Image Mining within Meteosat Data: A Case of Modeling Forest Fire

Umamaheshwaran Rajasekar

March, 2005

Image Mining within Meteosat Data: A Case of Modeling Forest Fire

by

Umamaheshwaran Rajasekar

Thesis submitted to the International Institute for Geo-information Science and Earth Observation in partial fulfilment of the requirements for the degree in Master of Science in *Geoinformatics*.

Degree Assessment Board

Thesis advisor	Prof. Dr. Alfred Stein Dr. Wietske Bijker
Thesis examiners	Chairman: Prof. Dr. Ir. M.G. Vosselman External examiner: Ms. M.J. Caldas Paulo

INTERNATIONAL INSTITUTE FOR GEO-INFORMATION SCIENCE AND EARTH OBSERVATION

ENSCHEDE, THE NETHERLANDS

Disclaimer

This document describes work undertaken as part of a programme of study at the International Institute for Geo-information Science and Earth Observation (ITC). All views and opinions expressed therein remain the sole responsibility of the author, and do not necessarily represent those of the institute.

Abstract

Remote Sensing Images are being collected nowadays every 15 minutes from satellites such as Meteosat, covering large areas of land. These repositories of images can be used for a range of different purposes. For the human mind, it may be hard to consider each image individually, analyze it as well as their relationships with the previous images of varying time steps. In order to address that issue, this research attempts to develop a simple, time efficient and effective generic model to facilitate the process of pattern discovery from series of remote sensing imageries.

The data used for the study were of Meteosat Second Generation. The focus of this study is on development of a model for monitoring and analyzing forest fires in space and time. As a case, a diurnal cycle of fire, which took place in Portugal, on 28th of July, 2004 was taken and analyzed. Kernel convolution method was used to characterize the hearth of the fire in space. The patterns of these fire objects in space were then extracted and tracked over time. These algorithms were automated to analyze for a series of imageries. The results thus obtained forms the knowledge gained about the fire in space and time. This knowledge was used for a further understanding of the behavior of fire with respect to vegetation and wind.

This mining model allows one to better understand the behavior of the fire in space and time. Such a model may then be useful for making predictions of hazards at an almost real time basis. The research is promising for data mining, possibly allowing other spatio-temporal phenomena to be modeled as well.

Keywords

Remote Sensing, Data Mining, Knowledge Discovery, Meteosat, Modeling fire, Pattern Analysis

Contents

List of Figures

List of Tables

Acknowledgements

I would like to sincerely thank Prof. Alfred Stein and Dr. Wietske Bijker for their support and guidance all through this work.

I would like to thank Boudewijn van Leeuwen for helping me in understanding the data and in its acquisition, Jan Hendrikse for helping me with the projections and Bas Retsios for his improvements to GDAL.

I would like to thank Valentyn, Arta, Ulanbek, Nimish and Jorge for all interesting discussions.

I would like to thank my mom and dad for their blessings.

Chapter 1

Modeling Forest Fire

1.1 Observations of Space in Time

Earth system science is an interdisciplinary science strongly coupled to remote sensing data. Satellite observations are often the only way to obtain the required information to understand the individual system processes as well as processes linking different systems. Great progress has been made in the last two decades with the advent of the space program and associated Earth observations [1]. The number of remote sensing imageries (RSI) that are being collected every day from satellites, aerial sensors, telescopes and other sensor platforms is large and plans of new missions are building up every day. Remotely sensed data, combined with additional data from the ecosystem models, offer an unprecedented opportunity for predicting and understanding the behavior of the Earth's ecosystem [2]. It is of importance to Earth system science to extract useful information from these data. These data have still not been fully explored and analyzed even with the advancement in techniques that became available recently [3]. The majority of the data are archived before information can be extracted from them [4].

A development in the field of information technology (IT) in terms of processing algorithms and data mining methods enable more a opportunistic use of data banks of remote sensing images [3]. The importance of developing such data driven inductive approaches to geographical analysis and modeling is to facilitate the creation of new knowledge and aid the processes of scientific discovery [5].

Geographical Data Mining can be regarded as a special type of data mining that seeks to perform similar generic functions as conventional data mining tools. It is modified to take into account the special features of geoinformation, the rather different styles and needs of modeling relevant to the world of GIS, and the peculiar nature of geographical explanation [5]. In this regard the development of special tools and techniques which would help the earth science research scientists to explore the massive archives of data to extract useful information in terms of patterns and knowledge from them becomes important.

1.2 Problem Statement

1.2.1 Need for image mining

"The world produces between 1 and 2 exabytes of unique information per year, that is roughly 250 megabytes for every man, woman, and child on earth. An exabyte is a billion gigabytes, or 10^{18} bytes" [6]. On this immense body of data, little is being done to analyze and to extract all possible information. This problem of massive increase in the size of the data and problems related to information extraction from them is akin to the field of Remote Sensing where there has been advancement in satellite technology over the past few years. Remote Sensing Images are being collected nowadays covering large areas of land. These repositories of images can be used for a range of different purposes. However, for the human mind, it may be hard to consider each image individually, analyze it and develop relationships with the previous images of varying time steps. Therefore there is a need to address this issue of accumulation of remotely sensed data using the techniques of data mining.

1.2.2 Need for information extraction from Meteosat images [7]

The primary mission of Meteosat Second Generation (MSG) is the continuous observation of the earth's full disk. This is achieved with the Spinning Enhanced Visible and Infrared Imagery (SEVIRI) imaging radiometer. SEVIRI is a 12-channel (3 visible channels, 8 infrared channels and 1 high resolution visible channel) imagery observing the earth-atmosphere system. Eleven channels observe the earth's full disk with a 15-min repeat cycle. A high-resolution visible (HRV) channel covers half of the full disk in the east-west direction and a full disk in the north-south direction. The characteristic of this sensor has a lot of potential in terms of temporal duration and spectral variation for the analysis of spatio-temporal phenomena. Some of the potential opportunities from the MSG for land surface research and applications have been elaborated in Meteosat's technical document [8]. But many of these potentials are yet to be exploited from the imagery. Moreover on a regular day basis 96 images are collected and each image is a composite of 12 channels. Most of these images are archived before the information could be extracted from them. Therefore a strong need exists to identify and exploit the potentials of Meteosat imagery.

1.2.3 Need for monitoring fire

At present, concern for environment is increasing the interest in monitoring and predicting ecosystem changes [9]. Forest fires are one such phenomenon that impairs biodiversity, influences climate on regional & global scales and promotes soil erosion. Fires also increase the load of soot and organic particles in the atmosphere, leading to more clouds, but less rain, and a reduction in the amount of solar irradiance reaching the surface of the earth.

Remote sensing satellites are currently widely used for fire prediction, detection, monitoring and assessment. Several satellites are operational like NOAA,

Landsat, TERRA, AQUA, ERS, DMSP, SPOT, BIRD, GOES, etc., which are being currently used for this purpose. Based on observations from these satellites and relevant ground truth much work has been done to predict, detect and model the characteristics of the fire. One of the major constraints in monitoring the forest fires of Europe and the Savanna fires of Africa is the lack of observation time. Even satellites such as BIRD which are designed for experimental monitoring of fire are polar orbiting in nature. Therefore continuous observation of the event to characterize the phenomenon becomes impossible.

To overcome this difficulty European Space Agency (ESA) is planning to come up with small scale missions that would help in continuous observation of the event of fire. In addition one of the potentials of MSG, SEVIRI sensor as identified in the pre-launch phase is for fire detection. Owing to the presence of suitable bands in SEVIRI sensor and the ability to continuously observe regions of Europe and Africa with a lag time of 15 minutes makes it one of the potential satellites for monitoring the characteristics of fire. Further the fire dynamics could be better studied along with the wind and climate data that are extracted from the MSG by weather departments. These data are disseminated to the users on request. Thirdly there have been cases of fire been clearly visible in the MSG images both in the infra red and visible bands.

This potential of MSG and the need for high temporal fire monitoring is explored in this research using techniques of image mining.

1.3 Research Objective

Fire is a dynamic phenomenon with important changes occurring on an hourly basis. In order to monitor these changes one needs remote sensing imagery of high temporal nature. Satellites such as MSG despite of having low spatial resolutions provides imageries of high temporal resolutions(every 15 mins). The main objective of this research is to exploit the potentials of MSG by developing a model to monitor and analyze fire patterns in time. The sub objectives of the research are that the developed model be

- scalable to multiple images
- is effective in terms of characterizing the fire and efficient in processing
- acts as a generic model to analyze other phenomena from Meteosat

1.4 Research Questions and Assumptions

The development of the model for monitoring and analysis of forest fire using high temporal resolution imagery has many issues attached to it. This research answers some of the questions related to modeling and spatio-temporal analysis, such as:

Modeling

- How to handle large amount of RSI

- What would be a suitable function to represent forest fire in a low spatial resolution imagery
- What would be a suitable method for fitting the data
- How could the fire be characterized over space and tracked over time

Spatio-temporal analysis

- What would be the effect of vegetation and wind on the behavior of fire

Some of the main assumptions that were considered in this research were

- Forest fire under study is large enough to be identified on the Meteosat imagery
- Due to the low spatial resolution of the Meteosat imagery, the radiometric value of the pixel is the mixture of all the elements within that space (3 km^2). This value of the pixel is therefore assigned to the center of pixel rather than being uniformly distributed within it.

1.5 Study Site

The forest fire which occurred in the south of Europe in Portugal during 28th of July, 2004 is taken as the study site. This fire is modeled over space and time using images from Meteosat SEVIRI imagery. The knowledge acquired along with additional data form the Meteosat Meteorological Products and CORINE Land cover data is then used for prediction/understanding the behavior of the phenomenon

1.6 Structure of the thesis

Chapter 2 discusses the background study on fire and the Meteosat data. It is divided into two sections. Section 2.2 of the chapter 2 discusses the issues attached to remote sensing data mining with relevance to Meteosat imagery. Section 2.3 discusses forest fire, various satellites available for its detection and models used for monitoring and predicting their movement. Chapter 3 describes the image mining model developed within this research and explains various steps that were undertaken in the process of knowledge discovery. Chapter 4 summarizes the results of the model. Chapter 5 discusses various other techniques that were experimented with in the process of modeling the same and their results. This chapter also briefs about limitation and extensibility of the developed model. Chapter 6 summarizes the research and concludes.

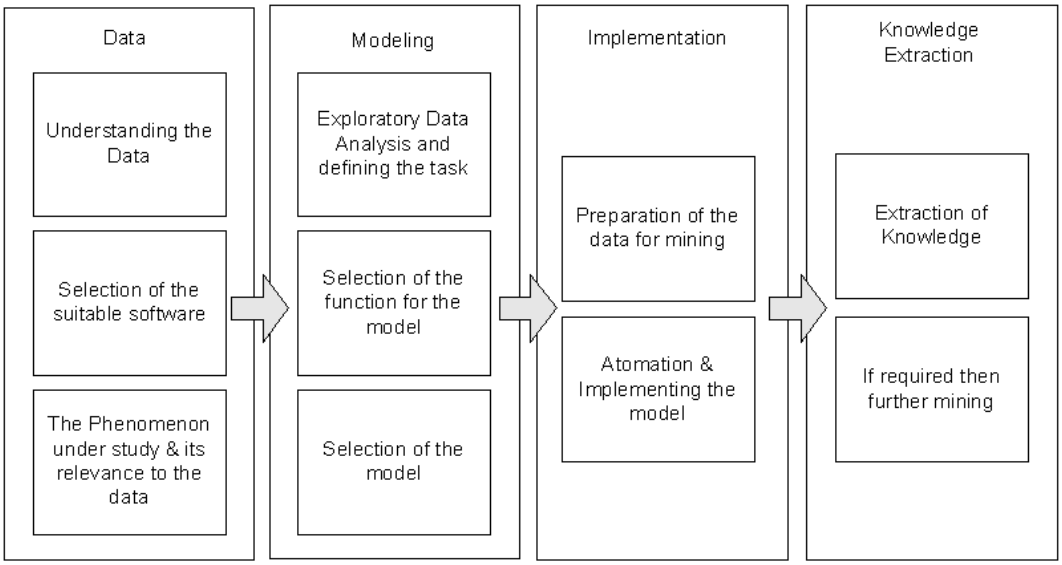

Figure 1.1: Architecture of the research with the steps undertaken

Chapter 2

Data Mining and Fire

2.1 Introduction

This chapter explains in brief about data mining and forest fire monitoring. Section 2.2 of this chapter describes data mining issues related to remote sensing imagery with importance to Meteosat imagery. This section also summarizes data mining models that are available and the spatial data mining software that could be used to implement those models. Section 2.3 gives a brief introduction to the characteristics of forest fire, remote sensors that are currently being used for fire detection and models developed for to predict, detect and model fire characteristics.

