric conditions. The presence of particulate aerosols means the available condensed water is shared over a larger the number of cloud condensation nuclei (CCN) during cloud formation, giving a higher number of droplets but reducing the average droplet size by 20-30%.

The larger number of droplets increases the scattering within the cloud, with an associated rise in the cloud albedo, providing a cooling affect by reflecting some of the incident solar radiation. This represents an indirect forcing affect due to the presence particulate aerosols.

Clouds with smaller droplets are also more persistent as they are less likely to cause rain, and can exert their increased cooling affect for longer, compared to regular cloud formations.

1.2.3 Deposition of Mineral Dust

The main source of iron in the oceans away from river outflows, is through deposition of particulate aerosols [Buseck and Psfai, 1999]. The amount of Iron present has been

linked to a lower presence of plankton than that expected when the abundance of nutrients in the ocean is considered. Therefore the deposition of mineral dust may have a major effects on life at bottom of the food chain, affecting the ecosystem for large areas of the ocean.

1.3 Observing Particulate Aerosols

Monitoring of particulate aerosols presents a more complex task than monitoring greenhouse gases, due to the high spatial variability and short lifetimes. For example, as early as 1960, global increases in CO_2 concentrations were measured using a single ground-based instrument, due to the homogeneous distributions and long lifetimes. In contrast, continuous in-situ or satellite measurements are needed to monitor particulate aerosols on a global scale [Kaufman et al., 2002a].

Remote sensing from space offers a greater scope for global monitoring due to its greater coverage. The importance of monitoring the Earth's atmosphere to quantify the affects of anthropogenic greenhouse gases and particulate aerosols has led to an increasing number of satellites in recent years for this purpose.

The recently launched A-Train [L'Ecuyer and Jiang, 2010] is one attempt to increase the capability for monitoring the Earth's atmosphere and including particulate aerosols. The A-Train is a constellation of satellites operated by National Aeronautics and Space Administration (NASA) and the French government space agency, Centre National d'Etudes Spatiales (CNES) and includes multiple complementary sensors for measuring aerosols, clouds, and temperature amongst others. This includes instruments for multi-band spectral imaging (MODerate resolution Imaging Spectrometer, MODIS on AQUA), cloud profiling radar (CloudSat) and a space borne lidar (CALIPSO), and the Ozone Monitoring Instrument (OMI).

Despite the multiple dedicated platforms, remote sensing of aerosols remains a difficult problem, and no one satellite or sensor can provide complete coverage in all conditions [King et al., 1999a]. One of the major problems is that aerosol plumes are semi-transparent in satellite images, and particulate aerosols such as smoke and mineral dust have similar spectral signatures so it is difficult to distinguish between them.

Shao and Dong [2006] identify four main challenges for satellite remote sensing of mineral dust, namely

1. to provide real-time monitoring through dust storm identification from satellite imagery;
2. to derive land-surface and atmospheric parameters for dust modeling;
3. to derive physical quantities for validation of dust predictions and for data assimilation; and
4. to derive long-term dust climatology.

Whilst these challenges refer specifically to mineral dust, they are also applicable to the task of remote sensing of particulate aerosols in general.

Currently there are many techniques that perform this task but with limitations such as the terrestrial surface type or the time of day or with a limited temporal or spatial resolution. This thesis presents work that addresses the first point, presenting a new approach to identify particulate aerosols in satellite imagery which works without modification over ocean and terrestrial surface, in both day and night time images which has the potential for global coverage at a higher resolution than is currently possible.

1.3.1 Structure of this Thesis

In the following chapters a new approach to identifying particulate aerosol using satellite imagery is discussed. This new technique addresses the problems in the existing methods for aerosol detection which are reviewed in Chapter 2. The new approach uses supervised texture classification, which has not been applied to the task of remote sensing of atmospheric aerosols before. This utilises data from two distinct satellite sources, to provide a linear training set to form the classifier. Chapter 3 discusses the use of texture classification and the advantages when classes cannot be resolved spectrally. Chapter 4 investigates the suitability of linear training sets for supervised texture classifications, which has not been explicitly investigated in literature. Chapter 5 describes the new supervised texture method to identify atmospheric aerosol distributions presented in this thesis, with example results from mineral dust test cases over. Chapter 6 introduces signature extension and chain classification methods for extending the coverage of the new approach, including example results using signature extension for first order temporal classification as step towards global coverage. Chapter 7 continues with a summary of the conclusions of the work and the thesis concludes with possible future work in chapter 8.

Chapter 2

Existing Methods of Detecting Tropospheric Aerosols

This aim of this chapter is to provide a summary of the existing platforms and current approaches used to detect tropospheric aerosol distributions, with a focus on particulate aerosols. The emphasis is on providing an overview of satellite remote sensing approaches, as these are capable of providing global coverage. The products that are measured, the coverage provided, and associated failure modes of the different techniques are examined to provide a snapshot of the current capabilities of monitoring tropospheric aerosols.

2.1 Platforms used for Monitoring Atmospheric Aerosols

As discussed in chapter 1, multiple platforms are used to monitor atmospheric aerosol distributions. Satellite-based sensors offer the only viable method to achieve global coverage, but other platforms are still important in particular to provide validation data for satellite techniques, or providing high resolution observations of specific events. Here the three main sensor platforms are introduced with examples of each approach and how each approach is likely to be employed.

2.1.1 Ground-Based Sensors

Ground-based networks normally consist of a set of federated sensors that monitor the aerosol content in immediate area surrounding each site and, as such, provide detailed information on aerosol content at sparse and irregular sample points. Individual projects usually impose standardization on the sensors and instruments, calibration and processing to provide quality assurance when comparing data from different sites within the same network. Due to the limited coverage of ground-based networks, one of their main functions is to validate satellite aerosol retrievals and as such they have a significant input into global monitoring.

The AErosol RObotic NETwork (AERONET) is an example of ground-based network, established by NASA, CNES and the University of Lille [NASA, 2010b]. AERONET uses a network of sunphotometers to measure the direct aerosol optical depth (AOD), as well as using radiance measurements to produce aerosol optical properties such as the particle size distributions, single scattering albedo, phase functions and the complex refractive index.

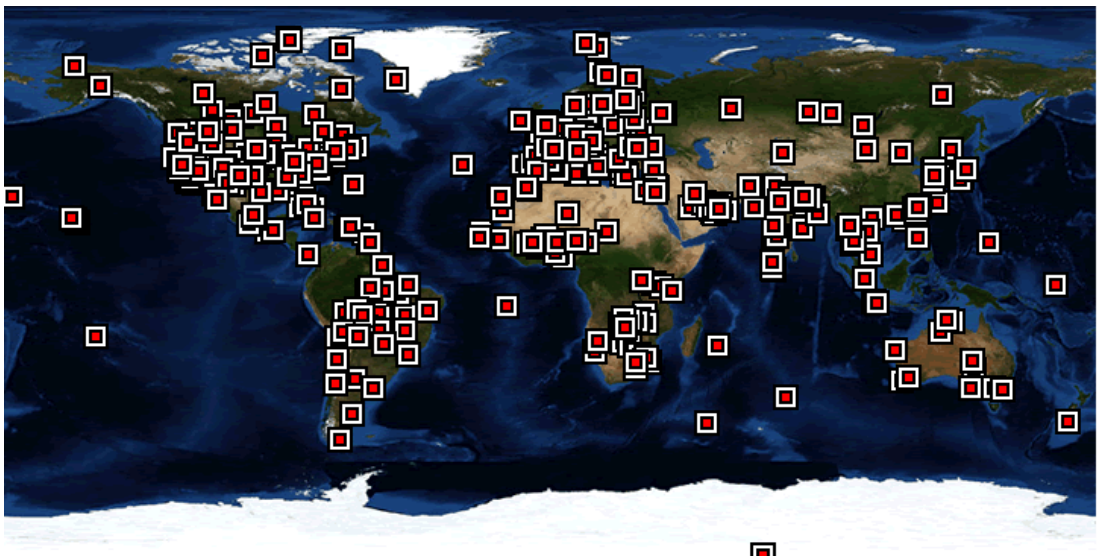


Figure 2.1: Distribution of AERONET sites around the World.

AERONET has established itself as the reference standard for measuring AOD due to it high accuracy, and the large number of sites [Liu et al., 2003] and as such is used to validate various satellite retrievals [Chu et al., 2002] [Remer et al., 2002] [Torres et al., 2002a].

Another example of a ground-based network is the European Aerosol Research LIdar NETwork (EARLINET), which consists of over 25 lidars stations distributed over Eu-

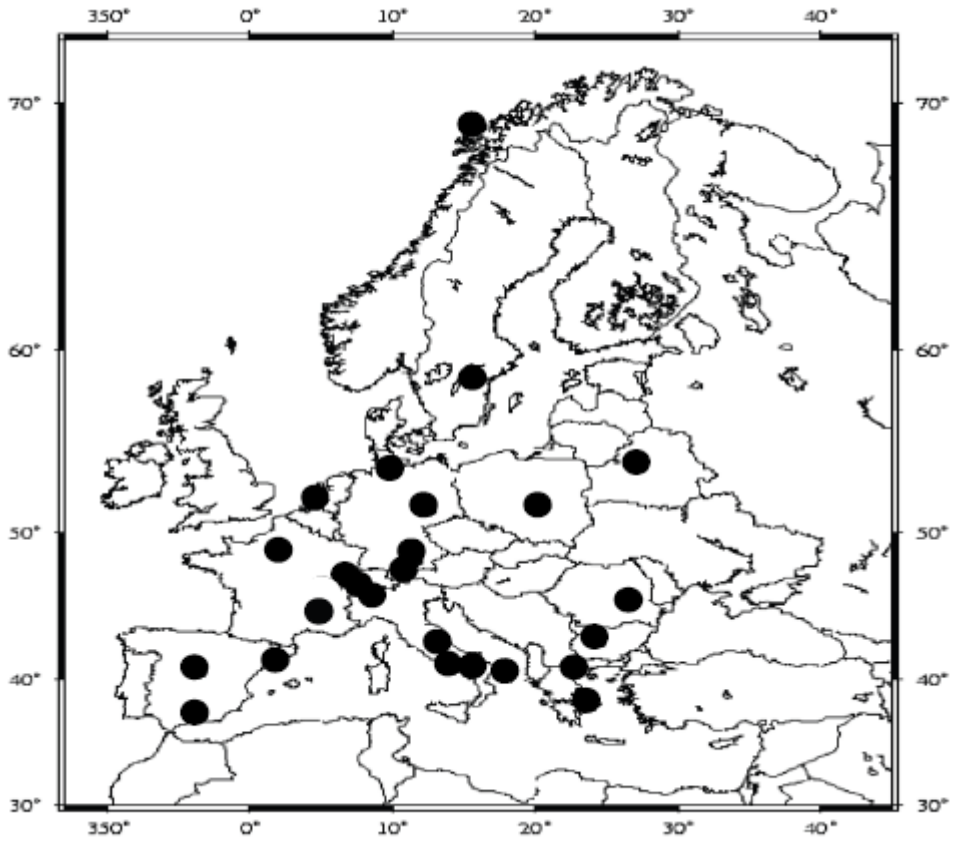


Figure 2.2: Distribution of EARLINET sites around Europe, reproduced from Pappalardo et al. [2009]

rope, see figure 2.2. EARLINET was established in 2000 to provide a comprehensive quantitative and statistically significant database for the aerosol distributions over Europe [EARLINET, 2011], and has recently contributed to the monitoring of volcanic ash from the Eyjafjallajökull eruptions [Wiegner et al., 2011].

2.1.2 Aircraft-Based Sensors

Tropospheric aerosol distributions can be examined using in-situ aircraft instruments. An example of which is the Facility for Airborne Atmospheric Measurements (FAAM) BAe-146 aircraft, which is a joint project of the Met Office and the Natural Environment Research Council (NERC). The aircraft is equipped with multiple core sensors to support the UK atmospheric research community on campaigns throughout the world, with the possibility of additional non-core to provide supplementary data. Of particular use for monitoring particulate aerosols are:

PCASP $(core)^{a,b}$**.** Passive Cavity Aerosol Spectrometer Probe , an optical probe used to measure aerosol particle sizes in the range of $0.1 - 3\mu$ m and records the concentration, mean volumetric ratios and the size spectrum.

The Nephelometer $(core)^{a,b}$**.** Rosemount pair aerosol inlet used measure the optical properties of aerosols by examining the scattering and back scattering coefficients at three wavelengths, 450nm , 550nm and 700nm.

PSAP $(core)^{a,b}$**.** Particle Soot Absorption Photometer, measures the aerosol Absorption coefficient.

CCN $(non - core)^{a,b}$**.** CCN instrument measures the number of Cloud Condensation Nuclei.

EZLidar $(non - core)^{a,c}$**.** Nadir viewing near-ultraviolet Lidar providing vertical profiles of atmospheric layers (clear air, aerosol, clouds) and allowing aerosol optical properties to be derived.

[a] [FAAM, 2010]; [b] [Highwood et al., 2007a]; [c] [Marenco, 2010]

Aircraft based sensors have been used for numerous campaigns to investigate atmospheric particulate aerosols, including DODO, DABEX [Formenti et al., 2008]; for measuring aerosol properties for use in radiative transfer calculations (ADRIEX) [Highwood et al., 2007b], validation of AERONET and satellite based aerosol retrievals [Christopher et al., 2009]. Most recently they have been employed for monitoring the volcanic ash fall out from Icelandic volcano Eyjafjallajoekull, determining the position, altitude and thickness of ash plume [Woolley, 2010].

Aircraft measurements provide detailed information on the physical and chemical properties of the atmospheric dust distributions. This allows fingerprinting of the dust to identify the source and also the impact their presence in the atmosphere and deposition will cause. Despite the detailed information aircraft measurements offer, they are costly and are limited by the area covered by the aircraft during its flight.

2.1.3 Satellite-Based Sensors

The existing methods employed for satellite-based aerosol detection and retrieval fall into three categories; radiative transfer model (RTM), brightness temperature difference (BTD) methods and space-borne lidar.

RTM and BTD approaches are similar in that they are employed on radiometric data and often both approaches are used to on data from the same multi-spectral sensor. These sensors can be geostationary, such as Meteosat Second Generation (MSG), or polar orbiting, such as Ozone Monitoring Instrument (OMI) on AURA and the Moderate-resolution Imaging Spectroradiometer (MODIS) on Terra (1999) and Aqua (2002).

The nadir viewing nature of these sensors measures the observed radiance at the top of the atmosphere over a relatively large field of view (FOV). As such, they offer good horizontal resolution, limited by the raw data from the sensor, but little information about the vertical profile of the atmospheric aerosols. Therefore, using RTM and BTM will yield 2D aerosol products with information in the horizontal plane.

Space-borne Lidar represents a significantly different approach to RTM and BTD methods. Lidars are active sensors, emitting laser pulses in the ultra violet (UV), visible (VIS), and near infra red (NIR) spectral range, which allows the range of objects to be determined by measuring the time delay between transmission and detection of a reflected signal. Space-borne Lidar can therefore examine the properties of aerosols via their interaction with the incident pulses through the observed backscatter, and combines this with the range to create a vertical profile of the atmosphere. Lidar outputs therefore contain excellent vertical resolution, but have a limited FOV, and are therefore mounted on polar orbiting platforms to give a continuous vertical profile directly underneath the satellite platform.

The specific techniques employed by these different satellite-based sensors are discussed in detail in the next section, and further information on the horizontal, vertical and temporal resolution of these outputs are discussed.

2.2 Methods of Observing Aerosols using Satellite-Based Sensors

The previous section introduced three main techniques used by satellite-based sensors to observe atmospheric aerosols. Details are given on how these techniques are described in the following section, including the sensor platforms where they are utilised and the strengths and weaknesses of the different approaches.

It is common for researchers to develop their own algorithms tailored to a specific sensors, although methods exist that can be applied to the raw sensor data and are

applicable to multiple platforms. The techniques included here are intended to give a snapshot of the current capabilities for the remote sensing of particulate aerosols from space.

2.2.1 Radiative Transfer Model Methods

These methods are based around estimating the surface contribution to the radiance observed at the top of the atmosphere and employ models of aerosols to explain the difference between the measured and expected clear sky radiance. This approach infers the presence and aerosol type, which leads to a dust mask, simultaneously with the retrieval of the aerosol properties such as aerosol optical thickness (AOT) and mass concentration. This is achieved by comparing the received and expected radiance and using a pre-calculated table to infer the type and amount of aerosol that could lead to the observed difference.

The accuracy of the results is dependent on the ability to calculate the proportion of the measured radiance that is due to aerosols and the ability of the aerosol models to represent the aerosols present. The key problem for retrievals performed from satellites is how to separate the observed spectral reflectance into atmospheric and surface contributions. Over the ocean the surface contribution is assumed to be negligible, as the surface reflects virtually none of the infra red wavelength. Over land, however, the reflectance varies with surface type and has a contribution similar in magnitude to the observed reflectance of the atmosphere. The methods for separating surface and atmospheric components of the measured radiance vary between algorithms and have differing levels of complexity.

Dark Target Method

Retrieval of the AOT from radiometric data is achieved by modeling the reflectance and scattering of solar radiation. The AOT can be inferred by accounting for radiation from surface reflectance, scattering by the atmosphere and absorption by gases and aerosols. The reflected radiation at the top of the atmosphere expressed as a reflectance function [King et al., 1999b]:

$$R(\tau_a, \omega_0; \mu, \mu_0, \phi) = \frac{\pi I(0, -\mu, \phi)}{\mu_0 F_0} \tag{2.1}$$

where, $I(0, -\mu, \phi)$ is the reflected intensity or radiance; τ_a is the aerosol optical depth; ω_0 is the single scattering albedo; μ is the absolute value of the cosine of the zenith angle, θ_0, measured with respect to the downward direction; ϕ is the relative azimuth angle between the emerging radiation and incident solar radiation; and μ_0 the cosine of the solar zenith angle, θ_0. This expression normalises the emerging radiation with the incident solar flux, F_0 , and so represents the albedo of the Earth-atmosphere for isotropic radiation.

For a cloud free vertically homogeneous atmosphere over a Lambertian surface with reflectance, A_g, the reflectance function in equation 2.1 becomes:

$$\begin{aligned} R(\tau_a, \omega_0; \mu, \mu_0, \phi) =& R_a tm(\tau_a, \omega_0; \mu, \mu_0, \phi) + \frac{A_g}{1 - A_g r_a \bar{t}m(\tau_a, \omega_0)} \\ & \cdot t_{atm}(\tau_a, \omega_0; \mu)\, t_a tm(\tau_a, \omega_0; \mu_0) \end{aligned} \tag{2.2}$$

where $R(\tau_a, \omega_0; \mu, \mu_0, \phi)$ is the reflectance function; $r_a \bar{t}m(\tau_a, \omega_0)$ is the spherical albedo; and $t_a tm(\tau_a, \omega_0; \mu_0)$ is the total transmission (diffuse plus direct) when $A_g = 0$. The Lambertian reflectance is often replaced by the bidirectional reflectance properties of the surface when equation 2.2 is used on satellite observations [Lee and Kaufman, 1986].

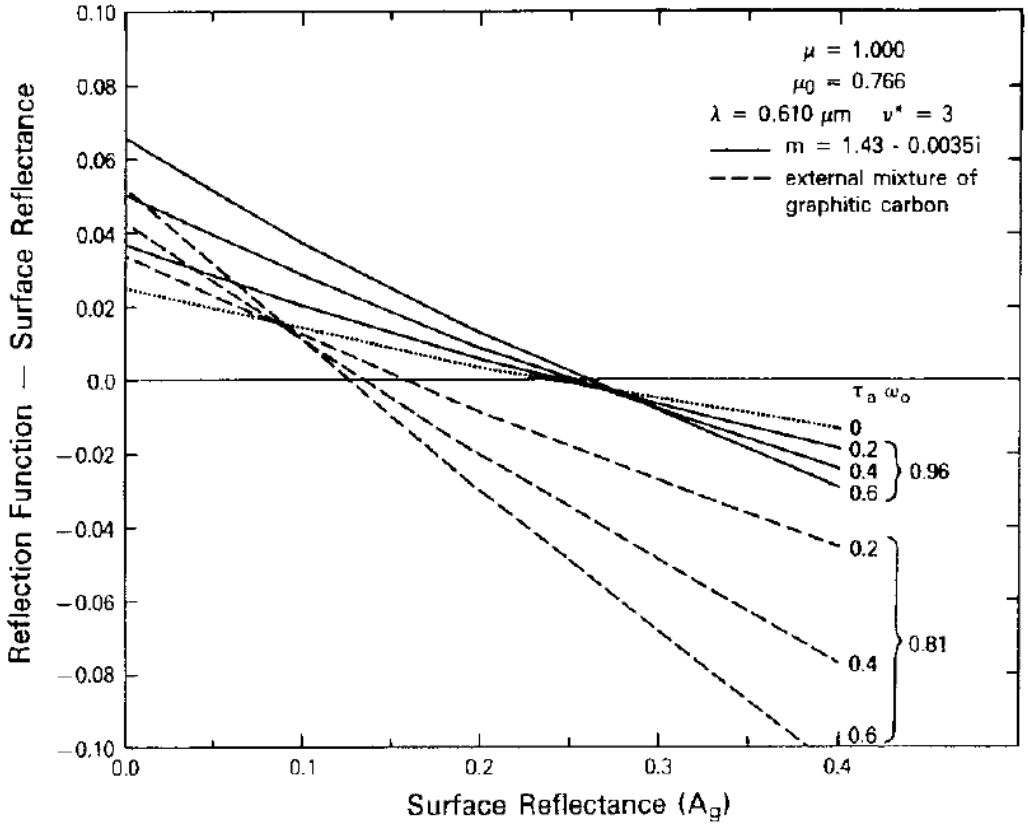


Figure 2.3: Difference between the reflectance function and the surface reflectance for various values the aerosol optical thickness τ_a and single scattering albedo ω_0 [King et al., 1999b]

The effect of aerosol optical depth and single scattering albedo on the reflection function is shown in figure 2.3. This shows the difference between the reflectance function and surface reflectance for four different values of aerosol optical thickness ($\tau_a = 0.0, 0.2, 0.4, 0.6$) and two values of the single scattering albedo ($\omega_0 = 0.81$ and 0.96). There is an increased sensitivity to aerosol optical depth when the surface reflectance, A_g, is less than 0.1. This implies that the aerosol optical depth τ_a can be measured over dark surfaces such as the ocean, and dark targets over land. Consequently, this approach is referred to as the dark target method.

The MODIS sensor (on Aqua and Terra) uses a dark target method as the main method for performing aerosol retrieval. The MODIS aerosol retrieval algorithm is split into two parts covering retrievals over land and ocean. The core idea of both approaches is to use a radiative transfer model to pre-compute a look up table (LUT), for a set of aerosols with log normal size distributions and various surface parameters. The algorithm assumes that the aerosol properties within a scene can be adequately described by combining one fine and one coarse aerosol mode, with appropriate weightings. The

observed spectral reflectance is then compared to the LUT and a least-squares fit is used to find the best solution [Remer et al., 2009].

Over the ocean, the surface contribution is assumed to negligible and retrievals can be directly performed on the observed radiance. Over the terrestrial surface, the surface contribution can not be ignored and needs to be removed from the observed radiance before performing an inversion. [Kaufman et al., 2002b] showed that the surface reflectance's over vegetation and dark soils in visible wavelengths are correlated with the reflectance in shortwave infrared (SWIR) channels. This enables the surface contribution to the reflectance to be found and then removed, allowing the atmospheric contribution to be inverted and the recovery of the aerosol information.

The MODIS aerosol products have a $10km \times 10km$ resolutions at nadir, compared to $250m$, $500m$, $1km$ resolution for the individual channels, as the mean radiance of the good pixels with the $10km \times 10km$ window is used to perform a retrieval. 'Good' pixels defined as being over a suitably dark background and free from other contaminants. Over land, no retrievals are performed if the $2.1\mu m$ surface reflectance is above 0.15, which prevents coverage for larger areas of land, especially over the desert surfaces which are a significant source of atmospheric dust [Hsu et al., 2004]. The MODIS cloud mask [Martins et al., 2002] and snow/ice mask [Li et al., 2005] are used to mask contaminated pixels, as the presence of water has a similar affect on the observed radiance as aerosols. Inland water bodies are also masked using the Normalized Difference Vegetation Index (NDVI) [Remer et al., 2005].

After masking contaminated pixels, and identifying dark pixels, a retrieval is only performed if there 12 or more pixels. This means that when dust clouds are interspersed with clouds, or over a bright surface, no information can be retrieved. This presents a problem when attempting to identify and track mineral dust over its bright surface desert source.

Deep Blue Method

The Deep Blue algorithm was developed to address the limitations of dark target model for aerosol retrievals over bright land surface such as deserts and urban areas [Hsu et al., 2004]. The algorithm has been designed to work with the Sea-viewing Wide Field-of-view Sensor (SeaWiFS) [Hooker, 1992] and both MODIS instruments currently in operation. Deep Blue utilises radiance measurements from blue channels (412 nm, 490 nm and 670 nm) and uses a polarized radiative transfer model to calculate the reflected intensity field at the top of the atmosphere from a database of precalculated surface

albedo measurements.

A cloud screening procedure based on the spatial variance is used to prevent retrievals with cloud contaminated pixels. Additionally, to distinguish between thick dust layers and clouds the Deep Blue aerosol index (DAI) is used, which is a similar measure to the Total Ozone Mapping Spectrometer (TOMS) aerosol index. The measured radiance is then compared to a LUT generated using smoke and dust models with various values for solar, satellite, and azimuth angles, the surface reflectance, AOT, and single-scattering albedo. A Maximum Likelihood method is used to find the best match of AOT and single-scattering albedo for the given radiance, and hence provide the spatial distribution of a given aerosol. Assumptions are placed on the dominant aerosol model used for the retrieval based on the geographical location and time of the year. Results over sites in Nigeria and Saudi Arabia have been compared to AERONET, showing good agreement with results for AOT consistently within 20% to those measured by sunphotometers [Hsu et al., 2004]].

The Deep Blue algorithm produces outputs at the resolution of the data from the satellite. MODIS channels 8 (412 nm) , 10 (490 nm), and 13 (670 nm), have a resolution of 1 km^2 at nadir, giving deep blue retrievals a higher resolution than the normal MODIS aerosol retrieval with the additional benefit of working over bright surfaces. However this approach is only be applicable over bright surfaces, e.g. arid, semiarid, urban and sparse vegetation surface, and so limits where the deep blue algorithm can be applied.

OMI aerosol algorithm

The objectives for the Ozone Monitoring Instrument (OMI) on the EOS Aura Satellite mission include monitoring atmospheric constituents that affect the Earth's climate and, as such, includes an aerosol retrieval algorithm [NASA, 2011]. OMI is the successor the Total Ozone Mapping Spectrometer (TOMS) series of satellites of satellites which has a proven contribution to monitoring global aerosol sources over the past 25 years. OMI provides daily global coverage with a spatial resolution of $13 \times 24\, km^2$, and is located in the A-train constellation with EOS-Aqua and CALIPSO amongst others. OMI provides hyper-spectral data in the visible (VIS) and ultraviolet (UV) spectrum, which is advantageous over infrared based methods as retrievals are possible over the whole of the terrestrial surface including arid desert regions.

In a manner similar to MODIS, radiative transfer calculations are used to generate a LUT for 24 different aerosol models drawn from 5 major aerosol types, covering

urban/industrial, biomass burning, desert dust, oceanic and volcanic sources. The LUT are generated for various solar, viewing and relative azimuth angles, surface pressures and aerosol concentrations. OMI has two different retrieval modes; the multi-wavelength method using 17 different spectral bands, and the near UV method which uses only two spectral bands [Stammes and Noordhoek, 2002].

The multi-wavelength method is predominantly used over ocean, while the near UV approach used over land, but overall the approach is fundamentally the same for both methods. Cloud screening is used to mask cloudy pixels, as retrievals are only possible for cloud free scenes. Geographical and spectral data is used to select a subset of candidate aerosol models used to fit the observed radiance to the LUT which is corrected using a surface reflectance database returning the AOT, single scattering albedo (SSA) and the dominant aerosol type.

OMI cannot perform retrievals when aerosols are interspersed with cloud. OMI and its predecessor instrument, TOMS, have proven track records for monitoring atmospheric aerosols and OMI does have a robust method for retrievals over bright surfaces [Ahn et al., 2008, Torres et al., 2002b]. The major drawback of OMI retrievals is the low resolution compared to other methods such as MODIS and Deep Blue retrievals, which only allows observations of large-scale trends such as daily and monthly average of the aerosol optical depth

Infrared Difference Dust index

The Infrared Difference Dust Index (IDDI) is a dust product derived from Satellite IR imagery. Originally, the IDDI was applied to first generation Meteosat Satellite IR images, but these have been replaced by images from the Spinning Enhanced Visible and Infrared Imager (SEVIRI) sensor on Meteosat Second Generation (MSG) satellites. The IDDI is the simplest of the RTM methods considered in this section, as the aerosol contribution to the measured radiance is not directly modeled. Instead, it is assumed to be solely due to the presence of mineral dust, which geographically limits where the IDDI can be applied.

The IDDI uses a quasi-linear relationship between satellite detected radiance and the shortwave optical depth of dust [Legrand et al., 2001]. The affect of dust on the radiance detected is examined by comparing an instantaneous IR image with a reference image representing the clear sky radiance. The reference image is built up over a period of days and assumes that the maximum radiance detected at each pixel equates to its clear sky response. The difference between the instantaneous and reference images can

then be used to determine the AOT.

The IDDI is calculated at midday as this maximizes the sensitivity of the outgoing IR radiance to dust. As water vapour has a similar radiative forcing effect as dust clouds it can produce false dust readings. This effect can be minimized by only calculating the IDDI over arid regions such as the Sahara and Sahel regions of North Africa and by using cloud masks to exclude affected parts of the image where clouds are present.

These requirements mean that the IDDI is a geographically limited metric, which can only be applied to arid regions away from the coast. Therefore, the IDDI does not have global coverage and its accuracy is limited by that of the cloud mask and the quality of the clear sky reference image.

2.2.2 Brightness Temperature Difference Methods

Brightness Temperature Difference methods use the Thermal IR (TIR) channels to retrieve information about optical thickness and particle size. The use of TIR channels means that these techniques can be applied in both day and night conditions and over bright surfaces. The 8-10m and 10-12m channels show a strong spectral variation in the transmittance for many aerosols [Ackerman, 1997], whilst the atmosphere is relatively transparent. Therefore, BTD techniques can be used to recover information about the aerosols present.

Ackerman used theoretical simulations of mineral dust using a spherical particle model to explain the satellite BTD observations. These simulations showed a quasi-linear relationship between $BT_{11} - BT_{12}$ and $BT_8 - BT_{11}$ and the AOT, where BT_x represents the equivalent brightness temperature of channel x, and $BT_x - BT_y$ represents the difference between the equivalent brightness temperature of channels x and y. Ackerman [1997] also suggested that negative values of $BT_{11} - BT_{12}$ are useful for observing dust storms over bright surfaces where radiative transfer methods traditionally fail.

MSG dust enhancement

The MSG dust enhancement is an example of a trispectral BTD method and is used to identify mineral dust outbreaks using images from the SEVIRI sensor on MSG satellites. The BTD of various channels is used in conjunction with range clipping and gamma correction to form an RGB image, see table 2.1.

Table 2.1: MSG dust RGB composite settings, where BT_x represents the equivalent brightness temperature for the spectral band centred on $x\,\mu m$ and $BT_x - BT_y$ is the difference between two brightness temperature images.

Beam	Channel	Range	Gamma
Red	$BT_{12} - BT_{10.8}$	$-4 - 2K$	1.0
Green	$BT_{10.8} - BT_{8.7}$	$0 - 15K$	2.5
Blue	$BT_{10.8}$	$261 - 289K$	1.0

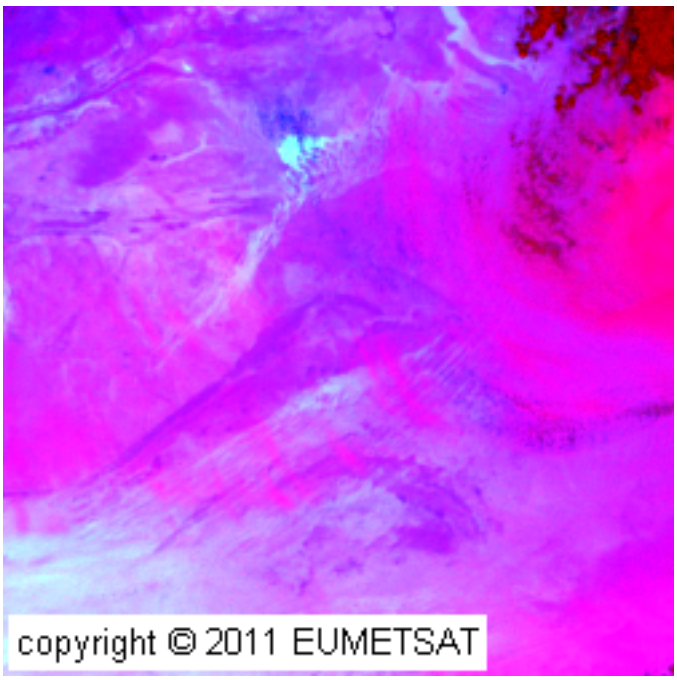


Figure 2.4: Example of MSG dust product. Dust is indicated as magenta in the RGB image

The mineral dust appears as magenta within the RGB image, with the tone varying depending on the time of day and the background surface. As such, it is only a qualitative measure of atmospheric dust distributions. An example of the MSG dust product is included in figure 2.4.

Zhang's BTD product

Zhang et al. [2006] built on the work by Ackerman [Ackerman, 1997] to develop a dust storm mask, which can be used to identify dust outbreaks. The observations used the discrete ordinates radiative transfer program (DISORT) coupled with Mie scattering code as a forward model to simulate behaviour observed in the BTD images. Furthermore, the modelling showed that the AOT and particle size can be retrieved using a pre-calculated LUT if the underlying surface temperature is known.

Table 2.2: Zhang BTD dust mask, reproduced from [Zhang et al., 2006]. BTD(x-y) is the difference between the equivalent brightness temperature images for the spectral band $x\,\mu m$ and $y\,\mu m$.

Threshold	Mask Flag	Description
BTD(11-12) >-0.5 and BTD(8-11) >0	1	Relative strong dust region
BTD(11-12) >-0.5 and BTD(8-11) <0	2	Relative weak dust region
BTD(11-12) >0 and BTD(8-11) >0	3	Ice cloud
BTD(11-12) >0 and BTD(8-11) <0	4	Low cloud or surface
0 >BTD(11-12)>-0.5	5	Uncertain region

The dust storm mask and retrieval algorithm was developed to be used on MODIS images but has been shown to work with MSG data [Li et al., 2007], giving results at resolutions of $1\,km^2$ and $3\,km^2$ at Nadir, respectively. This approach provides global coverage through MODIS and MSG data, and has the ability to detect dust storms during day and night. However the mask was developed on a specific severe dust storm in Northern China in April 2001 and so may not be globally applicable for example for African or Arabian dust events.

2.2.3 Space-Borne Lidar: CALIOP on CALIPSO

The primary payload of the Cloud-Aerosol Lidar and Infra red Pathfinder Satellite Observation (CALIPSO) platform [NASA, 2010a] is the Cloud-Aerosol Lidar with Orthoganol Polarization (CALIOP) sensor. CALIOP is equipped with two redundant Nd:YAG lasers, of which only one is used at a time, producing linearly polarized beams at wavelengths 532 nm and 1064 nm. CALIOP uses three receiving channels; 1 to measure the total backscatter intensity at 1064 nm and 2 channels to measure the orthogonally polarised components of 532 nm backscatter, which are used to provide vertical profiles of the atmosphere.

The primary products for CALIPSO are those derived from the CALIOP sensor. An introduction to the CALIPSO retrieval and data products can be found in [Vaughana et al., 2004], which breaks down the processing into three areas each with their own algorithm theoretical basis document. The three areas are:

Feature detection and Layer Properties [Vaughan et al., 2005] These algorithms are concerned with identification of "features"within the backscatter signals that correspond to targets of interest, namely clouds, aerosols and surface returns. Until the feature boundaries are identified, the subsequent tasks of feature identification and property retrieval cannot be achieved.

Scene Classification [Liu et al., 2005] The first task is to determine whether the feature is cloud or aerosol using the cloud-aerosol discriminations (CAD) algorithm, and then once this has been determined use the scene classification algorithms (SCA) to determine the cloud ice-water phase and the cloud and aerosol subtype.

Extinction Retrieval and Particle Properties [Young et al., 2008] Once the feature type has been determined, the optical properties can be retrieved. Multiple scattering effects cannot be ignored in space-borne lidar, and so feature-dependant corrections are applied, producing optical depths for each feature within the scene

Due to the mixture of strong and weak features, and the inherent noise in the signals, horizontal and vertical averaging is required to enhance feature detection and lead to accurate retrievals. This is achieved using a Selective, Iterated Boundary Locator (SIBYL) to identify feature boundaries, and the Hybrid Extinction Retrieval Algorithm (HERA) and only once these are complete can the cloud and aerosol results be extracted for use in the output products [Liu et al., 2005].

Outputs from CALIPSO are produced in three formats: the vertical feature mask (VFM) containing the "what" and "where" information on the vertical structure of the atmosphere along-track; cloud and aerosol layers providing statistical descriptions of the detected feature, and separate profile products mapping the vertical distribution of backscatter and extinction coefficients for clouds and aerosols [Vaughana et al., 2004].

The VFM is the key product for identification of atmospheric aerosols from CALIPSO, providing breakdown of the atmospheric layers including the aerosol subtype. Cloud and Aerosol discrimination utilises the different scattering properties of clouds and aerosol, using the backscatter colour ratio $\chi = \beta_{1064}/\beta_{532}$ where β_{1064} and β_{532} is the backscatter at 1064 nm and 532 nm respectively.

Clouds generally have large particle size with respect to the CALIOP wavelengths, which means there will be no substantial spectral variation in backscatter and hence $\chi_{cloud} \approx 1$. In contrast, for most aerosols the particles are much smaller and so the backscatter exhibits spectral dependence, giving $\chi_{aerosol} < 1$. The ability to discriminate between cloud and aerosols in this manner can be seen in figure 2.5.

Aerosol subtyping is achieved with model matching scheme using using the following characteristics:

- optical - $\chi, \beta_{532}, \delta_v$ the depolarization ratio, which is proportional to the hydration

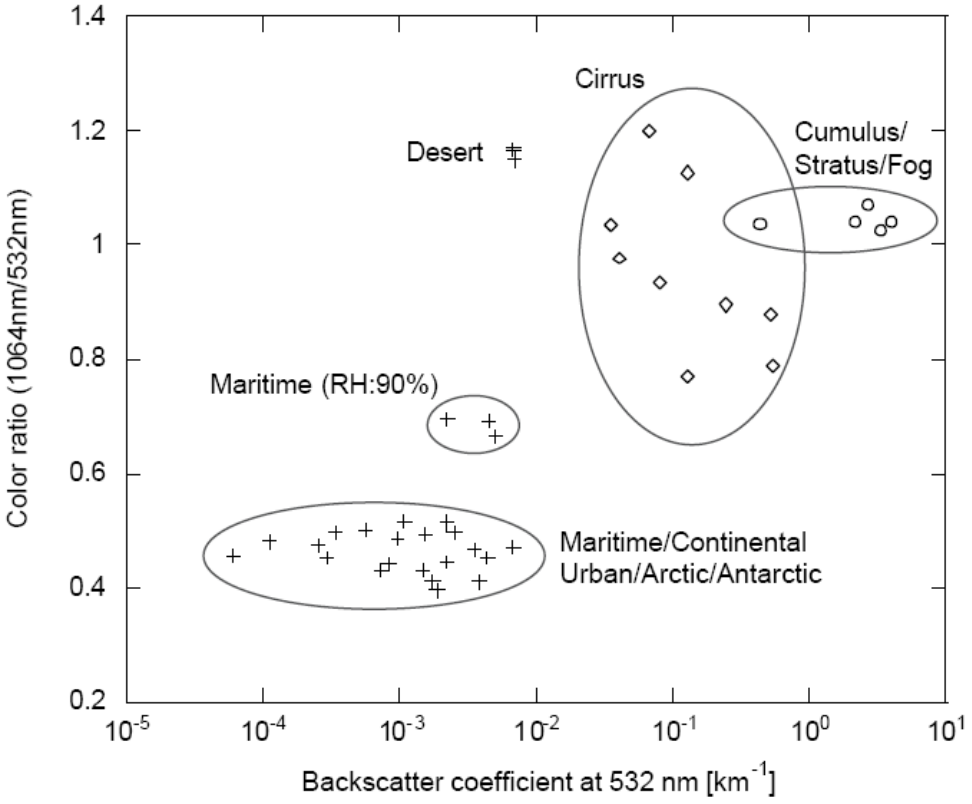


Figure 2.5: modelled scattering properties for aerosol and clouds, reproduced from [Liu et al., 2005]

state

- geophysical - latitude and longitude to indicate surface type
- temporal - to indicate season, as aerosol loading can be seasonally dependant

These characteristics are used to create decision points for selecting the most likely aerosol type for each layer of aerosol identified. Six different aerosol subtypes are defined; polluted continental, biomass burning, desert dust, polluted dust, clean continental and marine. The properties of the first four are derived from cluster analysis of AERONET data, and clean continental and marine aerosols were synthesized using long-range continental transport and sea salt observations using a backscatter nephelometer respectively.

Due to the active nature of the CALIOP sensor, the CALIPSO VFM has several advantages over passive imaging approaches for aerosol identification. The CALIPSO VFM is able to return the aerosol types over all surface types, and the range information from the LIDAR allows the vertical structure of aerosol distribution to be found, and allows for observations when aerosols are occluded preventing detection in passive imaging approaches.

The drawbacks of CALIPSO VFM are the limited FOV of $\approx 100\,m$ and the polar orbit, which limits the area that can be observed, which prevents CALIPSO from being able to provide near-continuous monitoring of particular aerosol events.

2.3 Summary of Existing Products for the detection of Tropospheric Aerosols

This section includes a table summary of methods and platforms discussed in the previous sections to indicate the current capabilities and to identify where there is scope for new approaches.

Whilst Aeronet and the FAAM BAe-146 offer the most reliable techniques for the detection of particulate aerosols, the nature of the platforms mean they provide limited coverage and in the case of the FAAM BAe-146, is expensive to operate for continuous monitoring. Satellite platforms therefore provide the only real opportunity for global monitoring and detection of particulate aerosols. MODIS and OMI provide the most robust radiometric approaches for aerosol retrieval but these have their limitations. The standard MODIS retrieval approach fails over bright surfaces and in the

presence of water vapour and cloud, the Deep blue algorithm proposed to counter this allows MODIS to provide retrievals over bright surfaces, but has limited applicability over other surfaces. The polar orbit of the MODIS platforms also prevent continuous monitoring of a specific event, such as volcanic ash cloud or smoke from a forest fire. OMI is similarly limited by its polar orbit, and fails with the present of cloud but the wavelengths used allows retrievals over a wider range of surfaces compared to MODIS. The resolution of the OMI products is also an issue, with OMI having the coarsest resolution of all the radiometric approaches.

The CALIPSO VFM is extremely useful product, with the active LIDAR giving the ability to provide vertical profile of the atmosphere and the ability to discriminate between aerosol types and clouds over all terrestrial surface. However the field of view and coverage is extremely limited, and is dependant on the overpass coinciding with a particulate event in order to provide data. Despite these limitations, the CALIPSO VFM can still be utilised to provide accurate aerosol classification along its overpass which can be combined with another satellite source to increase the coverage.

Table 2.3: Summary of Techniques for the Detection of Tropospheric Aerosols

Name	Sensor Type	Coverage	Resolution	Technique	Products	Comments
Aeronet [NASA, 2010b]	Ground based Sun photometer network	Sparse and irregular sample points based on sensors distribution	N/A	Sun Photometer	AOD_{500nm}, Size Distribution, fine and Coarse Aerosol Mode	Limited coverage but accurate results. Reference Standard for Satellite based techniques.
FAAM BAe-146 [FAAM, 2010] [Highwood et al., 2007a] [Marenco, 2010]	Multiple on Aircraft (see section 2.1.2)	Determined by flight plan	N/A	Multiple: Photometer, Multispectral Optical Probe, Lidar	Aerosol Particle size, concentration, volumetrci ratios, Scattering Properties, No. of CCN, Vertical Profile, layers and AOD from EZLidar	Used for detailed monitoring of specific events, does not provide global coverage.
MODIS Aerosol Product [Remer et al., 2009]	Multispectral Spectroradiometer on TERRA and AQUA satellites	2330km cross track on 16 day polar orbit	$10km \times 10km$ at nadir	Radiative Transfer Model, Dark Target Approach	AOD at multiple wavelengths $(470 - 2130\,nm)$, Aerosol type, CCN, Mass concentration	Requires sufficient pixels over a dark target to function. Unable to perform retrievals in presence of snow/ice and cloud or over bright surfaces.
MODIS Deep Blue [Hsu et al., 2004]	Multispectral Spectroradiometer on TERRA and AQUA satellites	2330km cross track on 16 day polar orbit	$1km \times 1km$ at nadir	Radiative Tranfer Model	AOD and SSA at multiple wavelengths $(412 - 659\,nm)$, Aerosol type.	Higher resolution than the normal MODIS aerosol but limited to bright surfaces.
OMI Aerosol [Stammes and Noordhoek, 2002]	OMI multispectral sensor on AURA satellite	2600km cross track	$13 \times 24\,km^2$	Radiative Transfer Model	AOT, SSA, Dominant aerosol type	Technique has proven track record on TOMS. Low resolution better suited for daily / monthly trends. Does not work when aerosols are interspersed with cloud

Table 2.4: Summary of Techniques for the Detection of Tropospheric Aerosols

Name	Sensor Type	Coverage	Resolution	Technique	Products	Comments
CALIPSO VFM [Vaughana et al., 2004]	CALIOP Lidar linearly polarized beams at wavelengths 532 nm and 1064 nm 3 receiving channels; backscatter intensity at 1064 2 channels orthogonally polarised components of 532 nm backscatter	FOV $\approx 100\,m$, 16 day polar orbit	$0 - 8.2km \approx 333m$ horizontal $30m$ vertical, $8.2 - 20.1km \approx 1km$ horizontal $60m$ vertical, $20.2 - 30.1km$ $\approx 1.667km$ horizontal $120m$ vertical	Model matching and clustering scheme	AOD, Backscatter, Aerosol type	Active sensor allows aerosol type retrieval over all surfaces, gives vertical structure of atmosphere, limited FOV and coverage prevents continuous monitoring of events
Zhang BTD product [Zhang et al., 2006]	Applied to Multispectral Radiometer Data (MODIS, MSG)	MSG: Geostationary, centered on prime meridian. 15 minute repeat cycle MODIS: 2330km cross track on 16 day polar orbit	$3km \times 3km$ at nadir $3km \times 3km$ at nadir	BTD	Dust mask, AOT, particle size	Works day and night due to use of IR channels, depending on satellite coverage. Designed for severe dust storms over Northern China, and may not be applicable at other locations

Chapter 3

Supervised Texture Classification

This chapter introduces the idea of image texture and its applications. Image texture can be used for classification problems which cannot be resolved spectrally. In these instances, image texture provides a method to separate the classes within feature space. Multiple different texture features schemes exist in literature, with two of the most widely used approaches, Grey Level Co-occurrence Features (GLCF) and Gabor Filter Banks (GFB) presented. These texture schemes have also previously been used in the analysis of remotely sensed imagery. This chapter also includes an introduction to supervised classification methods and compares this approach to unsupervised approaches. This includes the importance of representative training samples and the effect on the accuracy of supervised classification, and how the definition of representative samples differs depending on the classifier selected. Finally a method of using supervised classification for identifying atmospheric aerosol distributions in satellite imagery is presented.

3.1 Image Texture

In many image analysis tasks, it is common to assume a certain level of uniformity of spectral intensity over a local region within the image. This assumption does not always hold for real images and can therefore cause errors particularly for classification tasks. The local variation of spectral intensity for a region of a specific object or class is the image texture. The texture pattern may be due to the physical structure of the surface which directly relates a tactile pattern to an image texture. Image texture can also be be due to difference in the reflectance properties of a surface.

Image texture can be easily be interpreted by the eye, but is hard to define in machine learning tasks. The definitions of how to define a image region as belonging to a single texture varies throughout the literature, with a summary given in [Tuceryan and Jain, 1993]. The most useful definition for classification tasks is:

Definition 1 *Texture regions are defined as equivalent if the statistics or measures selected are the same, slowly varying or periodic within the given region. [Sklansky, 2007]*

Image texture is widely used in remote sensing applications for a diverse set of tasks including aerosol detection [Khazenie and Lee, 1992], oil slick detection [Brekke and Solberg, 2005], lithiological discrimination [Chica-Olmo and Abarca-Hernandez, 2000] and land cover in synthetic aperture radar (SAR) images [Fukuda and Hirosawa, 1999].

Texture features can be seen as a way of deriving multiple features from a single spectral band. When image texture analysis was first proposed, the number of spectral channels was small and so image texture feature provided additional features to improve the accuracy and reliability of machine learning outputs. With the advent of multi- and hyper-spectral sensors, the amount of spectral information available is greatly increased. Given the vast quantities of spectral data now available, texture analysis is frequently employed when different classes within an image share a similar spectral signature are not separable spectrally, thus demonstrating the continued importance of image texture analysis.

3.1.1 Grey Level Co-occurrence Features

Grey level co-occurrence features were one of the first attempts to model image texture explicitly [Haralick et al., 1973]. Haralick et al. [1973] proposed using neighbouring grey-tone spatial dependencies to create the grey level co-occurrence matrix (GLCM) from which texture features are derived. A 2D digital intensity image is represented at each pixel by one of N_G grey levels, with $G \in 0, 1, ...N_G - 1$ representing the set of grey levels in the image. Each pixel, excluding edge pixels, has 8 nearest neighbours, see figure 3.1.

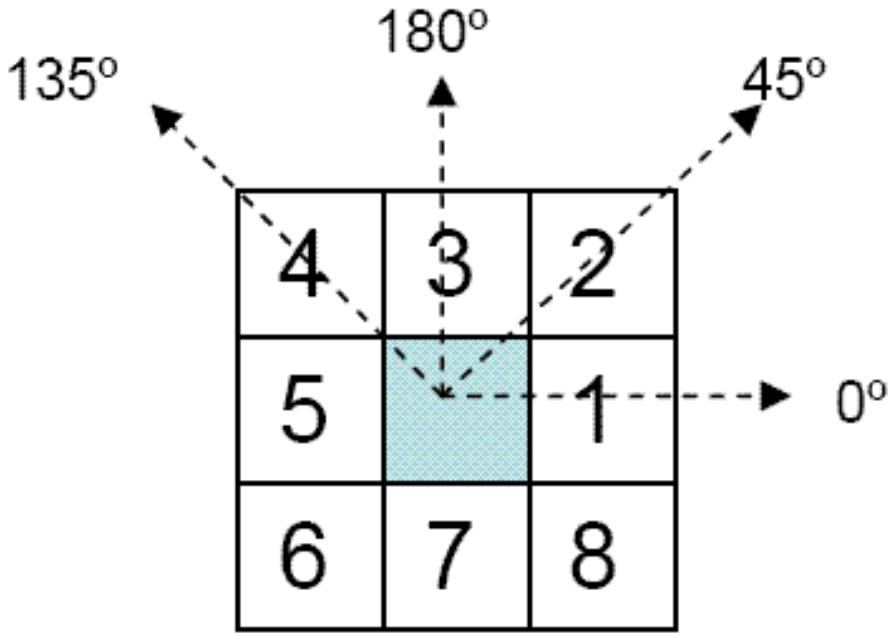


Figure 3.1: Indicating the 8 nearest neighbour pixels for a non-edge pixel

Haralick et al. [1973] proposed that texture information is contained in the average or overall spatial relationship the grey tones have with one another. Specifically this is captured by the relative frequency, $P(i, j)$, of which two pixels with grey levels i and j , that are separated by a distance, d , and angle, θ, occur within an image block. This results in a $N_G \times N_G$ matrix, with each entry at i, j indicating the probability of a pixel pair having the grey value of i and j . The total number of pixel pairs is used to normalize the matrix $P(i, j)$, and is referred to as the Grey Level Co-occurrence Matrix.

Haralick et al. [1973] only considered a single pixel spacing, which gives four possible orientations (figure 3.1). Despite 8 nearest neighbours, only 4 orientations are required to capture all the co-occurrence relationships. For example, the grey levels represented at the centre pixel and 6th pixel in figure 3.1 will be captured when 6 is the centre pixel along the 45° orientation. When the pixel spacing increases, the number of possible orientations increases, however only those in the range $0 - 180°$ need to be considered. Once the GLCM has been calculated, multiple textural features can be generated, see table 3.1.

Table 3.1: Texture Features for GLCM and grey level co-occurrence linked list (GLCLL) implementations, adapted from Clausi and Zhao [2003]

Feature	GLCM	GLCLL
Uniformity (UNI)	$\sum_{i=1}^{G}\sum_{j=1}^{G} P_{ij}^2$	$\sum_{n=1}^{L} P_n^2$
Entropy (ENT)	$-\sum_{i=1}^{G}\sum_{j=1}^{G} P_{ij} log P_{ij}$	$-\sum_{n=1}^{L} P_n log P_n$
Maximum Probability (MAX)	$max\{P_{ij}\}\forall(i,j)$	$max\{P_n\}\forall(n)$
Dissimilarity (DIS)	$\sum_{i=1}^{G}\sum_{j=1}^{G} P_{ij}\,\lvert i-j\rvert$	$-\sum_{n=1}^{L} P_n\,\lvert i_n-j_n\rvert$
Contrast (CON)	$\sum_{i=1}^{G}\sum_{j=1}^{G} P_{ij}(i-j)^2$	$sum_{n=1}^{L} P_n(i_n-j_n)^2$
Inverse difference moment (IDM)	$\sum_{i=1}^{G}\sum_{j=1}^{G} \frac{P_{ij}}{1+(i-j)^2}$	$\sum_{n=1}^{L} \frac{P_k}{1+(i_n-j_n)^2}$
Inverse difference (INV)	$\sum_{i=1}^{G}\sum_{j=1}^{G} \frac{P_{ij}}{1+(i-j)}$	$\sum_{n=1}^{L} \frac{P_n}{1+(i_n-j_n)}$
Correlation (COR)	$\sum_{i=1}^{G}\sum_{j=1}^{G} frac(i-\mu)(j-\mu)P_{ij}\sigma^2$ $\mu=\sum_{i=1}^{G}\sum_{j=1}^{G} P_{ij}$ $\sigma^2=\sum_{i=1}^{G}(i-\mu)^2\sum_{j=1}^{G} P_{ij}$	$\sum_{n=1}^{L} \frac{i_n-\mu)(j_n-\mu)P_{ij}}{\sigma^2}$ $\mu=\sum_{n=1}^{L} i_n P_n$ $\sigma^2=\sum_{n=1}^{L}(i_n-\mu)^2 P_n$

One of the problems of GLCM is the relatively high computation cost in terms of calculation time and memory storage required. A different GLCM is required for each pairing of pixel spacing, d and orientation, θ, within each image block. Clausi and Jernigan [2002] suggest multiple approaches to reduce the computation costs for calculating the GLCM:

Reduce N_G: Quantising the image to have fewer gray levels, reduces the size of each GLCM which in turn reduces the storage requirements and the number of iterations required to calculate. The disadvantage of this approach is the potential removal of useful information.

Use non-overlapping image blocks: Pixelwise segmentation and classification requires GLCM and subsequnet features to be calculated for an image block generated using the desired window size centred on each pixel. The use of nonoverlapping image blocks will reduce the number of GLCM that need to be calculated, but the resolution of the output will also be reduced and can become blocky.

Reducing the number of features and orientations: This approach limits the computation cost by only calculating features which provide the most information. This is better than the previous suggestions as no information is lost and the output will match the input resolution. However, the pertinent features and orientations are likely to change depending on the classification task, and will likely need to be determined by examining a "full set"of features.

Clausi and Jernigan [2002] conclude that a better approach would be to reduce the

unnecessary computational steps when calculating the GLCM. GLCM are recognised as being sparse, and therefore looping over all possible gray level pairing is inefficient when most of the entries will be zero. Clausi and Jernigan [2002] demonstrate that within a 20×20 image block, a maximum of $20 \times 19 \times 2 = 760$ different entries are possible. If $N_G = 256$, then $256 \times 256 = 65536$ iterations would be performed to capture a maximum of 760 non-zero entries.

A linked list implementation [Clausi and Zhao, 2003, Clausi and Jernigan, 2002] is one approach to reduce the storage and computation for calculating gray level co-occurrence probabilities. The grey level co-occurrence linked list (GLCLL) approach only stores the non-zero probabilities, greatly reducing the storage. At the nth node, linked list contains:

- Probability value, P_n
- Grey level pair, (i_n, j_n)

As the image block is processed and a particular (i, j) pairing identified, the list is searched to see if the grey-level pair already exists. If it is present in the list, the probability is updated; if not then then a new entry is created relating to the grey-level pair.

Features can be calculated by summations over the length of the list, L, which only contain non-zero probabilities and hence only processes information that contributes to the feature statistics. This requires slight changes to the how the feature statistics are calculated, and is included in table 3.1.

3.1.2 2D Gabor Filters

The most efficient way to process the large amounts of data associated with biological and computer vision is a repetition of similar localized operations, similar to those observed in studies of cortical arrangements [Porat and Zeevi, 2002]. Decomposing an image using multi-channel filtering allows the relevant texture feature to be used to classify the textures. This multi-channel approach mimics the Human visual system [Clausi and Jernigan, 2000], and can be achieved by utilising Gabor Filters, as the Gabor elementary functions are localized.

A 2D Gabor filter consists of a sinusoidal plane wave of a particular frequency and orientation modulated by 2D Gaussian Envelope, and have optimal resolution in both

the spatial and spatial frequency domains [Jain and Farrokhnia, 1991]. The output from a Gabor filter would ideally contain a discontinuity at a texture boundary and so detection of these discontinuities forms the basis of segmentation [Dunn and Higgins, 2002]. However the parameters of the Gabor filter would need to be selected to illicit a response to a particular texture, and so it is more common to apply a bank of Gabor filters.

A bank of 2D Gabor filters can be applied to image to provide texture information, using a parametric sweep to alter the filters size, orientation and frequency in order to generate responses to multiple texture types. The magnitude response to the Gabor Filters provides consistently good texture segmentation of the Brodatz test images and is simple to implement [Clausi and Jernigan, 2000]. Also more specific to task of identifying atmospheric aerosols, the magnitude response has already previously been used to identify dust and smoke in IR satellite imagery [Khazenie and Lee, 1992].

The 2D Gabor Filter is a Gaussian modulated by a complex sinusoid, of the form:

$$h(x,y) = \frac{1}{2\pi\sigma_x\sigma_y} exp\{-\frac{1}{2}\left[\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2}\right]\} exp\{j2\pi Fx\} \tag{3.1}$$

The x-axis of the Gaussian is aligned to the orientation of complex exponent, θ, and so rotation of the $x-y$ plane can generate arbitrary filter orientations. The filter bank parameters were set by selecting the number of orientations, n_θ, and setting the orientation bandwidth, B_θ, to the orientation spacing which provides reasonable coverage in the spatial-frequency domain [Clausi and Jernigan, 2000]. This allows the radial bandwidth,B_r, to be calculated that maximizes the coverage in the frequency domain whilst minimizing the overlap between filters [Jain and Healey, 2002, Bovik et al., 1990]:

$$B_r = log_2(\frac{1 + tan(B_\theta/2)}{1 - tan(B_\theta/2)}) \tag{3.2}$$

The extent of Gaussian envelope can then be calculated by setting the cut off in frequency and angular direction to -6db [Clausi and Jernigan, 2000]:

$$\sigma_x = \frac{\sqrt{ln2}(2^{B_r} + 1)}{\sqrt{2}\pi F(2^{B_r} - 1)} \tag{3.3}$$

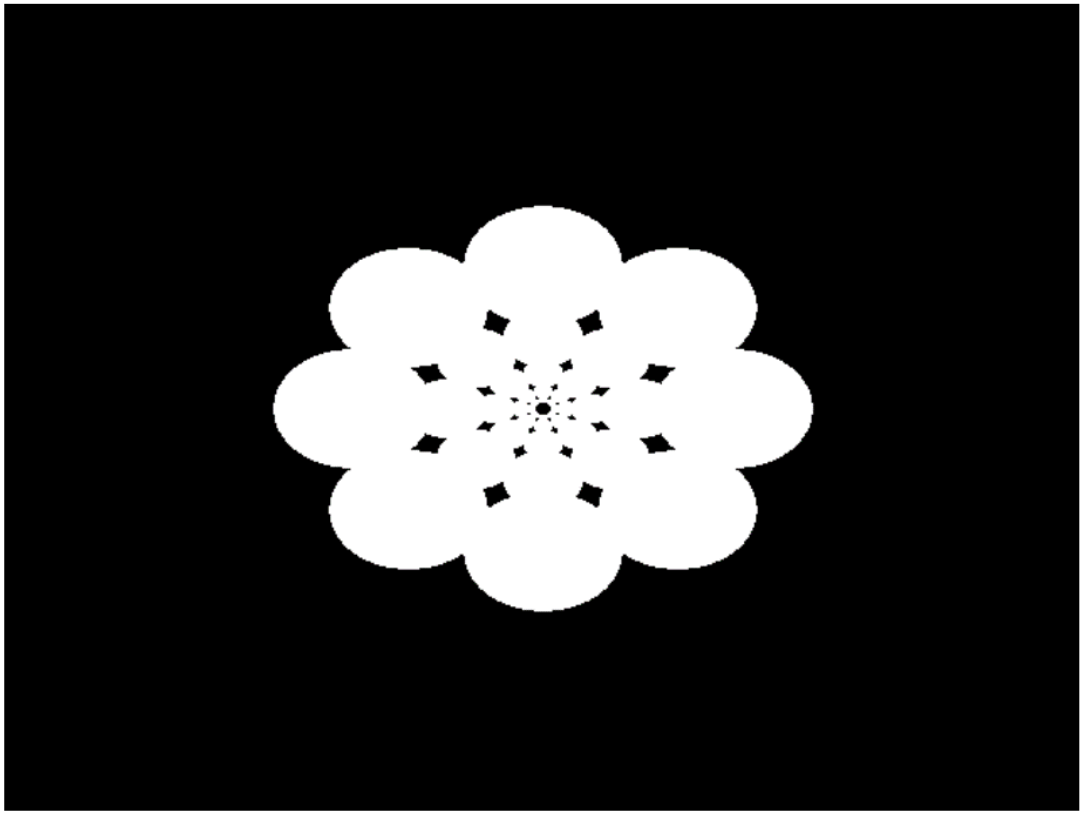


Figure 3.2: Half magnitude frequency response to a example Gabor filter bank with 4 orientations and 5 frequencies

$$\sigma_y = \frac{\sqrt{ln2}}{\sqrt{2}\pi F tan(B_\theta/2)} \tag{3.4}$$

The size of the filter is truncated to $3\sigma_m$ to reduce the computation cost during the convolution [Jain and Healey, 2002].

The filter bank frequencies can be determined by calculating the peak frequency, F_{peak}, and then setting further frequencies using a single octave spacing [Bovik et al., 1990].

$$F_{peak} = \frac{0.5}{1 + tan(B_\theta/2)} \tag{3.5}$$

$$F_m = 0.5F_{m-1} \tag{3.6}$$

Figure 3.2 shows the coverage in the frequency domain for a Gabor filter bank with four orientations ($n_\theta = 4$) and five frequencies ($n_f = 5$). The white areas show the

half magnitude extent of the GFB and represent points where the magnitude of the frequency response is greater than half of the peak magnitude. The x axis corresponds to $\theta = 0$ and increasing values of result in anti-clockwise rotations about the origin.

Lower values of frequency correspond to larger spatial variations, and eventually the spatial extent can become so large the filter response no longer captures textural information. Lower frequency values also increase the filter size (see equations 3.3 and 3.4), which leads to increased computation time. Careful selection of the number of frequencies is required to capture all the necessary texture information without capturing responses to features too large to be texture based.

Bovik et al. [1990] and Clausi and Jernigan [2000] suggest that the application of a Gaussian post filter to smooth the magnitude response to a Gabor filter. Textures that do not have a sufficiently narrow bandwidth leak into other filter responses and Gaussian post filtering can reduce the leakage whilst maintaining textural boundaries. The post filter is matched to the Gabor filter by using the Gaussian envelope but with greater spatial extent. The Gaussian envelope in equation 3.1 can be represented by:

$$g(x, y) = \frac{1}{2\pi\sigma_x\sigma_y} exp\{-\frac{1}{2}[\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2}]\} \tag{3.7}$$

The extent of the Gaussian post filter can be controlled by λ , such that the post filter is represented by:

$$g(\lambda x, \lambda y) \tag{3.8}$$

Smaller values of λ lead to greater smoothing. Bovik et al. [1990] and Clausi and Jernigan [2000] recommend a value of $\lambda = 2/3$.

3.2 Supervised Classification

Classification methods can be either classified as supervised or unsupervised, depending on whether or not training data is used during the construction of the classifier. Training data is a group of pixels within an image whose classes are known a priori and are used to estimate the texture properties of each class. Supervised classification uses this information to decide to which class to assign each pixel, based on the similarity of

the textural properties. Unsupervised classification uses clustering algorithms [Xu and Wunsch, 2005] to group pixels that exhibit similar features. It is an iterative approach that attempts to minimize an error measure between the texture features of the individual pixels in each class and the average features as represented by the mean values at the cluster centre. A commonly used example is the K-means clustering algorithm which attempts to minimise the variance within the cluster.

One of the advantages supervised classification methods have over their unsupervised counterparts is the ability of the user to define categories relating to meaningful classes, as opposed to those naturally occurring within the feature space. No further user input is required to match the natural categories with the information classes of interest to derive a thematic map [Campbell, 2006]. Selecting defined classes allows the user to identify single or multiple features of interest and also allows comparison between classifications produced at different times and locations.

However, a major limitation of supervised schemes is the need for suitable training data. In many remote sensing applications ground truth is collected by field work or manual image interpretation and one of the three main map sampling schemes, for example simple random sampling, used to provide appropriate training data [Baraldi et al., 2005]. Due to difficulty and expense it is not always practical to provide the desired number of appropriately distributed samples and this frequently leads to problems when using high-dimensional feature spaces [Jimenez and Landgrebe, 2002]. In order to produce reliable classifiers, a training set should ideally fully describe all of the classes present[Foody and Mathur, 2004b], and several rules of thumb exist which indicate the minimum number of samples per class required to achieve this [Jimenez and Landgrebe, 2002, Mather, 2004].

A commonly used limit for multivariate data sets is a minimum of $30p$ pixels per class, where p is the number of features and the samples are independent. This indicates that there are two failure methods of potential training data for supervised classification. Firstly, Small training sets relative to the number of features can lead to ill- and poorly-posed learning problems. Both of these introduce the curse of dimensionality and produce classifiers with poor generalization capability [Jimenez and Landgrebe, 2002].

The second issue is the independence samples, which relates to the sampling scheme used to obtain the training set. If the class of a particular pixel was known, then it is likely that a neighbouring pixel would also be of the same class. This is due to image region pertaining to the single class are likely to contain more pixels in the central bulk than at at an edge where the pixel borders another class. Thus it is nearby pixels and not just those that are adjacent that will not be statistically independent. This

correlation between nearby pixels is called spatial auto-correlation. Samples that are not independent, will over estimate the presence of classes in the training set compared to independent samples, which consequently causes the covariance of a particular class to be underestimated and implies the class is more homogeneous and distinct that it really is [Mather, 2004].

Random sampling ensures the independence of samples, and provides better classification results compared to using a contiguous block of samples [Campbell, 1981, Mather, 2004], and is generally accepted to be the best sampling scheme for generating training sets. This places a higher demand on the number of samples required in the ground truth, to allow a sufficiently sized random sampling to be drawn to provide a representative and independent training set.

However, classification techniques exist that do not require a full statistical description of each class in order to train the classify. These approaches make no assumptions about an underlying parametric model, and examples of these non-parametric classifiers include support vector machines (SVM) [Foody and Mathur, 2004a] and decision trees [Simard et al., 2002]. Instead, training samples near the classification hyperplane or decision boundaries that separate the classes are the most useful. Non-parametric classifiers therefore have the potential to be more robust to small training sets [Foody and Mathur, 2004b]. However, Huang et al. [2002] have shown that larger training sets still provide better classification for SVM. In their work, the training sets were selected by random sampling and, as such, larger sets have a higher probability of containing points near the hyperplane boundary. This implies that non-parametric classifiers will only perform well with small training sets, providing adequate border training samples are included. This introduces the additional problem of identifying border training samples from within a ground truth sample in order to be sub-sampled.

The definition of a useful training set will depend on the classifier being used, however all approaches require accurate ground truth in order to generate classifiers that can be readily generalized and applied to subsequent scenes and images. Due to the difficulty in obtaining useful training sets, many practical applications of supervised classification are undertaken with less than ideal training data, but are still capable of producing useful results.

3.2.1 Feature Space Reduction

Advances in computing power and the amount of data collected and stored has led to significantly larger number of features being considered for machine learning tasks in

recent years. 15 years ago few applications considered more 40 features whereas in the past decade feature numbers in the order of $10^2 - 10^4$ are not atypical [Guyon and Elisseeff, 2003]. High-dimensional feature spaces suffer from the well-known curse of dimensionality, which adversely affects many learning algorithms [Yang and Pedersen, 1997]. Therefore, reducing the dimensionality of the feature space has many potential benefits including a reduction the storage requirements, shorter training and utilization times, and importantly for small training sets, defying the curse of dimensionality to improve performance [Guyon and Elisseeff, 2003]. Feature space reduction can be broadly separated into two areas; feature selection and feature extraction.

Feature selection reduces the dimensions of the feature space by selecting a subset of the original features which, ideally, maximizes the classification ability whilst reducing the computational costs. Feature space reduction is a logical choice when dealing with small training sets and one such class of techniques is wrapper approaches, which use the classification method as a black box and rank the feature subsets based on the classification accuracy compared to the training data [Yusta, 2009]. Typically, an exhaustive search can not be used to find the optimal feature subset as it is too computationally expensive and, instead, sequential search methods can be applied. Forward sequential feature selection (FSFS) adds features one-by-one to the subset that maximises the classification accuracy. Similarly, backward sequential feature selection (BSFS) successively removes features to maximise the accuracy metric [Kohavi and John, 1997].

As an alternative to feature selection, feature extraction can be used to remap the feature space, for example by using a linear combination of the existing features, to a new feature space which is more favourable either through better separation of the classes or by concentrating the majority of information in a small number of features. Here, feature extraction methods are applied to the original texture features to provide a remapped feature set, with the aim of producing a higher classification rate. Principal component analysis (PCA) and linear discriminant analysis (LDA) are common examples of this approach. Ideally, the new feature set will provide better separation between the classes than the original features.

Principal component analysis (PCA) is one of the standard techniques for reducing a complex data set to a lower dimension, which can reveal underlying simplified structures. PCA is an orthogonal linear transformation that transforms the data to a new coordinate system such that the greatest variance by any projection of the data comes to lie on the first coordinate (the first principal component), the second greatest variance on the second coordinate etc. The underlying assumption of PCA is that the direction with the largest variance contains the data of interest, and so the new coordinate system is one that maximizes the signal to noise ratio (SNR). The outputs from

PCA frequently show the proportion of information represented by each coordinate, which can be used to reduce the dimensions of the data required to describe the underlying behaviour. One of the draw backs of using PCA is the inability to incorporate a priori information, such as a training set in supervised classification example, in order to select the coordinate transform. Kernel PCA methods are an extension to the PCA which can deal with non-linear transforms derived from prior information.

Linear discriminant analysis (LDA) is related to PCA, as they both look at linear combinations of the data in an attempt to understand the underlying behaviour. Where they differ is that LDA explicitly takes into account the the training data to create a linear combination of original features that yields the largest difference between the classes. LDA is intuitively more attractive than PCA as it considers the underlying class structure. This is supported by Cheriyadat and Bruce [2004], who showed that PCA is not the optimal solution for dimensionality reduction for classification applications. However, when the classes are under represented, PCA has been shown to outperform LDA [Martínez and Kak, 2002].

3.2.2 The Applicability of Supervised Texture Classification Using Linear Training Sets

As previously discussed in chapter 1, one of the difficulties in a observing particulate aerosols in satellite imagery is the similarity on the spectral response of multiple particulate aerosols, making it hard to distinguish between the different types [King et al., 1999a]. Furthermore, there is a wide variation in the spectral response, depending on the time of day, year and the underlying surface.

To address this, Khazenie and Lee [1992] proposed that texture information could be used to address these issues and consistently identify aerosol plumes from the underlying background and to separate aerosol types. They used three differently derived texture measures namely grey level co-occurrence features, Normalized Difference histograms and 2D Gabor Transforms, and found that 2D Gabor transforms were the most efficient and robust in detecting aerosol features in NOAA Advanced Very High Resolution Radiometer (AHVRR) images. Unsupervised clustering was used to identify 7 clusters, including deep dust and light dust. Diop et al. [2006] used a similar approach on Meteosat First Generation (MFG) images using unsupervised clustering to identify dust clouds in IR images using GLCF. The segmentation results showed a comparable performance to that of the IDDI for identifying dust clouds over North Africa.

These unsupervised approaches show the potential for texture based classification to be

able to identify particulate aerosol distributions in satellite imagery. Further improvements could be made by replacing the clustering approach with supervised classification, which would allow the a specific particulate aerosol to be selected and identified without any further user input.

A potential source for training data is the CALIPSO VFM. As discussed in section 2.2.3, the VFM provides discrimination between clear air, clouds and aerosols, and provides aerosol subtypes to allow for detection of specific particulate aerosols. The CALIPSO VFM can provide training data for satellite imagery, providing the data is co-located with the VFM and captured at the same time, which was first presented in Wiltshire et al. [2009].

MODIS on AQUA and SEVIRI on MSG are two good candidates to provide the imagery for supervised texture analysis. AQUA is located in the A-Train constellation with CALIPSO, and so will always provide imagery that is temporally and spatially matched to CALIPSO. In contrast, MSG is a geostationary satellite with a limited FOV centred over the prime meridian, and so will not always be spatially matched to CALIPSO. MSG does have a short repeat cycle, compared to MODIS on AQUA, which allows for better continual observation of an aerosol event and it is also ideally located to observe dust events generated over the Saharan desert.

The CALIPSO VFM may not provide useful training data due to the narrow FOV and the polar orbit of CALIPSO. This translates into a near linear path of training samples, which can lead to spatial auto-correlation effects introduced earlier. An example of a CALIPSO VFM and SEVIRI imagery is included in figure 3.3. In this instance the CALIPSO VFM is used to provide training data to identify a dust storm in a 10.6 IR SEVIRI image taken over North Africa on 21/02/2007. The SEVIRI image is shown with the CALIPSO overpass overlaid in blue and the dust locations shown in red.

The potential of supervised classification using linear sets has a wider applicability than the task presented in Baraldi et al. [2005]. As well as the ability to detect atmospheric mineral dust distributions, a simple extension would provide the ability to identify any of the other aerosol subtypes included in the CALIPSO VFM.

Additionally, linear training sets can be provided by other means. Texture classification techniques have also been used to identify oil spills in synthetic aperture radar (SAR) images. In particular, gray level co-occurrence features (GLCF) has been found useful for distinguishing between oil spills and naturally occurring dark patches on the sea surface [Brekke and Solberg, 2005]. In these instances, the classifiers were trained using data that was hand-classified by human experts. The infrequency of oil spills means that training data is scarce [Kubat et al., 1998] and the need for continuous global monitoring

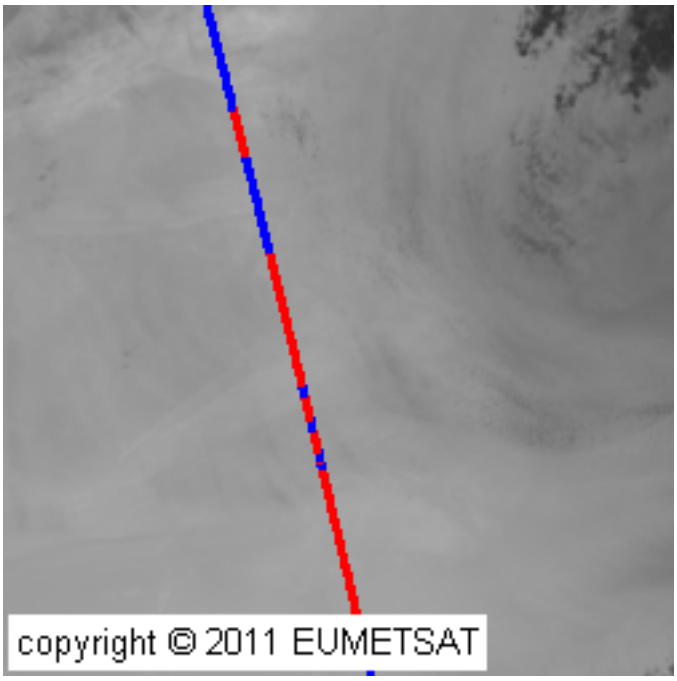


Figure 3.3: IR10.8 SEVIRI image from UT 13:45 on 21/02/2007 with the CALIPSO overpass marked in blue and the dust locations in red, reproduced from [Wiltshire et al., 2009]

places a high cost on expert classification. Here, training data provided along a linear tract could be useful in two instances. Firstly, ground truth provided by observations from a ship traversing the slick could prove useful for supervised classification of large scale spills caused by ecological disasters. Secondly, for smaller scale accidental or deliberate spills from ships, expert analysis can quickly and efficiently identify a linear path across the slick, allowing its spatial extent to be determined using a supervised classifier and hence reducing the workload of manual inspection.

Another example of where linear training data can be found is in the remote sensing of volcanic ash. The importance of this application has recently been highlighted by the disruption to air traffic over Europe caused by eruptions of the Icelandic volcano Eyjafjallajoekull. During the disruption a downward facing lidar and other instruments on-board an aircraft operated by the Facility for Airborne Atmospheric Monitoring were used to determine the altitude and thickness of the dust plume along the flight paths of a number of missions [Woolley, 2010]. The flight paths included significant linear portions and so the data recorded can provide a linear training set in order to identify the volcanic ash distribution over a wider area in satellite imagery.

The difficulty in obtaining ground truth for particulate aerosol distributions means that it is difficult to find a reliable, independent source to validate any outputs using the supervised texture classification with a linear training set. Therefore, in order to

determine whether the linear training set can produce reliable classification outputs, simulated data was used. As the ground truth was explicitly known, the performance of the linear training can be evaluated accurately to determine its suitability, and this is the focus of the next chapter.

Chapter 4

Performance Evaluation of Supervised Classification Schemes Using Linear Training Sets

This chapter investigates the applicability of linear training sets for supervised texture classification, and compares the performance of different texture features and representative list of different classifiers. As discussed in the chapter 3, the CALIPSO VFM can provide a linear training set to identify particulate aerosols in another satellite image, such as SEVIRI on MSG or MODIS. The lack of ground truth for particulate aerosol distributions prevents the evaluation of linear training sets on real data, and so instead simulated images are used. Simulated images are used to identify the best performing set of texture features and classification schemes for use with a linear training set, which has not been previously studied. The evaluation methodology is introduced including the generation of the test images, training sets and the texture features used. The classifiers used, and the performance evaluation techniques are described and finally the results and conclusions of the simulations are presented.

4.1 Simulated Texture Classification Methodology

4.1.1 The Suitability of Linear Training Sets

Linear training sets will be subject to spatial autocorrelation effects. The effect this leads to the variance of the classes within the image being underestimated, resulting in classes appearing more homogeneous than they really are and producing classifiers that generalize poorly.

The use of the CALIPSO VFM as a training set for supervised classification was first presented in Wiltshire et al. [2009] and was used to identify atmospheric mineral dust distributions. Based on this scenario, the task falls somewhere ill- and poorly- posed classification problems. This relates the number of features, or dimensions, of the feature space compared to the number of samples in the training set and is defined below [Baraldi et al., 2005].