2.2 Data Mining

2.2.1 Introduction

Data Mining is developing in fields where the collection of large amount of data occurs. For example in the field of Remote Sensing with its advancement in technology during the past few years the quantities of Remote Sensing Imagery (RSI) that are being collected every day from satellites, aerial sensors, telescopes and other sensor platforms are immense. Majority of these images are archived before information can be extracted [4]. The history of data mining goes back to 1970's by the development of many expert system applications basically in the field of medicine MYCIN (medical diagnosis [10]) and in the field of defense [11]. Even though the word data mining was coined in late 1980's and early 1990's, the process of finding useful information from the previous experiences (databases) is quite old. Data Mining is a field which has developed by encompassing principles and techniques from statistics, machine learning, pattern recognition, numeric search and scientific visualization to accommodate the new data types and data volumes being generated [12]. The tasks of data mining might vary but the premise about discovering unknown information from large database remains the same. In short data mining can be defined as "The analysis of (often large) observational data sets to find unsuspected relationships and to summarize the data in novel ways that are both understandable and useful to the data owners" [13].

Over the last few years the techniques of data mining have been pushed by three major technological factors which have advanced in parallel. First, the growth in the amount of data has led to the development of mass storage devices. Second, the problem of accessing these information has led to the development of advanced and improved processors. Third, the need for automating the tasks involved in data retrieval and processing, which led to the development advancements in statistic and machine learning algorithms. This section briefs about data mining models and techniques and also discusses the issues related to mining remote sensing images.

2.2.2 Data mining models & techniques

Data mining models

A model is a high-level, global description of a data set [13]. It takes a large sample perspective. The models can be categorized into two major divisions: one is a descriptive model which is used for summarizing the data in a convenient and concise way and second is an inferential or predictive models which allows one to make some statements about the population from which the data were drawn or about likely future data values [13]. Further data mining methods often do not focus on the appropriateness of the model of the data, namely the goodness of fit. Before finding the best model in a given class of models, it is important to determine the class of models that best fits the data [14]. In order to arrive at appropriate class of models, one need to understand the data. Table 2.1 gives an idea about the types of modeling techniques that are available.

Table 2.1: Types of modeling (modified from Hand et.al., (13))

Modeling		
Predictive		Descriptive
Classification	Regression	Probability distribution and density functions
The Perceptron	Linear Models	
Linear Discriminants	Generalised Linear Models	Probabilistic Model-Based Clustering Using Mixture Models
Nearest Neighbour Methods	Non-parametric "Memory Based" Local Models	
Tree Models		
Logistic Discriminant Annalysis	Artificial Neural Networks	Clustering Algorithms
Native Bayes		

In this research the combination of both predictive and descriptive modeling were used.

Table 2.2: Data mining Tasks and Techniques (modified from Miller. H. J. et.al., (12))

Segmentation	Clustering: It is the process to determine a finite set of implicit classes that describes the data Classification: The process of mapping data items into pre-defined classes	Cluster Analysis, Bayesian Classification, Decision or classification trees, Artificial Neural Networks
Dependency Analysis	Determining Rules to predict the value for some attribute based on the value of the other attribute	Bayesian Networks, Association Rules
Outlier Analysis	Finding data items that exhibit unusual deviations from expectations	Clustering and other data mining methods, outlier Detection
Trend Detection	The task of summarizing the database through lines and curves, often over time	Regression, Sequential pattern extraction
Characterization	Compact Descriptions of the data	Summary rules, Attribute-oriented induction

Data mining tasks

Data mining tasks can be broadly classified into five categories based on their tasks i.e. segmentation, dependency analysis, outlier analysis, trend detection and characterization. In order to do these tasks, various techniques such as cluster analysis, neural networks, genetic algorithms, Bayesian networks, decision trees, etc., are available. Some of these techniques are also good at executing more than one task and have their own advantages and disadvantages. Table 2.2 provides an overview on the list of data mining tasks and their techniques that are available.

In this research, the tasks defined are the trend detection and pattern analysis. The patterns of fire over space are defined, extracted and analyzed. Techniques of regression are then used to explore the trend over time.

2.2.3 Issues related to remote sensing data mining

Data mining in image databases is similar to automated image processing. One of the main differences is that, in the case of data mining a very large amount of data needs to be processed in order to retrieve information or hidden knowledge, while image processing usually concentrates on the analysis of a single or few images [15].

Processing methods, originally designed for relational data structure and image data cannot be used for remote sensing imageries. Characteristics of remote sensing imageries such as dependency within observations, data uncertainty, non-stationarity, non-linearity, high levels of multivariateness and time

interacting with space makes them different from general category of data [5]. Therefore, mining patterns from Earth Science data is a difficult task due to this very spatio-temporal nature of the data [2]. There is also a myth that data mining is universally useful [5] but building such a large, centralized system to serve diverse user communities is also expensive and difficult to implement [1]. There are various issues attached to the image processing in general such as nature of data, quality of data and location of data and software used for processing. These are briefly discussed with respect to Meteosat data in the subsequent sections.

Nature of the data

As in the case of Meteosat imagery, the primary mission is continuous observation of the earth's full disk. This is achieved with the Spinning Enhanced Visible and Infrared Imagery (SEVIRI) imaging radiometer. SEVIRI is a 12-channel imagery observing the earth-atmosphere system. Eleven channels observe the earth's full disk with a 15-min repeat cycle. A high resolution visible (HRV) channel covers half of the full disk in the east-west direction and a full disk in the north-south direction. The high resolution visible channel has a spatial resolution of 1.67 km. The spectral characteristics of the SEVIRI sensor is listed in the table 2.3.

Table 2.3: Number of bands and their spectral characteristics of the MSG SEVIRI sensor (7)

Channel No	**Name**	$\lambda_{central}$ μm	$\lambda_{minimum}$ μm	$\lambda_{maximum}$ μm	**Main gaseous absorber or window**
1	VISO 6	0.64	0.56	0.71	Window
2	VISO 8	0.81	0.74	0.88	Window
3	NIRI 6	1.64	1.50	1.78	Window
4	IR 3.9	3.90	3.48	4.36	Window
5	WV 6.2	6.25	5.35	7.15	Water vapor
6	WV 7.3	7.35	6.85	7.85	Water vapor
7	IR 8.7	8.70	8.30	9.10	Window
8	IR 9.7	9.66	9.38	9.94	Ozone
9	IR 10.8	10.80	9.80	11.80	Window
10	IR 12.0	12.00	11.00	13.00	Window
11	IR 13.4	13.40	12.40	14.40	Carbon dioxide
12	HRV Broad-band		0.4	1.1	Window Water Vapor

The Data Acquisition and Dissemination Facility (DADF) collects together the various types of data required for dissemination through the satellite via Low Rate Information Transmission (LRIT) and High Rate Information Transmission (HRIT) services. Further there are a number of meteorological products that are generated from the Meteosat data. Some of the products are cloud mask and cloud amount, cloud type (including fog), cloud top temperature / height, convective rainfall rate, high resolution wind vectors from HRVIS, rapidly developing thunderstorms, Air mass analysis etc.,

The quality and uncertainty of the data

The quality and the uncertainty involved in data plays an important role in the knowledge discovery process. Even though it has little to do with processing, it plays a major role in the interpretation of the results. The quality and uncertainty that is present in the data and the model results, need to be understood before any further decisions are taken. Moreover accuracy of results is not always important in case of data mining [11] but understanding it plays an important role.

Data quality deals with understanding the limitations of data and probable errors that are embedded in it. The quality of remote sensing imagery generally revolves around the scale. It is defined as the ratio of distance on map/image to distance on ground. It is also termed as 'spatial resolution', and relates directly to size of the smallest feature recorded on the ground [16]. In case of Meteosat imagery a high resolution visible (HRV) channel covers half of the full disk in east–west direction and a full disk in north–south direction. The high resolution visible channel has a spatial resolution of 1.67 km, as the oversampling factor is 1.67 and the sampling distance is 1 km at nadir. The corresponding values for eight thermal IR and three solar channels are 4.8-km spatial resolution, with an oversampling factor of 1.6 that corresponds to a sampling distance of 3 km for nadir view. The instantaneous field of view (IFOV) corresponds to the area of sensitivity for each picture element. Since the aperture angle for each IFOV is constant, it follows that the corresponding area at surface varies with satellite-viewing angle. A complete image, that is, full disk of the earth, consists of nominally $3712 * 3712$ pixels for channels 1–11. The HRV channel covers only half of full disk in E–W direction and therefore a complete image consists of $11136 * 5568$ pixels.

Omnipresence of uncertainty requires us to be able to cope with it: modeling uncertainty is a necessary component of almost all data analysis [5]. Indeed, in some cases our primary aim is to model the uncertain or random aspect of data. Failure to recognize uncertainty, whatever its source, may lead to erroneous and misleading interpretations [17].

There are three main factors of uncertainties involved in the process of modeling fire from Meteosat data. First is the uncertainty related to the selection of suitable bands. The studies by Chilar et.al., [8] and Giglo et.al., [19] state that that fire can be detected by using bands of low wavelength i.e. $3.9\mu m$ band 4 in the case of SEVIRI sensor but there is not much work done on the type of fire and its thermal reflectance for the SEVIRI sensor. Second factor is resolution of imagery. Figure 2.1 shows pixel resolution of SEVIRI sensor. The final product which is disseminated is a resampled imagery of 3km x 3km. Therefore even though a fire is detected in certain pixels, the exact location and spread cannot be quantified due to low resolution of the sensor. Let us consider the phenomenon occurring over a pixel at one point: in case of Meteosat resampling is done at 3km. Since fire is dynamic in nature, there is a possibility of spectral mixing happening within pixels and identification of center of fire or for that matter exact pixel within which it occurs involves uncertainty. Thirdly, it is uncertainty related to validation of results. There is no ground truth or any other equiva-

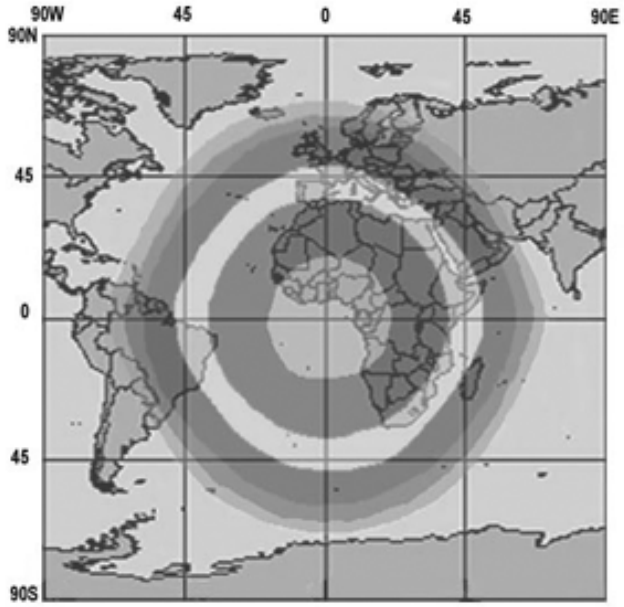

Figure 2.1: Illustration of SEVIRI multi-spectral image ground resolution (equivalent surface). The bands show the decrease in pixel resolution away from the sub-satellite point, 3.1km pixel resolution (inner circle), 4km, 5km, 6km, 8km and 11km (outer band) (18) (18)

lent source of information available for the case under study. This makes the validation of the movement of fire difficult.

Location of data

Location of the image data plays an important role in time taken for processing. Processing time for many machines are directly proportional to location of the data base [13]. Amount of time that would be taken by a computer to fetch data from the storage devices such as RAM (random Access Memory), hard disk, CD-ROM (Compact Disk Read Only Memory), LAN (Local Area Network), WAN (Wide Area Network), magnetic disks & tapes, etc and analyze it vary to a greater extent. Even though in terms of processing the best and efficient medium would be RAM it is practically impossible to store all data required for processing within it, since the availability of memory would be considerably less. This problem is aggravated while processing remote sensing imageries since its size and type demands more space. In this research, even though data are archived in servers, the primary source of information is assumed to be stored in a personal computer.

Software

Data mining software facilitates data miners in successful implementation of a desired method, however it doesn't help users in arriving at desired results.