Definition 2 *Ill- posed classification occurs when the number of data dimensions exceeds the total number of representative samples in the training set, and is much greater than than the number of per-class samples.*

Definition 3 *Poorly- posed classification occurs when the number of data dimensions is less than the total number of representative samples in the training set, but is much greater than or equal to the number of per-class samples.*

As well as the lack of independent samples, the overall number of samples is limited by the along track resolution of the VFM leading to a small training set which prevents further sub sampling as insufficient samples would be retained. The result is a small training set with unrepresentative samples. The work in this chapter examines whether a linear training set can produce good results, despite the poorly-posed nature of the classification task. Specifically, for the particular problem of identifying a single texture of interest within an image containing multiple textures, the reduction in classification performance resulting when the training data is sampled linearly, instead of randomly, is quantified for a number of commonly used texture features and classification techniques, allowing the best performing combinations to be identified.

4.1.2 Texture Feature Generation

Whilst the motivation for this work is to ascertain the feasibility of using linear training sets for supervised classification, and subsequently discover a best combination of textural features and classifiers, the overarching task is to apply these techniques to identify particulate aerosols. This application drives the type of texture features considered.

Previous work using unsupervised classification has used GLCF [Diop et al., 2006, Khazenie and Lee, 1992] and 2D GBF Khazenie and Lee [1992] to identify particulate aerosol distributions, and are a logical choice for inclusion. Furthermore both GLCF and 2D GFB are robust and mature texture approaches and are widely used in the remote sensing community.

GLCF were generated for pixel spacings $d = 1$ and $d = 2$ within a fixed 9×9 window. The orientations of the features are set by considering all possible spacings using the Chebyshev distance, resulting in 4 orientations for $d = 1(0°, 45°, 90°$, and $135°)$ and 8 orientations for $d = 2(0°, 26.6°, 45°, 63.4°, 90°, 116.6°, 135°$ and $153.4°)$. For each pair of orientation and spacing, the 8 features in table 3.1 were calculated using a linked list implementation, see Section 3. This gives two GLCF sets, of 32 or 64 features for $d = 1$ and $d = 2$ respectively.

Clausi [2002] proposed a preferred GLCF set containing only contrast, correlation and entropy which was determined by grouping the different features into three areas; smoothness, homogeneity and correlation based on the correlation of the individual features. The main motivations of the work were to reduced the computation cost with little or no loss of classification performance as well as reducing the number of features, to avoid the curse of dimensionality when using small training sets. Close examination of the results shows the performance of the preferred feature set was only significantly statistically better than all 8 features for one of data sets considered, and this case was explained by the small sample size leading to a poor estimate of the covariance.

The work presented in this chapter is concerned with the accuracy of the classification outputs and not the computational cost of the feature generation. Also, reducing the number of features considered, potentially removes information that provides better separability of the classes. Therefore, GLCF with all 8 features will be used. In order to limit the affect of the curse of dimensionality, feature selection and extraction techniques are applied to all GLCF sets, and these techniques are explained in the classification section later in this chapter. Should the extra features not provide any benefit, then contrast, correlation and entropy will dominate the feature selection and

weightings for feature extraction, and there will be no loss of classification accuracy.

Features from the 2D GBF were generated by calculating the magnitude response of when each filter is convolved with the input image. The GFB were calculated using the automated process outlined in section 3 using equations 3.1 - 3.6. This process requires the number of orientations, n_θ, to be set and the GFB parameters and thus the 2D GFB can be generated. Clausi and Jernigan [2000] suggest a filter spacing of 30°, which equates to $n_\theta = 6$, although $n_\theta = 4$ is often used in literature.

Here, a value of 4 or 6 is used for n_θ and the first six frequencies ($n_f = 6$), giving a GFB consisting of 24 or 36 filters respectively. Gaussian smoothing can be applied to the magnitude response of the GFB, by setting a parameter γ to control the spatial extent of the Gaussian envelope. Two different values of γ were used; $\gamma = 1$ to match the Gaussian envelope used in the corresponding 2D Gabor filter, and $\gamma = \frac{2}{3}$ which gives the best classification result in Bovik et al. [1990] and Clausi and Jernigan [2000]. This results in three post filtering schemes when no smoothing is also considered. For easy identification the 2D GFB are referred to using the triplet (n_θ, n_f, γ), where $\gamma = 0$ is used to denote no post filtering.

A third texture feature can be generated using a fused feature set combining GLCF and 2D GFB features. Clausi and Deng [2005] provided a design rationale for concatenating these features and compared the performance of the fused feature sets with those of the contributing GLCF and 2D GFB. The fused feature set was proposed to overcome the inconsistent performance of 2D GFB with higher frequencies. GFB are ideal for characterizing a pure sinusoid, and a tuned filter will return a flat magnitude response to sinusoid. A practical signal can be decomposed into a weighted sum of sinusoids, and hence GFB are well suited for texture analysis [Clausi and Jernigan, 2000].

Clausi and Deng [2005] demonstrated the effect of sinusoids with Gaussian noise by calculating the magnitude response of the sinusoid using a tuned GFB over a range of frequencies with the results showed the inconsistency of Gabor filters at higher frequencies. This is explained by the higher frequency filters having a larger spatial-frequency bandwidth, and as the noise is evenly distributed in the spatial-frequency domain, higher frequencies filters include more noise energy. Their solution was to replace the high frequency Gabor filters with GLCF statistics of contrast, entropy and correlation with pixel spacings of $d = 1$ and $d = 2$, which results in the capture of local high frequency information. Similar analysis using sinusoids with additive Gaussian noise has shown that the these GLCF show no frequency dependence.

Furthermore, Clausi and Deng [2005] suggest that GLCF are not suitable for capturing the low and mid-frequency information as process is highly parameterized, with multiple

values of grey-level quantisation, window size and pixel spacings. For each set of parameters, the co-occurrence frequencies need to be calculated to derive the features which is computationally expensive. 2D GFB provide a computationally more efficient means to capture the low and mid-frequency texture information.

The best performing texture set [Clausi and Deng, 2005] in terms of class separability and classification accuracy consisted of:

1. The GLCF statistics of contrast, entropy and correlation calculated for pixel spacings $d = 1$ and $d = 2$, at four orientations: $(0°, 45°, 90°$, and $135°)$ within a fixed 9×9 window.
2. The magnitude response of the GFB defined by the triplet $(6, 4, 2/3)$

This results in a fused feature set containing 24 GLCF (3 statistics, at 4 orientations and 2 pixel spacing) and 24 GFB magnitude response (4 frequencies at 6 orientations).

4.2 Classification Methods

Applying a representative range of classification techniques to the simulated images described above enables the most appropriate classification approach for use with linearly sampled training data to be determined. As previously discussed, feature selection and extraction techniques are used to reduce the number of features used by some of the classifiers, as this helps reduce the problems associated with small training sets and high dimensional feature spaces. Feature space reduction is not required for the increasingly popular SVM techniques. This is because SVM use a kernel function to provide a non-linear mapping of the training data, but as the inner product never needs to be evaluated, dimensionality issues are avoided [Oommen et al., 2008]. Due to the poorly-posed nature of the supervised classification task and the unrepresentative samples, an unsupervised approach is also applied to demonstrate the benefit of using linear training sets over no *a priori* knowledge. K-means clustering was selected as commonly used unsupervised learning algorithm, and has previously been used to identify the spatial extent of dust clouds in satellite imagery [Diop et al., 2006, Khazenie and Lee, 1992]. Three distinct types of classifiers were used, those based on feature selection, feature extraction, and non-parametric classifiers.

4.2.1 Feature Selection Based Classifiers

A minimum distance classifier was used in conjunction with wrapper approaches, FSFS and BSFS, for the feature selection based classifiers. Minimum distance classifiers work by calculating the distance between the features at each pixel in the input image and class mean of the corresponding features for the texture of interest, derived from the training data. Their performance can be improved when feature selection or feature extraction is used to reduce the dimensionality of the feature space.

FSFS adds features one-by-one to the subset to maximise a chosen classification metric. Initially, classification is performed using each feature on its own, with the best performing feature selected. The remaining unselected features are then tried in conjunction with the 1st feature, to find the best pair of features, and so on until the feature subset reaches a desired size. Conversely, backward sequential feature selection starts with the entire feature set in the subset and sequentially removes a feature one at a time that gives the most improvement in the classification of the training set.

The classification is performed by examining the distance between a pixel under test and the class centre. As the true statistics are unknown, the class centre is estimated from the training set, such that the mean vector μ_k for class k is found using:

$$\mu_k = \frac{1}{N}\sum_{n=1}^{N}(\overline{x}_n - \mu_k)(\overline{x}_n - \mu_k)^T \tag{4.1}$$

where $\overline{x}_n$ are the individual pixels within class k, of a total N.

The Euclidean distance $d_k^{Euclidean}$ is the the simplest distance measure but it does not take into account the distribution of the data around the class centre. Equation 4.2 calculates the Euclidean distance between the feature vector at a pixel, x, and the class mean, μ_k.

$$d_k^{Euclidean} = \sqrt{(x - \mu_k)^T(x - \mu_k)} \tag{4.2}$$

The Mahalanobis distance (MD) measure takes uses the covariance matrix which takes into account the shape of the distribution of the data around the class centre. The Mahalanobis distance measure is a weighted Euclidean distance, with the weighting determined by the variability of the sample point calculated using the by the covariance matrix [Wölfel and Ekenel, 2005].

$$d_k^{Mahalanobis} = \sqrt{(x-\mu)^T C_k^{-1}(x-\mu)} \tag{4.3}$$

Again, as the true class statistics are unknown the covariance matrix is estimated from the training data using Equation 4.4.

$$C_k = \frac{1}{N-1}\sum_{n=1}^{N}(\overline{x}_n - \mu_k)(\overline{x}_n - \mu_k)^T \tag{4.4}$$

To determine whether the samples are classified as belonging to a particular texture, an Upper Control Limit (UCL) is calculated based on an outlier detection method, which treats the square MD as T^2 values and applies the Hotelling T^2 tests [MacGregor and Kourti, 1995]. The upper control limit for MD calculated using estimate of the covariance matrix is found using:

$$T_{UCL}^2 = \frac{(N-1)(N+1)q}{N(N-q)}F_\alpha(q, N-q) \tag{4.5}$$

where q is the number of texture features currently in the subset and F_α is the upper $\alpha 100\%$ critical point of the F-distribution with q and $N-q$ degrees of freedom. A value of $\alpha = 0.95$ was used.

The classification is then performed pixel by pixel, with a pixel being assigned to the class of interest if $MD \leq T_{UCL}^2$. The classification performance is evaluated for all the feature sets by comparing the output to the training set. Accuracy is commonly used to evaluate machine learning tasks, however it is not a good metric when the classes are unequally represented in the training set. The simulations in this chapter model a real world situation where the training set is likely to be dominated by the class of interest, for example a particular type of particulate aerosol. For example, a classifier that only correctly identifies half of the pixels in the central band of the simulated images shown in figure 4.1 as belonging to the texture of interest will still achieve a classification accuracy of 77.78%. In these situations both the True Positive Rate (TPR) and the True Negative Rate (TNR) are expected to be high simultaneously and their geometric mean (GM), described by [Kubat et al., 1998] and calculated by:

$$GM = \sqrt{TPR.TNR} \tag{4.6}$$

provides a balanced measure of the performance between the classes. This measure is nonlinear, independent of the distribution of the classes and is robust when the distribution changes with time or is different in the training and test set

The GM scores calculated using the training set are used for comparison to identify which feature to add (FSFS) or remove (BSFS) from the feature subset.

4.2.2 Feature Extraction Based Classifiers

Feature extraction approaches used Principal component analysis (PCA) and linear discriminant analysis (LDA). PCA was applied to the original feature set to provide a new feature with the principal components used as features for the minimum distance classifier described above. As well as being a feature extraction technique, LDA is a commonly used classifier in its own right and is applied using the inbuilt functions in the Matlab statistics tool box [MATLAB, 2009].

4.2.3 Non-parametric Classifiers

Non-parametric classifiers include SVM and decision tree classifiers. LIBSVM [Chang and Lin, 2001] is used to implement soft margin SVM classifier using radial basis functions (RBF) as the kernel. Two parameters are required in this instance, γ, which controls the width of the RBF and C, the value of the soft margin. A reliable approach would be to conduct an exhaustive grid search over a suitable parameter space and select the best settings [Staelin, 2003]. Exhaustive searches can result in long run times due to the high number parameter combinations and grid search is widely used to reduce the computational burden. In this instance a log_2 parametric grid search is used as suggested in the LIBSVM documentation [Chang and Lin, 2001]. N-fold cross validation is commonly used to evaluate each parameter set it alleviates the problem of over-training by using independent training samples to evaulate the classifer [Staelin, 2003]. Here 10-fold cross validation is seleted to determine the optimal values for the soft margin, C, and the RBF, γ as it led to high classification accuracy. Decision tree classification is implemented using the inbuilt functions in the MATLAB statistics tool box [MATLAB, 2009] which uses the common "Gini "splitting function and allows pruning to simplify the final tree and prevent over-fitting.

Preliminary work using SVM has shown that support vectors are more likely to be drawn from near the boundary between different classes. A linear training set crosses this boundary and hence are likely to provide support vectors despite the small training

set and therefore achieve a good classification rate.

4.3 Classification Performance Evaluation

4.3.1 Texture Image and Training Data

Simulated images were used to model the situation where there is a single texture of interest amongst other, possibly multiple textures. To this end, two 256 × 256 pixel test images were created containing two and five separate textures respectively, as shown in figure 4.1. Both underlying masks contain a central band representing the texture of interest, analogous to dust in the SEVIRI images in Wiltshire et al. [2009] or an oil spill on the sea surface. The central band extends across the full width of the image to ensure that any sampling line passing down the image will always contain pixels from the texture of interest. The two-texture test image models the simple binary classification case while the five-texture image represents the more realistic situation of a single texture to be identified in an image containing multiple textures.

All texture boundaries in the test images are based on sinusoids so that they are not aligned with any mask-based texture features. As in Clausi and Jernigan [2000], the images were populated with Brodatz textures [Brodatz, 1999] that exhibit distinct regions of texture without regular or repeating pattern and hence mimic the natural textures that are observed in remotely sensed images of the Earth. The two texture image consists of D29 (beach sand) in the central band, surrounded by D9 (grass). The five textures image maintains D29 (beach sand) in the central band, and D57 (hand made paper), D12 (bark), D38 (water) were selected in addition to D9 (grass) for the surrounding textures.

A further test image was generated by populating the five texture mask with textures from the Outex texture database [Ojala et al., 2002a]. The Outex textures were selected to match test 000 from the Outex Supervised texture segmentation suite, namely Canvas 002 at 45, Canvas 001 at 15, Canvas 026 at 60, tile 005 at 60 and carpet 004 at 75. All the Outex images were captured using the "inca"illumination at 100dpi. The resulting synthetic texture image is shown in figure 4.2.

The linear training sets were generated of arbitrary vertical orientations by randomly selecting pairs of pixels from the top and bottom rows of the image and joining them with a straight line. This fixed the number of samples per training set to the image height of 256 pixels and 100 such sets were generated, see figure 4.3. Each random

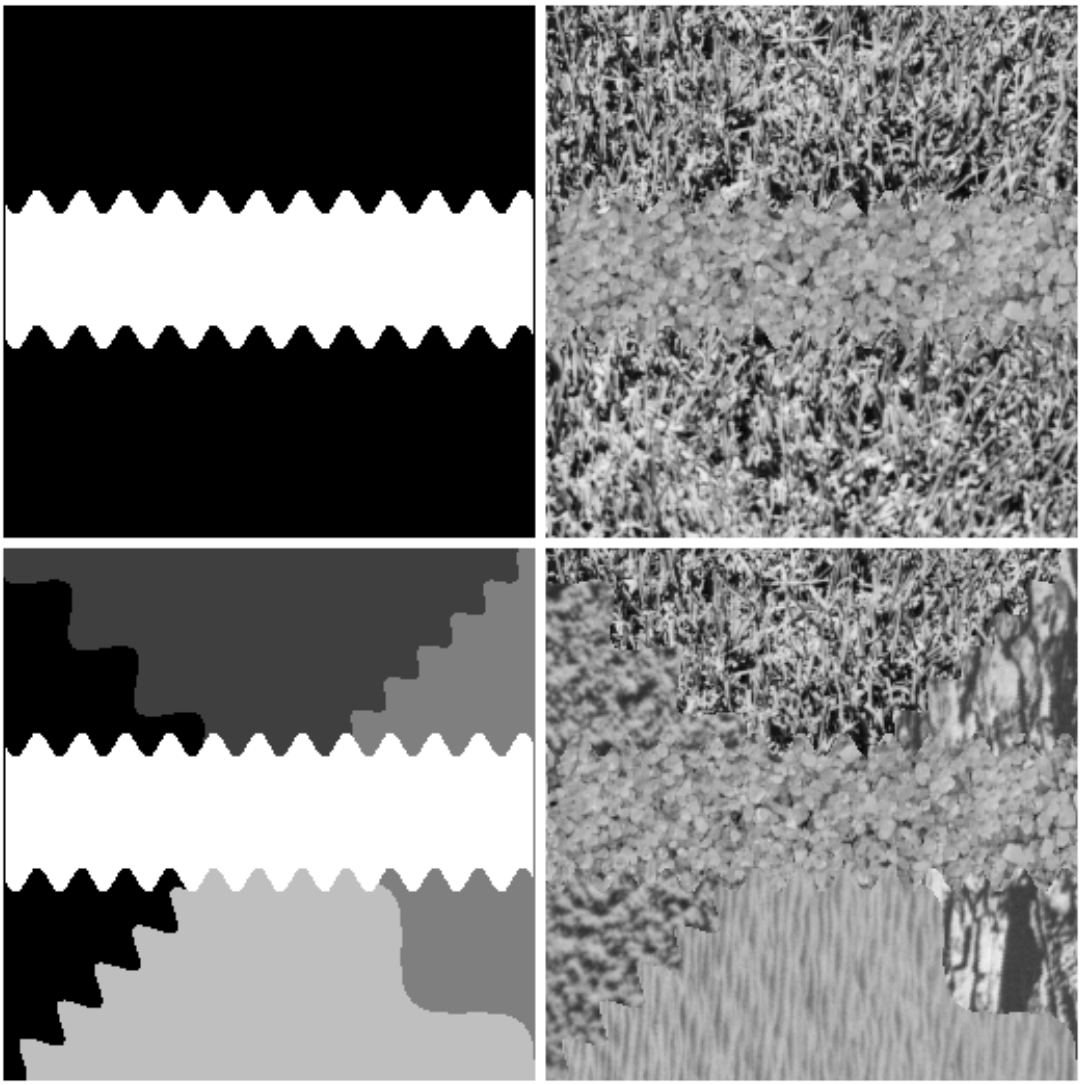


Figure 4.1: Texture masks (left) and the corresponding synthetic texture test images (right) for the two texture (top) and five texture (bottom) simulated images.

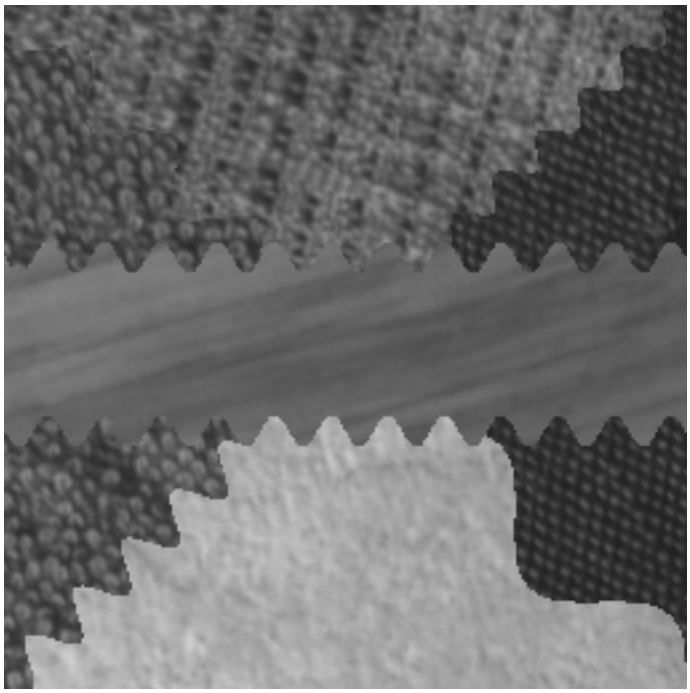


Figure 4.2: Five texture test image populated with the Outex textures

training set was generated by randomly selecting 256 pixel locations from within the collection of samples from the 100 linear training sets and 100 such sets were produced. This strategy results in two sets of training data that only differ in the sampling method used.

4.3.2 Performance Metrics

For each combination of feature set and classifier the overall classification performance is determined using the GM equation 4.6. This quantifies how well the class of interest is identified compared to the original masks (figure 4.1). For comparison, the classification performance using the randomly-sampled training sets is also found.

In addition, the Forstner covariance distance (FCD) is used to compare how well the random and linear training sets represent the true class statistics [Förstner and Moonen, 1999]. The FCD measures how different two covariance matrices are, in this case between those of the training set and the full texture. The FCD is an extension of the covariance analysis used in Campbell [1981], Mather [2004] and is given by

$$FCD = \sqrt{\sum \ln^2 \lambda_i(H, C)_{i=1}^{n}} \tag{4.7}$$

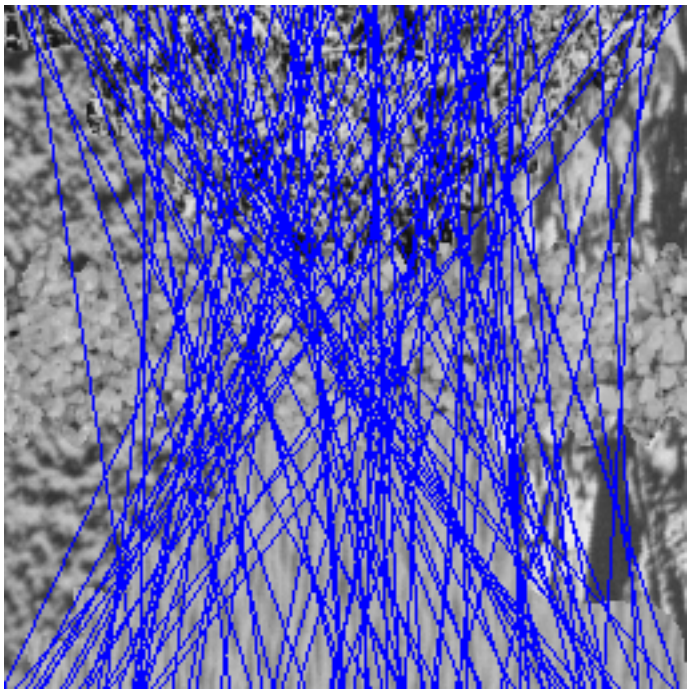


Figure 4.3: 5 texture test image with the 100 linear sample lines overlaid in blue.

Table 4.1: Binary Confusion matrix used for the McNemar's Test

	Classification 2	
Classification 1	Correct	Incorrect
Correct	f_{11}	f_{12}
Incorrect	f_{21}	f_{22}

where, H is the covariance matrix from the reference samples, C is the covariance matrix from the test samples, and λ_i is the i^th eigenvalue from the generalized eigenvalue problem $|-C|$. If the true class statistics are used to generate H, and the training set is used to generate C, then smaller values of FCD indicate a better representation of the underlying class. This provides a quantitative comparison of the ability the two sampling schemes to capture the true class statistics.

In order to determine whether the differences in the classification scores using different feature sets, classifiers or sampling schemes are significant, a McNemar's test was used [Foody, 2004]. The first task is to condense the confusion matrix to a 2 by 2 matrix, which determines at each pixel whether the classification is correct or not, see Table 4.1.

The test is based on the standardised normal test statistics (equation 4.8), and assumes z^2 follow a chi-squared, χ^2, distribution with one degree of freedom. This modifies the test equation from 4.8 to 4.9,

Table 4.2: Best mean GM performance (±2 standard deviations) for each classification technique and the texture features with which the best results was achieved, applied to the 2 class image. The GLCF are denoted Cn where n is the pixel spacing and the GFB are described by the triplet (n_θ, n_f, γ) and Fused is a combined GLCF and GFB

	2 class	
	Random	Linear
FSFS	0.905 ± 0.029 (6,6,0)	0.853 ± 0.077 (C1)
BSFS	0.913 ± 0.025 (6,6,0)	0.856 ± 0.072 (C1)
PCA	0.900 ± 0.066 (6,6,0)	0.899 ± 0.161 (C1)
LDA	0.952 ± 0.004 (6,6,2/3)	0.935 ± 0.010 (C1)
TREE	0.913 ± 0.040 (C1)	0.910 ± 0.077 (C1)
SVM	0.941 ± 0.022 (6,6,0)	0.919 ± 0.037 (C1)

$$z = \frac{f_{12} - f_{21}}{\sqrt{f_{12} + f_{21}}} \tag{4.8}$$

$$\chi^2 = \frac{(f_{12} - f_{21})^2}{f_{12} + f_{21}} \tag{4.9}$$

This allows the calculated z values to be compared to tabulated chi-squared values to indicate the statistical significance. For a significance level of 0.05, a value of $|z| > 1.96$ shows the difference in the classification score is significant.

4.4 Simulation Results

Every combination of texture features and classifier is applied to the 100 linear and 100 random training sets from the two simulated Brodatz images described previously and the average GM classification performances recorded. Table 4.2 presents the maximum GM results achieved by each of the classification schemes for both linearly and randomly sampled training sets with 95% confidence intervals, and also shows the texture features with which this performance is achieved. The tests were repeated with the position of the Brodatz textures in the 2 and 5 texture test images varied to give different textures of interest and these results confirmed that those in Tables 4.2 and 4.3 are representative.

When random sampling is used, GFB features provide the best classification results for 5 of the 6 classification techniques for both the 2 and 5 texture test images. The GFB (6,6,0) is the most popular texture feature set and when used by SVM provides the best classification performance for the 5 texture test image. The performance of the

Table 4.3: Best mean GM performance (±2 standard deviations) for each classification technique and the texture features with which the best results was achieved, applied to the 5 class image. The GLCF are denoted Cn where n is the pixel spacing and the GFB are described by the triplet (n_θ, n_f, γ) and Fused is a combined GLCF and GFB

	5 class	
	Random	Linear
FSFS	0.869 ± 0.018 (6,6,1)	0.823 ± 0.041(C1)
BSFS	0.864 ± 0.027 (Fused)	0.811 ± 0.073 (C1)
PCA	0.850 ± 0.048 (6,6,1)	0.809 ± 0.078 (Fused)
LDA	0.864 ± 0.021 (6,6,0)	0.731 ± 0.151 (6,6,0)
TREE	0.799 ± 0.053 (6,6,0)	0.752 ± 0.145 (C1)
SVM	0.900 ± 0.039 (6,6,0)	0.813 ± 0.071 (C1)

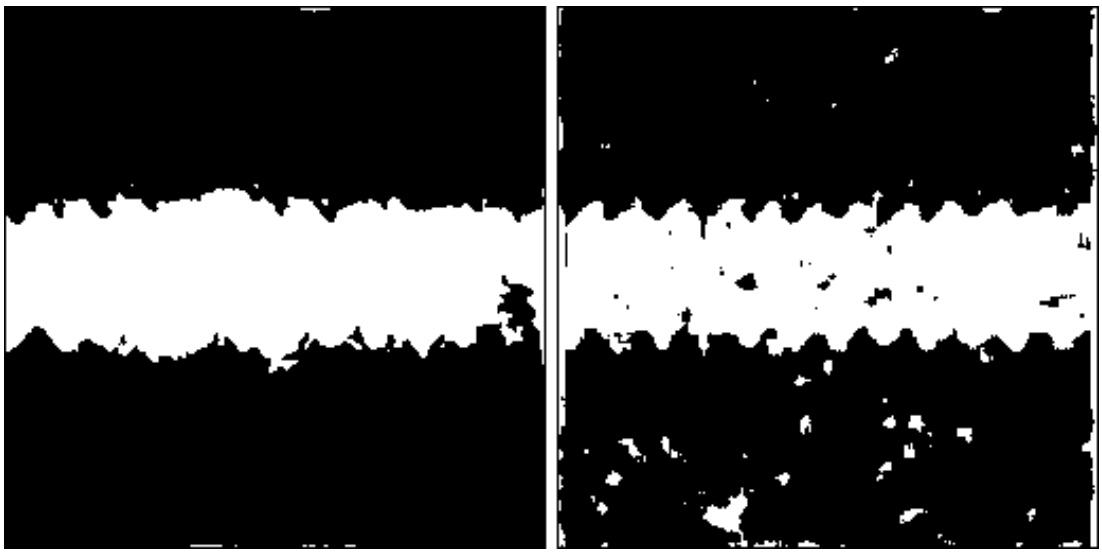


Figure 4.4: Example output images for best performing classification of the 2 texture image. The image on the left is representative of the mean GM for random sampling, using LDA on $(6, 6, 2/3)$. The image on the right is representative of the mean GM for linear sampling, using LDA on C1.

GFB texture features are in line with those reported elsewhere for texture classification schemes using random sampling [Clausi and Jernigan, 2000], when the reduced size of the training data is taken into account. Conversely, when linear sampling is used GLCF with a single pixel spacing consistently provide the best classification scores. They are the features preferred by all classification techniques for the two texture test image and by 4 out of 6 classifiers for the 5 texture test image. Over all experiments, the GLCF that were frequently selected were dissimilarity, contrast, inverse difference and correlation. This feature selection shows general agreement with the selection of one feature from each of the classes of smoothness, homogeneity and correlation, that has been shown to be the ideal scenario [Clausi, 2002].

The results for 2 texture test image show that LDA provides the best classification

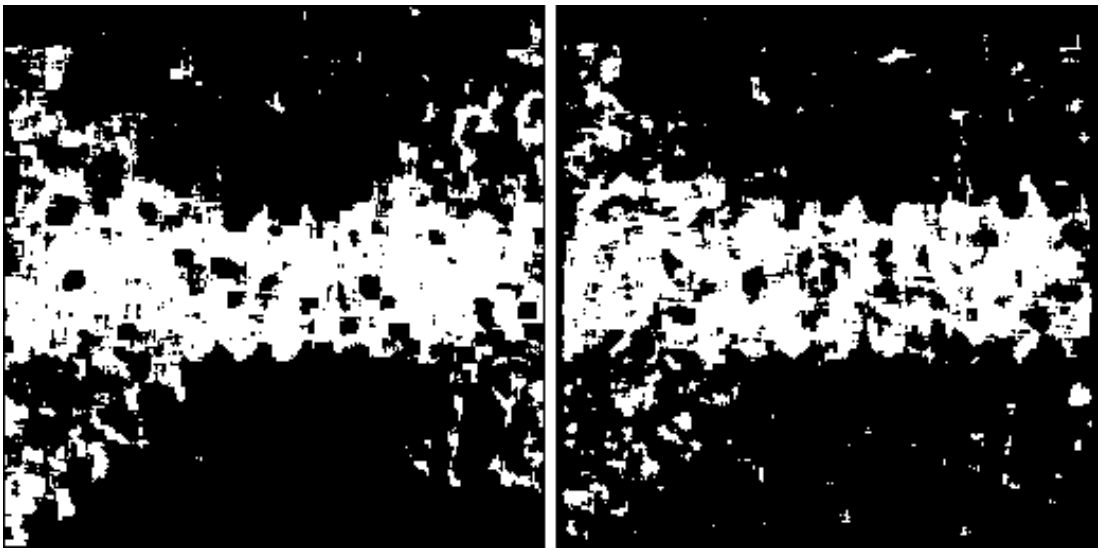


Figure 4.5: Example output images for classification of the 5 texture image. The image on the left is representative of the mean GM for linear sampling, using FSFS on C1. The image on the right is representative of the mean GM for linear sampling, using SVM on C1.

for both the linear and the randomly sampled training sets. Figure 4.4 shows output images with a GM close to the mean GM, when LDA is used on (6,6,2/3) for random sampling and C1 for linear sampling. LDA also exhibits the smallest variance in GM scores, with a very small confidence interval for both sampling schemes, indicating LDA gives consistently good results over all the runs. All the classification techniques perform well for randomly sampled training sets with GM scores in excess of 90%. As expected, the classification performance decreases when linearly sampled training sets are used. Whilst GM scores for LDA decrease by less than 2%, the wrapper techniques (FSFS and BSFS) show decreases in excess of 5%. Despite this, the results show that linearly sampled training sets still provide a high classification performance, with a GM above 85% for all approaches. The spread of the GM scores also increases when linear sampling is used, with the confidence intervals approximately doubling in size when compared to random sampling. This is expected, due to the spatial auto-correlation that affects the linear sampling. Should the linear tract occur over a region that is not useful to the classifier then multiple training samples will be affected, providing many fewer samples of use. This is unlike random sampling, where the distribution of useful samples within the training set will be approximately constant over all the training sets.

For all classifiers, the overall classification performance for the 5 texture image is lower than for the 2 texture image. This is to be expected as this image poses a more complicated classification problem and hence results in lower GM scores. In this more realistic scenario, linear sampling shows a larger decrease in GM scores than for the 2

texture image, with an average decrease of 6.8%. Unlike the 2 texture case, where it produced the best GM performance, LDA only out-performs PCA for random sampling and is the worst performing classifier for linear sampling. This decrease in performance is due to the introduction of additional textures in the input image, whilst fixing the number of classes at 2. LDA represents the class of interest as an individual class and groups all other classes together into a single class. While this has no effect for the 2 texture image, as the other class contains only one texture, for the 5 texture case the 4 other textures are grouped into a single class. This is an ill-posed problem for LDA as the classification boundaries will be found using one set of class statistics that, in reality, contains multiple classes. With the 5 texture test image, SVM provides the best classification using random sampling (90%) but drops significantly to 81.3% when linear sampling is used. Feature selection, in particular FSFS, provides the best classification using linear sampling (82.3%) although BSFS (81.1%), SVM (81.3%) and PCA (80.9%) also perform well. Figure 4.5 shows representative output images for the 5 texture example, using FSFS and SVM applied to C1. The confidence intervals for the GM scores are broadly comparable to those for the 2 texture test image. They also approximately double when moving from random to linear sampling.

The classification performances of the supervised techniques were also compared to that of a representative unsupervised classifier. To this end, the Matlab implementation of k-means clustering algorithm was applied to the 5 texture simulated image using FSFS and GLCF with single pixel spacing. Using the 100 linear tracks as the training sets the best individual run recorded a GM of 68.5% and the peak average GM performance was 60.5%. The average GM result is within 3% of the equivalent k-means results with randomly sampled training sets but is 13% lower than the worst performing supervised classifier and over 20% lower than the best performing supervised classifiers in 4.3. This demonstrates there is a significant advantage in using supervised classifiers, even when the linearly sampled training data is less than ideal.

Figures 4.6 and 4.7 show the mean GM and FCD versus the number of GLCF selected for the 5 texture image using the FSFS minimum distance classifier, the classifier and features that produced the best performance with linear sampling. Examination of the FCD shows that, in general, linear sampling provides a better representation of the texture of interest than random sampling, a result that at first sight appears to contradict the existing literature on training samples. The explanation of this lies in the number of samples needed to adequately represent the class of interest. On average, each linear and random training set contains 64 samples from the texture of interest.

Using $30p$ samples per class as a guideline for complete representation, this suggests that the training sets only provide an adequate statistical description when just 1 or 2 features are used. Random sampling does have a smaller FCD than linear but only

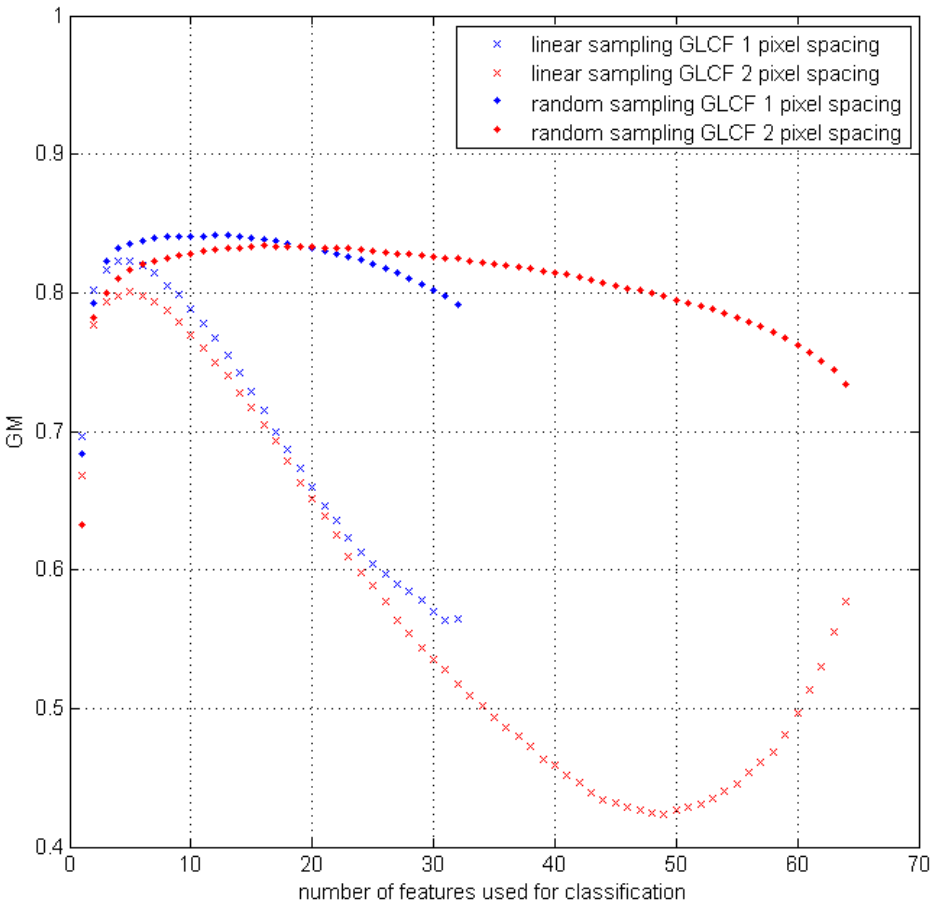


Figure 4.6: Average GM scores for 100 linear and 100 random sample sets for the 5 texture test image with FSFS using GLCF with pixel spacings of 1 and 2.

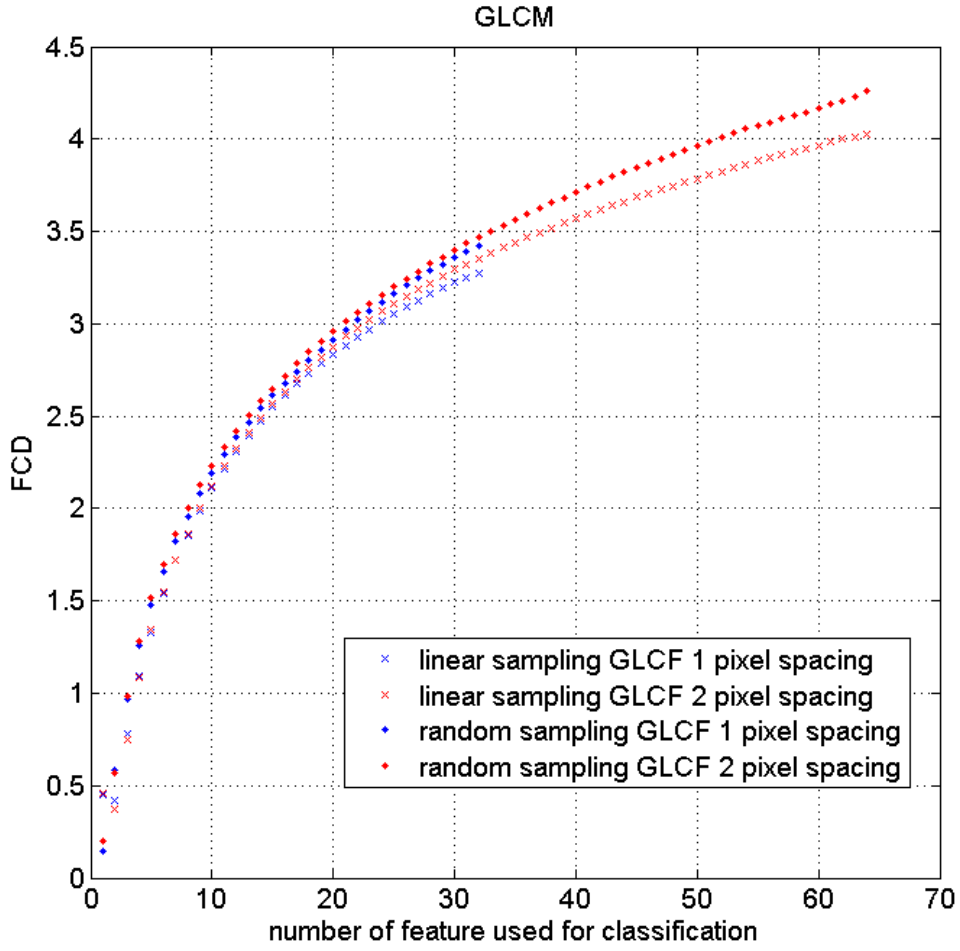


Figure 4.7: Average FCD scores for 100 linear and 100 random sample sets for the 5 texture test image with FSFS using GLCF with pixel spacings of 1 and 2.

when a single feature is used and the $30p$ samples per class criterion is met. When 2 features are used, the number of samples in the training sets is on the limit of the minimum required to fully describe the class and, with each subsequent feature added, the class of interest is increasingly under-represented in the training sets. In these circumstances, linear sampling can provide a better representation of the class providing the class is relatively homogeneous because its spatial autocorrelation effects correctly predict a lower variance, and hence a better representation of a relatively flat and homogeneous texture. Although the minimum FCD occurs when only 1 or 2 features are selected this does not always result in the maximum classification score; this implies that a good representation of the class does not guarantee a good classification performance. Indeed, the accuracy of the output map will depend on how well the training set represents the underlying class and how distinct the classes are within that feature subset.

The GM scores in Figure 4.6 show that feature selection has a greater affect when linear sampling is used. Initially, as features are added the GM increases for both random and linear sampling up to a peak value. For random sampling the GM score then slowly decreases with increasing features, with a maximum drop of 5-10% depending on the pixel spacing of the GLCF. The decrease for linear sampling happens at a much faster rate with the GM falling by over 25%, compared to the peak GM value. These results were also observed when GFB features and the fused feature set were used. For linear sampling, the highest GM values occur with GLCF for a single pixel spacing in the 4-6 features range, where the curve is relatively flat. This therefore provides a good choice for the number of features to use with FSFS minimum distance classifiers and linearly sampled training data in practical classification applications. Figure 4.6 also shows that the classification rate for linear sampling using GLCF with 2 pixel spacing begins to increases with the addition of features after the minimum at 49 features. The expected result would show an increase in the classification rate with additional features until a maximum is reached, and then the classification rate will level off or decline as observed for the random sampling. This behaviour was also observed using BSFS although the trend is reversed; the classification rate falling as features are removed before rising as expected to a maximum. This can be explained by the greedy nature of selection algorithms and as such it is possible to arrive a local minima.

Of the two wrapper techniques, FSFS and BSFS provide very similar GM scores, in particular for the difficult 5 texture image using linear sampling. Here, the choice between FSFS and BSFS may come down to computational efficiency which in turn depends on the number of features in the subset. When the subset size approaches the total number of features, BSFS will arrive at the best subset faster than FSFS. Conversely, with a small subset FSFS will be faster, although it is noted that both FSFS and BSFS are likely to be suboptimal due to the greedy nature of their algorithms.

Table 4.4: Peak GM scores (±2 standard deviations) using FSFS with GLCF with pixel spacings of 1 and 2. The corresponding number of features with which this performance is achieved is shown in brackets

		FSFS	
Pixel Spacing		1	2
2 textures	Random	0.876 ±0.449 (7)	0.888 ±0.024 (17)
	Linear	0.853 ±0.077 (5)	0.848 ±0.072 (5)
5 textures	Random	0.842 ±0.020(13)	0.834 ±0.020 (16)
	Linear	0.823 ±0.041 (5)	0.801 ±0.056 (5)

Table 4.5: Peak GM scores (±2 standard deviations) using BSFS with GLCF with pixel spacings of 1 and 2. The corresponding number of features with which this performance is achieved is shown in brackets

		BSFS	
Pixel Spacing		1	2
2 textures	Random	0.880 ±0.038 (8)	0.883 ±0.027 (13)
	Linear	0.856 ±0.072 (6)	0.840 ±0.066 (6)
5 textures	Random	0.842 ±0.024 (9)	0.825 ±0.027 (12)
	Linear	0.811 ±0.073 (6)	0.783 ±0.076 (7)

Tables 4.4 and 4.5 compares maximum GM values achieved using FSFS and BSFS and the number of features at which they occurred for both test images with linear and random sampling. The results show that over the 2 test images the average GM for linear sampling is only 3.2% below that of random sampling, with the difference for the 5 texture image being slightly less than for the 2 texture image. Linear sampling achieves its best performance with a feature subset that is typically less than half the size of the subset used by random sampling. Therefore, whilst random sampling provides the best classification performance of the feature selection approaches, it requires more features and hence is more computationally expensive. For linear training sets, the feature subset sizes for the peak GM scores show that a small subset of 7 or less features provides the best classification performance, and hence FSFS is more computationally efficient than BSFS.

Table 4.6: McNemar's test comparing FSFS with linear training data using GLCF with 1 pixel spacing to all other feature sets for the 5 texture test image. GLCF are denoted Cn where n is the pixel spacing and the GFB are described by the triplet (n_θ, n_f, γ) and Fused is a combined GLCF and GFB features

Feature Set	C2	$4,6,0$	$4,6,\frac{2}{3}$	$4,6,1$	$6,6,0$	$6,6,\frac{2}{3}$	$6,6,1$	Fused
Z mean	14.84	63.44	31.62	13.24	77.91	60.09	57.65	35.33
std	19.89	27.17	24.05	39.86	26.92	33.24	28.63	18.55
% of runs above significance threshold	74.1%	98.8%	89.1%	61.1%	99.8%	96.0%	97.4%	96.4%

Table 4.7: Best mean GM scores (±2 standard deviations) for each rotation of the Brodatz test image, using FSFS and SVM with GLCF with single pixel spacing . The corresponding number of features with which this performance is achieved is shown in brackets for FSFS

	Linear		Random	
Texture of interest	FSFS	SVM	FSFS	SVM
D29	0.823 ±0.041 (5)	0.813 ±0.071	0.842 ±0.020 (13)	0.871 ±0.036
D9	0.918 ±0.102 (1)	0.941 ±0.050	0.951 ±0.015 (1)	0.956 ±0.018
D38	0.881 ±0.091 (3)	0.899 ±0.068	0.931 ±0.029 (6)	0.940 ±0.019
D12	0.660 ±0.143 (7)	0.709 ±0.108	0.650 ±0.072 (15)	0.799 ±0.059
D57	0.788 ±0.088 (5)	0.805 ±0.084	0.812 ±0.036 (11)	0.851 ±0.039
Average	0.814 ±0.205	0.834 ±0.180	0.837 ±0.219	0.883 ±0.122

The GM scores for the more realistic 5 texture scenario suggest that FSFS using GLCF with single pixel spacing ($d = 1$) provide the best classification performance using linear training data. To determine whether this result is statistically significant, McNemar's Test is used to compare GLCF with d=1 to all other feature sets. Table 3 presents the mean z values and their standard deviations, with positive z values indicating that the GM classification performance of GLCF with d=1 is better than that of the feature set under comparison. For a significance level of 0.05, a value of $|z| > 1.96$ is statistically significant. The results in Table 4.6 show that the mean values of z are always positive and above the significance threshold. Using the standard deviations of z, Table 4.6 also shows the percentage of runs for which there is a 95% confidence that GLCF with $d = 1$ are significantly better. These results demonstrate that, even when compared to the next best feature set, GFB (4,6,1), GLCF with $d = 1$ are still significantly better for 61% of runs, and are significantly better for over 96% of runs for 5 of the other 7 feature sets compared. Similar analysis using SVM instead of FSFS shows that GLCF with d=1 again produce the highest GM results, although as all the mean values of z are in the range ±10 with standards deviations ≥ 20 this preference is not statistically significant.

The classification results presented above are for the Brodatz test images and to establish how well they can be generalized the textures in the five-texture Brodatz and Outex test images shown in Figure 4.1 and 4.2 are rotated such that the texture of interest to be identified is different in each image. This produces two sets of test images with five images in each. The two best two performing classifiers, FSFS and SVM, are then applied to each test image using the 100 linear and 100 random training sets shown in Figure 4.3. GLCF with a pixel spacing of 1 were used with both classifiers, as these feature have been shown to produce the best classification performance with linear training.

Table 4.8: Best mean GM scores (±2 standard deviations) for each rotation of the Outex test image, using FSFS and SVM with GLCF with single pixel spacing . The corresponding number of features with which this performance is achieved is shown in brackets for FSFS

	Linear		Random	
Texture of interest	FSFS	SVM	FSFS	SVM
Carpet 004	0.932 ±0.065 (2)	0.958 ±0.027	0.953 ±0.026 (4)	0.973 ±0.013
Tile 005	0.865 ±0.074 (6)	0.845 ±0.074	0.862 ±0.094 (7)	0.904 ±0.029
Canvas 026	0.786 ±0.080 (8)	0.840 ±0.070	0.785 ±0.060 (11)	0.883 ±0.025
Canvas 001	0.746 ±0.056 (4)	0.677 ±0.140	0.751 ±0.021 (15)	0.744 ±0.084
Canvas 002	0.779 ±0.057 (5)	0.791 ±0.092	0.772 ±0.026 (15)	0.837 ±0.028
Average	0.821 ±0.1516	0.822 ±0.204	0.825 ±0.159	0.865 ±0.159

The GM classification performances for each rotation of the Brodatz and Outex images and the averages of the 5 rotations are shown in Tables 4.7 and 4.8, respectively. The results for the five texture Brodatz image shown in Figure 4.1 that was used in the above evaluations are shown in the first row of Table 4.7, where D29 is the texture of interest. Comparison of the classification performance of D29 with the average performances for the Brodatz and the Outex textures shows that the D29 image is a very representative of the test images, as its performance for all classifiers and training sets is within a few percentage points of the average performances. The results in Tables 4.7 and 4.8 show that SVM outperforms FSFS for 7 of the 10 tests using linear sampling; and 9 out of 10 tests using random sampling. The peak GM scores for FSFS show that linear sampling uses a smaller feature set than random sampling, with a median value of 5 features for linear sampling compared to 11 features for random sampling, confirming the findings for the D29 Brodatz test image reported above. Overall, the best classification performance achieved with linear training data is 82.8%, an average of 4.6% lower than that achieved with random training sets.

The average GM score for the Brodatz image shows that SVM outperforms FSFS using linear (83.4% c.f. 81.4%) and random sampling (88.3% c.f. 83.7%). SVM also outperforms FSFS for the average GM score of the Outex test images, although the difference for linear sampling is only 0.1% (82.1% c.f. to 82.2%). To determine whether these results are statistically significant McNemar's Test is again used to compare the classification performance of SVM to that of FSFS. For the Brodatz images, the mean Z values were all negative and above the significance threshold of 1.96, which indicates that the better performance of SVM is significant. The one instance where FSFS has a higher GM than SVM is shown to be not significant. For the Outex images, the mean Z values show that only one of the two instances where FSFS yields a higher GM than SVM is significant. Overall, SVM is above the significance threshold for 80% of runs for the Brodatz images, and for 75% of the runs for the Outex images. This indicates

that SVM will generally outperform FSFS, although there will be some instances where FSFS will yield statistically significant and higher GM scores.

4.5 Conclusions

The performance of supervised texture classification schemes for the particular case where the training data is limited to a small number of samples from a linear path, a situation encountered in several important remotely sensing applications. In these circumstances, random sampling is not possible and the training data is also subject to spatial autocorrelation effects. Example applications of this type include the detection of atmospheric dust, volcanic ash and identifying oil spills. In these cases there is a single texture of interest to be identified within images containing multiple textures.

Using a simulated image of Brodatz textures, the classification performances of different textural features and 6 supervised classification techniques were quantified and compared with the results achieved using randomly sampled training sets. The best performing classification techniques for linearly sampled training data were found to be the FSFS selection wrapper approach and SVM. Both these techniques achieve their best classifications using GLCF with $d = 1$, although this was only statistically significant for the FSFS classifier. A range of 4-7 GLCF was also found to produce the best FSFS performance. The classification performance of 4 of the 6 classifiers was in excess of 80% whereas an unsupervised k-means classifier only achieved a GM of just over 60%; this underlines the benefits of a supervised approach.

When SVM and FSFS and GLCF with single pixel spacing were applied to the Brodatz and Outex test images with different rotations of the texture of interest, both classifiers achieve average GM scores in excess of 82% using linear sampling. The average performance drop was also less than 5%, when compared with random sampling. This implies that despite the limitations of the training data, supervised classification schemes are appropriate techniques for applications of this type. Despite the SVM classification score being negligibly higher than FSFS for the Outex images and only 2% higher for the Brodatz images, the differences are statistically significant indicating that SVM is the preferred classifier for linear training data.

The results provide an underpinning experimental basis for the use of supervised texture classification to identify atmospheric dust. This approach derives a linear training set from the CALIPSO VFM which can be used to train supervised classifier for texture feature generated from SEVIRI imagery [Wiltshire et al., 2009]. The next chapter presents a case study that uses linear training sets derived from the CALIPSO VFM

to identify particulate aerosol distributions in satellite imagery, comparing the results with existing methods for detecting particulate aerosols.

Chapter 5

Remote Sensing of Atmospheric Particulate Aerosols with Supervised Texture Classification Over Different Terrestrial Surfaces

This chapter describes the application of supervised texture classification using linear training sets for a real world remote sensing task. By applying the texture features and classifiers identified in the previous chapter to two orthogonal satellite sources, SEVIRI images from MSG and CALIPSO Lidar overpass data, an output mask identifying the spatial extent of a particulate dust event can be produced.

The algorithm broadly follows the approach used in the simulation, with the inclusion of some additional pre-processing steps to accommodate the satellite images and training data. The algorithm is split into three stages:

1. Data selection and image registration: This enables the CALIPSO Vertical Feature Mask (VFM) to be used to identify the training pixels in the SEVIRI image
2. Generation of texture features: Using the best performing texture feature set findings from the simulations, texture features are generated for the SEVIRI image
3. Supervised Classification: Using the linear training set from the positions iden-

tified by the CALIPSO VFM, the preferred classifiers from the simulations are applied to generate a output dust mask.

Two example events are included; one over land and one over water, to demonstrate that the supervised texture approach can be used without modification over multiple surfaces. The example events are fully described in Section 5.4. The resulting output dust masks are evaluated against other existing methods for particulate dust detection. The choice of the technique for comparison is dependent on multiple factors including the underlying surface type and the time. Details on the when existing techniques are most applicable are discussed in Section 2.2.

5.1 Data Selection and Image Registration

The latest MSG satellite, designated Met-9, is a geostationary satellite positioned over prime meridian, and is ideally situated to monitor Saharan dust storms and their movement over the Atlantic Ocean. The SEVIRI sensor on MSG provides images with good latitudinal, longitudinal and temporal resolution over 12 spectral bands between $0.75\,\mu m$ and $13.4\,\mu m$, with a 15 minutes repeat cycle, but no vertical information.

As described in chapter 2, the CALIPSO satellite follows a 16 day polar orbit with the CALIOP sensor providing a continuous, along track vertical cross-section of the atmosphere. The CALIPSO overpass provides excellent vertical and along-track resolution but poor cross-track and temporal resolution. The CALIPSO VFM classifies the atmosphere into 8 classes, of which one is aerosol. The aerosol class is further subdivided in 8 subclasses, including mineral dust. Once the CALIPSO VFM is registered with the SEVIRI image, dust locations can be identified provide the training data required for supervised texture classification algorithms.

The first task is to identify a dust event that coincides with a CALIPSO overpass. CALIPSO archive data is available online [NASA, 2008] from the Langley atmospheric sciences data centre. A Java tool can be used to browse the data products from the CALIPSO project, and is searchable by data and location. The data product of interest is the level 2 VFM v2.01 which is latest release at the time this thesis was authored. The data is available in HDF format and can be directly imported into Matlab, but needs to be reconstructed before use [ASDC, 2008]. The CALIPSO VFM data packing and reconstruction method is included in Appendix C.

In order to register the two data sources, the SEVIRI image that most closely matches

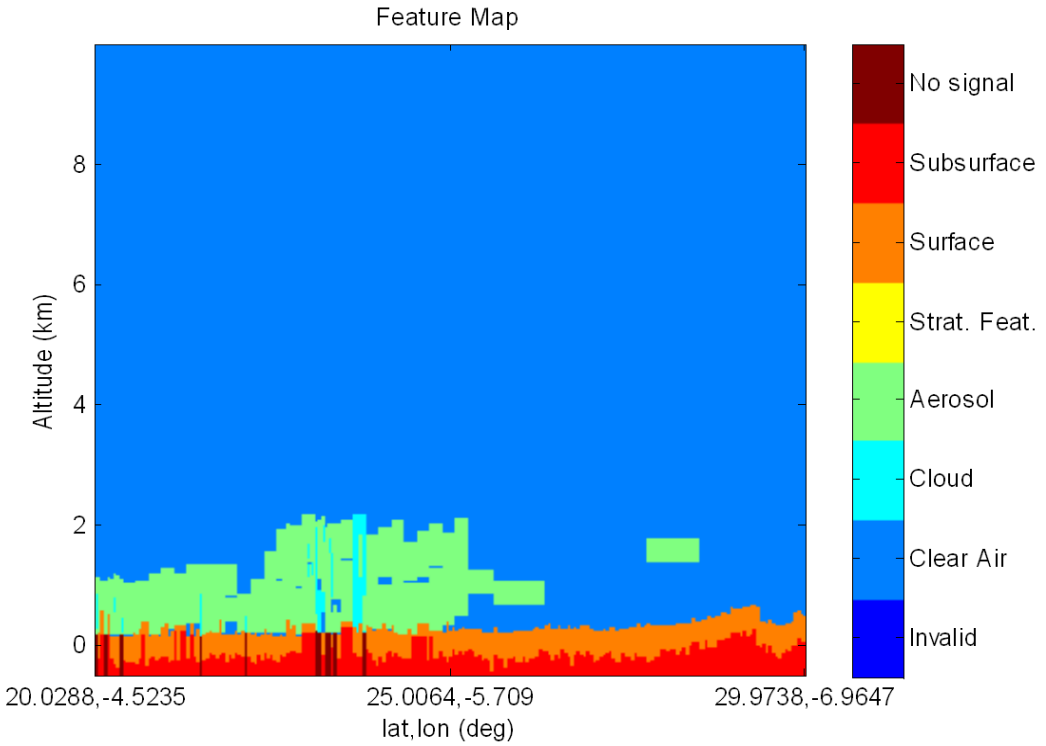


Figure 5.1: Saharan dust outbreak on 21/02/2007. Reconstructed vertical feature mask from the CALIPSO data for an overpass of North Africa from UT 13:50

the time and location of the CALIPSO overpass is obtained. SEVIRI images are generated at 15 minute intervals, although the image is formed over a 12 minute scan. The CALIPSO VFM is tied to its polar orbit, and hence time varies along-track. The approach used here is to assume that the overpass time is locally constant and to select the closest SEVIRI image to the CALIPSO overpass time. This is a reasonable assumption as in the worst case the error will be 7.5 minutes.

SEVIRI images are available from the EUMETSAT online archive [EUMETSAT, 2008] and are provided as a full disk image of the Earth centred over 0° longitude in an orthographic projection. Each image pixel has an associated latitude and longitude tag. To reduce the processing time required to generate the texture features, a region of interest is extracted. Here a $10° \times 10°$ region of interest is defined, centred on the dust event. The dust event itself is selected by using the CALIPSO VFM to identify an edge of dust cloud, so that region of interest will contain known dust and non-dust samples. The latitude and longitude of the CALIPSO is used to centre the region of interest, and this section is extracted from the full disk and is remapped so that the pixels are linear with latitude and longitude.

The SEVIRI data is delivered representing the counts and not radiance data. Counts

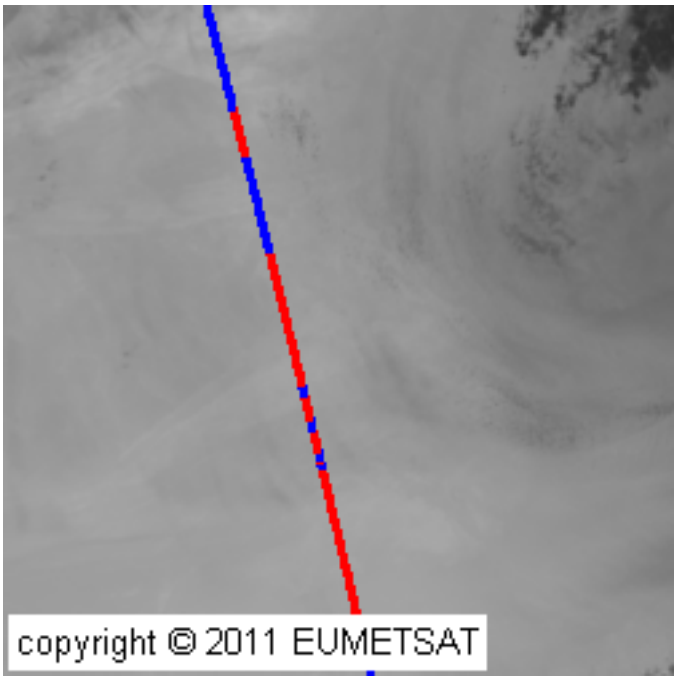


Figure 5.2: IR10.8 MSG image from UT 13:45 on 21/02/2007 with the corresponding CALIPSO overpass marked in blue and dust location in red.

are used by EUMETSAT to match the full range of the 10 bit integer representation to the range of radiances observed in each of the 10 MSG channel [Müller, 2007]. Counts can easily be converted to radiance by a simple linear scaling and offset, which is included in Appendix B.1. Here, SEVIRI channel 9 is used as it corresponds to the $10.6\mu m$ spectral band in the IR range. The visible range is not useful over land surfaces as airborne dust has similar reflective properties to the underlying surface. IR is selected as is provides good contrast for wind raised dust when compared to heated land surfaces [Lee, 1989]. Another advantage of using IR is the availability of imagery at night, enabling 24 hour coverage to be provided.

With the region of interest extracted and remapped, the task of spatially registering the two data sources only requires a simple solution. By defining the resolution of the pixels in the region of interest in terms of latitude (res_{lat}) and longitude (res_{lon}) the CALIPSO overpass can be mapped on the SEVERI image by.

$$res_{lat} = \frac{range_{lat}}{rows} \tag{5.1}$$

and

$$res_{lon} = \frac{range_{lon}}{cols} \tag{5.2}$$

This provides a measure of the degrees per pixel. As the latitude and longitude is known for each of the pixels in the SEVERI image, any latitude and longitude can be mapped to the image. For simplicity, the origin is selected as the bottom left of the SEVERI image, and the pixel locations for a given latitude and longitude can be found using:

$$x = \frac{longitude - lon_0}{res_{lon}} \tag{5.3}$$

and

$$y = \frac{latitude - lat_0}{res_{lat}} \tag{5.4}$$

where longitude and latitude are the input coordinates to map to the region of interest, x and y are the output pixel locations within the remapped SEVIRI image, and lat_0 and lon_0 are the latitude and longitude of the selected origin.

Each VFM data entry has geographical latitude and longitude tags, which allow the data to be mapped to the corresponding pixel within the reprojected SEVIRI image. Each entry in the VFM is represented by a 16 bit integer which contains information on the feature type and subtypes specific information. Table C.1 contains a description of the feature classification flags to interpret the VFM. The first task in identifying the training set is to find the entries classified as aerosol, with sub feature type of interest, in this instance dust. If dust is present and is not occluded by cloud or another aerosol type then it is marked as dust, all other entries are marked as non-dust. This achieved in practice by examining each column of the VFM, if all altitudes higher than the dust are identified as clear air then the training sample is labelled as dust. The thickness of the dust layer is not considered here, so a single dust entry in the VFM will be classified as dust providing it is not occluded. The minimum thickness for a dust layer to be enter the training set is therfore set to the vertical resolution of the VFM. This varies depending on the height of the layer, with a resolution of 30m under 8.2km, 60m between 8.2 to 20.1km and 180m for over 20.1km (see Figure C.1).

Figure 5.2 shows an example SEVIRI IR10.8 μm image for a Saharan dust outbreak from 21/02/2007 that corresponds to the CALIPSO VFM of figure 5.1. This is taken

from the land test case used later to evaluate the performance of the supervised texture dust mask.

5.2 Generation of Texture Features

The simulation results in the chapter 4 indicate that GLCF with a single pixel spacing provides the best texture features for supervised classification using linear training set. However, the application to real world imagery and textures presents a more complex problem than the two simulation images considered. In this case, GLCF may not produce the best results with a linear training set.

Also, the previous work using unsupervised texture classification should not be ignored. Khazenie and Lee [1992] compared GLCF, Normalised Difference histograms and 2D GFB, concluding that 2D GFB were the most efficient and robust in detecting aerosol features. Despite the unsupervised approach used, this demonstrates that in some circumstances 2D GFB produce features that provide good separation of aerosols in the feature space, which lead to good classification results. Therefore, a full range of GLCF and 2D GFB were applied to see if the conclusions resulting from the simulation hold for real applications of remote sensing of atmospheric aerosols.

Before the texture features are calculated, the underlying background representing the spectral response of the Earth's surface can be removed. This leaves the spectral response of the atmosphere only, which intuitively suggests it will produce better results as the texture features relate to the atmospheric content and not the Earth's surface. Another advantage of removing the background image is that it allows the classification to be applied across land-ocean boundaries. This approach is adopted by Diop et al. [2006] in their unsupervised texture classification, and for the IDDI [Legrand et al., 2001] and other radiative transfer model approaches for detecting atmospheric aerosols over land surfaces. When these approaches are applied over ocean as well as terrestrial surfaces, the removal of the background reference image is not deemed necessary as the ocean is sufficiently flat and dark as to have minimal effect on the overall spectral response.

The underlying albedo reference image is built up by examining other SEVIRI images from the same time of day over 15 day window centred on the day of interest. For each pixel, a probability density is generated and the most likely value over the 15 day period is selected for the reference image (see Figure 5.3). This approach relies on clear sky responses being prevalent with the 15 day window. The underlying image is removed from the SEVERI image using an exclusive OR (XOR) operation [Bayoko et al., 1996,

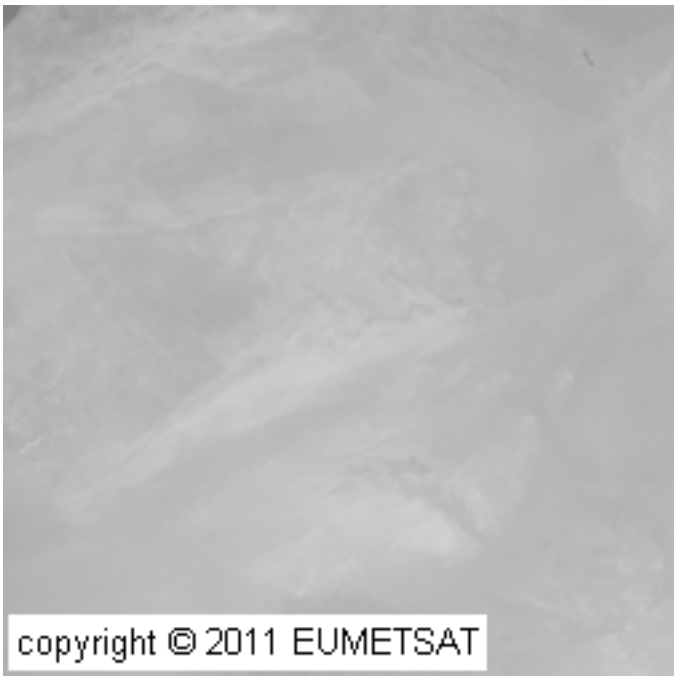


Figure 5.3: 15 day background reference image centred on 21/02/2007 at 13:45UT.

Diop et al., 2006] (see figure 5.4). This approach is preferred to a simple-element wise subtraction as the latter requires different ordering over land and ocean due to the differing surface brightness.

As in the simulations, GLCF were generated for pixel spacings $d = 1$ and $d = 2$ within a fixed 9×9 window. The orientations of the features are set by considering all possible spacings using the Chebyshev distance. For each pair of orientation and spacing, 8 features were calculated (see Table 3.1) giving two GLCF sets, of 32 or 64 features for $d = 1$ and $d = 2$ respectively.

Features from the 2D GBF were generated using the magnitude response of when each filter is convolved with the input image. The GFB were calculated using equations 3.1 - 3.6 as described in Chapter 3. GFB were generated using, $n_\theta = 4$ *or* 6, $n_f = 6$, and $\lambda = 0, \frac{2}{3}$ or 1. This results in 24 or 36 features, depending on the value of n_θ, and gives 6 different feature sets when the three post filtering schemes are considered.

5.3 Supervised Texture Classification and Evaluation

The simulations identified that the wrapper approaches and SVM consistently provided the best classification results for linear training sets. Small feature sets of in the range of 4-7 features provided the best classification for the wrapper approaches, and so FSFS

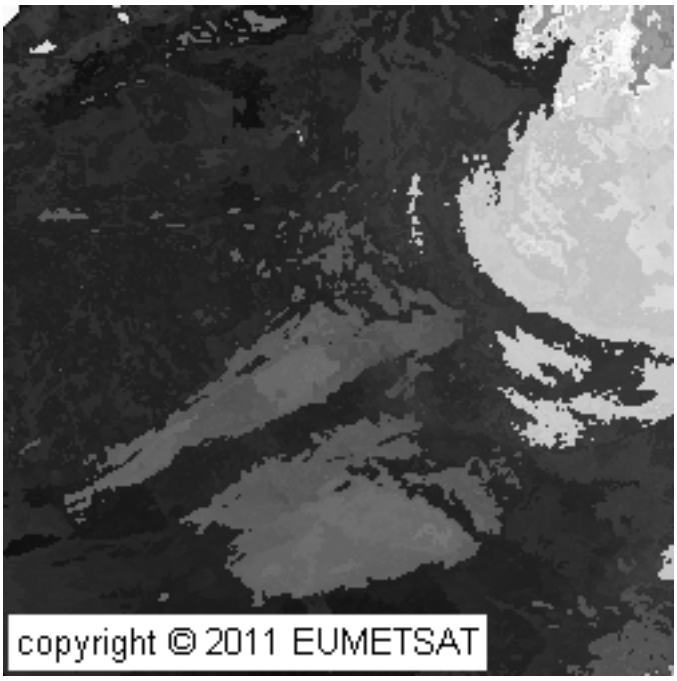


Figure 5.4: IR10.8 SEVIRI image from 21/02/2007 UT13:45 with background reference image removed by XOR operation.

is computationally more efficient than BSFS. For this reason FSFS with a maximum feature subset of 5 and SVM were selected to generate output masks.

The classification approaches were applied to texture features generated for the SEVIRI input image with and without background reference image removed. This allows the effect of removing the spectral response from Earth's surface to be determined.

The output masks from the supervised texture classification algorithm were compared to dust flags from OMI on AURA (ocean and land) and BTD (land). The OMI dust product was selected due to its proven long history of contribution to monitoring of atmospheric aerosols [Ahn et al., 2008]. In particular the near-UV aerosol sensing technique (See Chapter 2) was used for comparison due to its ability to detect desert dust over terrestrial background. Due to different resolutions of the input SEVIRI images and the OMI dust flag, the supervised texture dust flag is down sampled and remapped to match the OMI resolution and projection before comparison.

The OMI near-UV aerosol product (OMAERUV) is capable of distinguishing between smoke, dust and sulfates and also includes entries for when the aerosol type cannot be determined. This OMAERUV also has an associated quality assessment flag which indicates the confidence in the outputs and failure mechanisms when no retrievals are possible. These flags are used to indicate the reliable entries in OMAERUV, and only these regions are used in the comparison to the output masks generated by the

supervised texture classification method. Reliable entries in the OMAERUV *Final Algorithms Flag* can be classified as most reliable (0), reliable (1) and less reliable (2), and these values are used to identify the appropriate data for comparison.

The BTD dust mask was proposed by Zhang et al. [2006] for MODIS imagery and was used as an additional comparison technique for the test case over land. The output mask consists of 5 classes, including strong and weak dust. This output is recast as a binary dust mask, combining strong and weak region as the supervised texture approach does not distinguish between the relative strengths. The higher resolution of the BTD dust mask provides a better indication of the accuracy of the supervised texture classification at the borders with cloud, which can be missed in the larger footprint of OMI pixels.

For both the OMI dust flag, and the BTD dust mask, the Geometric mean (Equation 4.6) was used to rate the accuracy of the dust mask from the supervised texture algorithm relative to the other dust products.

5.4 Evaluation of Dust Event Test Cases

The supervised classification algorithm was applied to two dust events over; one over ocean and one over land. The European organisation for exploitation of METereological SATellites (EUMETSAT), who operate the MSG satellites with the European space Agency (ESA) provide a selection of topical test cases for study on their website, and the two examples were selected from the list of dust test cases observed with the SEVIRI sensor on MSG EUMETSAT [2010].

The CALIPSO VFM were registered to the IR10.6$\mu\,m$ region of interest for the corresponding SEVIRI image, and the training sets generated, illustrated in figures 5.7 and 5.14. The region of interests for both cases were selected to include an apparent edge in the dust storm, as indicated in the CALIPSO VFM (see figures 5.6 and 5.13, for the ocean and land test cases, respectively.

5.4.1 Ocean Test Case

The ocean test case is part of a major transport of desert dust from West Africa to the Lesser Antilles during 20th -26th June 2007. The specific images used were taken from 21st June 2007 at 16:15 UTC. Figure 5.5 shows the MODIS false colour RGB image

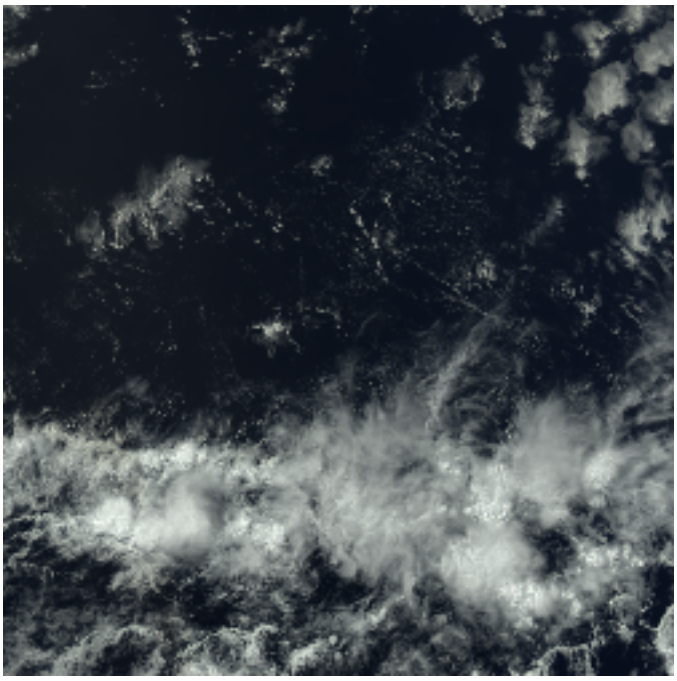


Figure 5.5: False Colour RGB of region of interest for the Ocean test case, generated using MODIS on AQUA data from 16:15UTC 21/06/2007

of the region of interest for the ocean test case. Figure 5.6 shows the corresponding section of the CALIPSO VFM, with all of the region classified as aerosol also having the subclassification of mineral dust. Figure 5.7 shows the SEVIRI IR10.6μm over the region of interest, with the CALIPSO VFM overpass indicated in blue, and dust locations indicated in red.

Table 5.1 shows the best GM scores comparing the supervised texture dust flag to the OMI dust flag for the the Ocean test case. The results show high correspondence with the OMI dust flag, with a GM of 92.8% using a single feature which drops by 6.5% to 86.3% when 5 features were used (see 5.8). A qualitative comparison of the output mask with the false colour RGB of the scene indicates that a single feature may not provide the best results.

Figure 5.9 shows the supervised texture dust flag for the ocean test case, using a both 1 and 5 features. Whilst the GM score is higher when compared to OMI for 1 feature, the RGB image (see Figure 5.5) shows that dense cloud regions in the lower half of the image were identified as dust. When 5 features were used, the regions of cloud misclassified as dust were reduced.

The GM is higher for 1 feature due to the dense cloud regions being misclassified in the OMI data as dust, which indicates the inherent problem of comparing the results to data that itself can contain errors.

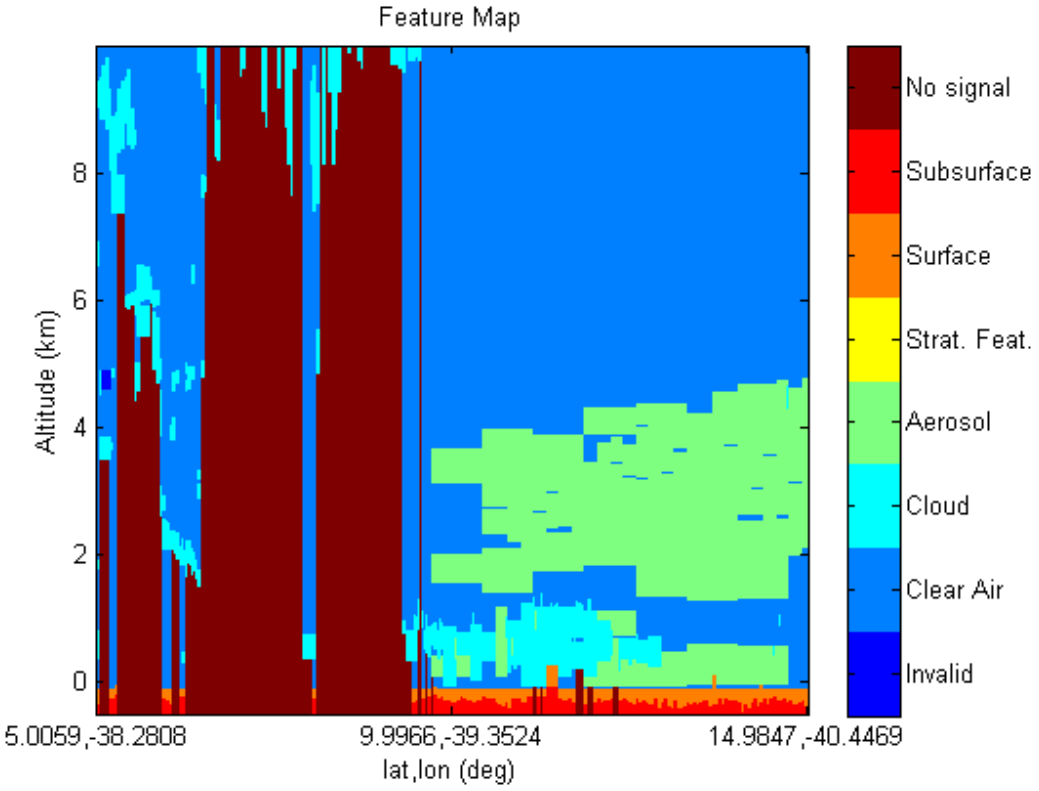


Figure 5.6: CALIPSO VFM for ocean test case on 21/06/2007

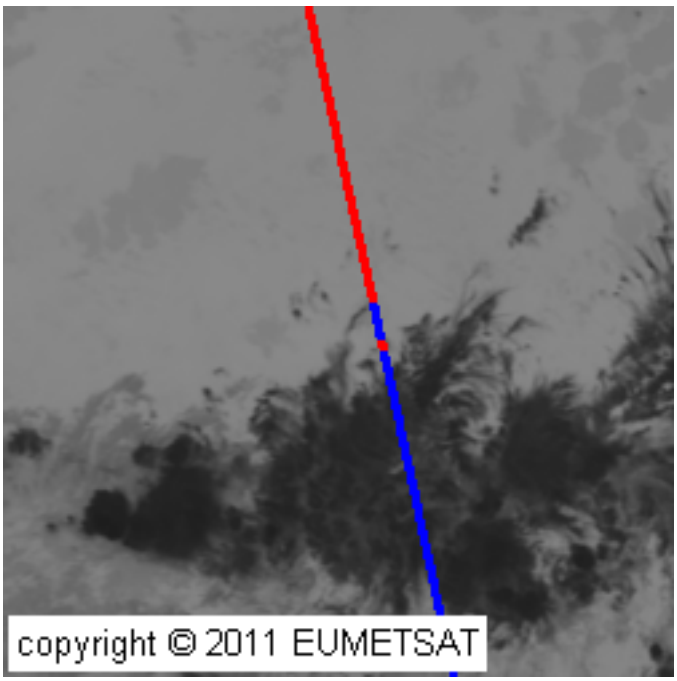


Figure 5.7: IR10.8 SEVIRI image for the ocean test case at UT 16:30 on 21/06/2007 with the CALIPSO overpass marked in blue and the dust locations in red

Table 5.1: Best geometric mean scores for the Ocean test case comparing supervised texture dust flag to the near UV OMI dust flag for multiple texture types . For FSFS, the number of features that produced the best GM are included in brackets.