Many software's are designed for specific kind of data base and data structure. Therefore selection of a suitable software should be based on, data structure and task at hand. Some of the spatial data mining software are listed below

- ENVI [20] is commercial software developed by Research systems Inc. It is a software for visualization, analysis, and presentation of digital imagery. The IDL data miner is the extension of these facilities provided by the software to address the issue of data mining
- S-PLUS [21] software is good at exploratory data mining and statistical analysis. The spatial package S+SpatialStats is good for analyzing spatially correlated data
- R [22] is an open source freeware available for linux, mac and windows. It has a range of packages contributed by various people. In general R has many of the functionalities of S-PLUS except for a graphical user interface. Recent addition of spatial packages and connectivity through GDAL has made R capable of handling remote sensing imageries
- Spatial Statistics toolbox for Matlab [23]. This is a toolbox developed for advanced spatial analysis. This toolbox contains public domain spatial software written in Matlab (Matlab Spatial Statistics Toolbox 2.0) capable of estimating very large spatial autoregressions.

There are also other software's such as Terraseer, NEM, GeoMiner, SPIN, etc., available for spatial data mining. In this research 'ENVI' was used for visualization and exploratory data analysis, 'R' was used for mining and 'ArcScene' was used for presentation of results. 'R' was selected since it was an open source and also provides a strong platform for executing statistical and mathematical functions.

2.3 The Fire

2.3.1 Introduction

Concerns for environment are increasing interest in monitoring and predicting ecosystem changes [9]. Southern European climates are predominantly Mediterranean, i.e. characterized by hot dry summers and cool wet winters. Areas of rugged terrain are abundant, and its natural vegetation is typically evergreen, resistant to drought. These sets of environmental conditions makes the region prone to wild land fires [24]. Socioeconomic and demographic trends that have prevailed in rural areas of this region during last four decades further reinforces its susceptibility to fire. Many rural areas have experienced substantial population decreases during second half of the 20th century, leading to abandonment of agricultural lands, decrease in sizes of herds and reduction in consumption of forest fuels [24].

During the year 1980-2000 there have been many fires in Europe. These fires have led to devastation of many hectares of land [25] (refer graphs 2.2).

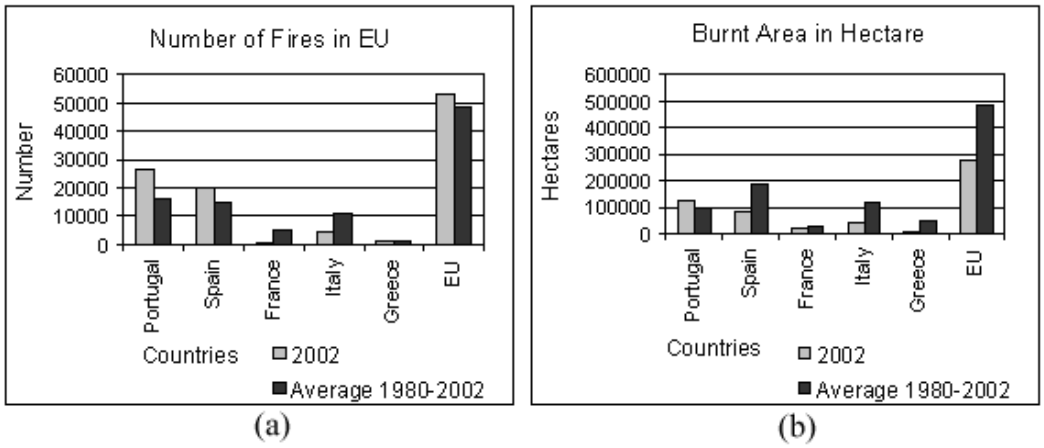

Figure 2.2: Number of fires and the burnt area in hectares for the years 1980-2002 in the EU Mediterranean countries (modified from forest fire in Europe, bulletin) (25)

2.3.2 Effects of fire

Extensive burning of forests impairs biodiversity and encourages soil erosion [26]. Fires influence climate on regional and global scales. Changes in land use have significant effects on heat and water inventories of ground and of lower levels of the atmosphere. For example, a change in land use affects reflection of solar irradiance, temperature and emissivity [26]. Further greenhouse and trace gases, causes of changes in the chemical processes taking place in the atmosphere, are products of combustion [27] [28]. Fires also increase load of soot and organic particles in the atmosphere leading to more clouds but less rain and a reduction in amount of solar irradiance reaching the surface of the earth. Vegetation fires have effects similar to those of large-scale coal fires on the chemistry of the atmosphere (for example, in China) [29]. To understand these effects of fire there is a need for monitoring of fire behavior. This requires proper detection methods and may lead to a reduction of impacts. Magnitude and importance of ecological damages and economic losses caused by wildfires in Europe has led to the development of a large and expensive infrastructure designed to support prevention, detection and fighting against fires [24]. Since vegetation fires are associated with various ecological and physical processes responsible for radiometric signals these can be remotely sensed [24].

2.3.3 Remote sensing and fire

Remote sensing satellites are widely used for fire prediction, detection, monitoring and assessment. Many satellites like NOAA, Landsat, TERRA, AQUA, ERS, DMSP, SPOT, BIRD, GOES, etc., are being currently used for this purpose. The satellites AQUA & TERRA carry the MODIS (Moderate Resolution Imaging Spectroradiometer) sensor are used for collecting Thermal Anomalies data. This includes fire occurrence (day/night), fire location, the logical criteria used for the fire selection, and an energy calculation for each fire. Over the past

few years there have been many models and algorithms developed to detect, predict and describe the fire characteristics. Some of them are discussed in the forthcoming sections.

Detection

Detection addresses location of forest fire. Almost 90% of forest fires are detected by local people [30]. Fires that occur in unpopulated landscapes are difficult to detect. These fires can be most devastating since they are detected at a later stage. By the time fire fighters arrive, a large devastation has already taken place. One method for detecting fires in such isolated places is through use of remote sensing. For example, in Finland the fire service crosses the country twice everyday to detect unusual thermal activities. Most of fires however go undetected due to low frequency of their survey [30]. Furher this method of surveying may not be possible for countries like Spain and Portugal as it is too costly. Therefore these countries rely upon the satellite sensors for detection of fire. The Fire Monitoring, Mapping, and Modeling System (Fire M3) [31] is an initiative of the Canada Centre for Remote Sensing (CCRS) and the Canadian Forest Service (CFS), agencies of Natural Resources Canada to tackle similar issues. Goals of Fire M3 are to use low-resolution satellite imagery to identify and locate actively burning fires on a daily basis; to estimate annual area burned; and to model fire behavior, biomass consumption, and carbon emissions from fires.

Similarly, the experimental Wildfire Automated Biomass Burning Algorithm (WFABBA) is a product by University of Wisconsin [32]. It is currently generating half-hourly fire data for the Western Hemisphere. Geostationary NOAA weather satellite GOES-12 provides coverage for North and South America while GOES-10, covers North America only. Results from the WFABBA are typically available within 90 minutes of satellite scan time. Data from the GOES satellites and a landcover map derived from 1-km resolution Advanced Very High Resolution Radiometer (AVHRR) data are used to produce two categories of images. Continental view images of plotted locations of fires and regional view images indicating individual satellite fire pixels. Fires from the WFABBA are divided into six categories: processed fire, saturated fire pixel, cloudy fire pixel, high possibility fire pixel, medium possibility fire pixel, and low possibility fire pixel. Data noise, extremely hot surfaces and cloud shadows do give false fire alarms.

The Advanced Fire Information System (AFIS) [33] is currently being developed as a service module of the Wide Area Monitoring Information System (WAMIS). This is expected to deliver fire information products to the Fire Protection Agencies and Disaster managers all over Southern Africa in support of effective decision-making in the monitoring of natural and man made fires over the Southern African Development Community region. The model uses MODIS and MSG SEVIRI sensors for hot spots detection, thereby providing a continuous picture of active fires in Southern Africa. Each fire detected is represented spatially within 30 minutes of a satellite overpass on a Web GIS system. The MODIS sensor provides fire locations every 5–6 hours while the MSG data will

be updated every 30 minutes covering all of Africa.

Predicting

The most important aspect after fire detection is predicting possible movement of the fire. This becomes very important in evacuating people who reside in possible danger zones and in helping fire fighters in putting out the fires. Ameghino et.al., [34], have developed Cell-DEVS (based on cellular automata) for studying the spread of fire. This model takes into consideration many complex parameters for calculating spread of fire. Muzy et al [9] have compared the simulation methods of Cell-DEVS and a comparable method for fire spreading across a fuel bed and concluded that Cell-DEVS is more safe and cost effective. There are also other models such as ALOFT-FT^{AM} (A Large Outdoor Fire plume Trajectory model - Flat Terrain) developed by the Building and Fire Research Laboratory [35] to help in predicting fire movement.

Fire characteristics

Lambin et al [36], tried to analyze the remotely sensed indicators of burning efficiency of savanna and forest fires. In this study they tried to analyze timing of fire with respect to vegetation senescence. This study concluded that, rate of post-fire recovery of vegetation and spatial pattern of burning are potential indicators of burning efficiency. They also analyzed NDVI of areas before and after fire occurrence and discovered that fires affect areas with lowest NDVI and highest surface temperature. Giglio and Justice [37] studied effects of wavelength selection on characterization of fire size and temperature. They demonstrated that for realistic wildfires (composed of both flaming and smoldering components) the location of 4 and 11 μm channels can cause large differences in fire temperatures.

Cheikh Mbow et al [38], tried to use the spectral indices and simulation of savanna burning to assess risk of intensive fire propagation within a national park. They developed a model comprising of Fire Risk Assessment Algorithm (FIRA) and Fire Simulation algorithm (FARSITE). FIRA helped in predicting potential danger zones where probability of fire occurrence during a season is high and low. They used Land cover map and Topography along with other information such as fuel model to analyze effect of fire and to study its propagation. Koutsias and Karteris [39], have tried to study fuel complexes that favor fire occurrences and spread in the Mediterranean-type climate. Forest fuel mapping was done using the Landsat-TM and GIS layers to arrive at particular delineated forest types leading to a particular fire behavior.

ESA missions

Most of the sensors and satellites that are used for detection of fires and identification of their characteristics have some drawbacks since they are not designed for hot spot (fire) investigation. Moreover except for GOES and MSG the rest of the satellites are polar orbiting satellites that do not provide high temporal resolution data needed for active fire monitoring. In order to overcome these

limitations, experimental missions such as BIRD are implemented for better understanding of fire through remote sensing. Based on the BIRD, mission in near future ESA is planning to define new missions such as CHABLIS (for estimating forest damage at 1 hectare accuracy using SAR coherence products combined with Land Cover classification derived from SPOT images), REMSAT (to demonstrate use of real-time satellite communications, position determination, Earth Observation and meteorological services during Forest Fire Emergency situations), FUEGO (to develop a space-based system for service of forest fire fighting management teams by means of a dedicated constellation of mini satellites) and Fire Alarm system in Finland (to prototype an operational satellite-based real-time system to observe and alert of forest fires in Finland) for real time monitoring of fire activities in Europe.

In short, there are no existing satellites apart from MSG that could be used for continuous monitoring of fire activities in Europe and Africa. Furthermore high temporal nature of MSG could help in characterizing the dynamic nature of forest fires. This research exemplifies upon this rationale and explores the potential of MSG system for monitoring forest fires using techniques of data mining.

2.4 Summary

To summarize, problem of fire still exists and is of serious concern. Even though many algorithms have been developed for fire, lack of timely monitoring and detection using remote sensing is still a concern. Based on inspiration from GOES detection algorithm and Canada's M3 project, this research aims at developing a model using high temporal and low spatial resolution imagery i.e. Meteosat Second Generation (MSG) for monitoring and analyzing the behavior of fire in space and time. Temporal resolution of Meteosat is high (an image every 15 minutes leading to 96 images a day). Owing to this high frequency of data it becomes humanly impossible to analyze Meteosat images manually. At parallel front there are new tools and techniques being developed to address problems of handling large databases. A synergic merger of fire pattern extraction from MSG using techniques of image mining could prove to be fruitful. This research, hence exemplifies upon the concept of developing Meteosat image mining model for forest fire monitoring and fire behavior analysis.

The forthcoming chapter describes the instrumentalization of a image mining model. Forest fire that occurred in Portugal on July 28th, 2004 was taken as a case and diurnal cycle of this event was analyzed using imageries from MSG.

Chapter 3

Mining Fire from Meteosat Imageries

3.1 Introduction

This chapter explains steps that were involved in development of image mining model for forest fire monitoring and analysis from Meteosat Imagery. There were two major events of fire witnessed by the Meteosat imagery in the Iberian peninsula by the end of July 2004. Both fires were devastating, causing a huge loss of vegetation. Forest fire that occurred in Portugal during 28^{th} July, 2004 was taken as a case and its event analyzed. Since more than 96 images were used, many processes within the model had to be automated to facilitate in processing. This chapter describes the steps implemented for monitoring and analyzing patterns of fire in space and time. This chapter is divided into five sections based on sequence of steps involved in the mining. Section 3.2 is exploratory data analysis. This section describes the process of understanding the data and the event. Section 3.3 is selection of function & model. This section details the process of characterizing fire over space for monitoring. Section 3.4 is extraction of fire objects. This section explains the algorithm used for detecting & extracting the characterized patterns. Section 3.5 explains the procedure involved in tracking of fire i.e. process of correlating patterns extracted in space over time. Section 3.6 details the space-time analysis of knowledge extracted from the previous processes.

3.2 Exploratory Data Analysis

Exploratory data analysis can be described as data-driven hypothesis generation where an examination of the data is performed in search of structures that may indicate deeper relationships between cases or variables. This is in contrast to hypothesis testing, which begins with a proposed model or a hypothesis and undertakes statistical manipulations to determine the likelihood that the data arose from such a model [13]

The process of data analysis is divided into two subsections. First part deals with collection or retrieval of data and second part deals with its visual analysis.

3.2.1 Data acquisition

In this research GDAL (Geospatial Data Abstraction Library) was used to access files from main server (ITC Meteosat Image server, where the images are received from the satellite and are archived). GDAL is a translator library for raster geospatial data formats that is released under open source license. It consists of a series of utility programs for remote sensing file handling such as reporting information, copying, controlling output format, warping an image, etc. Desired subset of images were converted to ".img" format. Conversion of images also included relative geometric correction and corresponding radiometric corrections (to radiance $mW/m^2/sr/(cm^{-1})^{-1}$ for bands 1, 2 & 3 and to temperature in kelvin for bands 4 to 11). A program in Microsoft Visual Basic incorporating GDAL library was developed to automate this process of retrieval and conversion. Figure 3.1 shows the interface that was developed in Visual Basic to ease the task of converting images. the interface that was developed in Visual Basic to ease the task of converting images.

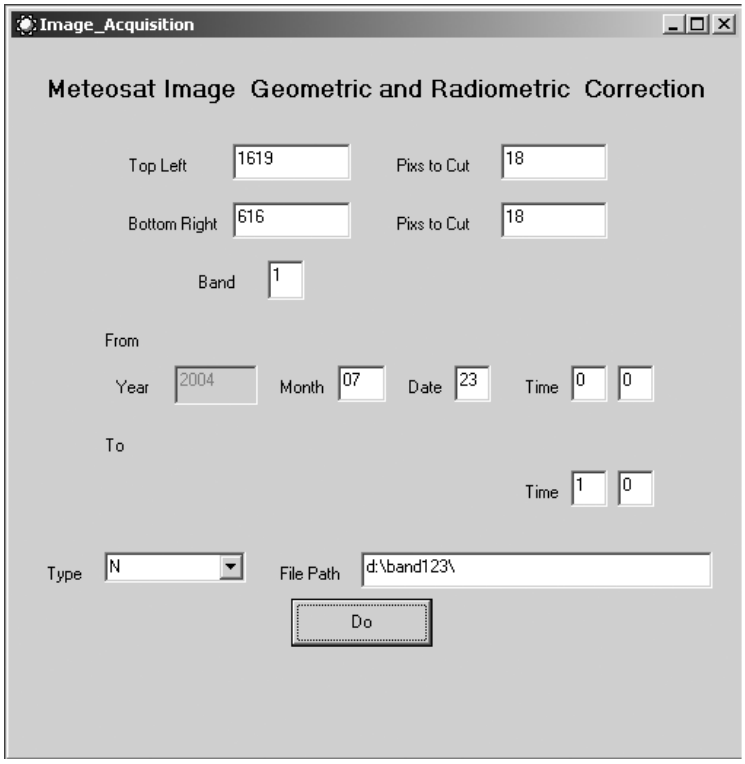

Figure 3.1: Software interface developed in Visual Basic using GDAL for the automation of data acquisition

3.2.2 Data analysis

Exploration of imageries

The image data from the satellites Meteosat, Aqua and Terra were used for the visual analysis. The event of fire was spotted by the MODIS fire detecting Algo-

rithm from the images of Terra and Aqua. These images are shown in figure 3.2. The area that was under fire in Portugal was southern Algarve province. There were two large fires that had burned in a central mountain area. These fires joined into one large fire near the city of Loule during afternoon [40]. There were also four other small fires which were reported in the same region on 28^{th}. These fires were not visible from Meteosat imagery but were spotted by the MODIS fire detection algorithm. Fires in Spain occurred in a large swath of nature reserve in eastern Spain. Fires in Spain and Portugal led to severe loss in vegetation and forced evacuation of people in nearby villages.

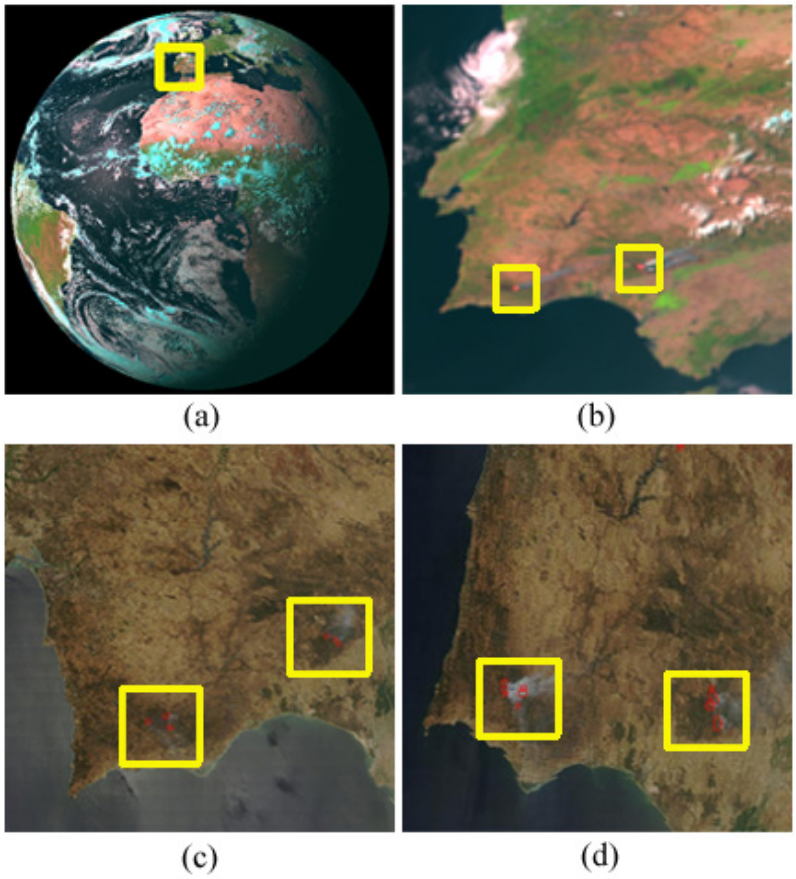

Figure 3.2: (a) & (b) Images acquired by Meteosat in visible bands (3,2,1) at 15:30 hours. (b) & (c) are images acquired by Terra at 11:35 Universal Time (UTC) and Aqua at 13:15 UTC respectively. These images were acquired on 28^{th} of July 2004. The yellow boxes on the image shows the area of active fires that had been detected by the Meteosat, Terra and Aqua

Fires that took place in Spain and Portugal were spotted by the Meteosat satellites. Fire that occurred in Portugal was selected for this study. A subset of $18 * 18$ pixels of Meteosat image covering a region of approximately $54\ km^2$ was selected for this modeling. Care was taken that this event of fire was well contained within this subset for its entire diurnal period (00:00 hours to 23:45 hours). A subset was mainly taken to increase efficiency during process-

ing. Subsequent analysis in this study were done using this subset. A main assumption at this stage was that the developed model would have little or no effect on size of the subset.

Exploration of bands

Spectral profiles of bands were analyzed to identify information or patterns in the data set. Figure 3.3 illustrates spectral profile of thermal bands 4 to 11 of meteosat. From figure 3.3 it is clear that the saturation limit of band 4 (IR window of 3.9μm), band 7 (IR window of 8.7μm) and band 9 (IR window of 10.8μm) are higher than the remaining.

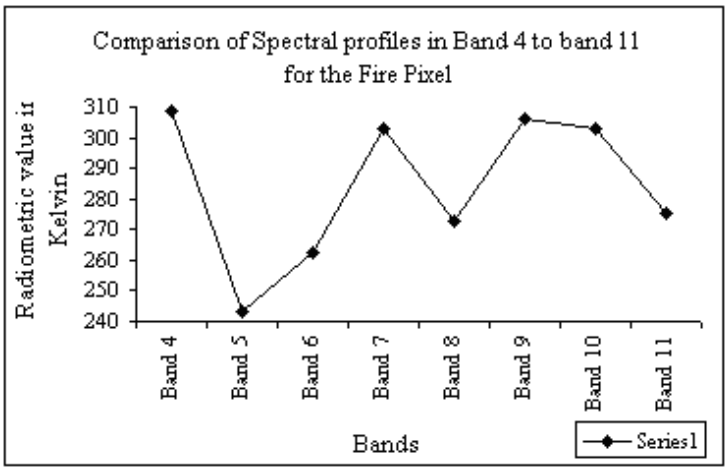

Figure 3.3: Spectral profile of the thermal bands 4 to 11 of meteosat

Figure 3.4 (a) illustrates the image subset of band 4 of area under study. This was an image acquired on 28th July, 2004 at noon 12:15 hours. The bright white regions in the middle of the figure 3.4(a) denote pixels with high temperature. Figure 3.4(b) illustrates change within spectral profile of the pixels for a 24 hour period. The pixels used were from the same area but different dates i.e. 23^{rd} (a non fire day) and 28^{th} (a fire day) of July. Sudden abrupt dips in figure 3.4(b) from the normal are due to errors. These errors were neither normal nor continuous. Furthermore from figure 3.4(b) one can also see variations between non-fire and fire pixels over time. The variation in a non-fire pixel follows a Gaussian distribution with its temperature being low during early hours of morning, gradually increasing and reaching a maximum by noon and then gradually decreasing by night. The variation of fire pixel can be seen to be different. The sudden changes in temperature of the fire pixel were due to the presence and the absence of fire. One can see the sudden drop in temperature during early evening (around 18:00 hour). This is mainly due to the movement of fire from that pixel (pixel which was under study) after sufficient burning.

The patterns in band 4 were characterized in space for monitoring. The next section describes in detail the process using the method of kernel convolution.