Ocean Test Case		
Feature Set	Classifier	
	FSFS	SVM
C1	0.9282(1)	0.8788
C2	0.9234(1)	0.8425
4,6,0	0.8634(1)	0.8634
4,6,$\frac{2}{3}$	0.8788(1)	0.8530
4,6,1	0.8814(1)	0.8399
6,6,0	0.8557(4)	0.8345
6,6,$\frac{2}{3}$	0.8989(1)	0.8073
6,6,1	0.9063(1)	0.8425

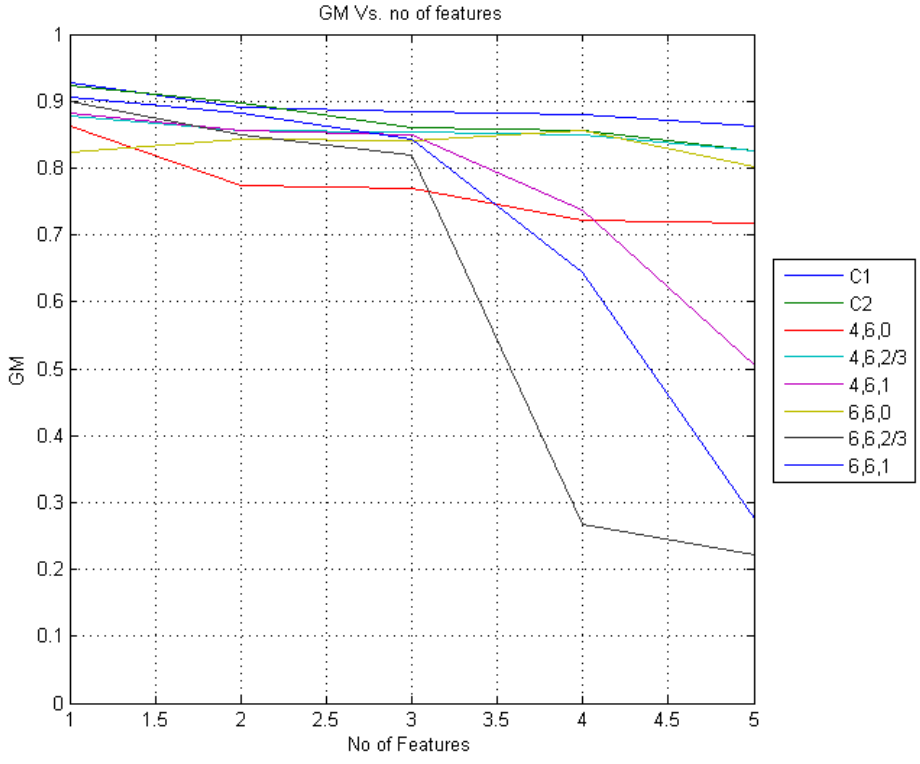


Figure 5.8: GM versus features using multiple Feature Sets for the Ocean test case

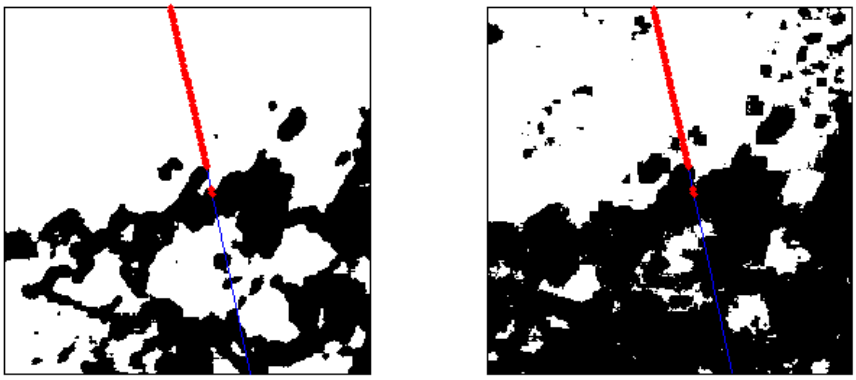


Figure 5.9: Supervised texture dust flag using GLCF with d=1 and FSFS with 1 (left) and 5 (right) features for the ocean test case with dust indicated in white. The CALIPSO overpass is overlaid in blue and dust location indicated with red.

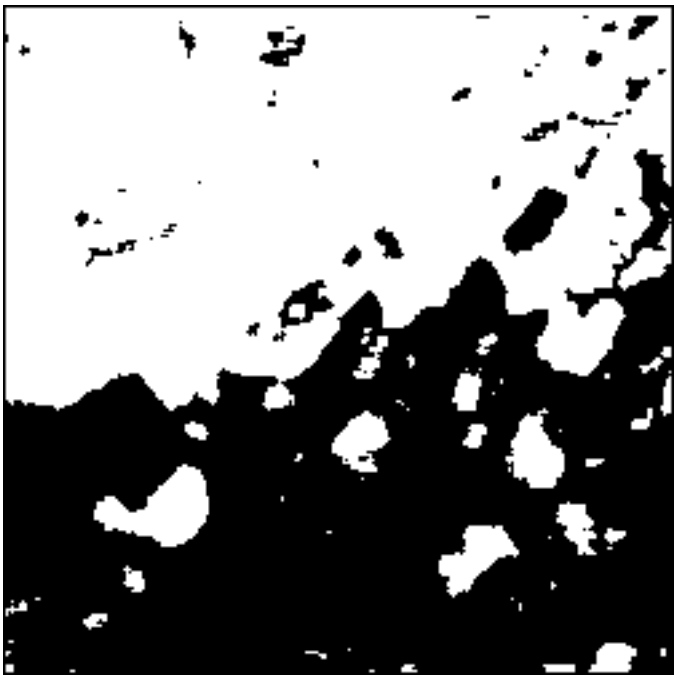


Figure 5.10: Supervised texture dust flag for the ocean test case using SVM on GLCF with d=1. Dust indicated in white and the CALIPSO overpass is overlaid in blue and dust location indicated with red.

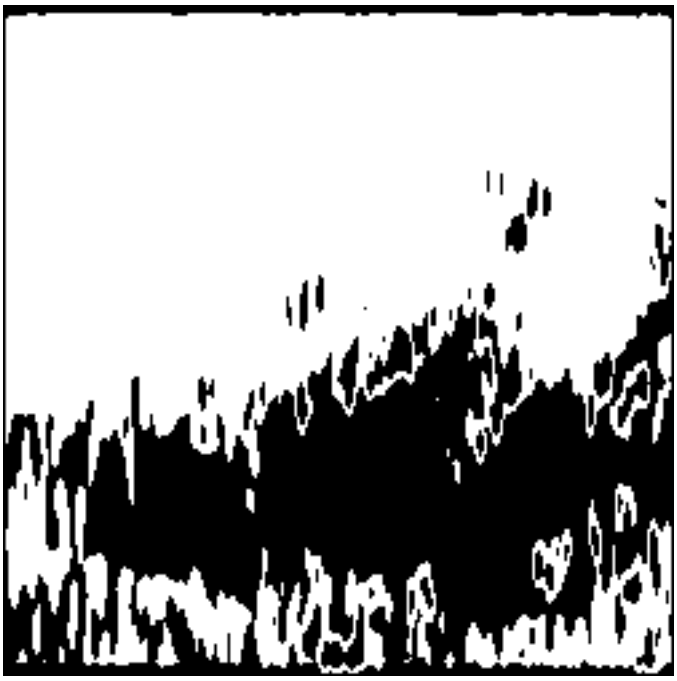


Figure 5.11: Output mask for the Ocean test case using FSFS with 1 feature and 2D GFB (6,6,1)

The ocean test case shows high GM scores for SVM (Table 5.1), again with the best results occurring for GLCF with $d = 1$. In this case, the resultant output mask produces very similar results to FSFS and can be seen in figure 5.10. The good performance of SVM for the ocean test case supports the idea that SVM will work providing the non-dust class does not contain a high number of textures. Over ocean, there is little background response from the ocean, and the CALIPSO VFM indicates the scene is dominated by Dust and Cloud. This scenario matches the simulations in the previous chapter and hence SVM yields a good performance.

The GM scores for all the 2D GFB indicate a good level of performance for the ocean test case ($> 80\%$), although GLCF with d=1 provides a higher GM. Examination of the output masks however shows that whilst 2D GFB indicate the bulk area where dust is present, the performance at the dust / non-dust boundary is poor. The best performing 2D GFB (6,6,1) is included as an example in figure 5.11, which demonstrates the smoothing and poorer performance at the boundaries.

5.4.2 Land Test Case

The land test case is taken from a large dust swirl over Algeria which occurred during the 20th - 22nd February 2007, with images taken at 13:45 UTC from the 21st February

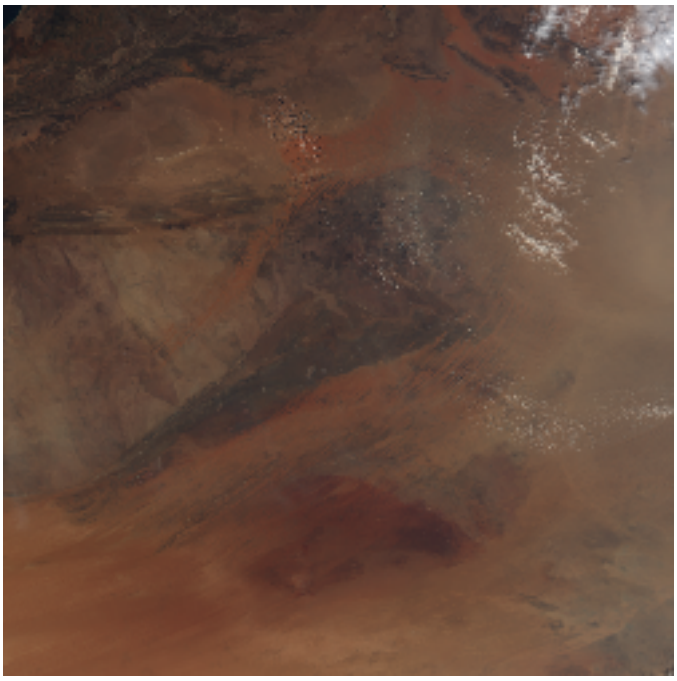


Figure 5.12: False Colour RGB of region of interest for the Land test case, generated using MODIS on AQUA data from 13:45UTC on 21/02/2007.

(see figure 5.12. For the land test case, the background reference image was generated with the option to remove it using the XOR operation (see figure 5.4).

Comparison with OMI dust product

Also, these dust event examples present a much more difficult problem to the classifiers, in particular to SVM. The reduction of the CALIPSO VFM to a binary set including 'dust' and 'non-dust', is an over simplification that can impact on the classification performance. Whereas in the simulations, non-dust could be one of 4 different texture, in the real world examples non-dust can contain a much larger range textures including other aerosols, clouds, different surface types and combinations of all these. This combination of textures in the non-dust class can make defining the hyperplane between the dust and non-dust class difficult, which affects the classification performance. FSFS is less susceptible to this as this is a true binary classifier and only considers the dust entries in the training set in order to identify dust throughout the whole image.

The land test case shows that the supervised texture approach can provide a good correspondence ($> 80\%$) with the OMI dust flag. The overall performance is generally lower than the ocean test case, which can have multiple causes. The OMI retrieval is a harder task over terrestrial surface, which can affect the accuracy of the OMI data

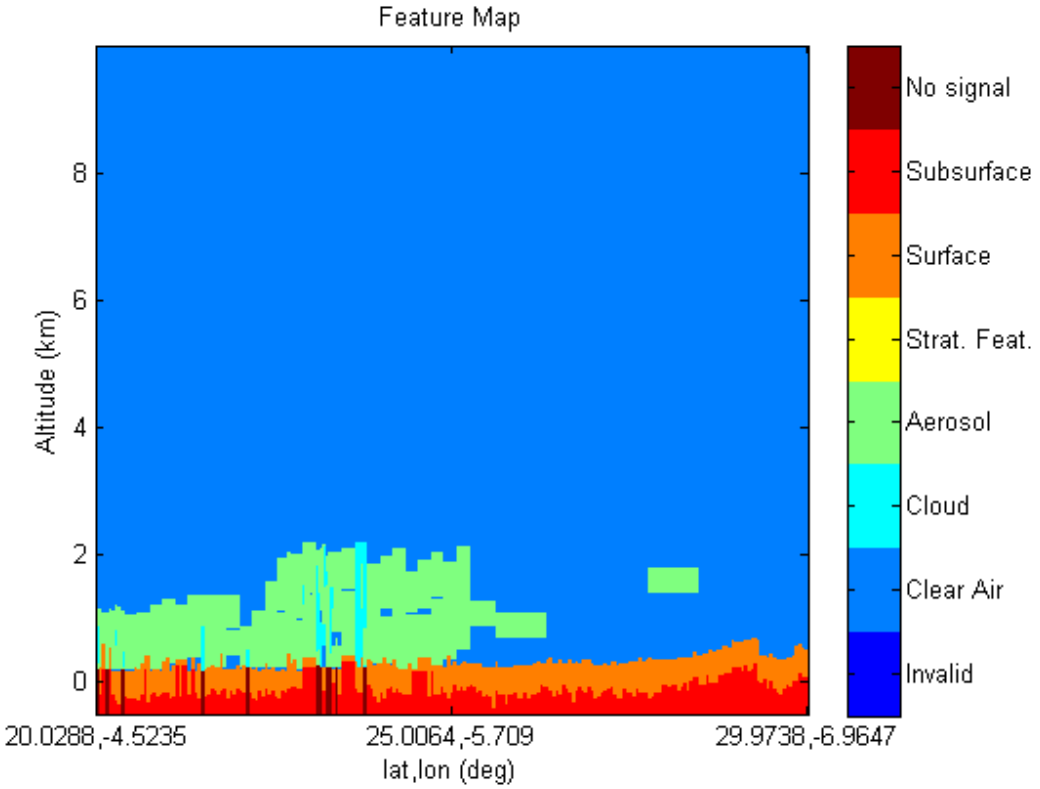


Figure 5.13: CALIPSO VFM for land test case on 21/02/2007 over North Africa from UT 13:50.

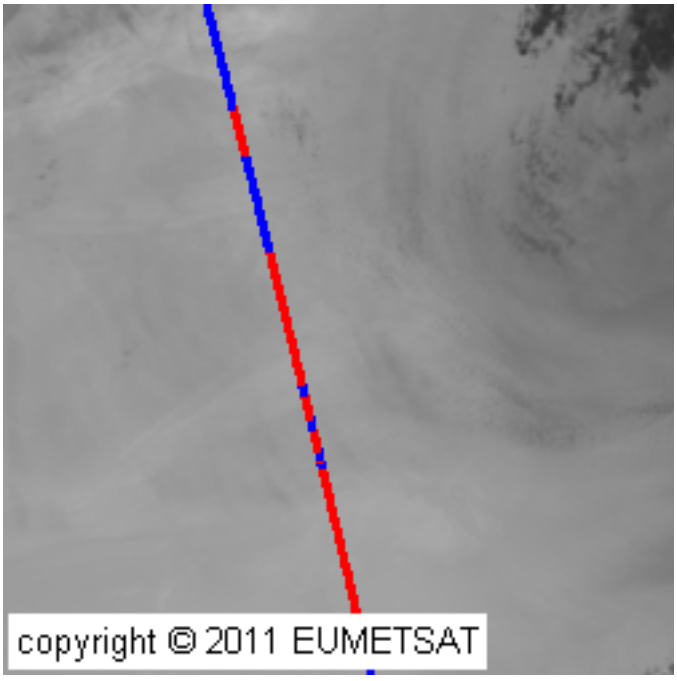


Figure 5.14: IR10.8 SEVIRI image for the land test case at UT 13:45 on 21/02/2007 with the CALIPSO overpass marked in blue and the dust locations in red.

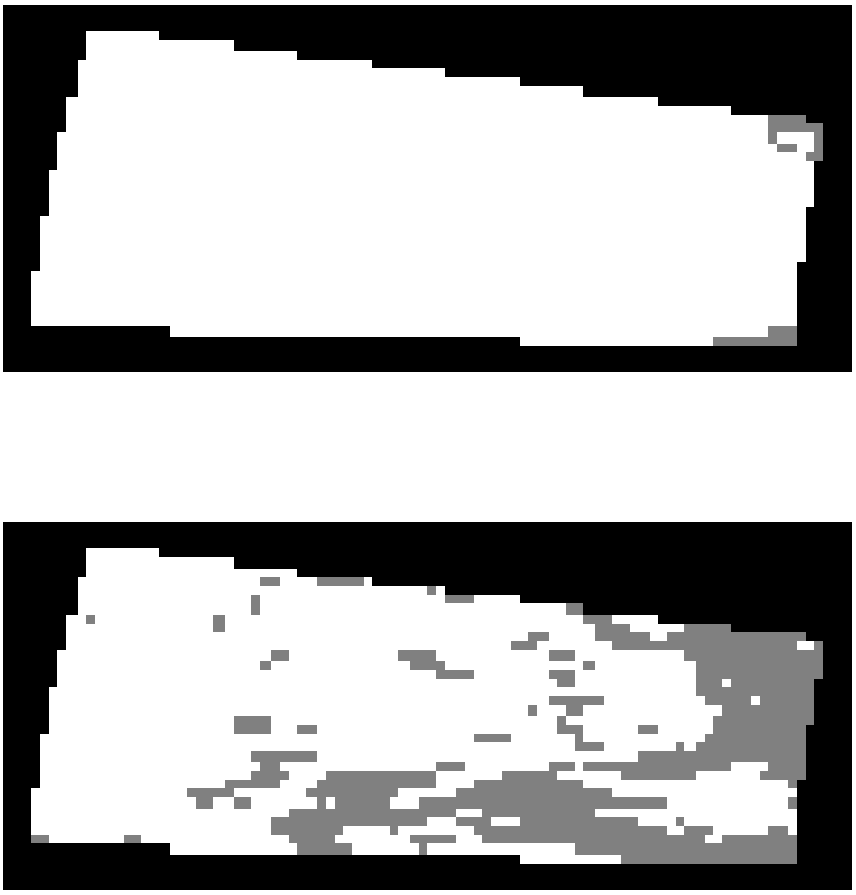


Figure 5.15: OMI dust flag (top) and the supervised texture dust flag (bottom) for the land test case, down sampled and projected to match OMI. Dust is shown in white, Non-dust in grey, and black signifies the region outside of the area of interest.

Table 5.2: Best geometric mean scores for the Land test case comparing supervised texture dust flag to the near UV OMI dust flag for multiple texture types . For FSFS, the number of features that produced the best GM are included in brackets.

Land Test Case OMI				
Feature Set	With background		Background removed	
	FSFS	SVM	FSFS	SVM
C1	0.8008(1)	0.6398	0.7247 (4)	0.1985
C2	0.7798(1)	0.6734	0.7429(2)	0.4646
4,6,0	0.8183(1)	0.3276	0.5346(4)	0.3993
4,6,$\frac{2}{3}$	0.8130(1)	0.1963	0.7890(1)	0.6876
4,6,1	0.7153(1)	0.6439	0.8092(1)	0.6860
6,6,0	0.6763(5)	0.5909	0.8536(3)	0.000
6,6,$\frac{2}{3}$	0.7830(2)	0.6371	0.7004(1)	0.2617
6,6,1	0.8044(2)	0.7529	0.6361(1)	0.1891

used for comparison. Also, the non-dust class is likely to contain multiple textures, which poses a more difficult classification task.

The evaluation is further complicated, as the OMI dust flag indicates that the scene is nearly completely dominated by dust apart from a small region of cloud in the top right corner. This allows an otherwise poor output mask which classifies everything as dust to appear to closely match the OMI dust flag. Figure 5.15 shows the OMI dust flag compared to the supervised classification dust mask for FSFS with 1 feature using GLCF with d=1, and indicates this issue. For this reason, the BTD dust mask is also used. As the BTD approach works on the resolution on the input data, results will match the resolution of supervised texture approach as opposed to the $13 \times 24\,km$ resolution of the OMI dust flag. As such, comparison to the BTD dust mask provide a better indication of the performance, particularly at dust cloud boundaries.

In most cases, removing the background image results in a lower GM. However the highest GM of 85.4% occurs with the background removed using FSFS on 2D GFB (6,6,0). This implies that removing the background can produce a better supervised classification result. Figure 5.16 shows the output mask from supervised texture classification using FSFS for 1 feature with 2D GFB(6,6,0). Comparison of this with the false colour RGB image scene (see 5.12), shows that the mask misclassifies obvious cloud regions, with predominantly only the borders between dust and cloud regions not classified as dust. This presents a rather unphysical dust distribution and only performs well in comparison due to the low resolution of the OMI dust flag coupled with the scene being dominated by dust.

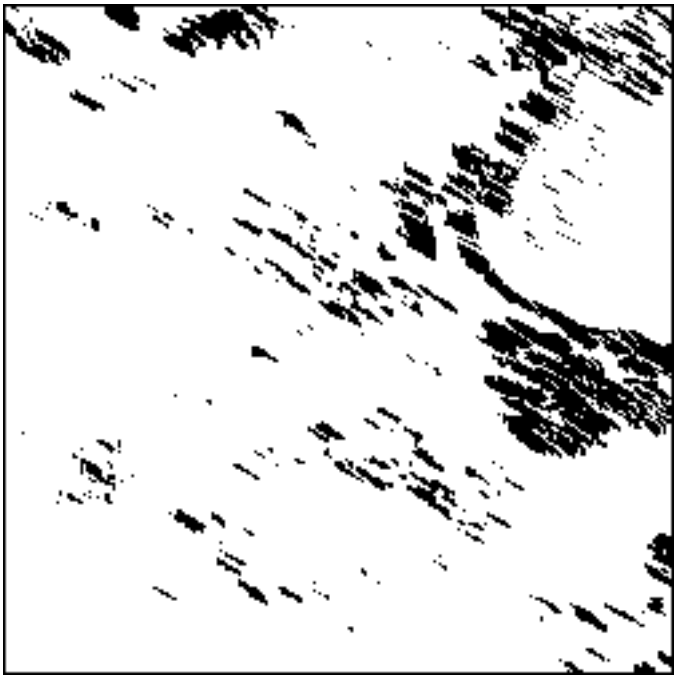


Figure 5.16: Output mask for the Land test case with the background removed using FSFS with 1 feature and 2D GFB (6,6,0).

Table 5.3: Best geometric mean scores comparing supervised texture dust flag to the BTD dust flag for multiple texture types. For FSFS, the number of features related to best GM is included in brackets.

Land Test Case BTD				
Feature Set	With background		Background removed	
	FSFS	SVM	FSFS	SVM
C1	0.7739(1)	0.6189	0.6258(4)	0.5061
C2	0.7604(1)	0.6393	0.5370(5)	0.5252
4,6,0	0.6491(2)	0.4424	0.5680(3)	0.5332
4,6,$\frac{2}{3}$	0.6228(1)	0.4515	0.5585(3)	0.6108
4,6,1	0.5790(1)	0.5171	0.5797(2)	0.6164
6,6,0	0.6647(3)	0.5583	0.6166(5)	0.4028
6,6,$\frac{2}{3}$	0.6200(3)	0.5643	0.5656(2)	0.4346
6,6,1	0.6050(2)	0.6159	0.4509(1)	0.5086

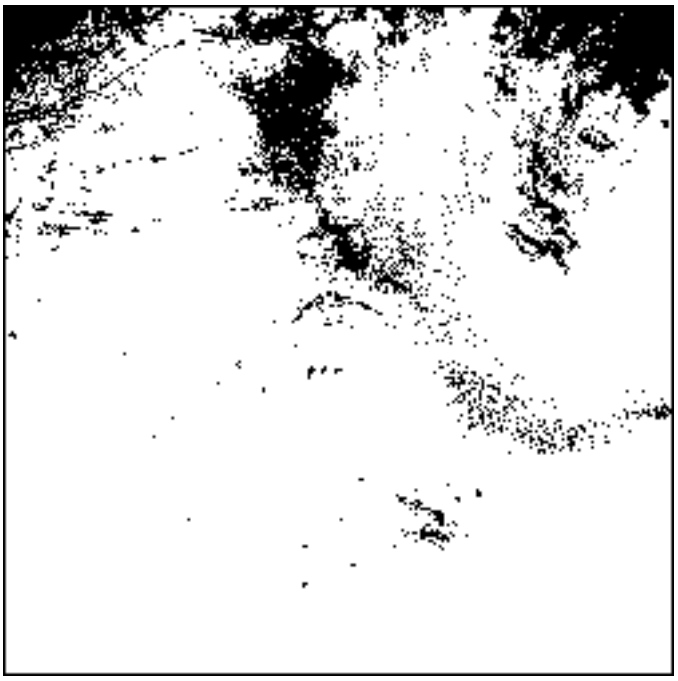


Figure 5.17: BTD dust mask generated using MODIS imagery for the land test case.

Comparison with BTD dust mask using MODIS imagery

Table 5.3 shows the GM for the land test case compared to the BTD dust mask generated using MODIS imagery (Figure 5.17). The results show that the best performance is achieved using 1 feature, achieving a GM of 77.4%. The GM score is lower when compared to BTD dust mask as opposes to the OMI dust flag, due to the fixed window sizes used to calculate the texture features. The features are calculated using a 9×9 window, which allows regions up to 4 pixels from the centre to affect the value. Therefore, whilst the window may be centred on a dust pixel, cloud can be present within the window which can affect the texture feature enough such that the supervised texture algorithm does not classify it as dust. This can be seen by comparing the supervised texture dust flag with the false colour RGB (Figure 5.12). For the land test case, a dense dust region can be seen in the upper right of the image interspersed with clouds. The supervised texture dust flag identifies this dust region, and rejects the cloud, although the texture approach is conservative at the dust cloud boundary, leading to dust being misclassified.

The down-sampling and remapping of the supervised texture dust mask for comparison with the lower resolution OMI dust flag has an averaging effect which reduces the impact of the conservative classification, and the OMI dust flag contains less boundaries, as demonstrated in Figure 5.15. The BTD dust mask has significantly more borders and operates at the same resolution as the supervised texture dust flag, and hence the GM

Figure 5.18: Output mask for the Land test case using FSFS with 1 feature and 2D GFB (4,6,0)

is lower.

Comparison of the feature sets show that GLCF provide better GM scores by at least 10% compared to 2D GFB evaluated against the BTD dust mask. This is not evident when the OMI dust mask is used for comparison, which indicates that 2D GFB can perform better than GLCF. An example supervised dust mask is included in Figure 5.18, which indicates the dust mask for the land test case with the background present using FSFS with 1 feature and 2D GFB (4,6,0). The poor performance is indicated by the lack of dust identified in the right had side of the image. This error is compounded by tha fact the dust mask does not identify a dense region of dust, indicated by as strong dust by the BTD dust mask, and is so dense is visible in the false colour RGB in Figure 5.12.

When the background is present, FSFS outperforms SVM for all the feature sets excluding the 2D GFB (6,6,1) for the land test case compared to the BTD dust mask. When compared to the OMI dust flag, SVM produces lower GM scores than FSFS for all the feature sets. The drop in the GM scores for the land test case compared to the OMI dust flag averaged over all the feature sets is $> 20\%$, compared to an average reduction in the SVM GM score of $< 5\%$. This larger drop for the land test case supports the idea that SVM will perform poorly when the non-dust class is comprised of multiple texture which is more likely over land surfaces.

On average, better results are obtained for the land test case when the background is present when compared to the BTD dust mask. Individual cases exist when the GM scores are better when the background is removed, for a constant feature set and classifier, for example SVM on 2D GFB (4,6,0), (4,6,$\frac{2}{3}$), and (4,6,1). However, the best results for the land test case occur when the background is present, using FSFS on GLCF with $d = 1$.

Despite the lower GM scores, feature sets and classifiers that perform well compared to the BTD dusk mask are thought to indicate a more reliable metric for the land test case than a comparison with the OMI dust flag. Qualitative comparison of the BTD dusk mask with the the false colour RGB indicate that it is better at rejecting cloud regions, and whilst this does not mean that the BTD approach is better at detecting dust, it does imply that it has a lower false positive rate. On this basis the most reliable classification occur when the background in present in the input SEVIRI image.

This may seem counter-intuitive, considering how a majority of existing dust detection methods attempt to remove the spectral response of the Earth's surface to aid in the detection and retrieval. However, the supervised classification uses the textural response, and not the spectral information. Removing the background from the SEVIRI image significantly alters the texture in the image, in particular the area corresponding to the liner training set. This can be seen in Figure 5.4 which shows stark difference between the dense dust region in the top right of the image, and the area where the training set is taken. This textural mismatch between the dense and weak dust regions results in a dust mask that does perform well.

5.4.3 Conclusions

A new supervised texture classification method for identifying a specific type of particulate aerosol is described. This work built on the analysis of simulated data that indicated good classification results can be obtained despite small linear training sets. The new approach has the potential to provide 24 hour coverage over multiple surface types, which is not available from any of the existing techniques. The use of IR imagery allows the detection of aerosols in day and night time conditions, and the use of SEVIRI imagery on the MSG platform provides data with a 15 minute repeat cycle allowing near-continuous observation. The methods for extending the coverage to achieve greater coverage is the focus of the next chapter.

The supervised classification algorithm was applied to a two test cases for atmospheric dust distributions, one over ocean and one over land, with the output dust masks com-

pared to existing dust detection methods. The results showed a good agreement with the OMI dust flag, with a correspondence of over 90% over ocean, and 80% over land in the test cases. Comparison to the Zhang et al. BTD dust mask and qualitative comparison to false colour RGB images of the for the land test case shows show successful rejection of cloud without the use of a cloud mask, and the successful identification of the dense dust regions observed in the visible RGB image. This is despite the training data only corresponding to the weaker areas of the dust distribution. The supervised texture dust image successfully rejects regions of cloud despite having no prior knowledge of the cloud texture, which implies that the textural response of cloud and dust are sufficiently different to enable successful separation.

The fidelity of classifications at the cloud/dust border were degraded due to the fixed window size used to generate the texture feature, which allows a few outlier pixels to affect the features in a much larger area, leading to conservative estimates of the extent of the dust distribution at the edges. Adaptive window sizes using a similarity metric to prevent outlier pixels affecting the texture features should help to address the problem and reduce the conservative classification at the edges of the dust distribution.

The most reliable classifications, when compared to the existing dust detection methods and qualatative analysis of the the false colour RGB, occurs using FSFS with small feature sets drawn from GLCF with $d = 1$. This closely matches the observations from the simulations, although the best performance occurs with a single feature, as opposed to the 4-6 range indicated by the simulated data. The difference could be due to smaller number of training samples available in the real world examples. The classification problem has become significantly harder, with larger number of texture present in the input scene. It is likely then that a higher number of samples are needed to distinguish between the classes, but as the number training samples is smaller, the curse of dimensionality affects larger feature sets leading to the best classification occurring with fewer features.

Removing the surface response from the SEVIRI image prior to generating the texture features did not have the expected benefit and lead to a poor performance of the supervised texture output masks. The perceived benefit of removing the surface response is based on isolating the spectral response from atmospheric aerosols, which is frequently used in existing particulate aerosol detection approaches. Removing the surface response significantly alters the textural response, and so when texture is used as the discriminant adversely affects the performance.

The supervised texture dust flag has demonstrated an ability to identify dust distributions over ocean and terrestrial surfaces, at higher resolutions than existing methods. The resolution is limited by the input images, and thus can provide outputs that match

SEVIRI on MSG ($3km^2$ at nadir) and MODIS on AQUA ($1km^2$ at nadir) compared to the lower resolutions of established techniques such as OMI on AURA ($13 \times 24km^2$) and MODIS dust products ($10km^2$). BTD approaches, in particular the technique described by Zhang et al. [2006], have the potential to offer dust masks at the same resolution as source images. However this approach does not appear to be robust over ocean as it failed to identify any dust when applied to ocean test case.

The algorithm can be easily adapted and applied to identify other types of atmospheric aerosols such as smoke from wild fires, and volcanic ash. These categories exist within the CALIPSO VFM, and as such only minimal changes to the algorithm are required.

The supervised texture approach has demonstrated an ability to detect atmospheric aerosol distributions in a local area around the CALIPSO overpass that is used to provide the training data. This provides only a limited coverage that is linked to the polar orbit of the CALIPSO satellite. The next chapter considers how to extend the temporal and spatial coverage away from the initial classification, with the aim of being able to monitor and track aerosol distributions when it is not co-located with the CALIPSO overpass.

Chapter 6

Continuous Observation of Particulate Aerosols using Signature Extension

The previous chapter demonstrated that supervised texture classification can be used to identify particulate aerosol distributions in SEVIRI images, using the CALIPSO VFM to derive the training set. One of the advantages of using SEVIRI imagery is the 15 minute repeat cycle that allows near continuous monitoring. Also, as previously stated, SEVIRI on MSG is ideally located to observe Saharan dust events, and can provide imagery to potentially track the dust event over its lifecycle.

6.1 Applicability of Supervised Texture Classification for Detection of Particulate Aerosols

The supervised texture classification approach can only be applied when a CALIPSO overpass is co-located with the scene, which prevents continual monitoring of the same event. In effect this limits the supervised texture approach to the same 16 day repeat cycle observed by the A-Train Satellite Constellation, which contains the CALIPSO and MODIS on AQUA platforms.

An illustration of the time scales for the repeat cycle of the CALIPSO VFM is provided for the land dust test case from the previous chapter 5.1. Figure 6.1 shows the day- and night-time CALIPSO overpass locations on 21/02/2007. A $10^\circ \times 10^\circ$ region of

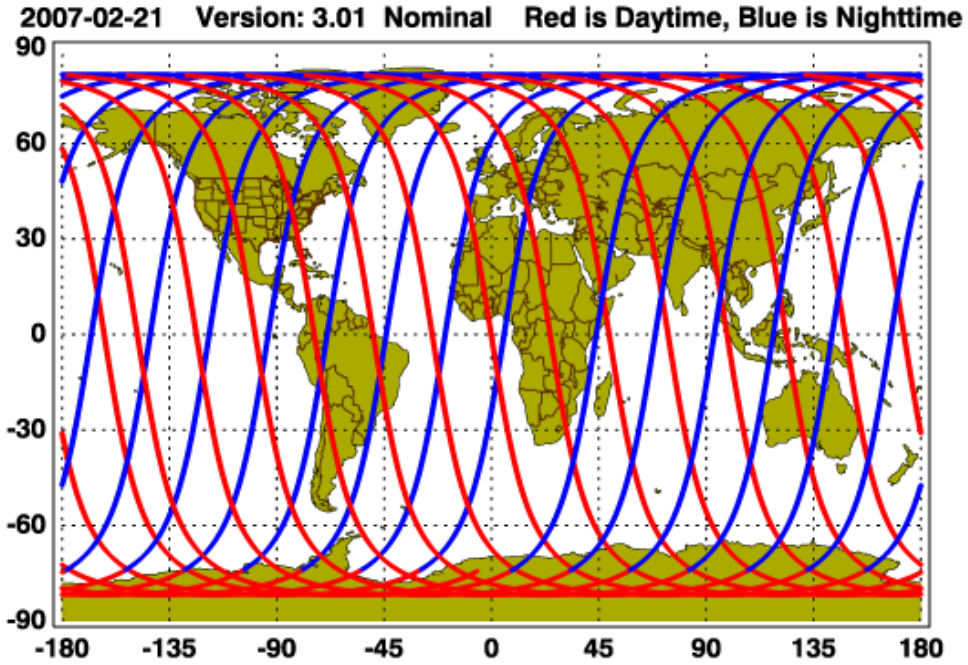


Figure 6.1: Showing the Daytime (red) and nighttime (blue) CALIPSO overpass locations for 21/02/2007

interest centred on $(-5°, 25°)$ is extracted from the SEVERI image at UT13:45 on 21/02/2007 which provides the nearest temporal match to the CALIPSO VFM. This daytime overpass can be seen in centre of Figure 6.1, crossing over North East African. The next daytime overpass of CALIPSO is approximately 25° west and 90 minutes after the original when compared at North African latitudes where the region of interest is located. In order for this to useful for supervised texture classification, the dust event needs to be moving west at a rate that allows a $10° \times 10°$ to contain a significant proportion of the dust event, which is unlikely to regularly occur.

Figure 6.1 shows the original overpass crosses a nighttime CALIPSO overpass just off the coast of Portugal. This overpass occurs approximately 11.5 hours before the original overpass. There are multiple crossing points of day- and nighttime overpasses which can be used to monitor particulate aerosol events over the regions where multiple overpasses intersect. The repeat cycle of 11.5 hours is significant when compared to the lifetime of particulate aerosol events which is of the order of days. Significant changes can happen within this time frame, so a shorter repeat cycle would be needed to accurately track the dust event.

6.2 Signature Extension and Chain Classification

The availability of training data limits when supervised texture classification can be used to monitor particular events. CALIPSO VFM are available over the same region at best every 11.5 hours, however the SEVIRI images that are used for the classification used are available every 15 minutes.

This problem is somewhat similar to the Land-Cover classification problem using satellite data. Land-cover classification using Landsat data is typically undertaken one scene at a time, which requires ground truth for each individual scene and limits rapid analysis. In order to address this several different approaches have been presented, although the most applicable to the classification of particulate aerosols are Signature Extension [Pax-Lenney et al., 2001] and chain classification [Knorn et al., 2009].

Signature extension refers to the process of using a classifier trained on one scene, and the resulting signatures are applied to other scenes at different locations and times. Pax-Lenney et al. [2001] successfully used signature extension to identify conifer forest type in Landsat Thematic Mapper (LTM). The results showed no statistically significant loss in accuracy when the initial training set was extended to other Landsat images from the same scene. This is akin to a temporal extension, looking at the same geographical region.

When the same methods were applied to subsequent images that differed over in location and time, the classification accuracy dropped by 8-13% depending on the atmospheric correction used to match the spectral signature between images. Additional work by [Olthof et al., 2005], showed that the ability to used signature extension is greatly affected by geographical distance from the original training scene, with the drop of approximately 50% for a 1500 km separation between the training data and the test image. This highlights the inherent problem with signature extension; the need to account for varying topography, illumination and atmospheric effects.

Chain classification, proposed by Knorn et al. [2009], is a technique to address the geographical dependence of the signature stability, and uses overlap region between subsequent satellite images to update the training set and produce a new classifier for the next image. The advantage of chain classification is that no atmospheric correction is needed to account for spectral differences between images, and the continual update of the training set should reduce the affect of geographical distance on the classification accuracy.

The results in [Knorn et al., 2009] show a loss of 5.1% in classification accuracy over

a chain of 6 LandSat scenes. Each scene is nominally 185 × 170 km, with the overlap between each scene varies depending on the latitude. Chain classification therefore shows a significantly smaller decrease in classification accuracy compared to signature extension over geographical ranges of 1000km.

Signature extension and chain classification can be used to extend the temporal and geographical coverage of the supervised texture classification method for identifying atmospheric dust. Signature extension is the simplest to implement, and is capable of providing monitoring over the same region in consecutive SEVIRI images. In order to track dust events as they move, chain classification would be more suitable as it continually updates the training set.

6.3 Detection of particulate Aerosols in Subsequent SEVIRI Images using Signature Extension

To demonstrate the continual monitoring of particulate aerosol events, a simple scenario was constructed to monitor the same geographical region over time for the land dust case. This allows SEVIRI imagery to be used, and provides imagery at 15 minute intervals. As the scenario requires only temporal extension, signature extension was selected due as it provides good results for temporal extension, and is simpler to implement than chain classification.

The statistics derived from the CALIPSO VFM are used to generate a dust mask on the original image, and additionally same scene extracted from the SEVIRI images from 07:45 to 19:45. This represents a ±6 hour temporal window centred on the original SEVIRI image at 13:45. This temporal range is selected as it represents the maximum temporal extension required when monitoring a region at or near the intersections of a day and nighttime CALIPSO overpass. Temporal extension can be applied forward from one overpass and backward from the other, thus an extension of 6 hours would be required for continuous monitoring. In this instance, the land dust case does not correspond to a region where day and nighttime overpasses intersect and is used to show the feasibility of the approach.