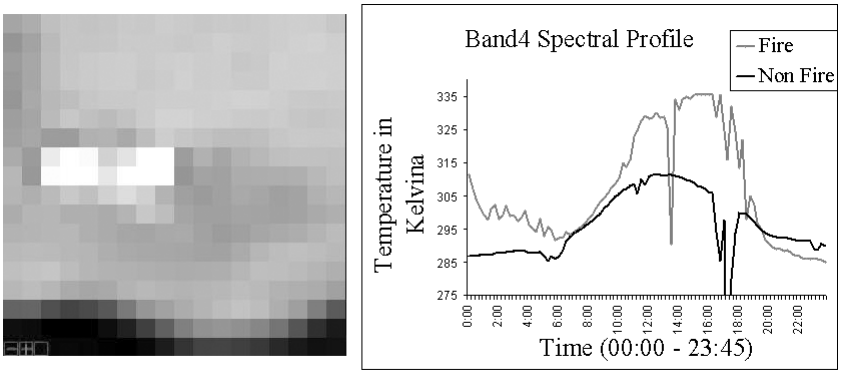

Figure 3.4: Image and spectral profile of band 4

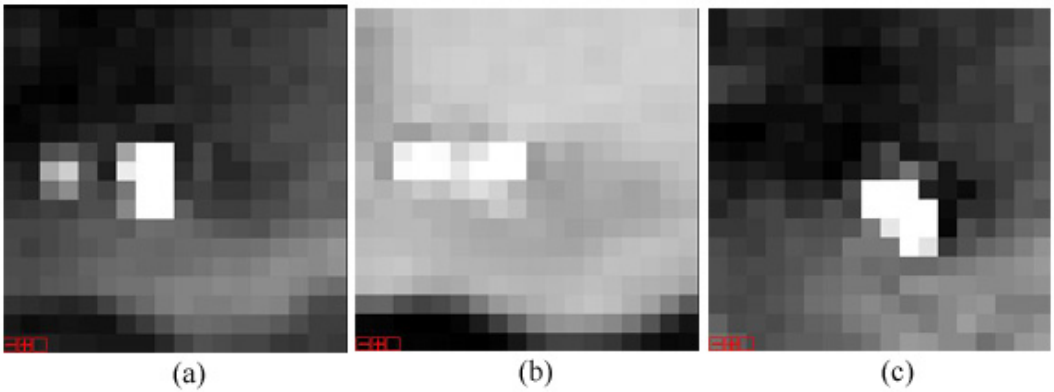

Figure 3.5: Transition of the fire over the space during the diurnal cycle. (a) is the subset of the image taken at time t=1 (00:00 hrs),(b) at time t=49 (12:15 hrs) and (c) at time t=96 (23:45 hrs). The bright white areas in all three images denote the possible areas of fire

3.3 Characterizing Patterns over Space

Pattern Analysis is a means to identify or detect patterns from the given set of data using statistical or non-statistical models. It also deals with the problem of detecting and characterizing relations in the data [41].

Figure 3.5 shows the images taken by the Meteosat in the thermal band 4 on 28th of July 2004. The temperature in Kelvin for the possible areas of fire were relatively high as compared to their surrounding. This can be clearly seen from the images in figure 3.5. The change in location of thermal activity is due to movement of fire over 24 hour period. It is also clear from the images that regions under thermal activity are not clearly defined. This is due to low resolution of the imagery. Furthermore there was no detecting technique available for Meteosat to separate these fire pixels from non-fire pixels. These factors lead to characterize fire pixels over space and classify them based on a functional definition. This was done in two phases, first was selection of a suitable function to describe the characteristics of fire. A Gaussian function was selected to

achieve this task. Second was to describe the patterns of variation over space using that function. This task was achieved using kernel convolution method.

3.3.1 Function

The process of selection of a method and a suitable function can be done either through the experience with various methods or through exploratory analysis of the data. In this research a function was selected based on our understanding of the nature of fire. There were two main assumptions made at this stage. First, heat dissemination by the process of fire is continuous in nature. Therefore the thermal reflectance of one pixel has influence on on its neighboring pixels and vice versa. Second, this influence is Gaussian in nature i.e. temperature is assumed to be transferred from one point to another gradually and characteristics of this dissemination would be Gaussian in nature having its peak at the center of the pixel. Furthermore the rationale was to use a simple function which could characterize fire effectively and would have less cost in terms of processing. Based on above mentioned assumptions and rationale Gaussian function was found to be appropriate. It was adapted to represent location and spread of fire in space. Figure 3.6 schematically illustrates the effect of Gaussian function on pixel characteristics. Instead of the radiance value of a pixel being assumed uniformly throughout pixel area, it is distributed in a Gaussian manner. This leads to gradual dissemination of intensity within pixels.

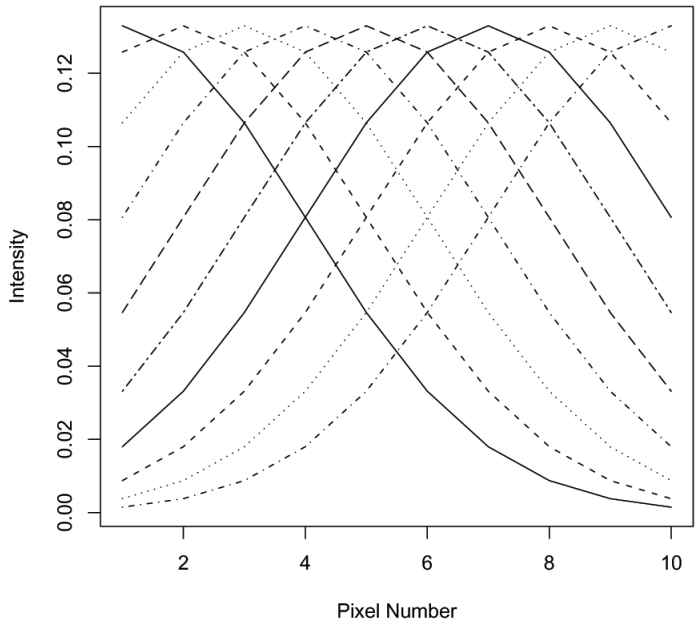

Figure 3.6: Illustration on the effects of values of pixels been considered as Gaussian rather than being uniformly distributed

3.3.2 Method

Methods such as splines, kernels, loess and kernel convolutions were explored to represent the data (these methods and their results are described in detail in subsection 5.3.2 on page 48). From analysis of results from these methods, method using kernel convolution were able to better represent the data and the phenomenon. Therefore kernel convolution methods were selected for describing patterns over space.

Kernels

Kernel smoothing refers to a general class of techniques for non-parametric estimation of functions. The kernel is a smooth positive function $w(z, h)$ which peaks at 0 and decreases monotonically as z increases in size. The smoothing parameter h controls the width of the kernel function hence the degree of smoothing applied to the data. As the smoothing parameter increases, the resulting estimate misses some details in the curvature of the data. As the smoothing parameter decreases, the estimator begins to track the data too closely and will end up interpolating the observed points [42].

One can define Kernel as a function k that for all $x, z \in X$ satisfies [41]

$$k(x, z) = \langle \phi(x), \phi(z) \rangle, \tag{3.1}$$

where ϕ is a mapping from X to an (inner product) feature space F

$$\phi : x \longrightarrow \phi(x) \in F \tag{3.2}$$

The degree of smoothing to be performed by the kernel is defined by its bandwidth h. The value of h increases as the degree of smoothing increases and vice versa. When the value of h starts to decrease beyond a certain threshold the degree of smoothing starts to overfit the data points leading to interpolation between the values. Similarly when it starts to increase beyond a certain threshold it tends to over fit the data, finally leading to a straight line. Therefore the selection of a suitable bandwidth is very important for describing the patterns. In this research the bandwidth was selected after several trials. The effect of bandwidth h on the data points is illustrated in figure 3.7

Convolution [43]

The convolution of f and g is written as $f * g$. It is defined as the integral of the product of the two functions after one is reversed and shifted.

$$(f * g)(t) = \int f(\tau)\, g(t - \tau)\, dt \tag{3.3}$$

If X and Y are two independent variables with probability densities f and g, then the probability density of the sum $X+Y$ is given by the convolution $f*g$. For discrete functions, one can use a discrete version of the convolution . It is given by the equation 3.4. The process of convolution is illustrated graphically in the figure 3.8

$$(f * g)(m) = \sum_{n} f(n)\, g(m - n) \tag{3.4}$$

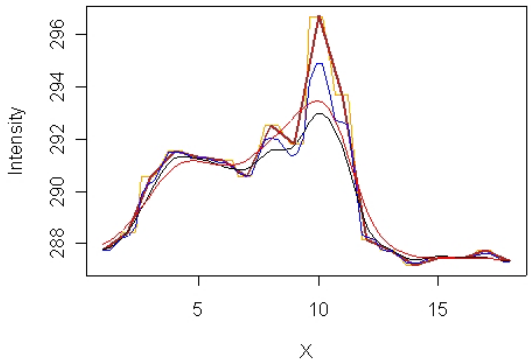

Figure 3.7: Illustration of the effect of bandwidth on a single dimensional data frame

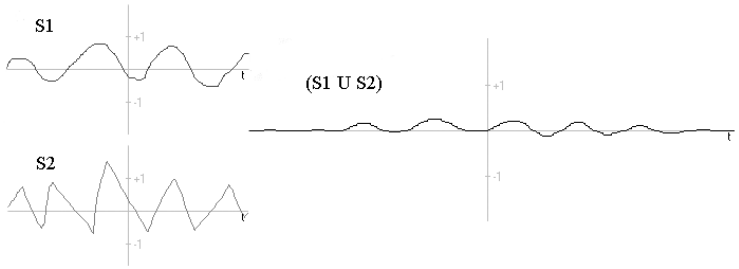

Figure 3.8: Illustration of the process of convolution

Kernel convolution modified from higdon [44]

A Gaussian process over R^d is to take i.i.d. Gaussian random variables on a lattice in R^d and convolve them with a kernel. Figure 3.9 shows an example using a Gaussian kernel to convolve i.i.d. Gaussian noise. The process involves successive increase in the density of the lattice by a factor of 2 in each dimension and reducing the variance of the variates by a factor of 2^d leads to a continuous Gaussian white noise process over R^d. The convolution of this process can be equivalently defined using some covariogram in R^d. The process of convolution gives very similar results to defining a process by the covariogram. Nevertheless the convolution construction can be readily extended to allow for non-standard features such as non-stationarity, edge effects, dimension reduction, non-Gaussian fields, and alternative space-time models.

Let $y_{(1,1)}, \ldots, y_{(i,j)}$ (where q is a two dimensional matrix of $(1,1), \ldots, (i,j)$) be data recorded over the two dimensional spatial locations $s_{(1,1)}, \ldots, s_{(i,j)}$ in S. In this research the spatial method represents the data as the sum of an overall mean μ, a spatial process $z = (z_{(1,1)}, \ldots, z_{(i,j)})^T$, and Gaussian white noise $\epsilon = (\epsilon_{(1,1)}, \ldots, \epsilon(i,j))^T$ with variance σ^2_ϵ,

$$y = s + z + \epsilon \tag{3.5}$$

Where the elements of z are the restriction of the spatial process $z(s)$ to the two dimensional data locations $s_{(1,1)}, \ldots, s_{(i,j)}$. $z(s)$ is defined to be a mean zero Gaussian process. But rather than specify $z(s)$ through its covariance function, it is determined by the latent process $x(s)$ and the smoothing kernel $k(s)$. The

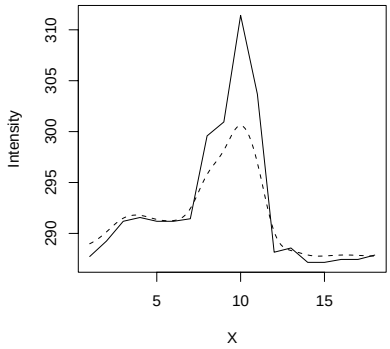

Figure 3.9: Result of smoothing performed using kernels for single dimensional data. The dotted line shows the result of the smoothing

latent process $x(s)$ is restricted to be nonzero at the two dimensional spatial sites $\omega_{(1,1)}, \ldots, \omega_{(a,b)}$, also in S and define $x = (x_{(1,1)}, \ldots, x_{(a,b)})^T$ where $x\omega_p = x(\omega_p); p = (1,1), \ldots, (a,b)$. Each x_p is then modeled as independent draws from a $N(0; \sigma_\epsilon^2)$ distribution. The resulting continuous Gaussian process is then

$$Z(S) = \sum_{p=(1,1)}^{(a,b)} x_j k(s - \omega_p) \tag{3.6}$$

Where $k(\bullet, -\omega_p)$ is a kernel centered at ω_p. This gives a linear model

$$y = \mu l(i,j) + Kx + \epsilon \tag{3.7}$$

Where $l(i,j)$ is the $(i,j)^{th}$ vector of $l's$, the elements of K are given by

$$K_{pq} = k(s_p - \omega_q) x_q, \tag{3.8}$$

$$x \sim N\left(0, \sigma_x^2 I_{(a,b)}\right) and \tag{3.9}$$

$$\epsilon \sim N\left(0, \sigma_\epsilon^2 I_{(i,j)}\right) \tag{3.10}$$

This results in a basic mixed effect model. The same method could also be extended to incorporate further dimensions. This research concentrates on only the two dimensional processes. The figure 3.10 shows the result of the kernel convolution. Figure 3.10(a) shows the image of $18 * 18$ pixels before convolution and figure 3.10(b) shows the image after convolving the image using kernels to a lattice grid of $144 * 144$.

3.3.3 Optimizing processing efficiency [45]

Discrete kernel estimation of $\psi_{(a,b)}$ requires $O(n^2)$ kernel evaluations which makes its computation very expensive for large sample sizes. One could also

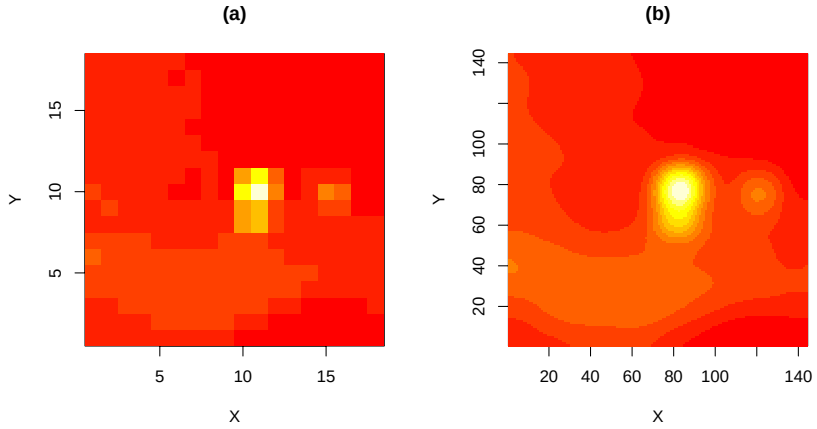

Figure 3.10: The image before and after convolution

use Fourier transform methods to compute the required convolution. The discrete Fourier transform of a complex vector $z = (z_0, \ldots, z_{n-1})$ is the vector of $Z = (Z_0, \ldots, Z_{N-1})$ where,

$$Z_j = \sum_{l=0}^{N-1} zle^{2\pi ilj/N}, j = 0, \ldots, N-1 \tag{3.11}$$

The vector z can be recovered from its Fourier transform Z by applying the inverse discrete Fourier transform formula.

$$z_l = N^{-1} \sum_{l=0}^{N-1} zle^{2\pi ilj/N}, l = 0, \ldots, N-1 \tag{3.12}$$

Discrete Fourier transforms and their inverse can be computed in $O(NlogN)$ operations using the fast Fourier transform(FFT) algorithm. The algorithm is fastest when N is highly composite such as a power of 2. The discrete convolution of two vectors can be computed quickly using the FFT by appealing to the discrete convolution theorem: multiply the Fourier transforms of two vectors element-by-element and then invert the result to obtain the convolution vector.

3.3.4 Automation

The process of kernel convolution using Gaussian function and FFT was applied for all the 96 images (00:00 hours to 23:45 hours) of the study area. This process of automation was done in 'R' environment. A function was developed and implemented to extract the remote sensing imagery, convert them to arrays and then implement the technique of kernel convolution using FFT. The results obtained were then stored in the form of matrices along with the relative coordinates of the images. The function of 'rgdal' and 'fields' were used to facilitate this process. The rationale behind conversion of image to matrix has two facets firstly mathematical and statistical functions could be easily extended to matrices and secondly the efficiency of processing also increases.

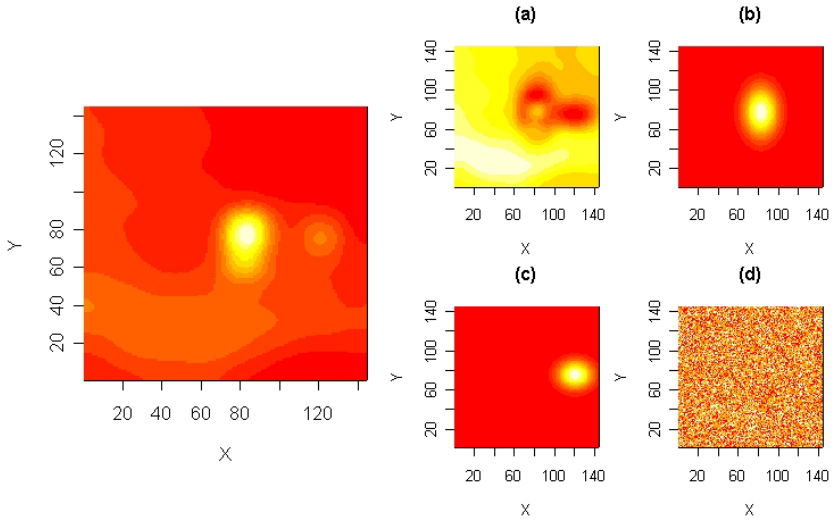

Figure 3.11: From left to right: The actual image and the breakup of the image into background (a), series of functions (b & c) and an error map (d)

3.4 Extraction of Fire Objects

One of the research objectives was to analyze the behavior of fire over a diurnal period. There are not much classification or identification techniques available for detection of fire from the Meteosat imagery. This led to, in defining possible fire pixels as objects and then extracting them from the imageries. In this research fire objects were defined as pixels with relatively maximum thermal radiance and were characterized by a function. The function used was a Gaussian bivariate function with its center to be the center of pixels with maximum thermal radiance.

A separate algorithm was developed and implemented in 'R' to automate this process of defining fire objects and extracting them. This algorithm is similar to that of steepest descent method. It starts at a maximum x & y and starts descending down in directions around it looking for a local minima. The algorithm terminates once these minimas are reached in their respective directions. The results are then stored. From these results, minimum of the value along x and y directions are taken as function of spread for the fire along x & y respectively. The center from which the descending algorithm starts becomes the mean μ in x & y and the maximum temperature becomes the height h. Based on these values a Gaussian bivariate function characterizing the fire is generated. The obtained function is then stretched between the minimum value (i.e. the global minima) and the height h. This resultant function is then subtracted from the main image leading to an image I_{sub} and a function f. Figure 3.11 shows the image before and after extraction. Figure 3.11(d) shows the error obtained while validating the left hand side of the equation 3.13 to its right hand side. The root mean square of the error values were of the order of e^{-14}. Since they were relatively small the error obtained was neglected. This process is summarized in the form of equation 3.13

$$I = I_{sub1} + f(x, y, \sigma_x, \sigma_y, \mu_x, \mu_y, h)_1 - c + \epsilon \tag{3.13}$$

Where C is a constant to be deducted. It is equal to the global minimum of the image and ϵ is the error. The function f is a Gaussian bivariate function. The function is explained in the equation 3.14

$$\frac{e^{\frac{\frac{-1}{2}\left(\frac{x-\mu_x}{\sigma_x}\right)^2 - 2(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y + \left(\frac{y-\mu_y}{\sigma_y}\right)^2}}}{2\pi\sigma_x\sigma_y} \tag{3.14}$$

This process is implemented for n iterations (where $N = 1, \ldots, n$). The value of n is decided by the user based on his/ her assumption or knowledge on the number of fires within a sample grid. In this research this value was restricted to 10 iterations. The final results after I iterations is summarized in the equation 3.15.

$$I = I_{sub_n} + f(k)_1 + \ldots + f(k)_n - C + \epsilon \tag{3.15}$$

Where $C = c_1 + \ldots + c_n$, $\epsilon = \epsilon_1, \ldots, \epsilon_n$, $k = x, y, \sigma_x, \sigma_y, \mu_x, \mu_y, h$. Each function $f(k)$ represents an object of "fire".

Table 3.1: Structure of the database created by the extracting algorithm

Order	**Time**	**X**	**Y**	σ_x	σ_y	**Intensity**
1	1	83	77	12	18	303.7802923
2	1	121	75	12	10	288.7962599
3	1	30	28	14	14	295.7822759
4	1	71	22	20	10	295.0587012
5	1	113	41	12	15	295.0325527
6	1	24	66	12	14	295.0759196
7	1	59	48	16	13	294.920269
8	1	25	124	12	10	296.3451497
9	1	58	117	14	14	295.2642629
10	1	107	121	17	12	296.5363322
1	2	84	78	10	18	300.6093024
2	2	122	77	11	12	291.7070526
3	2	32	27	16	13	294.5862264
4	2	74	21	22	10	293.7112652
5	2	109	44	12	14	293.7975799
6	2	62	50	12	14	294.1885672
7	2	26	61	12	13	294.2938686
8	2	56	86	12	16	294.6569482
9	2	109	114	17	15	295.3650647
10	2	126	24	9	12	295.3029823
1	3	84	77	10	18	298.6431007
2	3	122	76	11	10	292.3857322
3	3	32	24	14	12	293.5848398
4	3	72	22	20	10	293.1329065
5	3	111	43	12	16	292.4199837
•	•	•	•	•	•	•
•	•	•	•	•	•	•
•	•	•	•	•	•	•

This algorithm was implemented to extract fire objects automatically from all 96 images. Table 3.1 shows the structure of database created by the algorithm for a few instances. In the database the objects of fire pixels for a particular time are arranged in descending order of their intensities.

3.5 Tracking of Fire Objects

The extracted objects of fire pixels were then tracked through time for space time analysis. Since ordering of the objects were done based on their intensity, they change with time. There are also instances where the objects split

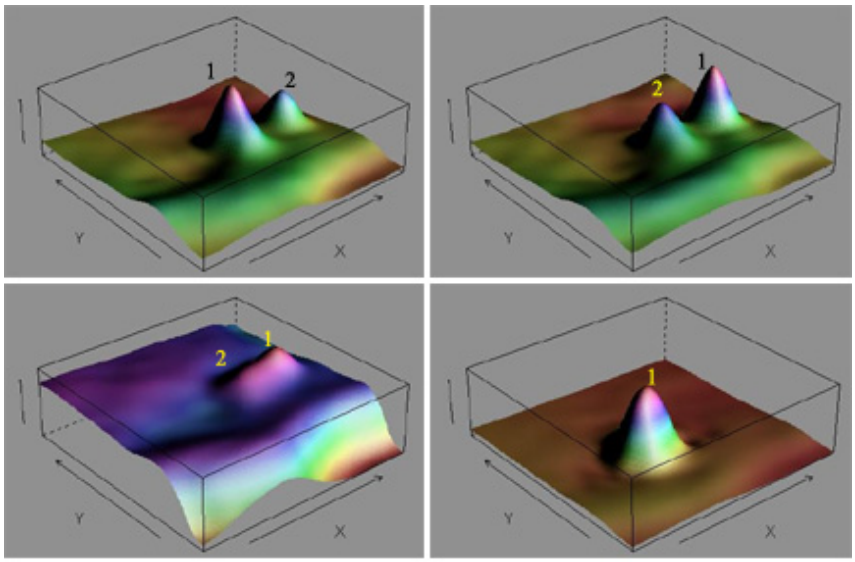

Figure 3.12: Illustration of the order of fire objects varying in time

Table 3.2: Result of the tracking algorithm. The objects marked gray are the result of one of the fire objects that was found to have continuity in time

Time	Obj1 T_n	Obj1 T_{n+1}	Obj2 T_n	Obj2 T_{n+1}	Obj3 T_n	Obj3 T_{n+1}	Obj4 T_n	Obj4 T_{n+1}	Obj5 T_n	Obj5 T_{n+1}
T_1	1	1	2	2	3	3	4	4	0	0
T_2	1	1	2	2	3	3	4	4	5	4
T_3	1	1	2	2	3	3	4	0	5	0
T_4	1	2	2	1	3	3	4	0	0	0
T_5	1	1	2	2	3	0	0	0	0	0
T_6	1	1	2	2	3	0	0	0	0	0
T_7	1	1	2	2	0	0	0	0	0	0
T_8	1	1	2	2	3	3	4	3	0	0
T_9	1	1	2	2	3	3	4	0	5	0
T_{10}	1	2	2	1	3	4	4	4	0	0
T_{11}	1	2	2	1	3	3	4	4	0	0
T_{12}	1	1	2	2	3	3	4	4	5	4
T_{13}	1	1	2	2	3	3	4	4	5	0
T_{14}	1	2	2	1	3	3	4	0	5	0
T_{15}	1	1	2	2	3	3	4	0	5	0
•	•	•	•	•	•	•	•	•	•	•
•	•	•	•	•	•	•	•	•	•	•
•	•	•	•	•	•	•	•	•	•	•

and merge thereby making the process of tracking complicated. Figure 3.12 illustrates this change of order of the objects between various time intervals.

To facilitate initial sorting of the objects from the group of extracted objects, an algorithm was developed and implemented. This algorithm performed tracking based on spatial structure and location of the objects between adjacent time frame. This concept is very similar to that of the nearest neighbor. This algorithm looks for least distance between various objects in adjacent time frames. This brings connectivity of objects in time. The main rationale for considering only the space in this tracking was, due to the nature of the imagery i.e. high resolution of Meteosat in time and low resolution in space. The objects that are having continuity in time based on their location were tracked and their results obtained were stored. This algorithm was made to function for n number of iterations based on user's input on the number of possible fires that could be present within a given time frame. Table 3.2 illustrates the results provided by

this tracking algorithm for one of the objects of fire. These results were then analyzed in 'ArcGIS'.