No attempt is made to correct any differences in the spectral differences between the subsequent images. The best performing classifier from the previous chapter was used to generate the images for the signature extension scenario, namely FSFS with 1 feature drawn from GCLF with $d = 1$.

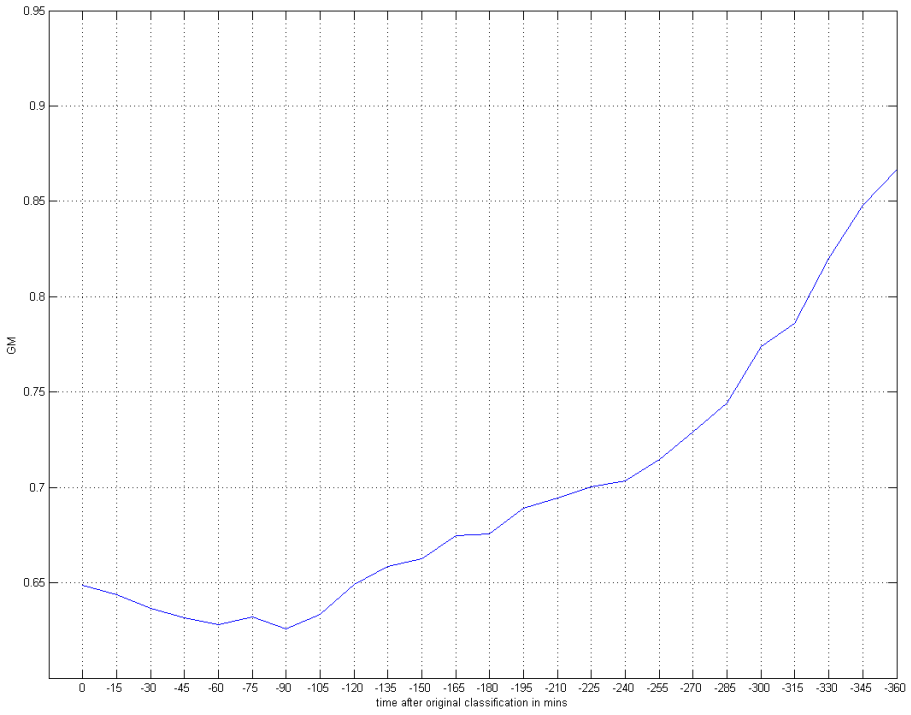


Figure 6.2: GM scores for backward temporal signature extension using FSFS and 1 feature for the land test case

The outputs from the signature extension texture classification are compared to the BTD dust mask [Zhang et al., 2006] generated using SEVIRI imagery, as all the other products previously used are on satellite platform in polar orbits and so can not provide continuous monitoring of the same scene (See tables 2.3and 2.4). Again the geometric mean (4.6) was used to rate the accuracy of the dust mask of the supervised texture algorithm using signature extension to the BTD dust mask.

Figure 6.3 shows the GM of the dust mask compared to the BTD dust mask using signature extension FSFS with 1 feature drawn from GLCF with $d = 1$. The result at $t = 0$ relates to the classification which corresponds to the registered CALIPSO VFM for the test case. Results at subsequent times relate to output masks generated using signature extension on subsequent SEVIRI images.

Figures 6.2 and 6.3 shows the GM score for backward and forward temporal extension for the land test case using FSFS with 1 feature. The key metric for measuring the applicability of signature extension is the decrease in classification performance as time increases. The BTD dust mask used for comparison is itself susceptible to errors, and does not represent a ground truth. Furthermore, the temporal range covered in

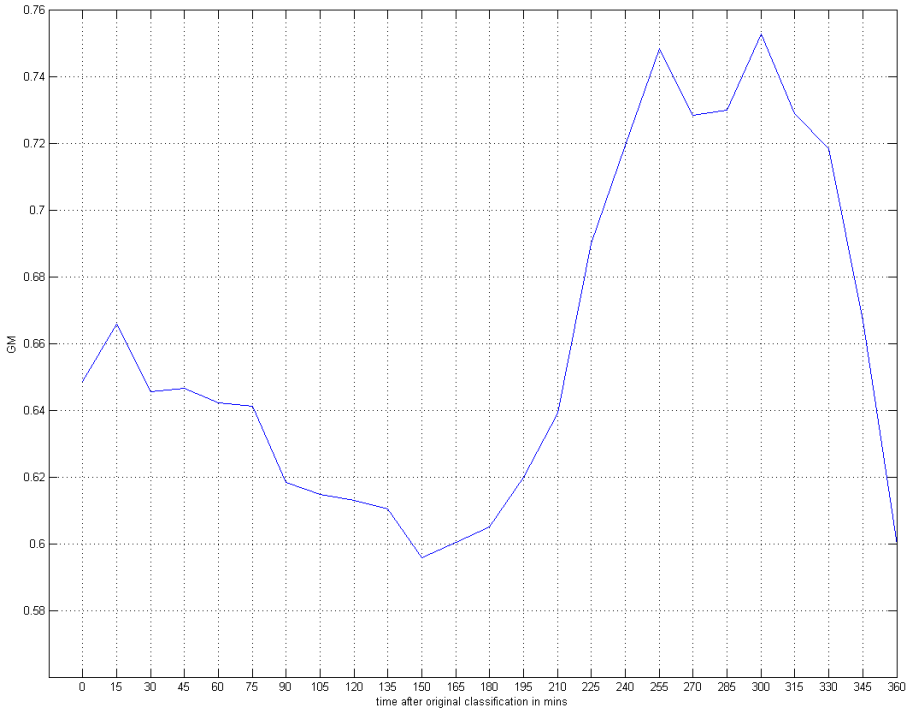


Figure 6.3: GM scores for forward temporal signature extension using FSFS and 1 feature for the land test case

this analysis covers day and night time images which can impact on the brightness temperature of dust, and hence the errors in the BTD may fluctuate with time. The GM scores only provide an indication of performance relative to the BTD dust mask, whose performance is unknown due to the lack of coverage from other techniques.

The initial GM is not particularly high and is a result of the supervised classification approach and data available to evaluate the results against and not signature extension as discussed in section 5.4.2. If the performance of the initial classification can be improved, then the classification using signature extension will also be higher.

The results for the backwards temporal extension shows a small general decrease in the GM score with subsequent images over 90 minutes, reducing by 2.3%. Intuitively, this is expected with the decrease in performance increases over longer temporal extension. Despite little change in the correspondance between the BTD mask and the supervised texture masks using signature extension, the true dust distribution could vary more significantly. However, the results correspond well with the work by Olthof et al. [2005] which showed that temporal signature extension causes little loss in classification accuracy.

Over the subsequent 4.5 hours, the GM score increases to a maximum of 86.7%. Should the BTD mask be a true representation of the dust present, the GM score would be be expected show a continual degradation as observed over the first 90 minutes. Instead, the increase implies that error in the BTD approach is changing. This is most likely due to the differences between BTD responses between day and night time images, that are covered in the time span examined.

Examination of the false colour RGB images, along with the BTD mask and the outputs from the supervised texture classification (see D), shows that over the period of study, there is overall movement in the of the visible part of the dust cloud in the RGB, which is also reflected in little bulk change in the BTD mask. Comparison of the supervised texture dust mask to the BTD mask shows that cloud continues to be successfully rejected as it moves within the scene. As with the supervised classification in the previous chapter, the output mask from signature extension are overly pessimistic at the cloud/dust boundary, again due to the fixed 9×9 window used to generate the GLCF.

Similar results are observed for the forward temporal extension. Over the first 10 images, representing 150 minutes, the GM scores shows a reduction of 4% from 65% to 61%. After this, the GM score increases upto a peak of 75% after 255 and 300 minutes. The start of the increase of the GM score corresponds to the transistion from day to night time images, reflected in the false colour RGB images.

The GM scores for the backward and forward temporal extension over the ±6 hour range indicate that BTD is not the best data for comparison. However, there is no other basis of comparison due to lack of coverage of other satellites and methods. Over shorter time spans ($-90, +150$ minutes) the comparison to the BTD shows the expected behaviour with the GM decreasing slighter as the period for the temporal extension increases. This correspond to a 4 hour window from 1200 to 1600 UT, over which time error in BTD would be expected to be reasonably constant due to similar factors such as temperature and the intensity of the incident light.

Despite the naive assumptions of this simplified signature extension example, the results indicate that it is a valid technique to monitor atmospheric particulate aerosols away from a CALIPSO overpass, thus extending the coverage of the supervised texture approach in the temporal domain by approximately ±2 hours. Due to the variable error on the only reference data from the BTD approach, no conclusions can be drawn about the validity of signature extension over the full ±6 hour range, but the initial work has shown that this approach is plausible.

6.4 Future work for the Near Continuous Monitoring of Particulate Aerosols

The results above indicate that is is possible to monitor particulate aerosols over the same location using temporal signature extension. This allows particulate aerosols to be monitored over a specific region, but does not allow the continuous monitoring of a single event as it position varies over time.

As previously mentioned, it is unlikely that the an event of interest will correspond with the orbital path the the A-Train, which means continuous monitoring using CALIPSO, OMI on AURA and MODIS on AQUA is not possible. SEVIRI provides a full disk image, and so can be used to continuously track an event providing it occurs in its field of view. Assuming a CALIPSO VFM can be found that intercepts the event, then supervised classification can be used to generate an output mask as demonstrated in the previous chapter.

The event will be moving in space and time relative to the initial supervised classification output mask, and so signature extension may not provide a good method to extend the coverage, although it may be possible to utilise chain classification. Due to the short repeat cycle of 15 minutes, there is likely to be significant overlap between regions of dust in subsequent images. Figure 6.4 shows the dust mask from the su-

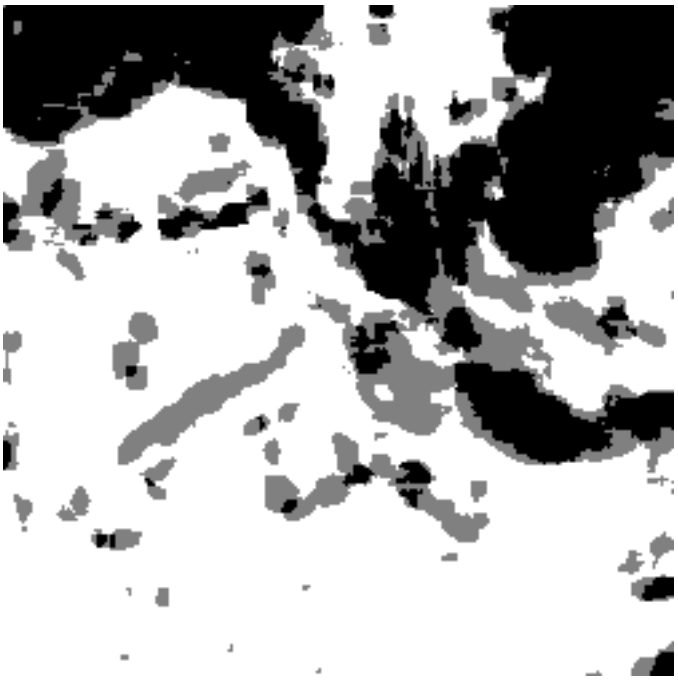


Figure 6.4: Combined dust mask for supervised texture classification of the Land test case at UT 1345 and the dust mask using signature extension at UT 1400. The white regions show where both dust masks overlap,grey showing where only one mask identifies dust, and black showing where no dust is identified in either mask.

pervised classification of the land test case combined with the dust mask derive from signature extension. The white regions show where both dust masks overlap, the grey showing where only one mask identifies dust, and black showing where no dust is identified in either mask. These images are taken when the results from signature extension match the expected performance, and when the error in BTD approach is expected to be stable.

Figure 6.4 indicates there is significant overlap between dust regions in subsequent SEVIRI images. This overlap can be used to derive a new set of class statistics for dust as described in chain classification, and be applied to a moving region of interest that tracks the event. As new classifiers are derived for each image, there is no need to account for variation in topography, illumination or atmospheric effects. The overlapping regions can be derived empirically, using a similar process to that used in Figure 6.4. Multiple images can be examined to derive a distance from edges that is likely to be dust in the next SEVIRI image, which can be used for chain classification. Alternatively, the dust regions can be derived using atmospheric models that take into account the local conditions including the wind vector that drives the movement of the dust.

A larger number of samples will be available as training data during chain classifica-

tion, compared to the limited number of samples provided by the CALIPSO VFM in the original supervised texture classification. Larger training sets may alter the most suitable texture features and classifiers for identifying particulate aerosols. In particular local binary patterns (LBP) would be of interest due to the rotational invariance and robustness gray-scale variation [Ojala et al., 2002b]. These properties are not necessarily required in the original classification as the textures in the classification and training sets are predominately at the same orientation desirable, but may be more relevant during chain classification as the texture orientation and illumination will change with time. Classification with LBP is normally achieved by comparing histograms of sample and known textures, with classes assigned using a nonparametric dissimilarity measure. With the increased number of samples available when chain classification is applied, LBP could be a feasible texture measure to use.

This potentially allows events to be tracked from a CALIPSO overpass either forwards or backwards in time. Therefore an output mask can be generated until the event corresponds with another CALIPSO overpass, and the process can be restarted. Theoretically this approach can provide near continuous monitoring of events in the SEVIRI on MSG field of view.

6.5 Conclusions

The work in the chapter 5 showed the benefits of using supervised texture classification to produce a 2D distribution of particulate aerosols in the atmosphere. The algorithm was applied to two test cases involving atmospheric dust distributions over land and ocean surfaces, showing its ability identify distributions over different terrestrial surfaces.

Despite the advantages of this approach, the applicability is still limited by the availability of ground truth data, which is provided by the CALIPSO VFM data. The polar orbit of CALIPSO platform meant that an overpass would not always intercept with a particular event that required monitoring.

Signature extension been used to extend the temporal coverage of the supervised classification approach over a fixed geographical location. The approach was applied to the land test case, which showed a 3% drop in classification ability over a 90 minute window, which matches published work using signature extension for first order temporal classification.

Further work can look at using chain classification to further extend the coverage over

time and space, which would allow specific events to be tracked. This would allow events to be tracked until the next ground truth is available, for example another CALIPSO overpass, which can be used to generate a new supervised classification mask to which chain classification can again be applied.

Chapter 7

Conclusions and Further Work

This chapter provides an overview key conclusions of this thesis and a brief discussion of where the technique can be improved, as well as other areas of remote sensing where this technique can be exploited.

7.1 Conclusions

In the preceding chapters a novel approach to detecting atmospheric particulate aerosols has been demonstrated. This approach differs from existing techniques by using supervised texture classification, which has not previously been used to address this problem.

The current techniques for remote sensing of particulate aerosols can be split into three categories; radiative transfer models, brightness temperature difference approaches and Space-borne LIDAR, radiative transfer models being the most prevalent and established approach. All of these approaches have limitations, ranging from inability to work over various terrestrial surfaces, when particulate aerosols are interspersed with cloud or water vapour, and limitations on coverage relating to repeat cycle and resolution. All of these contribute to preventing continual and global monitoring of atmospheric particulate aerosols.

Some previous work exists in the literature that used unsupervised texture classification to identify smoke and dust particulates in satellite imagery. Texture approaches present an advantage over a majority of the existing techniques as they are capable of separating classes that cannot be resolved in the spectral domain, which is the case with particulate aerosols.

In this thesis a new approach to identifying particulate aerosol using satellite imagery has been proposed. This approach differs from other techniques as it uses supervised texture classification approach, which aimed to improve on the existing unsupervised methods with the addition of training data. The proposed approach fuses data from two distinct satellite sources which is used to provide the training set to form the classifier, and the imagery to which the classifier is applied. The CALIPSO VFM was selected to provide a linear tract through imagery from other satellites to provide a training set.

A linear training set provides a non-ideal supervised classification task, with a small training set of potentially unrepresentative samples. This scenario was investigated on a set of simulated texture images using the Brodatz and Outex databases, due to the lack of ground truth data for satellite remote sensing scenarios. These simulations compared GLCF and 2D GFB texture feature and 6 supervised classification techniques using random and linear sampled training sets. The best result for linear sampling was achieved using FSFS and SVM, using GLCF with d = 1. The performance of the linear training sets was compared to the random sampling, with random sampling provided the best classification rates as expected. Linear sampling demonstrated that a high classification rate in excess of 80% can be achieved (see Table 4.8) whereas an unsupervised k-means classifier achieved a GM of just over 60%. This demonstrates the advantages of using a linear training set over unsupervised approaches, and provided an experimental basis for the supervised texture approach to be applied to real world examples for the detection of particulate aerosols.

The supervised texture classification approach was applied to two test cases over land and ocean, aiming to identify a 2D mask showing the extent of particulate dust. The best performing classifiers and feature sets identified on the simulation were applied to both test cases. The CALIPSO VFM was used to identify a training set withing IR SEVIRI imagery from the MSG platform. The results were compared with existing techniques and showed good correspondence with the OMI dust flag with rates of greater than 90% for the ocean test case and over 80% for the land test case, see tables 5.1 and 5.2. The resulting mask is also at a higher resolution than the existing approaches, producing an output of $3km^2$ compared to $13km \times 24km$ for MODIS, and $10km^2$ for OMI and demonstrated that no alogrithm changes were needed to produce dust masks for different terrestrial surfaces.

Signature extension was then applied to the supervised texture approach to provide temporal extension to the coverage over the same geographical region. Due to the limited satellite coverage, only the BTD approach was available to evaluate the results of signature extension. These results show a decrease of 3% in the classification rate for temporal extensions of ± 90 minutes from the original classification (see Figures 6.2 and

6.3). Results for a longer extension of ±6 hours were generated, but the results for the BTD vary over these timescales which prevents meaningful quantative comparison. The results of the signature extension indicate that it is possible to extend the coverage away from the initial CALIPSO overpass and provide a monitoring over a fixed geogrpahical region, but the length of the temporal extension could not be determined.

7.2 Improvements to the Supervised Texture Classification Algorithm

The main thrust of the work presented in this thesis is use of supervised texture classification with a linear training set to identify a texture of interest. The different elements of this technique such as the texture measures and the classification techniques used demonstrated a good level of performance for the test imagery, although only a limited set were considered in this initial investigation. As such, a number of the elements within the method could be improved and further study would be needed to investigate any performance enhancements.

As discussed in chapter 3, two different texture schemes were investigated as they have previously been applied in similar remote sensing tasks. However other texture schemes exist and may provide better performance with linear training sets wither on their own or in conjunction with other texture measures. In particular, LBP may be of interest for use with signature extension and chain classification approaches to extend the coverage of results.

Similarly, a representative set of popular classifiers were examined, but other approaches or variants may yield better results. One area of further investigation could be the performance metric used to evaluate the feature selection approaches such as FSFS and BSFS. The geometric mean (4.6) was selected due to the inability of accuracy to quantify classification creates in imbalanced data sets. Other metrics exist which have this capability too, such as balanced accuracy, the F measure and combinations of precision and recall which could lead to improved performance.

The performance of linear training sets was investigated on simulated imagery in chapter 4 and demonstrated it was a valid approach that can yield good classifications rates. Further analysis could be undertaken to investigate the affect of the number of samples in the training set on the classification rate. This could provide the limiting behaviour of linear training sets and indicate when linear training sets are a valid approach.

7.3 Improvements to Continuous Observations

The continuous observations using signature extension demonstrated the ability for temporal extension away the initial supervised classification but this approach can not track events as their location changes. Temporal extension is useful for applications such as monitoring volcanic ash plumes though fixed regions of airspace. As discussed in chapter 6, chain classification offers a more robust method to extend the coverage away from the original classification and would be better suited for the continuous monitoring of specific event in both the temporal and spatial domains.

7.4 Other applications of Supervised Texture Classification Approach

The supervised texture classification approach was demonstrated on examples of mineral dust deflated from the Saharan desert, but the algorithm is more widely applicable to other areas of particulate aerosols such as smoke from forest fires, and volcanic ash from eruptions. All of these aerosol distributions have a discernible affect on the Earth's atmosphere and ecology which is difficult to assess due to the problems in reliably detecting and identifying particulate aerosol distributions. These subclasses of aerosol are included in the CALIPSO VFM, so the algorithm can be adapted to generate 2D masks for other aerosol types.

The technique could be extended further by deriving training data from other sources. The most readily available example for remote sensing of atmospheric aerosols would be FAAM data. As discussed in chapter 3, FAAM flight paths often include significant linear portions, and can be used to provide linear training sets to generate classifiers. As with the CALIPSO VFM, FAAM has the ability to detect and identify multiple aerosol types, which can be used to further increase the coverage and scope of this approach.

7.5 Aerosol Strength and Texture Response

Currently, the training data for the class of interest is added to the training set providing there is a unoccluded view of the texture of interest in the test imagery. For the real world examples of particulate aerosol detection, the strength or concentration of the dust is likely to affect the texture response. Dense regions of particulate aerosols are

easily observed in spectral imagery, whereas weaker concentrations are optically thin. Therefore it is possible for the concentration to affect the texture response.

Further investigation on the concentration of the aerosol of interest compared to the texture recorded may show that it is possible to identify strong and weak distributions within the 2D mask produced and so textural response could be used to inform about the aerosol concentrations for specific events.

7.6 Further Verification

Underpinning all the suggested areas of further study is the need for independent data for comparison and evaluation. The supervised texture approached was proposed to address the deficiencies in the existing methods for the remote sensing of particulate aerosols, and as such often lacks data for verification. The most reliable and robust of the existing methods and platforms are MODIS, OMI and CALIPSO products, which are all within the A-Train satellite constellation. Whilst these can be used to verify the initial classifications using the CALIPSO VFM as the trainings set (see section 5), there are differences in output resolutions which prevent direct comparison.

The A-Train follows a 16 day polar orbit, so it can not provide data to evaluate signature extension or chain classification as there is insufficient coverage. Similarly, if other sources were used to derive the training data such as FAAM aircraft, there would need to be careful consideration of when and where to study so that suitable data was available for verification.

It may prove necessary to combine data sources in order to provide coverage for verification, especially when attempting to extend the coverage away from the initial classification. In this instance satellite data can be extended using wind vectors to model where known particulate aerosols move to [Govindan, 2007], which can be used to verify the results of signature extension and chain classification.

There are multiple applications and extensions to the concepts described in this thesis, and providing data for verification is likely to be the most difficult challenge.

Appendix A

MODIS channels and resolutions

Table A.1: Spectral Channels and Resolutions for the MODIS sensor.

Channel	**Bandwidth**	**Resolution**
1	620 - 670	$250m^2$
2	841 - 876	
3	459 - 479	$500m^2$
4	545 - 565	
5	1230 - 1250	
6	1628 - 1652	
7	2105 - 2155	
8	405 - 420	$1000m^2$
9	438 - 448	
10	483 - 493	
11	526 - 536	
12	546 - 556	
13	662 - 672	
14	673 - 683	
15	743 - 753	
16	862 - 877	
17	890 - 920	
18	931 - 941	
19	3.660 - 3.840	
20	3.929 - 3.989	
21	3.929 - 3.989	
22	3.929 - 3.989	
23	4.020 - 4.080	

Table A.1: (continued)

Channel	**Bandwidth**	**Resolution**
24	4.433 - 4.498	
25	4.482 - 4.549	
26	1.360 - 1.390	
27	6.535 - 6.895	
28	7.175 - 7.475	
29	8.400 - 8.700	
30	9.580 - 9.880	
31	10.780 - 11.280	
32	11.770 - 12.270	
33	13.185 - 13.485	
34	13.485 - 13.785	
35	13.785 - 14.085	
36	14.085 - 14.385	

Appendix B

Converting Counts to Radiance for SEVIRI Imagery

Table B.1: The scaling and offset applied to the counts to convert to radiance for each SEVIRI channel, reproduced from [Müller, 2007]

MSG channel	Scale Factor	Offset
1	0.020135499537	1.026910476387
2	0.025922000408	-0.322022020817
3	0.022258499637	-1.135183481500
4	0.003658666667	-0.186592000000
5	0.008318107869	-0.424223501303
6	0.038621983914	-1.969721179625
7	0.126743582444	-6.463922704667
8	0.103961229947	-5.302022727273
9	0.205034453663	-10.456757136830
10	0.222311421249	-11.337882483713
11	0.157606930632	-8.037953462244
12	0.029498599470	-1.504428572953

Appendix C

CALIPSO VFM Data Format

Figure C.1 shows how the format of the CALIPSO VFM data extracted from HDF file representing the vertical cross section of the atmosphere along the CALIPSO overpass. Additionally, the vertical and horizontal resolution of the data is shown for three different altitude bands. During the reconstruction of the data, the relevant bands are upscaled using nearest neighbour interpolation, so that the data matches the horizontal and vertical resolution of the -0.5 to 8.2km band.

Table C.1: Description of the feature classification flags for interpretation of the VFM, reproduced from ASDC [2008]

Feature Classification Flag Description		
Bits	Field Description	Bit interpretation
1-3	Feature Type	0 = invalid (bad or missing data) 1 = clear air 2 = cloud 3 = aerosol 4 = stratospheric feature 5 = surface 6 = subsurface 7 = no signal (totally attenuated)
4-5	Feature type QA	0 = none 1 = low 2 = medium 3 = high
6-7	Ice/Water Phase	0 = unknown / not determined 1 = ice 2 = water

Table C.1: (continued)

		3 = mixed phase
8-9	Ice/Water Phase QA	0 = none 1 = low 2 = medium 3 = high
10-12	Feature Subtype	
	If feature type = aerosol, bits 10-12 will specify the aerosol type	0 = not determined 1 = clean marine 2 = dust 3 = polluted continental 4 = clean continental 5 = polluted dust 6 = smoke 7 = other
	If feature type = cloud, bits 10-12 will specify the cloud type.	0 = low overcast, transparent 1 = low overcast, opaque 2 = transition stratocumulus 3 = low, broken cumulus 4 = altocumulus (transparent) 5 = altostratus (opaque) 6 = cirrus (transparent) 7 = deep convective (opaque)
	If feature type = Polar Stratospheric Cloud, bits 10-12 will specify PSC classification.	0 = not determined 1 = non-depolarizing PSC 2 = depolarizing PSC 3 = non-depolarizing aerosol 4 = depolarizing aerosol 5 = spare 6 = spare 7 = other
13	Cloud / Aerosol /PSC Type QA	0 = not confident 1 = confident
14-16	Horizontal averaging required for detection (provides a course measure of feature backscatter intensity)	0 = not applicable 1 = 1/3 km 2 = 1 km 3 = 5 km 4 = 20 km 5 = 80 km

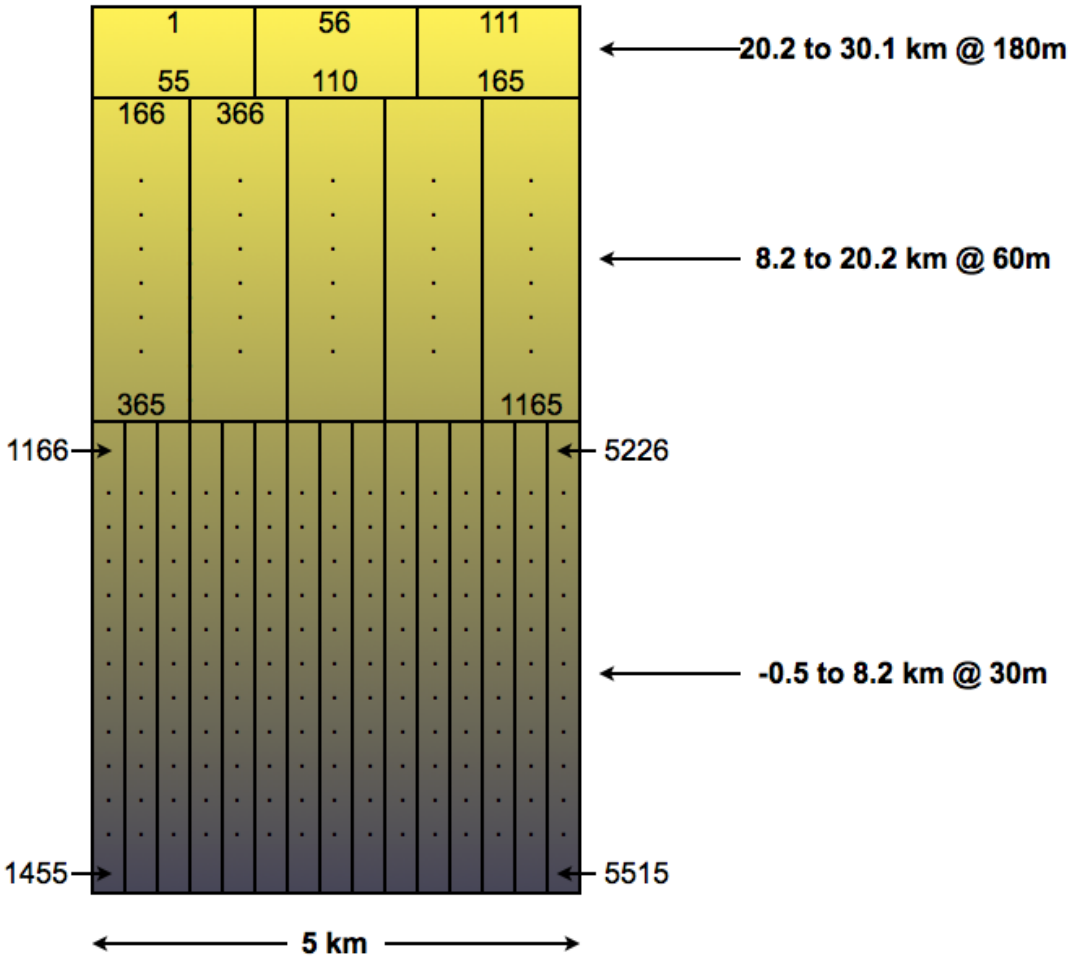
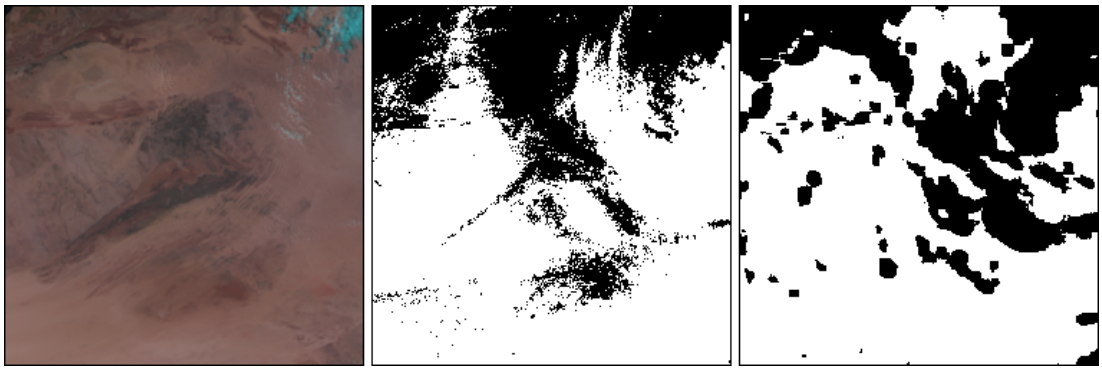


Figure C.1: The layout of the VFM values are stored as an 5515 element array (as rows in the HDF file) for a 5 km section of the overass

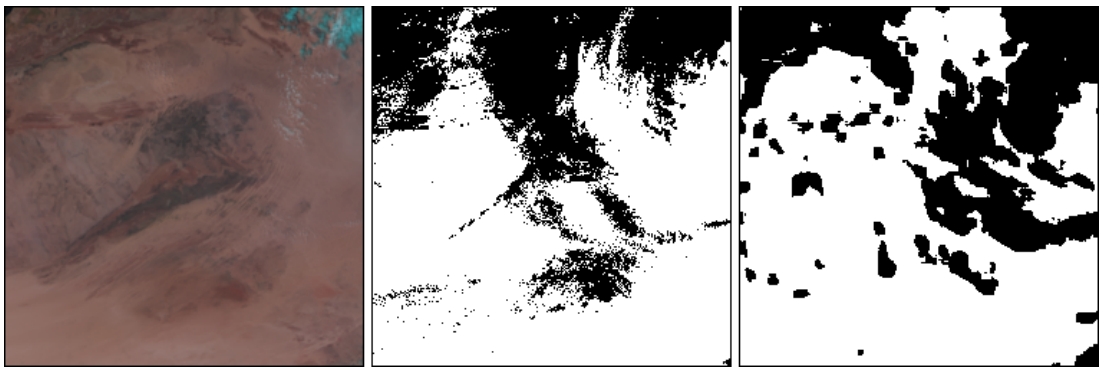
Appendix D

Images for Supervised Texture Classification Using Signature Extension

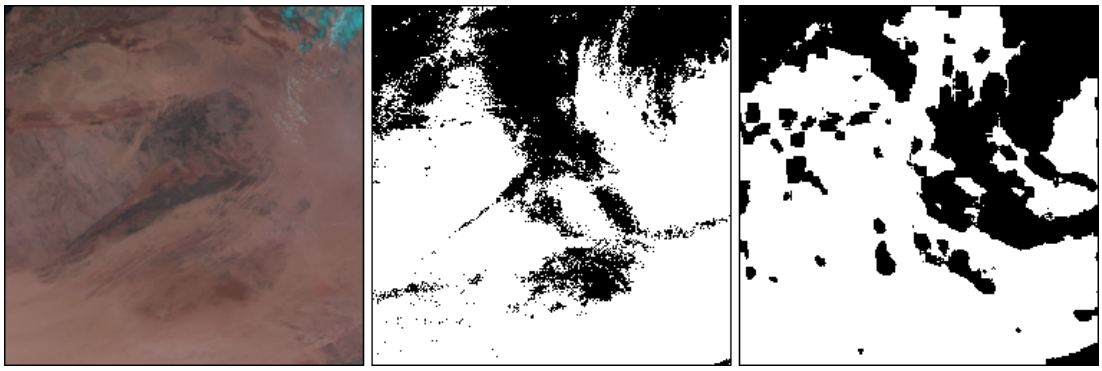
D.1 Backward Signature Extension



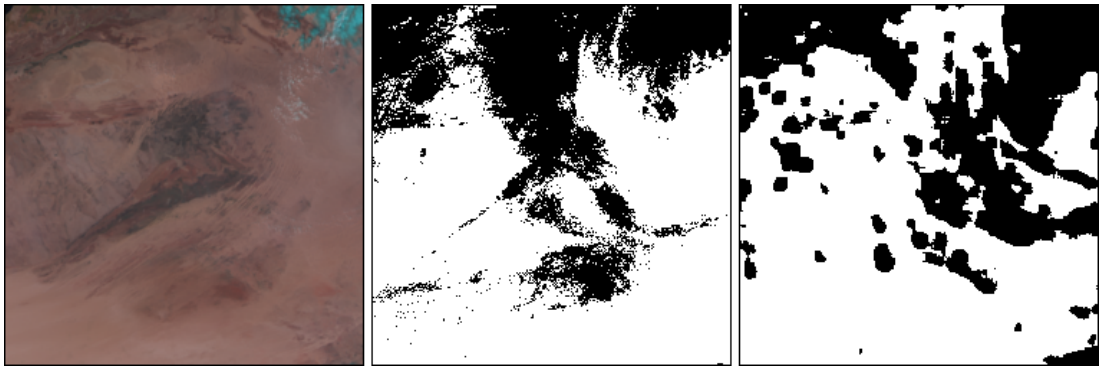
(a) Initial Classification at UT13:45



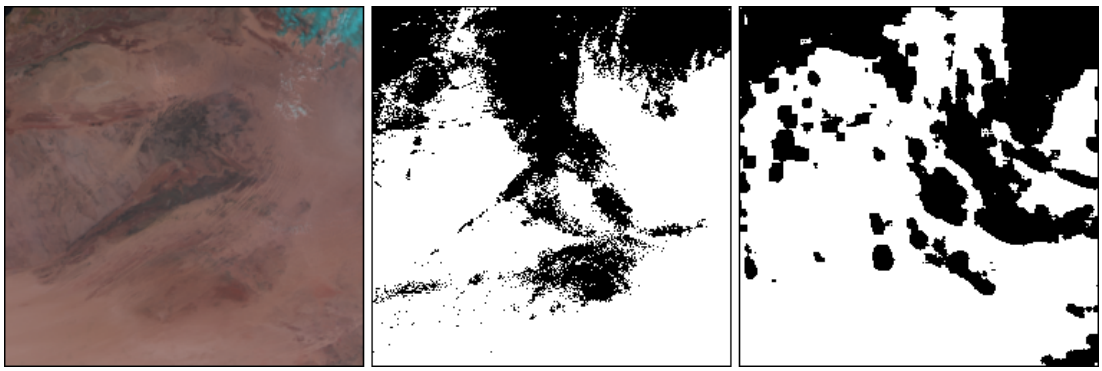
(b) Backward Signature Extension at UT13:30



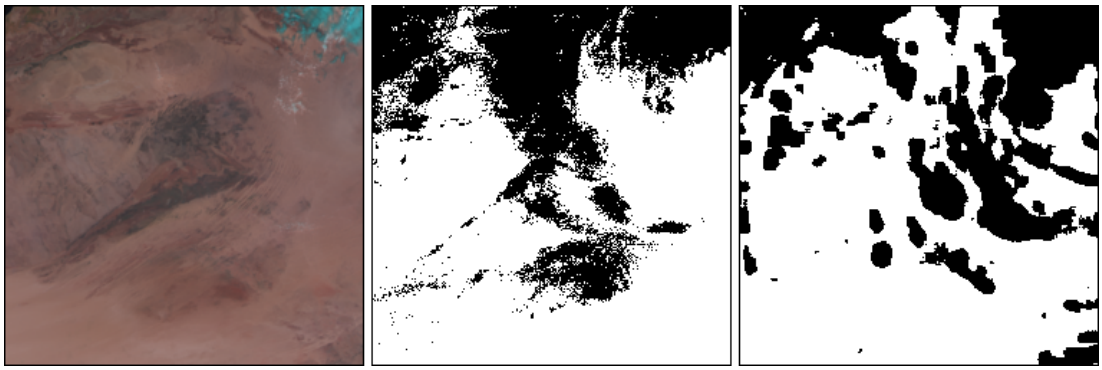
(c) Backward Signature Extension at UT13:15



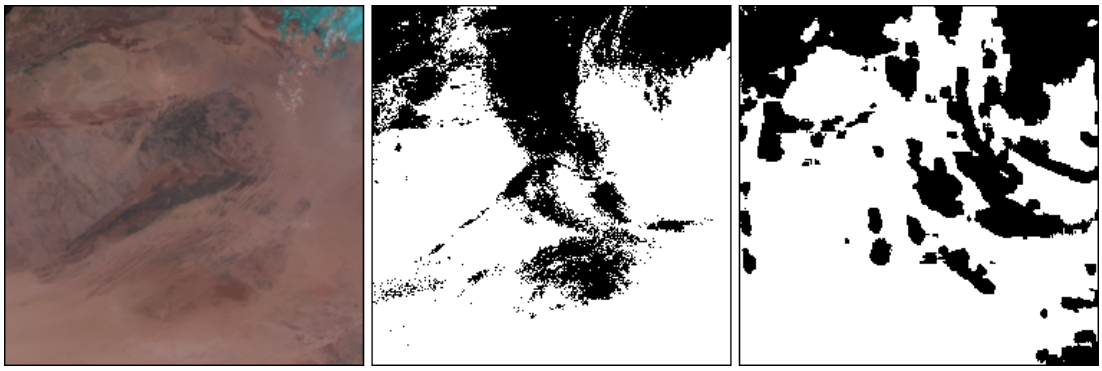
(d) Backward Signature Extension at UT13:00



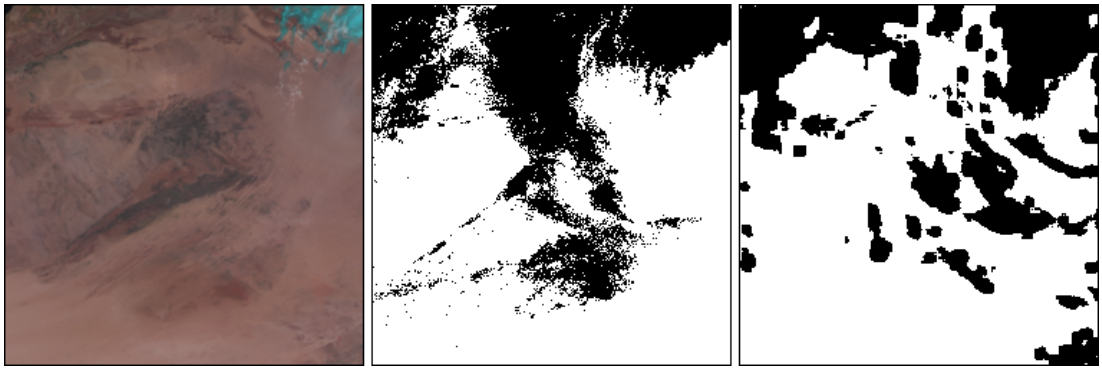
(e) Backward Signature Extension at UT12:45



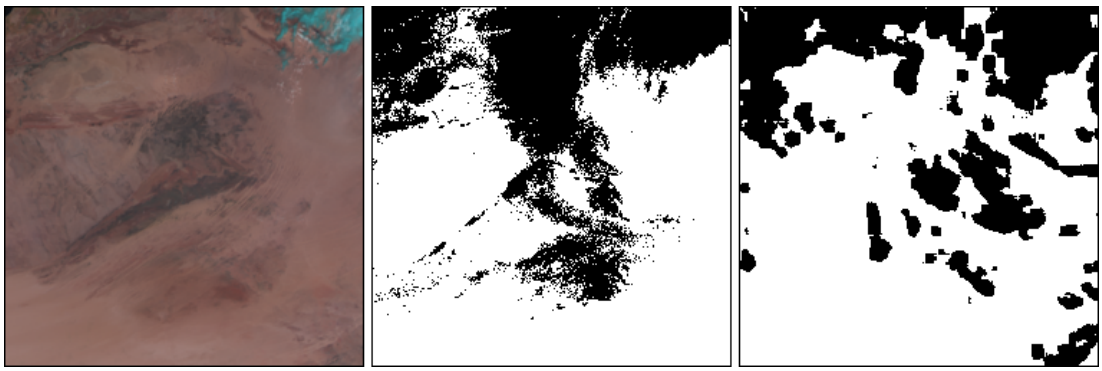
(f) Backward Signature Extension at UT12:30



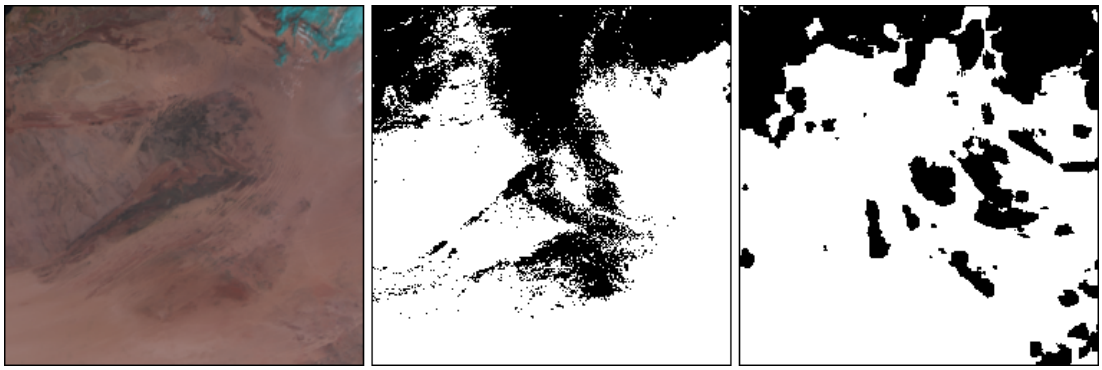
(g) Backward Signature Extension at UT12:15



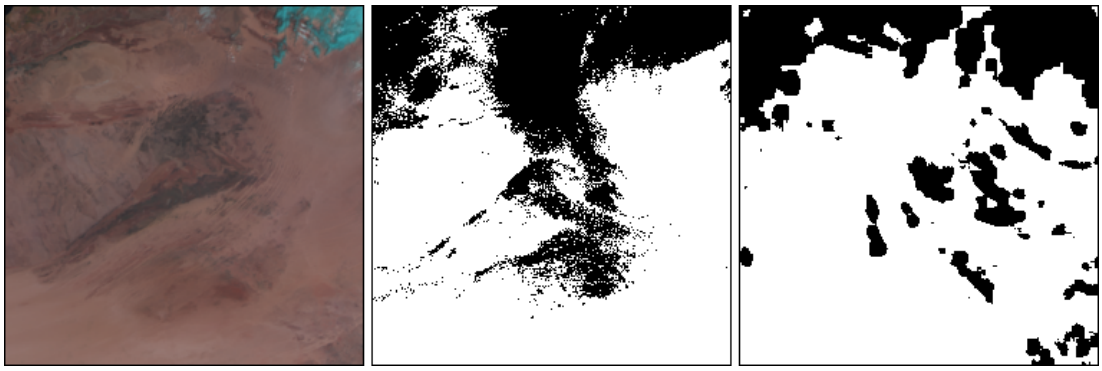
(h) Backward Signature Extension at UT12:00



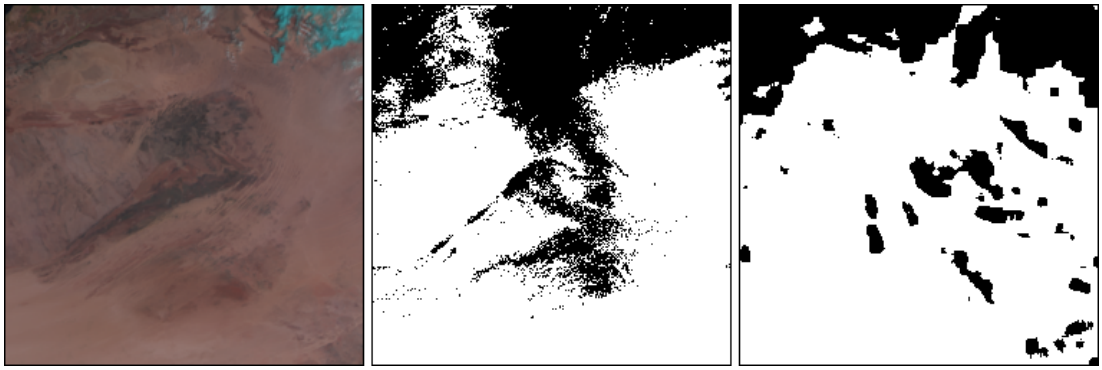
(i) Backward Signature Extension at UT11:45



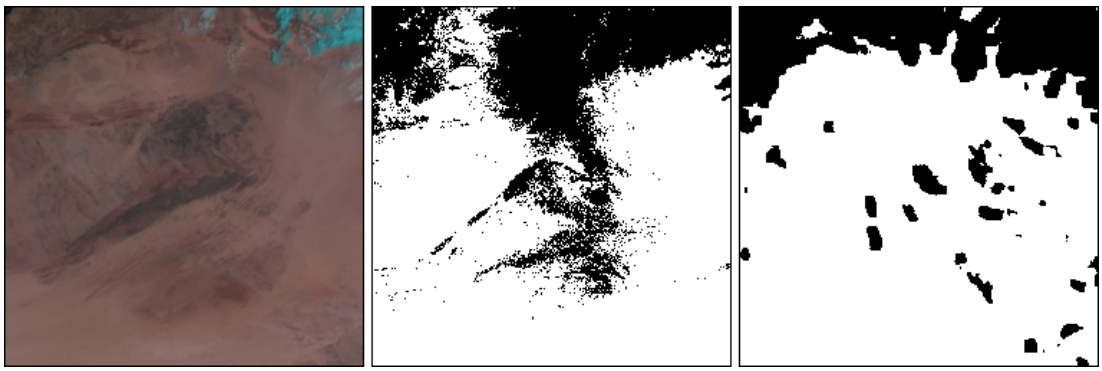
(j) Backward Signature Extension at UT11:30



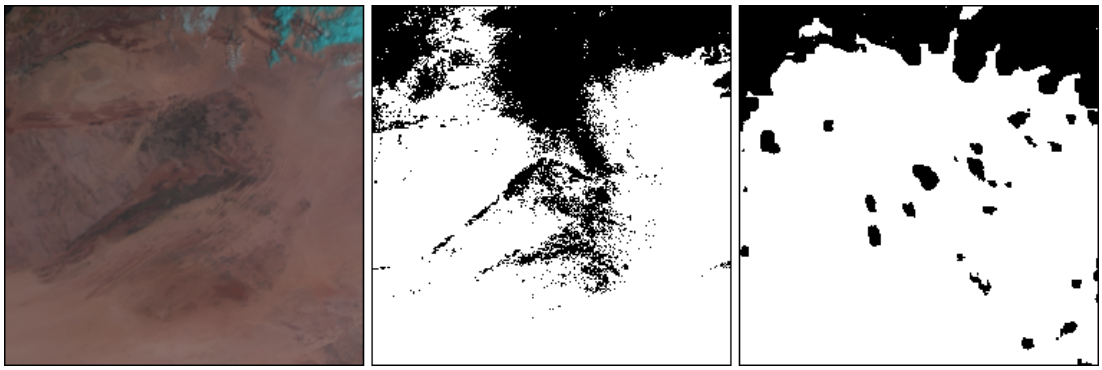
(k) Backward Signature Extension at UT11:15



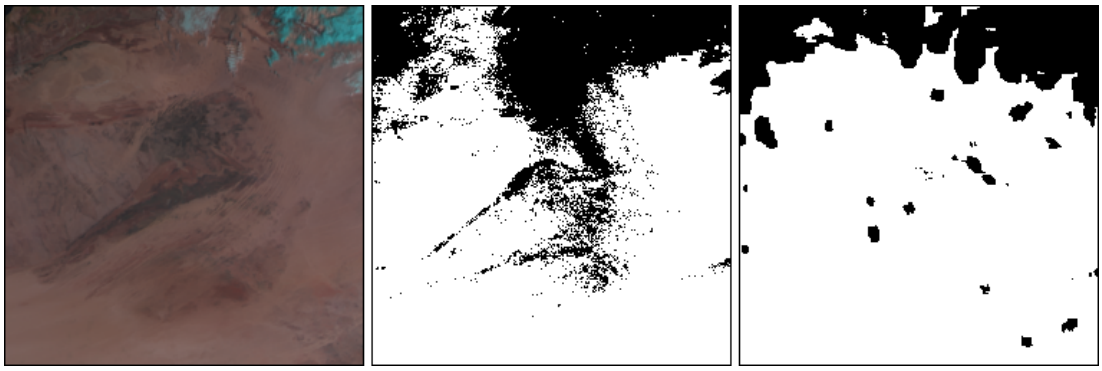
(l) Backward Signature Extension at UT11:00



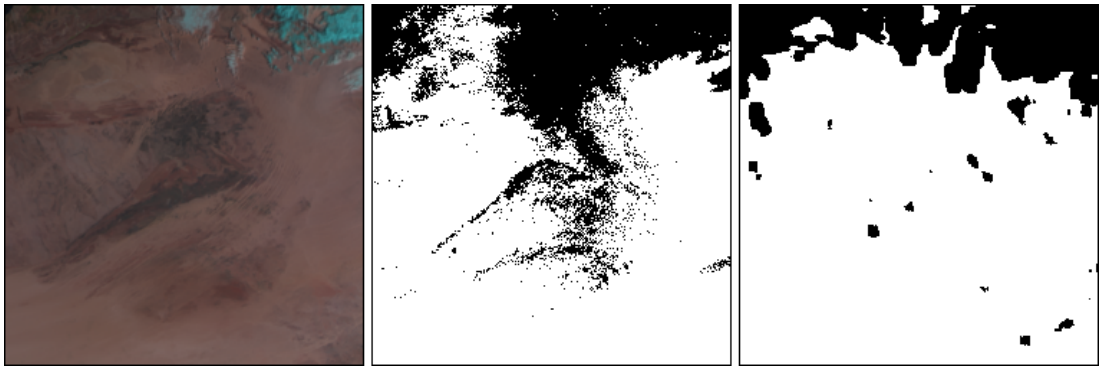
(m) Backward Signature Extension at UT10:45



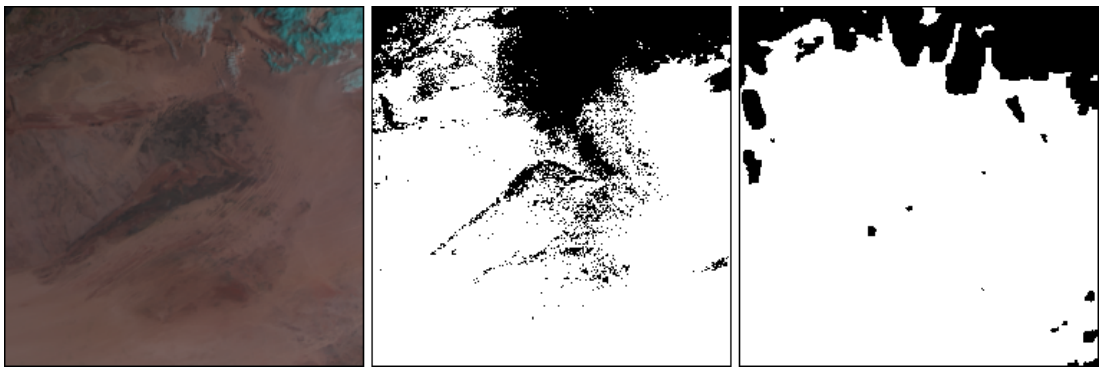
(n) Backward Signature Extension at UT10:30



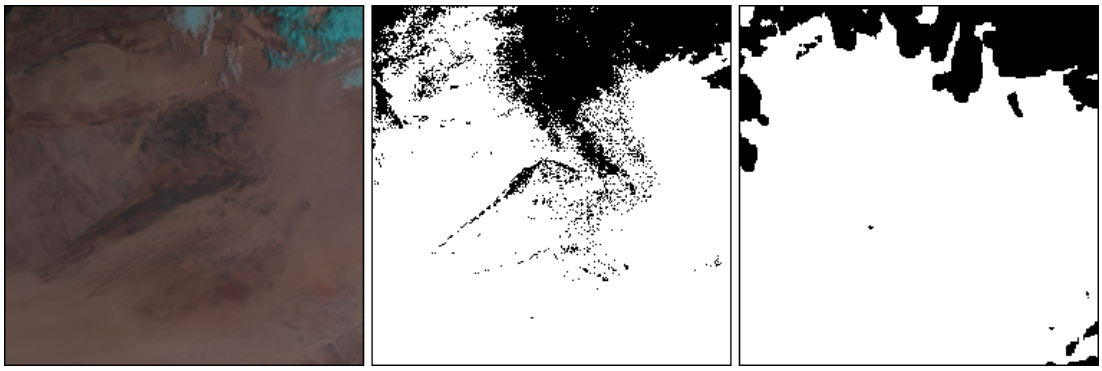
(o) Backward Signature Extension at UT10:15



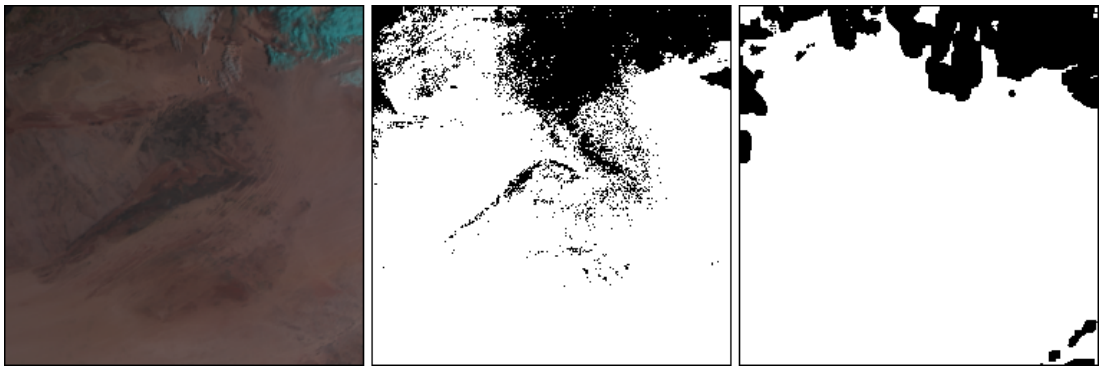
(p) Backward Signature Extension at UT10:00



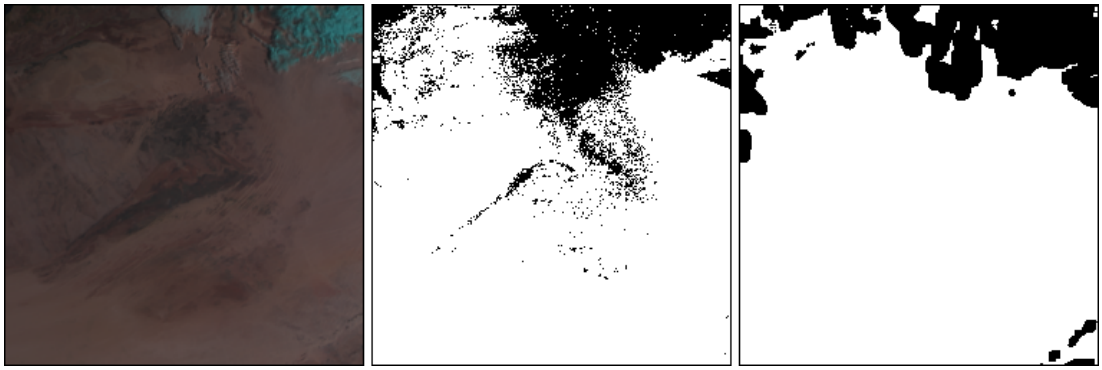
(q) Backward Signature Extension at UT09:45



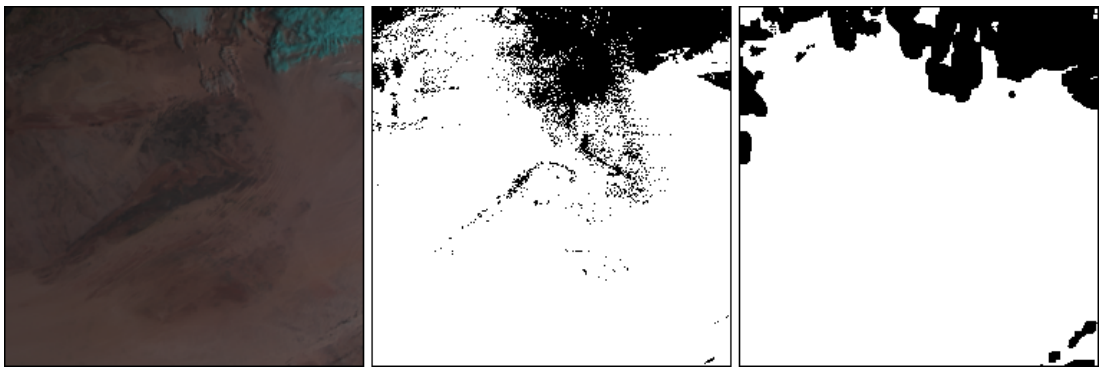
(r) Backward Signature Extension at UT09:30



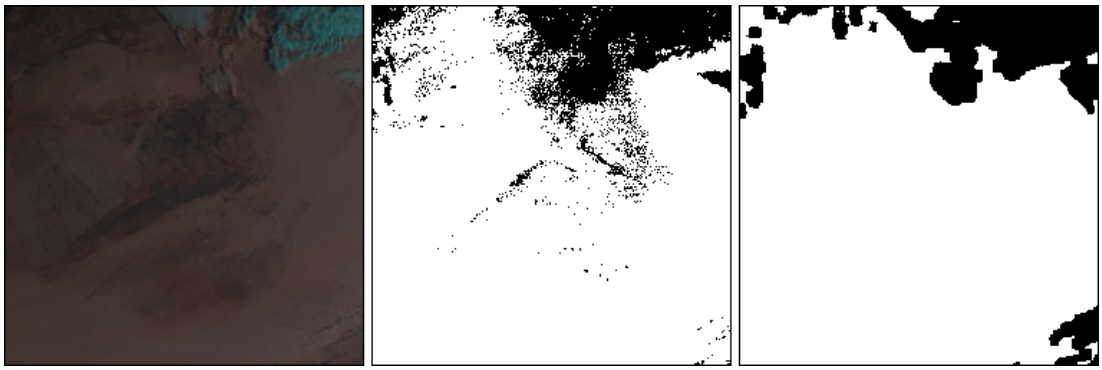
(s) Backward Signature Extension at UT09:15



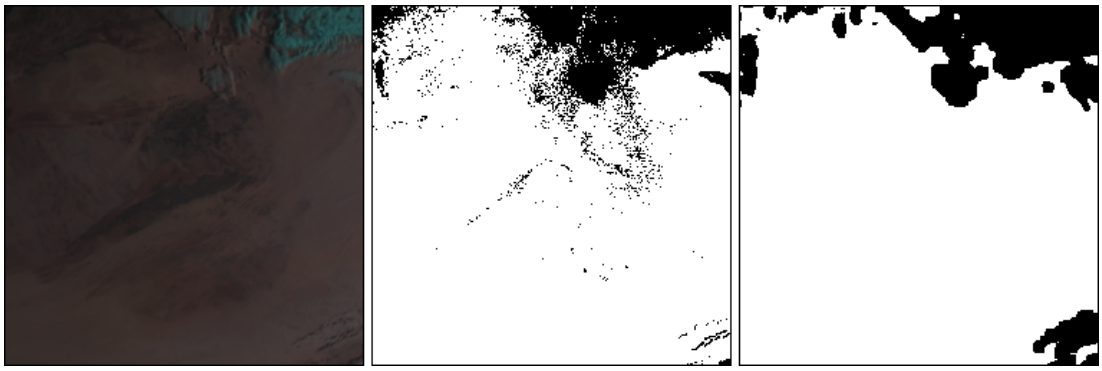
(t) Backward Signature Extension at UT09:00



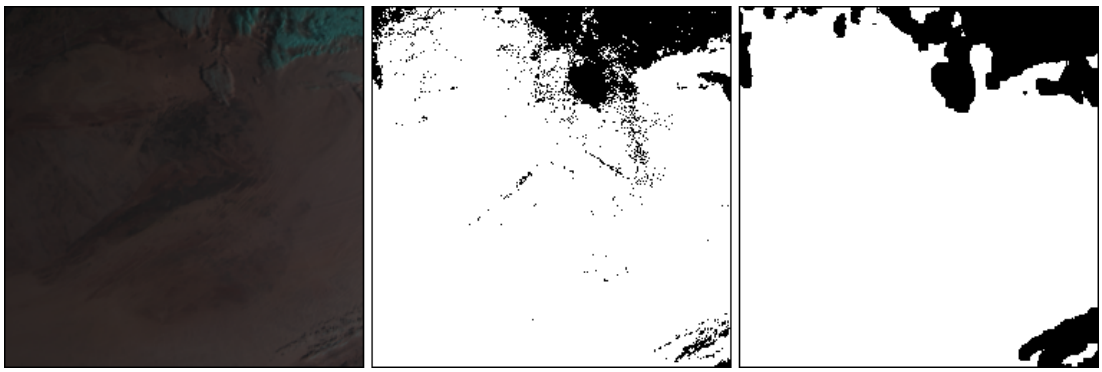
(u) Backward Signature Extension at UT08:45



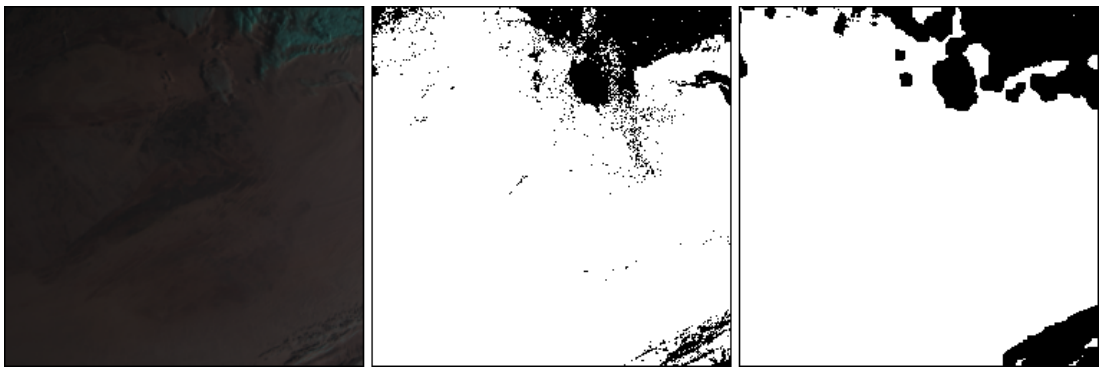
(v) Backward Signature Extension at UT08:30



(w) Backward Signature Extension at UT08:15



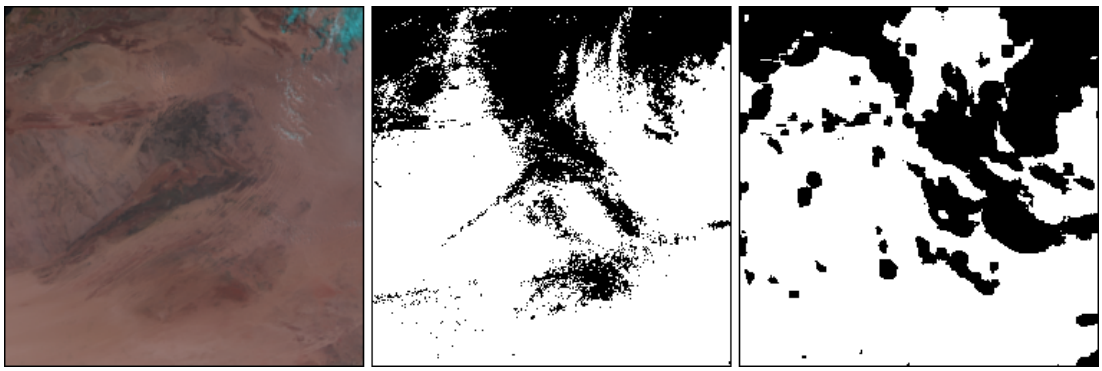
(x) Backward Signature Extension at UT08:00



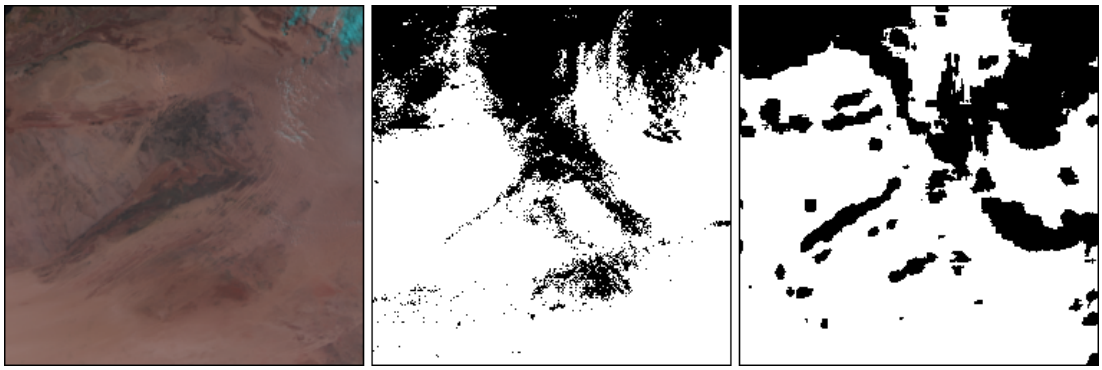
(y) Backward Signature Extension at UT07:45

Figure D.1: MODIS RGB image (left), BTD dust mask (centre) and Resulting masks from backward temporal extension using signature extension on the land test case (right). Dust is shown as white in the BTD mask and the resultant signature extension images

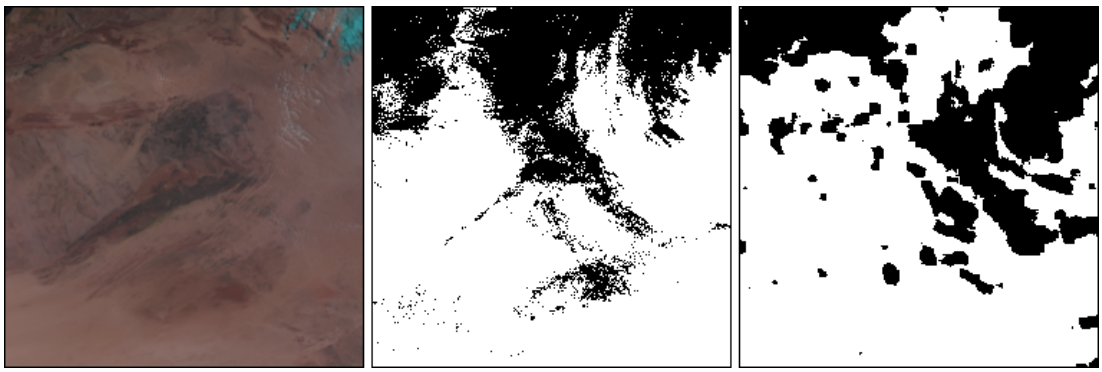
D.2 Forward Signature Extension



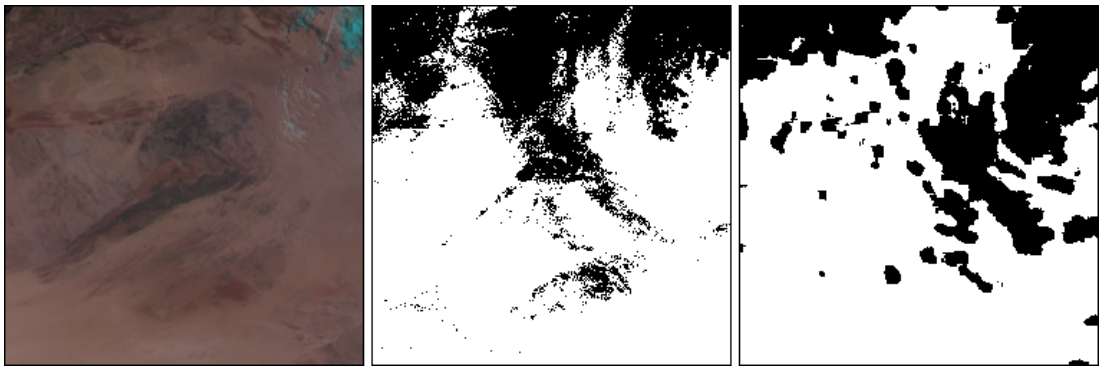
(a) Initial Classification at UT13:45



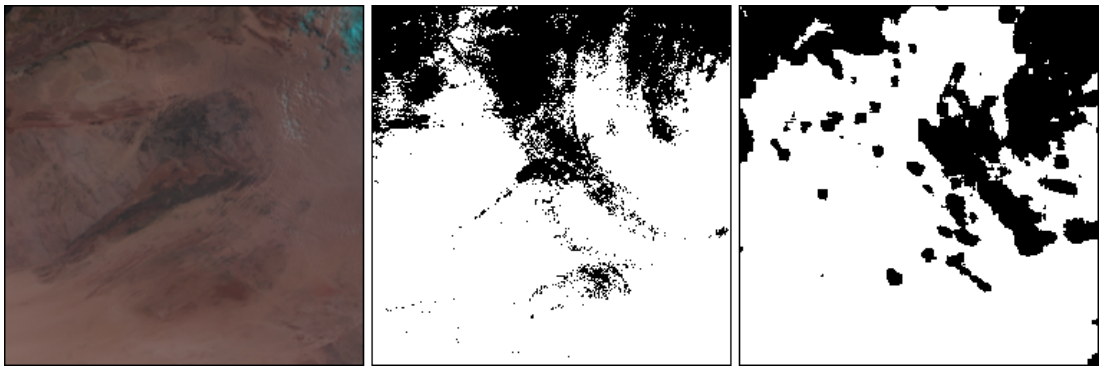
(b) Forward Signature Extension at UT14:00



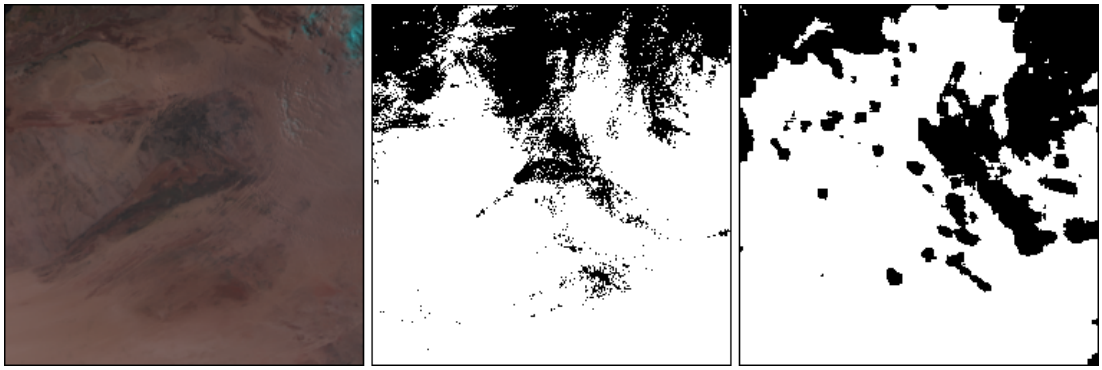
(c) Forward Signature Extension at UT14:15



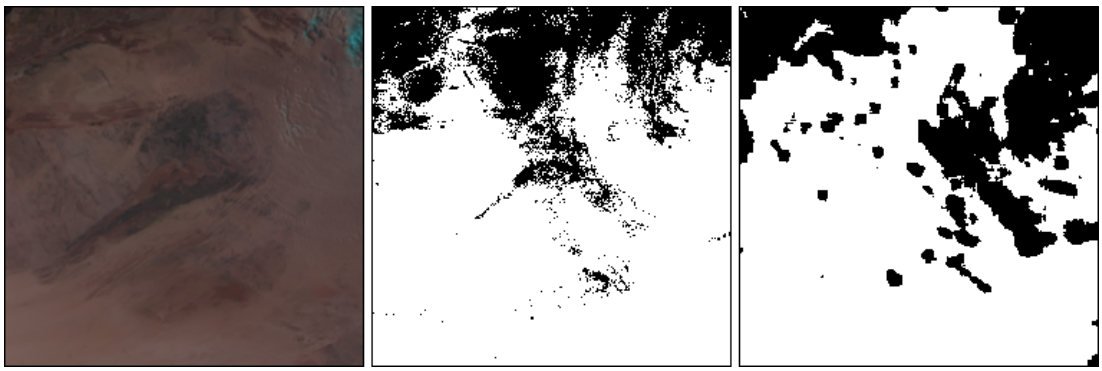
(d) Forward Signature Extension at UT14:30



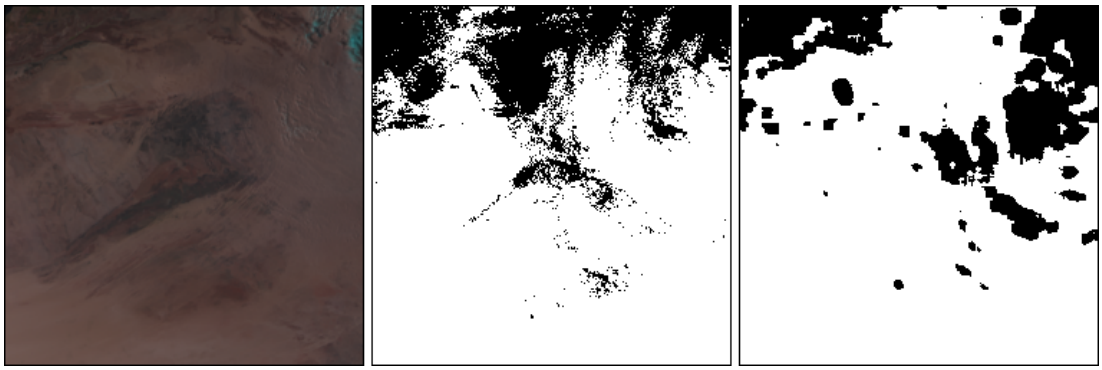
(e) Forward Signature Extension at UT14:45



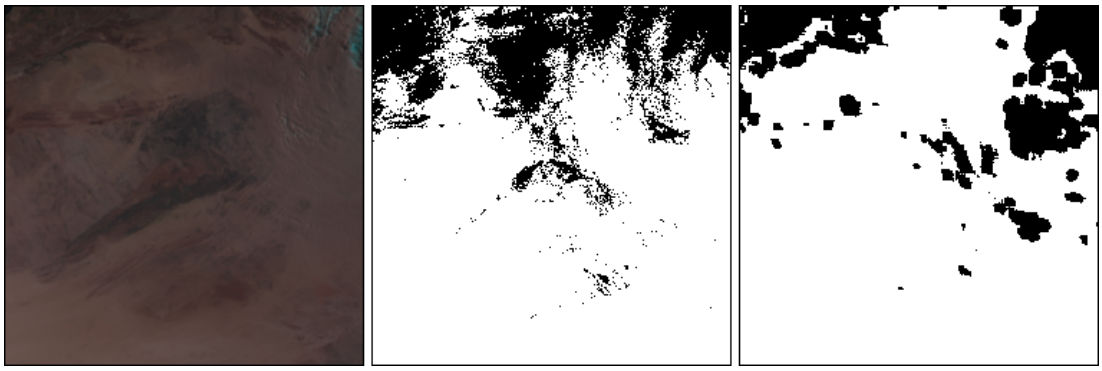
(f) Forward Signature Extension at UT15:00



(g) Forward Signature Extension at UT15:15

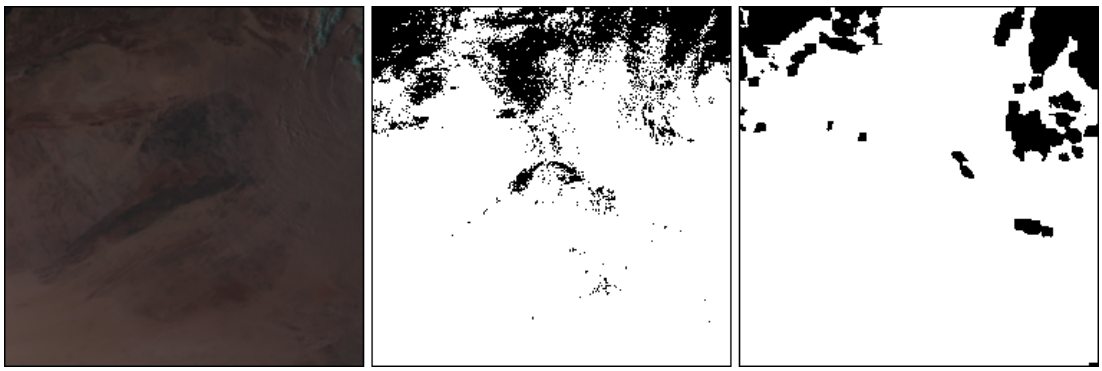


(h) Forward Signature Extension atUT15:30

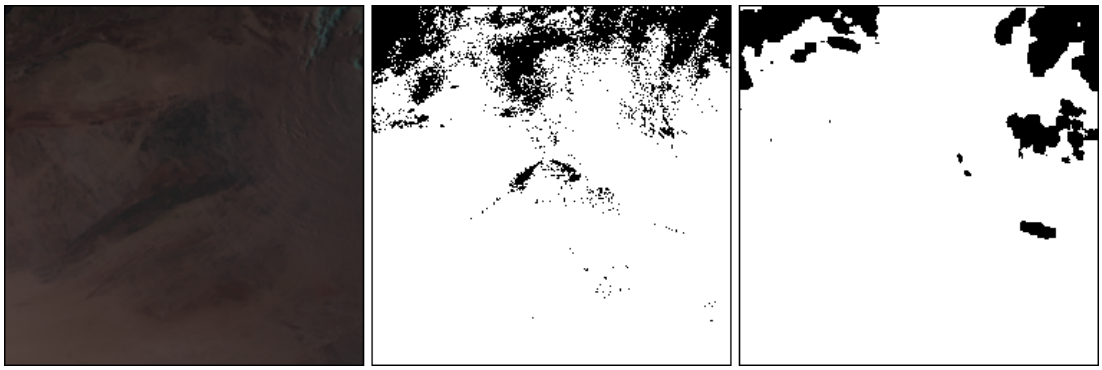


(i) Forward Signature Extension at UT15:45

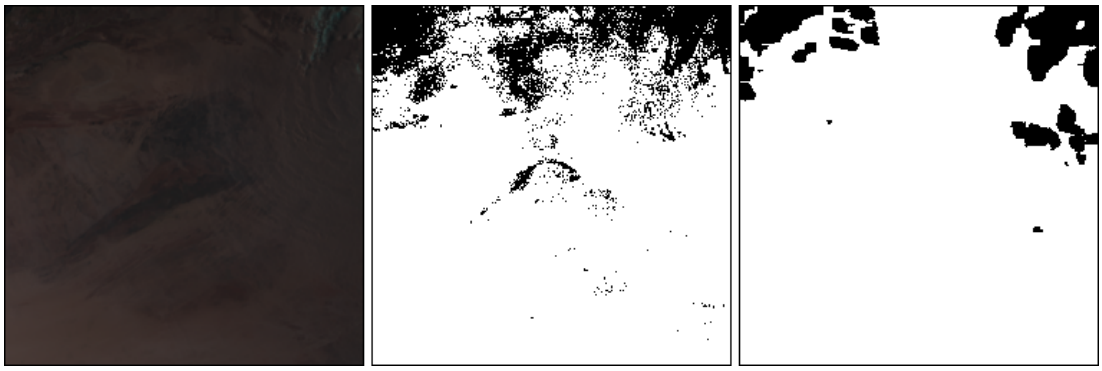
(j) Forward Signature Extension at UT16:00