3.6 Space-Time Analysis

The patterns from the extracted and tracked objects over time represents the knowledge gained from this modeling. This knowledge is in terms of location and characteristics of fire objects over time. These results obtained were then used for further understanding of behavior of fire. The questions that were considered for analysis were,

- how did fire behave in space over time,
- what was the effect of vegetation on fire,
- based on the available data can movement of fire be predicted.

In order to understand the behavior of fire over space and time, the tracked objects were plotted in a space-time cube/ cuboid. The locations of the center of fire objects were plotted against x & y axis and the time against the z axis. Size was used as a visual variable for representing the intensity of fire during various time period. From that resulting trajectory of the objects, interaction between the objects of fire in space and time was analyzed.

The effect of vegetation on the movement of the fire was studied using Normalized Difference Vegetation Index (NDVI) data and landcover data. The NDVI data was generated from the Meteosat imagery using bands 1 ($0.71\mu m$) and band 2 ($0.88\mu m$). The formula used to calculate the NDVI was $(band\ 2 - band\ 1)/(band\ 2 + band\ 1)$. Average of the NDVI values for dates 23^{rd}, 24^{th} & 25^{th} were taken into consideration to reduce uncertainty which could be involved by considering a single imagery. The images at time 12:30 hours were used for NDVI calculation. Apart from the NDVI data, further analysis on movement of fire with respect to vegetation was done using the Corine Landcover Data of Iberian peninsula of the year 2000. Movement of the center of fire over time was superimposed over these vegetation data sets (NDVI and landcover). Patterns of vegetation before fire to patterns of movement of fire was then analyzed.

Finally a linear regression model to predict movement of fire was developed. Additional data of wind for 28^{th} of July, 2004 was also considered in this model. The data of wind was acquired from wind motion vectors (WMV, a product of Meteosat). The WMV data are distributed in BUFR format (Binary Universal Form for the Representation of Meteorological Data). Ciper SoftBUFR software was used to convert these WMV data from BUFR format to text format. A program in Visual Basic was developed and implemented to automate this process of conversion. Figure 3.13 shows the interface developed to extract required information from the BUFR data. The wind and other relevant information for required spatial location from collections of files generated were further extracted. Based on the data of time, location, spread, intensity, NDVI, wind direction and wind speed at time t a simple linear regression model was then

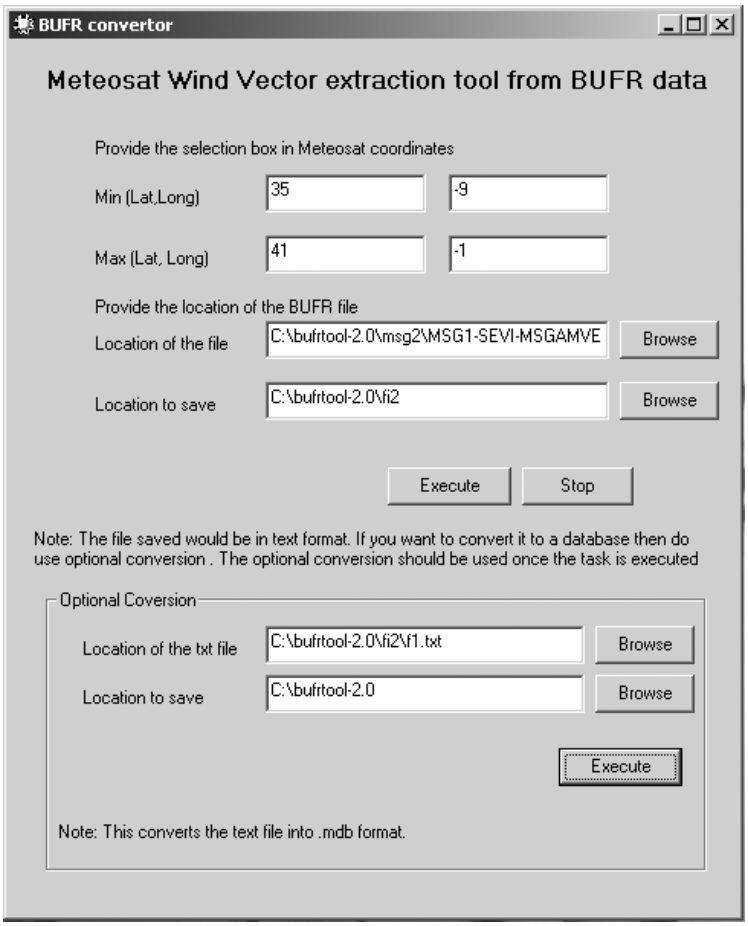

Figure 3.13: Interface of the data extraction software developed in VB

developed to predict the possible location of fire at time t_n. The rationale behind development of this model was to test the fact, whether the dynamics of fire could be modeled.

3.7 Summary

This chapter explained in detail the process of exploratory analysis, characterization of patterns, defining and extraction of fire objects, tracking of the center of fire over time and spatio-temporal analysis of those results with respect to other factors.

Various properties of the Meteosat Image data were visually analyzed with respect to the phenomenon under study. Based on this analysis band 4 was selected since it was found to be suitable for monitoring the event of fire as compared to rest of the bands. The assumption at this point was that if fire could be modeled using one band then methods for inclusion of further bands could be built upon it. This chapter then explained how this selected band was characterized over space using kernel convolution. FFT was used to increase the efficiency of the processing during convolution. The patterns developed were

then fed as an input for extracting algorithm to extract fire objects. The fire objects were first defined as a Gaussian bivariate function. Then possible fire pixels were extracted from the image using image decomposition, where the image is separated into a series of functions (objects) and a background. These extracted objects were then considered as an input for tracking over time. This was facilitated by a tracking algorithm. This algorithm correlated objects with least spatial distance within adjoining time frame leading to a series of objects in time. The patterns of these objects over space and time represent the knowledge gained from the system. This knowledge was then used for further understanding the characteristic behavior of fire. The process of movement of fire with respect to the change in its structure based on time, vegetation and wind were later analyzed.

The task of converting the image into matrices, characterizing, extraction and tracking were completely automated. Separate function were developed and implemented in the environment R to ease the processes. Source code for these algorithms are provided in the annex. Forthcoming chapter presents and explains the results obtained from above mentioned processes.

Chapter 4

Results

4.1 Introduction

This chapter elaborates upon the results obtained from the processes involved in pattern mining. The presentation of the results is divided into four sections. Section 4.2 describes the results of kernel convolution. Section 4.3 describes the results acquired by the extraction algorithm. Section 4.4 describes the results acquired by the tracking algorithm. Section 4.5 describes the results of space-time analysis.

4.2 Results of Characterizing Patterns over Space

Various models such as splines, loess and kernels were initially experimented with, before arriving at the method of kernel convolution for representing the Meteosat data. The results from splines and loess are discussed in the forth coming chapter. This chapter briefly describes the results obtained from the various kernel methods for the Meteosat data under consideration. Four kernels were explored for fitting the data. The sample grid ($18 * 18$ subset), latent grid (same as the sample grid, $18 * 18$) and bandwidth ($h = 0.75$) were kept the same to make the comparison between the methods easier. Figure 4.1 illustrates the actual data. Figure 4.2(a) illustrates the results acquired by the kernel smoothing without convolution. Figure 4.2(b) illustrates the results of kernel convolution method as defined in the function smooth.2d under the package lattice. Figure 4.2(c) illustrates the results of kernel convolution method defined by Bowman [42]. Figure 4.2(d) illustrates the results of the kernel convolution method defined by Higdon [44]. From figure 4.2(a) it was concluded that the kernel method without convolution was not able to characterize the data effectively. Therefore it was not included in the validation. The results of validation for rest of the methods are listed in the table 4.1.

Based on the RMSE results presented in table 4.1 it was concluded that methods using kernel convolution were performing sufficiently well. Even though these methods are different from each other in various ways (such as type and amount of input parameters, design of the kernel, processing time, flexibility, etc) the underlying concept of kernel convolution behind all of them remains

Table 4.1: The error obtained from various kernel convolution methods

Methods	**RMSE**
Kernel convolution from Fields	0.0981
Kernel convolution by Andrian Bowman	0.009
Kernel convolution by Dave Higdon	0.0028

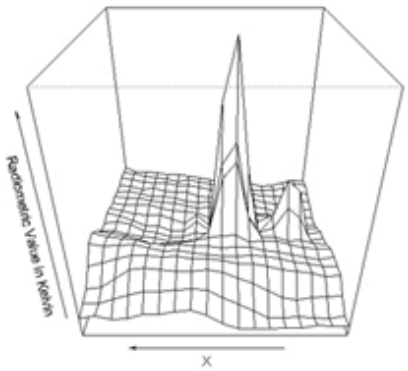

Figure 4.1: Perspective view of the study area. In this view the longitude and the latitude are projected along the x & y axis respectively and the temperature in Kelvin is projected in the z axis

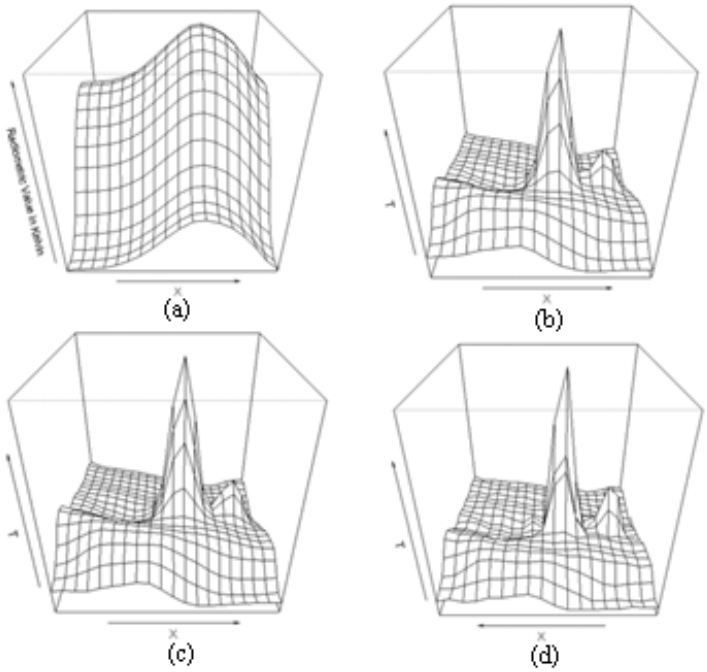

Figure 4.2: Results obtained by convolving the data with four different kernels. (a) is the result of kernel method without convolution, (b), (c) and (d) are results of kernel convolution methods from fields, Bowman and Higdon respectively

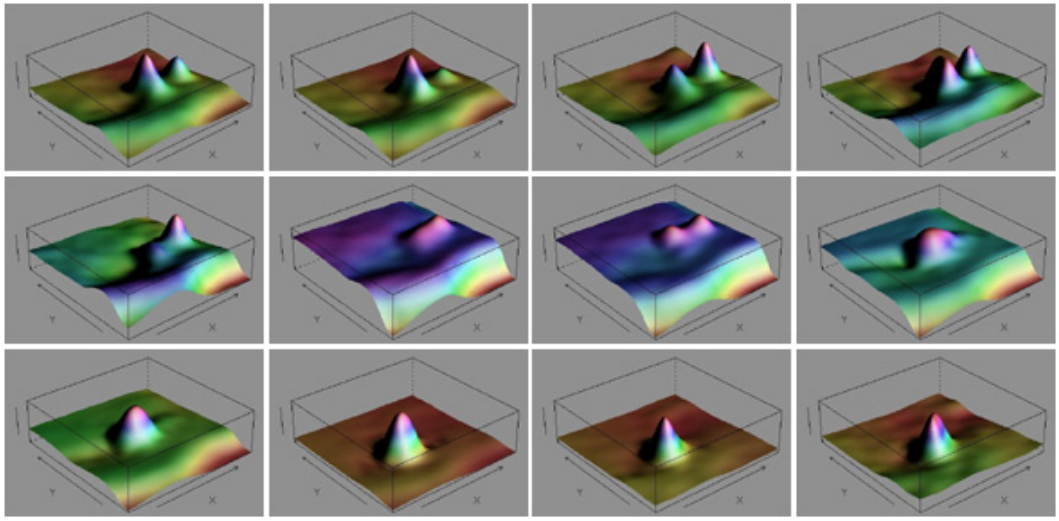

Figure 4.3: Results of kernel convolution applied to a series of images. From top to bottom and left to right the images are from time 00:00, 00:15, 02:30, 05:00, 07:30, 10:00, 12:30, 15:00, 17:30, 20:00, 22:30, 23:45 hours respectively

the same. From these three methods smooth.2d from 'fields' which implements FFT (offers increased efficiency in processing, refer section 3.3.3 on page 27) was selected for characterizing the entire data set.