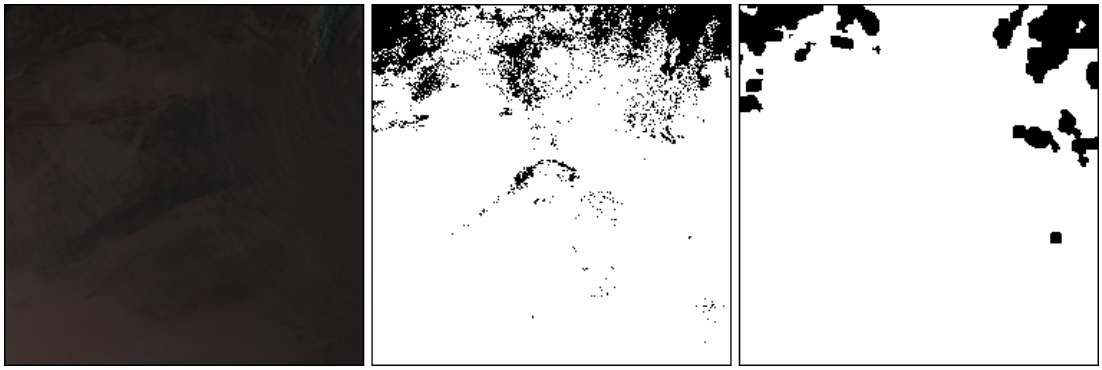
(k) Forward Signature Extension at UT16:15



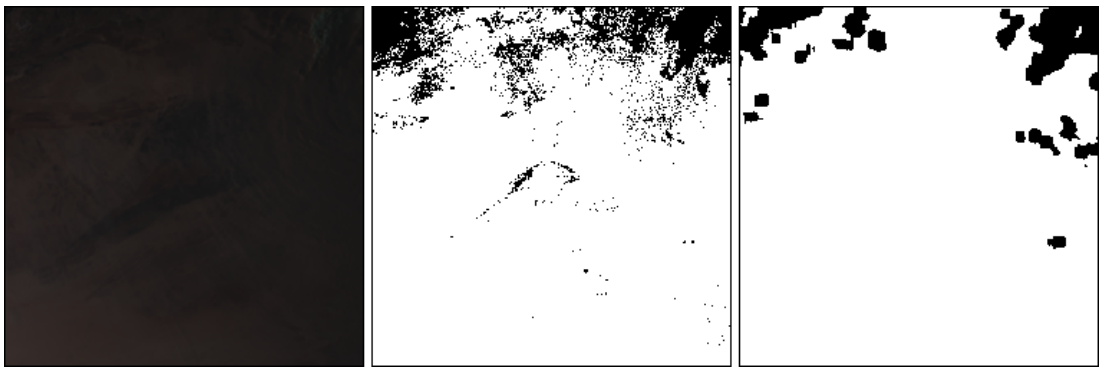
(l) Forward Signature Extension at UT16:30



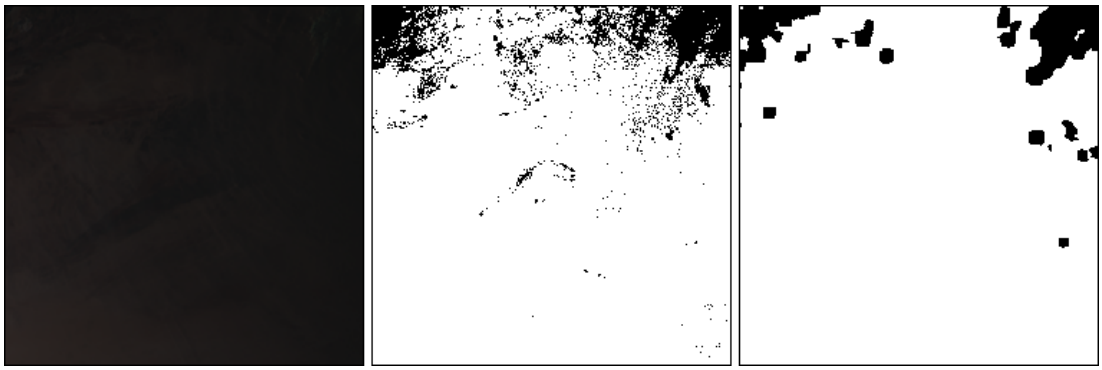
(m) Forward Signature Extension at UT16:45



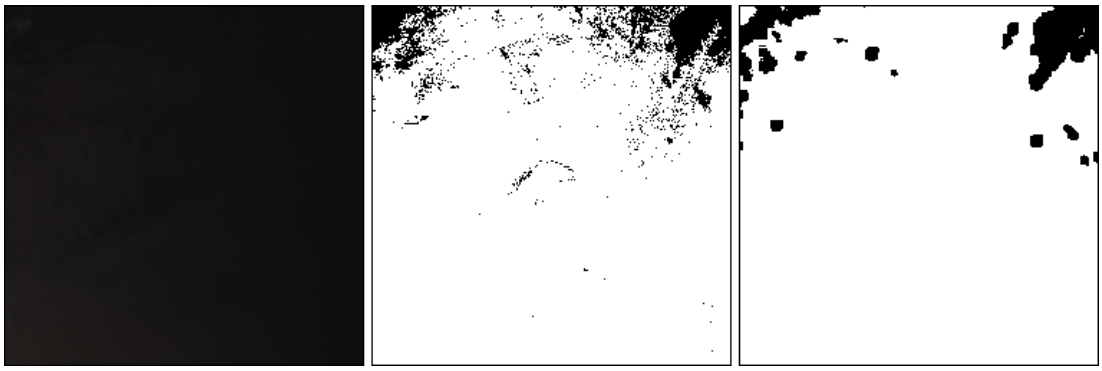
(n) Forward Signature Extension at UT17:00



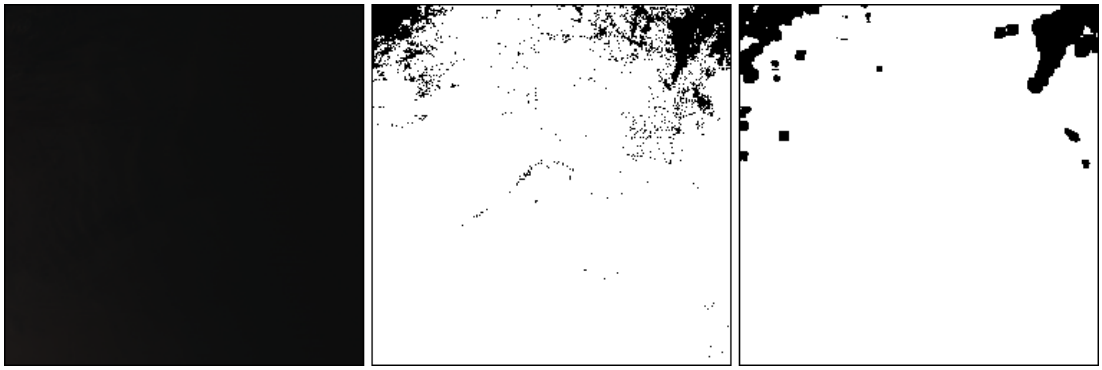
(o) Forward Signature Extension at UT17:15



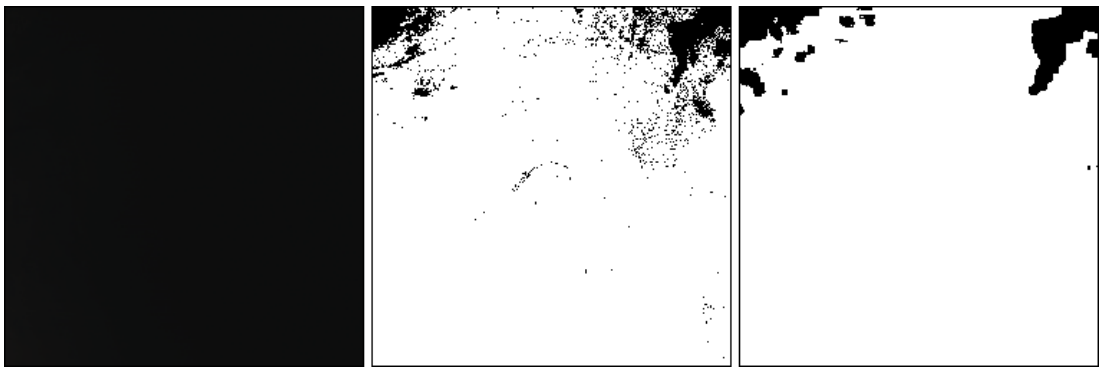
(p) Forward Signature Extension at UT17:30



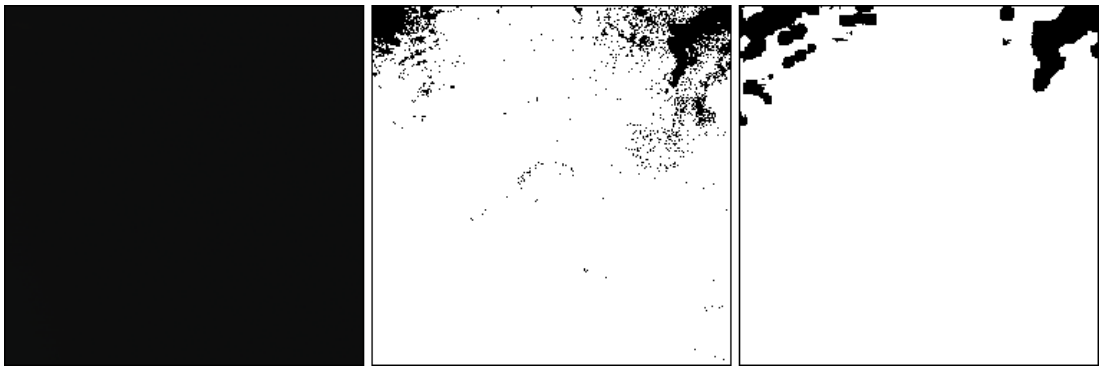
(q) Forward Signature Extension at UT17:45



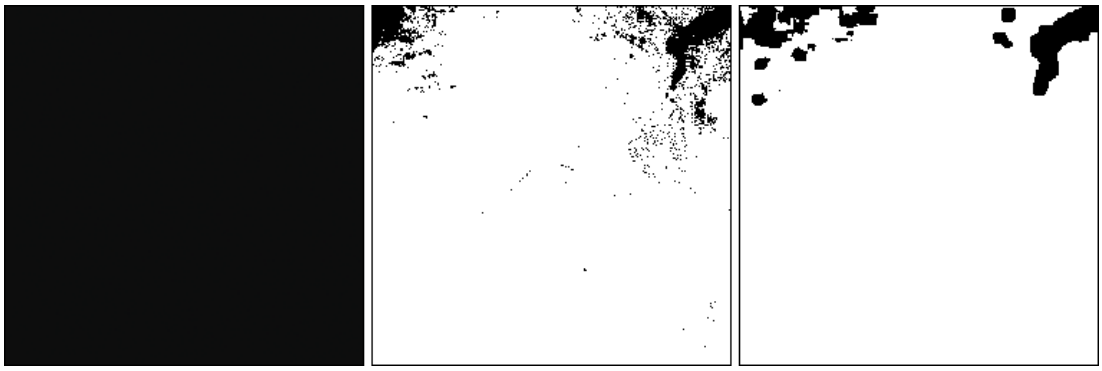
(r) Forward Signature Extension at UT18:00



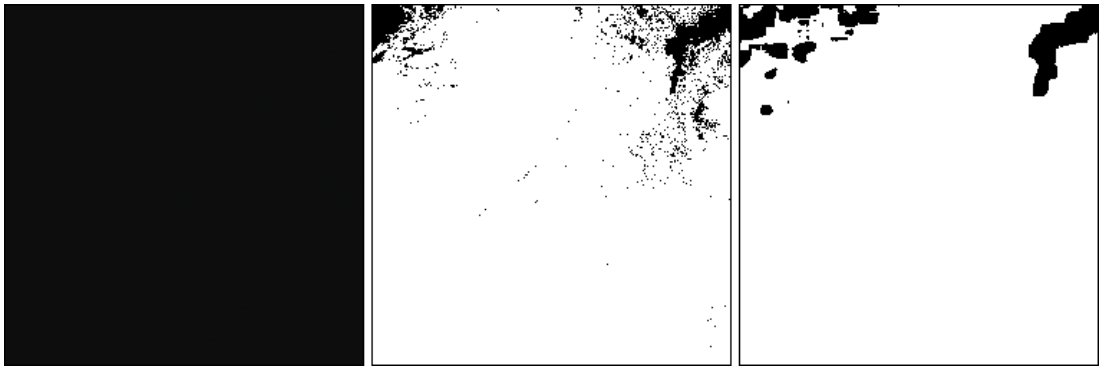
(s) Forward Signature Extension at UT18:15



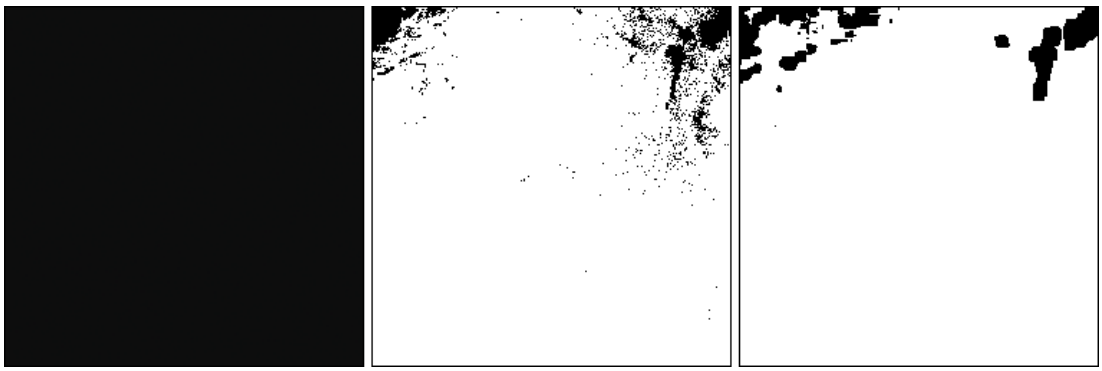
(t) Forward Signature Extension at UT18:30



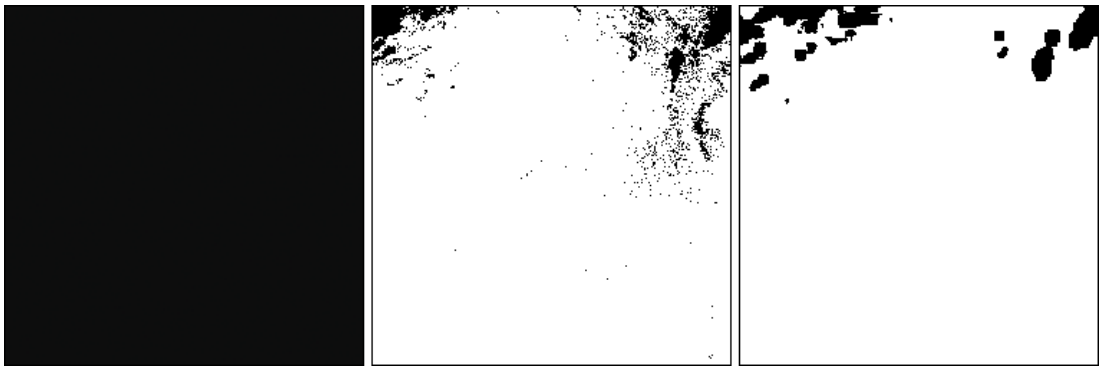
(u) Forward Signature Extension at UT18:45



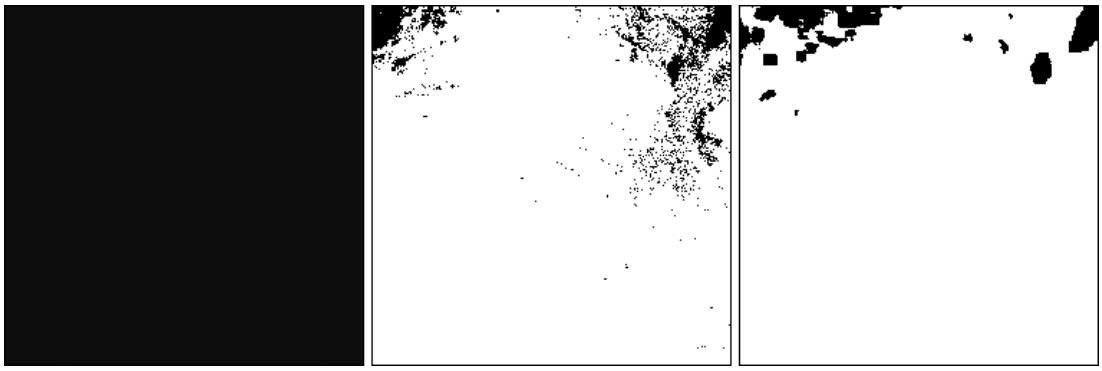
(v) Forward Signature Extension at UT19:00



(w) Forward Signature Extension at UT19:15



(x) Forward Signature Extension at UT19:30



(y) Forward Signature Extension at UT19:45

Figure D.2: MODIS RGB image (left), BTD dust mask (centre) and Resulting masks from forward temporal extension using signature extension on the land test case (right). Dust is shown as white in BTD mask and the resultant signature extension images

References

S.A. Ackerman. Remote sensing aerosols using satellite infrared observations. *Journal of geophysical research*, 102(D14):17069, 1997.

C. Ahn, O. Torres, and P.K. Bhartia. Comparison of ozone monitoring instrument UV aerosol products with Aqua/Moderate Resolution Imaging Spectroradiometer and Multiangle Imaging Spectroradiometer observations in 2006. *J. Geophys. Res*, 113, 2008.

NASA ASDC. Calipso quality statements: Lidar level 2 vertical feature mask overall quality summary: 2.01, 2008. URL `http://eosweb.larc.nasa.gov/PRODOCS/calipso/Quality_Summaries/CALIOP_L2VFMProducts_2.01.html`.

A. Baraldi, L. Bruzzone, and P. Blonda. Quality assessment of classification and cluster maps without ground truth knowledge. *Geoscience and Remote Sensing, IEEE Transactions on*, 43(4):857–873, 2005. ISSN 0196-2892.

A. Bayoko, S. Konate, F. Traore, and S. Fongang. Détection des nuages de poussières sur les images satellitaires Météosat et leur utilisation comme traceur pour l'identification des zones de déflation des sols par érosion éolienne. In *Surveillance des sols dans l'environnement par télédétection et systèmes d'information géographiques = Monitoring soils in the environment with remote sensing and GIS*, Colloques et Séminaires, pages 557–566. ORSTOM, 1996. ISBN 2-7099-1331-3. URL `http://www.documentation.ird.fr/hor/fdi:010008434`.

A.C. Bovik, M. Clark, and W.S. Geisler. Multichannel texture analysis using localized spatial filters. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 55–73, 1990. ISSN 0162-8828.

C. Brekke and A.H.S. Solberg. Oil spill detection by satellite remote sensing. *Remote Sensing of Environment*, 95(1):1–13, 2005. ISSN 0034-4257.

P. Brodatz. *Textures: a photographic album for artists and designers*. Dover Publications New York, 1999. ISBN 0486406997.

Peter R. Buseck and Mihly Psfai. Airborne minerals and related aerosol particles: Effects on climate and the environment. *Proceedings of the National Academy of Sciences of the United States of America*, 96(7):3372–3379, 1999. URL `http://www.pnas.org/content/96/7/3372.abstract`.

JB Campbell. Spatial correlation effects upon accuracy of supervised classification of land cover. *Photogrammetric Engineering and Remote Sensing*, 47:355–363, 1981. ISSN 0099-1112.

J.B. Campbell. *Introduction to remote sensing*. Guilford Press, 2006. ISBN 159385319X.

Chih-Chung Chang and Chih-Jen Lin. *LIBSVM: a library for support vector machines*, 2001. Software available at `http://www.csie.ntu.edu.tw/~cjlin/libsvm`.

A. Cheriyadat and L.M. Bruce. Why principal component analysis is not an appropriate feature extraction method for hyperspectral data. In *Geoscience and Remote Sensing Symposium, 2003. IGARSS'03. Proceedings. 2003 IEEE International*, volume 6, pages 3420–3422. IEEE, 2004. ISBN 0780379292.

M. Chica-Olmo and F. Abarca-Hernandez. Computing geostatistical image texture for remotely sensed data classification. *Computers & Geosciences*, 26(4):373–383, 2000. ISSN 0098-3004.

S.A. Christopher, B. Johnson, T.A. Jones, and J. Haywood. Vertical and spatial distribution of dust from aircraft and satellite measurements during the gerbils field campaign. *Geophysical Research Letters*, 36(6):L06806, 2009.

DA Chu, YJ Kaufman, C. Ichoku, LA Remer, D. Tanré, and BN Holben. Validation of MODIS aerosol optical depth retrieval over land. *Geophys. Res. Lett*, 29(12):8007, 2002.

D.A. Clausi. An analysis of co-occurrence texture statistics as a function of grey level quantization. *Canadian Journal of remote sensing*, 28(1):45–62, 2002. ISSN 0703-8992.

D.A. Clausi and H. Deng. Design-based texture feature fusion using Gabor filters and co-occurrence probabilities. *Image Processing, IEEE Transactions on*, 14(7):925–936, 2005. ISSN 1057-7149.

D.A. Clausi and M. Jernigan. Designing Gabor filters for optimal texture separability. *Pattern Recognition*, 33(11):1835–1849, 2000. ISSN 0031-3203.

D.A. Clausi and M.E. Jernigan. A fast method to determine co-occurrence texture features. *Geoscience and Remote Sensing, IEEE Transactions on*, 36(1):298–300, 2002. ISSN 0196-2892.

D.A. Clausi and Y. Zhao. Grey level co-occurrence integrated algorithm (GLCIA): a superior computational method to rapidly determine co-occurrence probability texture features* 1. *Computers & Geosciences*, 29(7):837–850, 2003. ISSN 0098-3004.

O. Diop, K. Kpalma, and J. Ronsin. Textural analysis for the detection of dust clouds from infrared satellite images. In *Proceedings of SPIE*, volume 6406, page 64061R, 2006.

R.A. Duce, R.J. Charlson, and J. Heintzberg. *Aerosol Forcing of Climate*. J.Wiley, New York, 1995.

D. Dunn and WE Higgins. Optimal Gabor filters for texture segmentation. *Image Processing, IEEE Transactions on*, 4(7):947–964, 2002. ISSN 1057-7149.

EARLINET. European aerosol research lidar network, Apr 2011. URL `http://www.earlinet.org/`.

EUMETSAT. Eumetsat archive services, 2008. URL `http://www.eumetsat.eu/Home/Main/Access_to_Data/Archive_Service/index.htm`.

EUMETSAT. Eumetsat case studies, 2010. URL `http://www.eumetsat.int/Home/Main/Image_Gallery/Case_Studies/Gallery/index.htm?l=en`.

FAAM. Facility for airborne atmospheric measurements, 2010. URL `http://www.faam.ac.uk/index.php/science-instruments/aerosol`.

G.M. Foody. Thematic map comparison: evaluating the statistical significance of differences in classification accuracy. *Photogrammetric Engineering and Remote Sensing*, 70(5):627–634, 2004. ISSN 0099-1112.

G.M. Foody and A. Mathur. A relative evaluation of multiclass image classification by support vector machines. *Geoscience and Remote Sensing, IEEE Transactions on*, 42(6):1335–1343, 2004a. ISSN 0196-2892.

G.M. Foody and A. Mathur. Toward intelligent training of supervised image classifications: directing training data acquisition for SVM classification. *Remote Sensing of Environment*, 93(1-2):107–117, 2004b. ISSN 0034-4257.

P. Formenti, JL Rajot, K. Desboeufs, S. Caquineau, S. Chevaillier, S. Nava, A. Gaudichet, E. Journet, S. Triquet, S. Alfaro, et al. Regional variability of the composition of mineral dust from western africa: Results from the amma sop0/dabex and dodo field campaigns. *Journal of Geophysical Research*, 113(D23):D00C13, 2008.

W. Förstner and B. Moonen. A metric for covariance matrices. *Quo vadis geodesia*, pages 113–128, 1999.

S. Fukuda and H. Hirosawa. A wavelet-based texture feature set applied to classification of multifrequency polarimetric sar images. *Geoscience and Remote Sensing, IEEE Transactions on*, 37(5):2282–2286, 1999.

R. Govindan. Three dimensional profiling of dust distribution in the earths atmosphere using horizontal and vertical satellite imageries. Master's thesis, University of Bath, 2007.

I. Guyon and A. Elisseeff. An introduction to variable and feature selection. *The Journal of Machine Learning Research*, 3:1157–1182, 2003. ISSN 1532-4435.

R.M. Haralick, K. Shanmugam, and I.H. Dinstein. Textural features for image classification. *IEEE Transactions on systems, man and cybernetics*, 3(6):610–621, 1973. ISSN 0018-9472.

EJ Highwood, JM Haywood, H. Coe, J. Cook, S. Osborne, P. Williams, J. Crosier, K. Bower, P. Formenti, J. McQuaid, et al. Aerosol Direct Radiative Impact Experiment (ADRIEX) overview. *Quarterly Journal of the Royal Meteorological Society*, 133(S1):3–15, 2007a.

E.J. Highwood, JM Haywood, H. Coe, J. Cook, S. Osborne, P. Williams, J. Crosier, K. Bower, P. Formenti, J. McQuaid, et al. Aerosol direct radiative impact experiment (adriex) overview. *Quarterly Journal of the Royal Meteorological Society*, 133(S1): 3–15, 2007b.

S.B. Hooker. *SeaWiFS technical report series: An overview of SeaWiFS and ocean color*, volume 1. National Aeronautics and Space Administration, Goddard Space Flight Center, 1992.

N.C. Hsu, S.C. Tsay, M.D. King, and J.R. Herman. Aerosol properties over bright-reflecting source regions. *Geoscience and Remote Sensing, IEEE Transactions on*, 42(3):557–569, 2004.

C. Huang, LS Davis, and JRG Townshend. An assessment of support vector machines for land cover classification. *International Journal of Remote Sensing*, 23(4):725–749, 2002. ISSN 0143-1161.

IPCC. Climate change 2007: Synthesis report, Nov 2007. URL `http://www.ipcc.ch/pdf/assessment-report/ar4/syr/ar4_syr.pdf`.

A. Jain and G. Healey. A multiscale representation including opponent color features for texture recognition. *Image Processing, IEEE Transactions on*, 7(1):124–128, 2002. ISSN 1057-7149.

A.K. Jain and F. Farrokhnia. Unsupervised texture segmentation using Gabor filters. *Pattern recognition*, 24(12):1167–1186, 1991. ISSN 0031-3203.

L.O. Jimenez and D.A. Landgrebe. Supervised classification in high-dimensional space: geometrical, statistical, and asymptotical properties of multivariate data. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, 28 (1):39–54, 2002. ISSN 1094-6977.

Y.J. Kaufman, D. Tanré, and O. Boucher. A satellite view of aerosols in the climate system. *Nature*, 419(6903):215–223, 2002a.

Y.J. Kaufman, A.E. Wald, L.A. Remer, B.C. Gao, R.R. Li, and L. Flynn. The MODIS 2.1-μm channel-correlation with visible reflectance for use in remote sensing of aerosol. *Geoscience and Remote Sensing, IEEE Transactions on*, 35(5):1286–1298, 2002b.

N. Khazenie and T. Lee. Identification of aerosol features such as smoke and dust, in NOAA-AVHRR data using spatial textures. In *IGARSS'92*, pages 726–730, 1992.

M.D. King, Y.J. Kaufman, D. Tanré, and T. Nakajima. Remote sensing of tropospheric aerosols from space: Past, present, and future. *BULLETIN-AMERICAN METEOROLOGICAL SOCIETY*, 80:2229–2260, 1999a.

M.D. King, Y.J. Kaufman, D. Tanré, and T. Nakajima. Remote sensing of tropospheric aerosols from space: Past, present, and future. *Bulletin of the American Meteorological Society*, 80(11):2229–2259, 1999b.

J. Knorn, A. Rabe, V.C. Radeloff, T. Kuemmerle, J. Kozak, and P. Hostert. Land cover mapping of large areas using chain classification of neighboring Landsat satellite images. *Remote Sensing of Environment*, 113(5):957–964, 2009. ISSN 0034-4257.

R. Kohavi and G.H. John. Wrappers for feature subset selection. *Artificial intelligence*, 97(1-2):273–324, 1997. ISSN 0004-3702.

M. Kubat, R.C. Holte, and S. Matwin. Machine learning for the detection of oil spills in satellite radar images. *Machine Learning*, 30(2):195–215, 1998. ISSN 0885-6125.

T.S. L'Ecuyer and J.H. Jiang. Touring the atmosphere aboard the a-train. *Physics Today*, 63(7):36, 2010.

T.A.E.Y. Lee and YJ Kaufman. Non-lamberlian effects on remote sensing of surface reflectance and vegetation index. *IEEE transactions on geoscience and remote sensing*, 24(5):699–708, 1986.

T.F. Lee. Dust tracking using composite visible/IR images: A case study. *Weather and Forecasting*, 4(2):258–263, 1989. ISSN 1520-0434.

M. Legrand, A. Plana-Fattori, and C. N'doumé. Satellite detection of dust using the IR imagery of Meteosat 1. Infrared difference dust index. *Journal of Geophysical Research*, 106(D16):18251, 2001.

J. Li, P. Zhang, TJ Schmit, J. Schmetz, and WP Menzel. Quantitative monitoring of a Saharan dust event with SEVIRI on Meteosat-8. *International journal of remote sensing*, 28(10):2181–2186, 2007.

R.R. Li, L. Remer, Y.J. Kaufman, S. Mattoo, B.C. Gao, and E. Vermote. Snow and ice mask for the MODIS aerosol products. *Geoscience and Remote Sensing Letters, IEEE*, 2(3):306–310, 2005.

Y. Liu, J.A. Sarnat, B.A. Coull, P. Koutrakis, and D.J. Jacob. Validation of MISR Aerosol Optical Thickness measurements using AERONET observations over the contiguous United States. *Journal of Geophysical. Research*, 2003.

Z. Liu, A.H. Omar, Yongxiang Hu, M.A. Vaughan, and D.M. Winker. *CALIOP Algorithm Theoretical Basis Document: Part 3: Scene Classification Algorithms*, Oct 2005.

JF MacGregor and T. Kourti. Statistical process control of multivariate processes. *Control Engineering Practice*, 3(3):403–414, 1995. ISSN 0967-0661.

Franco Marenco. *Capabilities and operations of the FAAM EZlidar*, Mar 2010.

A.M. Martínez and A.C. Kak. Pca versus lda. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 23(2):228–233, 2002. ISSN 0162-8828.

J.V. Martins, D. Tanré, L. Remer, Y. Kaufman, S. Mattoo, and R. Levy. MODIS cloud screening for remote sensing of aerosols over oceans using spatial variability. *Geophys. Res. Lett*, 29(12):8009, 2002.

P.M. Mather. *Computer processing of remotely sensed images: an introduction*, chapter Classification. Wiley, 2004. ISBN 0470849185.

MATLAB. *version 7.9.0.529 (R2009b)*. The MathWorks Inc., Natick, Massachusetts, 2009.

J. Müller. *MSG Level 1.5 Image Data Format Description*. EUMETSAT, Am Kavalleriesand 31, D-64295 Darmstadt, Germany, eum/msg/icd/105 v5a edition, August 2007.

NASA. Calipso data sets, 2008. URL `http://eosweb.larc.nasa.gov/PRODOCS/calipso/table_calipso.html`.

NASA. Nasa calipso website, 2010a. URL `http://www-calipso.larc.nasa.gov/`.

NASA. Aerosol robotic network, Apr 2010b. URL `http://aeronet.gsfc.nasa.gov/new_web/index.html`.

NASA. Nasa omi instrument on aura homepage, Apr 2011. URL `http://aura.gsfc.nasa.gov/instruments/omi.html`.

T. Ojala, T. Mäenpää, M. Pietikäinen, J. Viertola, J. Kyllönen, and S. Huovinen. Outex-new framework for empirical evaluation of texture analysis algorithms. *Pattern Recognition*, 1:10701, 2002a.

T. Ojala, M. Pietikäinen, and T. Mäenpää. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on pattern analysis and machine intelligence*, pages 971–987, 2002b.

I. Olthof, C. Butson, and R. Fraser. Signature extension through space for northern landcover classification: A comparison of radiometric correction methods. *Remote sensing of environment*, 95(3):290–302, 2005. ISSN 0034-4257.

T. Oommen, D. Misra, N.K.C. Twarakavi, A. Prakash, B. Sahoo, and S. Bandopadhyay. An objective analysis of support vector machine based classification for remote sensing. *Mathematical Geosciences*, 40(4):409–424, 2008. ISSN 1874-8961.

G. Pappalardo, A. Papayannis, J. Bösenberg, A. Ansmann, A. Apituley, L.A. Arboledas, D. Balis, C. Böckmann, A. Chaikovsky, A. Comeron, et al. EARLINET coordinated lidar observations of Saharan dust events on continental scale. In *IOP Conference Series: Earth and Environmental Science*, volume 7, page 012002. IOP Publishing, 2009.

M. Pax-Lenney, C.E. Woodcock, S.A. Macomber, S. Gopal, and C. Song. Forest mapping with a generalized classifier and Landsat TM data. *Remote Sensing of Environment*, 77(3):241–250, 2001. ISSN 0034-4257.

M. Porat and YY Zeevi. The generalized Gabor scheme of image representation in biological and machine vision. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 10(4):452–468, 2002. ISSN 0162-8828.

J.M. Prospero. Long-range transport of mineral dust in the global atmosphere: Impact of African dust on the environment of the southeastern United States. *Proceedings of the National Academy of Sciences of the United States of America*, 96(7):3396–3403, 1999. URL `http://www.pnas.org/content/96/7/3396.abstract`.

LA Remer, D. Tanré, YJ Kaufman, C. Ichoku, S. Mattoo, R. Levy, DA Chu, B. Holben, O. Dubovik, A. Smirnov, et al. Validation of MODIS aerosol retrieval over ocean. *Geophys. Res. Lett*, 29(12):8008, 2002.

LA Remer, YJ Kaufman, D. Tanre, S. Mattoo, DA Chu, JV Martins, R.R. LI, C. Ichoku, RC Levy, RG Kleidman, et al. The MODIS Aerosol algorithm, products, and validation. *Journal of the atmospheric sciences*, 62(4):947–973, 2005.

L.A. Remer, D. Tanré, Y.J. Kaufman, R. Levy, and S. Mattoo. Algorithm for remote sensing of tropospheric aerosol from MODIS: Collection 005. *National Aeronautics and Space Administration*, 2009.

S.E. Schwartz and M.O. Andreae. Uncertainty in Climate Change Caused by Aerosols. *Science*, 272(5265):1121–0, 1996. doi: 10.1126/science.272.5265.1121. URL `http://www.sciencemag.org`.

Y. Shao and CH Dong. A review on East Asian dust storm climate, modelling and monitoring. *Global and Planetary Change*, 52(1-4):1–22, 2006.

M. Simard, S.S. Saatchi, and G. De Grandi. The use of decision tree and multiscale texture for classification of JERS-1 SAR data over tropical forest. *Geoscience and Remote Sensing, IEEE Transactions on*, 38(5):2310–2321, 2002. ISSN 0196-2892.

J. Sklansky. Image segmentation and feature extraction. *Systems, Man and Cybernetics, IEEE Transactions on*, 8(4):237–247, 2007. ISSN 0018-9472.

C. Staelin. Parameter selection for support vector machines. *Hewlett-Packard Company, Tech. Rep. HPL-2002-354R1*, 2003.

P. Stammes and R. Noordhoek. *OMI Algorithm Theoretical Basis Document: Clouds, Aerosols, and Surface UV Irradiance*, Aug 2002.

I. Tegen, A.A. Lacis, and I. Fung. The influence on climate forcing of mineral aerosols from disturbed soils. *Nature*, 380(6573):419–422, 1996. ISSN 0028-0836.

O. Torres, PK Bhartia, JR Herman, A. Sinyuk, P. Ginoux, and B. Holben. A long-term record of aerosol optical depth from TOMS observations and comparison to AERONET measurements. *Journal of the Atmospheric Sciences*, 59(3):398–413, 2002a.

O. Torres, PK Bhartia, JR Herman, A. Sinyuk, P. Ginoux, and B. Holben. A long-term record of aerosol optical depth from TOMS observations and comparison to AERONET measurements. *Journal of the Atmospheric Sciences*, 59(3):398–413, 2002b.

M. Tuceryan and A.K. Jain. Texture analysis. *Handbook of pattern recognition and computer vision*, 276, 1993.

UNFCCC. Kyoto convention definition, Dec 1997. URL `http://unfccc.int/essential_background/convention/background/items/1353.php`.

M.A. Vaughan, D.M. Winker, and K.A. Powell. *CALIOP Algorithm Theoretical Basis Document: Part 2: Feature Detection and Layer Properties Algorithms*, Sep 2005.

M. Vaughana, S. Youngb, D. Winkerc, K. Powella, A. Omarc, Z. Liud, Y. Huc, and C. Hostetlerc. Fully automated analysis of space-based lidar data: An overview of the CALIPSO retrieval algorithms and data products. In *Proc. of SPIE Vol*, volume 5575, pages 16–30, 2004.

M. Wiegner, J. Gasteiger, et al. Characterization of the Eyjafjallajökull ash-plume: Potential of lidar remote sensing. *Physics and Chemistry of the Earth, Parts A/B/C*, 2011. ISSN 1474-7065.

B. Wiltshire, R. Govindan, I. Astin, and A.N. Evans. Combining CALIPSO and Meteosat Images to Study the Distribution of Atmospheric Dust. In *Geoscience and Remote Sensing Symposium, 2008. IGARSS 2008. IEEE International*, volume 3. IEEE, 2009.

M. Wölfel and H.K. Ekenel. Feature Weighted Mahalanobis Distance: Improved Robustness for Gaussian Classifiers. In *13th European Signal Processing Conference, EUSIPCO*, volume 2005. Citeseer, 2005.

A. Woolley. Volcanic ash over uk, Apr 2010. URL `http://www.faam.ac.uk/index.php/news/275-volcanic-ash-event`.

R. Xu and D. Wunsch. Survey of clustering algorithms. *IEEE Transactions on neural networks*, 16(3):645–678, 2005. ISSN 1045-9227.

Y. Yang and J.O. Pedersen. A comparative study on feature selection in text categorization. In *Machine Learning-International Workshop Then Conference-*, pages 412–420. Citeseer, 1997.

S.A. Young, D.M. Winker, M.A. Vaughan, Yongxiang Hu, and R.E. Kuehn. *CALIOP Algorithm Theoretical Basis Document: Part 4: Extinction Retrieval Algorithms*, 2008.

S.C. Yusta. Different metaheuristic strategies to solve the feature selection problem. *Pattern Recognition Letters*, 30(5):525–534, 2009. ISSN 0167-8655.

P. Zhang, N. Lu, X. Hu, and C. Dong. Identification and physical retrieval of dust storm using three MODIS thermal IR channels. *Global and Planetary Change*, 52 (1-4):197–206, 2006.