The parameters that were selected for the final method were; sample grid size $18 * 18$ (size of subset), latent grid size $144 * 144$ (size of the resultant grid) and smoothing function to be Gaussian. Figure 4.3 shows the 3-dimensional view of results of kernel convolution applied to series of images. From these images we can observe the following;

- Fire objects are characterized as Gaussian functions (these objects are seen as peaks with respect to the background).
- Fire objects change in intensity and location with respect to time.
- Two objects of fire are visible during the beginning of the day (00:00 hours). These fires could be later seen merging into one huge fire (15:00 hours).
- Background temperature slowly increases, reaches a maximum around noon (10:00–12:30) and then gradually decreases.

4.3 Results of Extraction of Fire Objects

The extraction algorithm facilitates in automation of extracting possible fire objects from the imagery. They characterize fire as a Gaussian bivariate function. The algorithm starts at global maximum and looks in area around it for a change in slope (from negative to positive). Once that desired limit is achieved the algorithm recreates the fire object using a Gaussian bivariate function. This

function is created using location of starting point $x\&y$ as center, intensity at that point as height h and values of minimum distance at which change in slope was evaluated as spread $\sigma_x\&\sigma_y$. Since fire behaves in a complex manner over plane, a merging and splitting occurs. This behavior of fire was classified into three cases. Case 1 is an instance where there are more than one fire in an image. Case 2 is an instance where, fires are splitting or merging. Case 3 is an instance where a single fire is defined by more than one function. The results of the algorithm for these three different cases are presented in figures 4.4–fig:extractioncase3.

Extracting algorithm was able to characterize the fire objects in terms of functions successfully. This process was carried out based on distance and spread of the functions. Intensity values represented in images are relative within the image (refer equation 3.13 on page 30). From figures 4.4–4.6 it can also be seen that the function do exaggerate the structure of fire objects to a minimal extent this is mainly due to nature of the function used (this is explained further under section 5.4 on page 50). The extracting algorithm extracted 960 objects as functions in space out of 96 images (refer section 3.4 on page 29).

4.4 Results of Tracking of Fire Objects

The center of extracted objects were then tracked over time based on their continuity and spatial correlation. Out of 960 objects extracted by the algorithm (i.e. 10 objects for every time stamp) 185 objects were tracked over time (refer section 3.5 on page 30). Within 185 objects that were tracked 103 objects were classified as fire 1, 61 objects were classified as fire 2 and 21 objects were classified as fire 3. This classification was based on spatial structure of the objects in space-time cube/ cuboid. Figure 4.7 illustrates these objects through a space-time cube. In a detailed analysis of the tracked objects, there were 26 objects missing in between the time intervals. This is mainly due to proximity of the objects within space at same time which led to pair of objects being grouped into a single object by the extracting algorithm.

Figure 4.7 shows objects over space and time. Band 4 image at time 00:00 hrs is used as a background. Size of the objects are used as visual variables for representing varying intensity of the objects over time. From the figure 4.7 one can see that at time 00:00 hours there are two objects in space. These objects merge at approximately 18:00 hours to form a single object. Objects 1 & 2 are fires and this can be easily identified from the imagery, being used as a base map.

The main problem faced by the tracking algorithm was its initialization. The starting point of every fire needed to be fed into the algorithm to track effectively. Therefore the algorithm had to be initialized several times for tracking a total of 185 objects in time. Based on the results a total of 103 objects were classified as fire objects 1, 61 as fire objects 2 and 21 as fire objects 3.

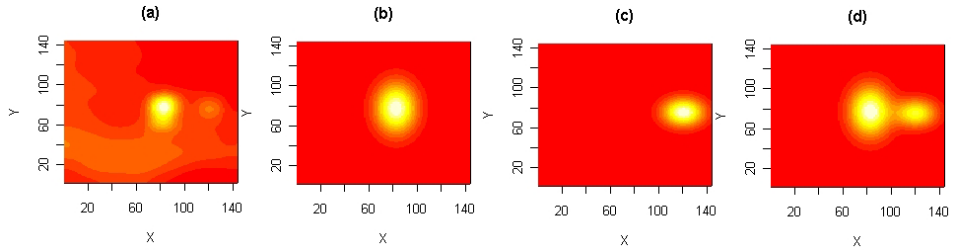

Figure 4.4: Results of extraction algorithm for Case 1 (more than one fire). Color red represents minimum intensity and white the maximum intensity. image (a) shows result of kernel convolution (input for the extracting algorithm). Images (b) and (c) shows the case where two fire objects are represented as Gaussian bivariate functions and image (d) shows result of two Gaussian functions summed together to recreate the fire object

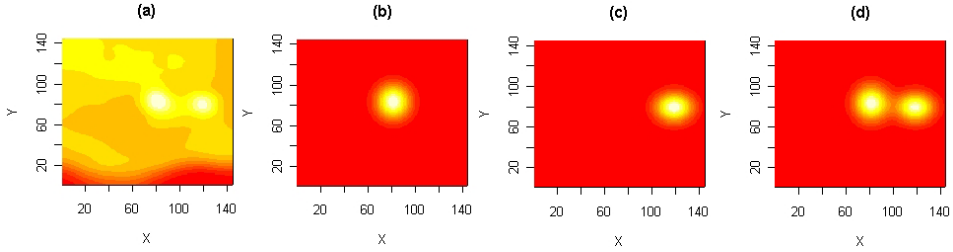

Figure 4.5: Results of extraction algorithm for Case 2 (merging of fire). Color red represents minimum intensity and white the maximum intensity. Image (a) shows result of kernel convolution (input for the extracting algorithm). Images (b) and (c) shows the case where two fire objects are represented as Gaussian bivariate functions and image (d) shows result of two Gaussian functions summed together to recreate the fire object

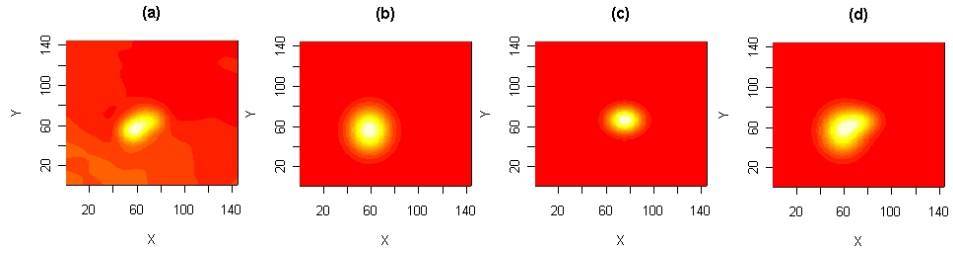

Figure 4.6: Results of extraction algorithm for Case 3 (single fire represented as multiple functions). Color red represents minimum intensity and white the maximum intensity. Image (a) shows result of kernel convolution (input for the extracting algorithm). Images (b) and (c) shows the case where two fire objects are represented as Gaussian bivariate functions and image (d) shows result of two Gaussian functions summed together to recreate the fire object

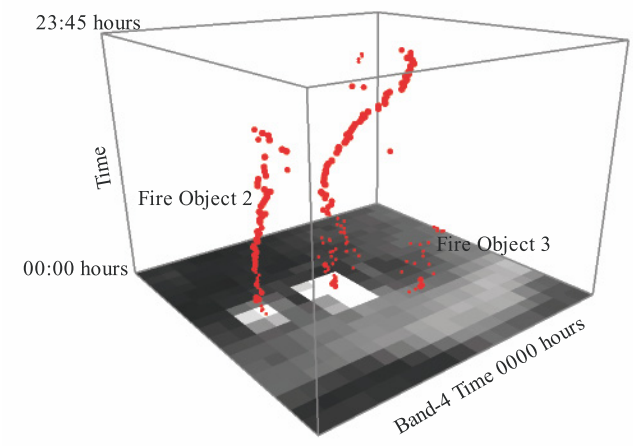

Figure 4.7: Illustration of fire objects tracked by the tracking mechanism

4.5 Results of the Space-Time Analysis

4.5.1 Behavior of fire objects in space and time

Figure 4.8 illustrate movement of fire in time. From this figure one can see a stronger continuity for objects 1 & 2 than for object 3. One can also observe that object 3 is discontinuous in time. Furthermore the swirling pattern of the fire objects in time is evident (it is most evident for object 1). This effect may be mainly due to vegetation, wind or topography.

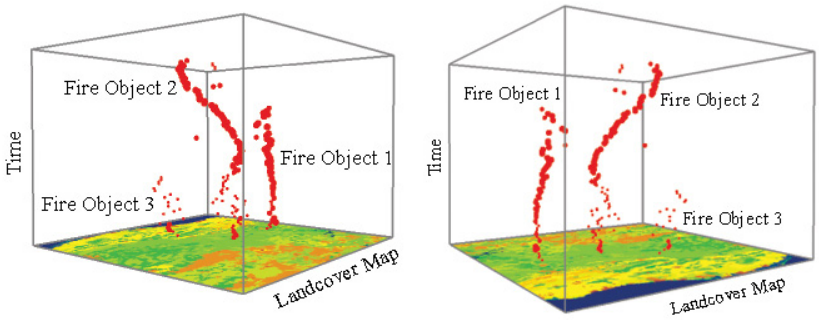

Figure 4.8: Illustration of movement of fire in space & time

Figure 4.9 illustrates the movement of fire objects 1 and 2 (represented as blue and red) over time. The ellipse delineates clustered fire objects. Clustering of fire objects was observed during the early hours in the morning (around 08:00 hrs). This might mainly be due to the building up of the fire i.e., the attainment of threshold before the fire started moving. The figure also shows that objects 1 and 2 are merging which is represented by the dotted lines. Furthermore, we could infer that movement of the objects were more along y-axis than along x-axis. This feature is more evident in figure 4.9 of the x-axis and time.

Table 4.2: Summary of NDVI values for the location of fire pixels

Total Number of fire objects	185
mean NDVI for those locations	0.290
Std.Dev of NDVI for those locations	0.014
min NDVI for those locations	0.262
First Quantile of NDVI for those locations	0.281
median	0.286
Third Quantile of NDVI for those locations	0.300
max NDVI from those locations	0.314
missing values	0.000

4.5.2 Effect of vegetation on fire

Figures 4.10 and 4.11 illustrate the effects of movement of fire over vegetation. Figure 4.10 shows the effect of movement of fire over time on NDVI and figure 4.11 the effect on landcover.

Most of the area under fire had a relatively high NDVI value with a mean of 0.290 (see table 4.2). According to the classification by Willams, 1995 [46] the fires were in the areas between medium and dense vegetation. From the table 4.2 it is clear that 95% of the fire objects were found in areas with relatively dense vegetation ($>= 0.281$ with the maximum of 0.314).

From the image 4.11 one can see that most of the fire pixels lie within forest and semi natural areas. Out of 185 fire objects identified by the algorithm over time, 168 objects were lying on areas that were forests and 17 objects were on agricultural area. All of the fire objects that were on the agricultural areas were of object 3.

4.5.3 Predicting the motion of the fire

The prediction of possible movement of fire was broken up into two parts to make it simpler. The first part was for predicting possible movement in x direction and the second part for predicting possible movement in y direction. This model is described in the equation 4.1 and 4.2.

$$\hat{x}_{t_{n+4}} \sim f\left(x, \sigma_x, I, WD, WS, NDVI\right)_{t_n} + e \tag{4.1}$$

$$\hat{y}_{t_{n+4}} \sim f\left(y, \sigma_y, I, WD, WS, NDVI\right)_{t_n} + e \tag{4.2}$$

Where, x & y are the location in latitude and longitude, σ_x & σ_y are the spread along x & y respective;y, I is the intensity of the fire, WD is the wind direction in true degrees, WS is the wind speed in pascal, t is the time and n is the instance.

The prediction was made for the possible movement of the fire within the next one hour from the given instance n. The results obtained from the model are listed in the table 4.3. From the results one can see that the R^2 is very high. This shows the high predicting ability of the model at a local level.

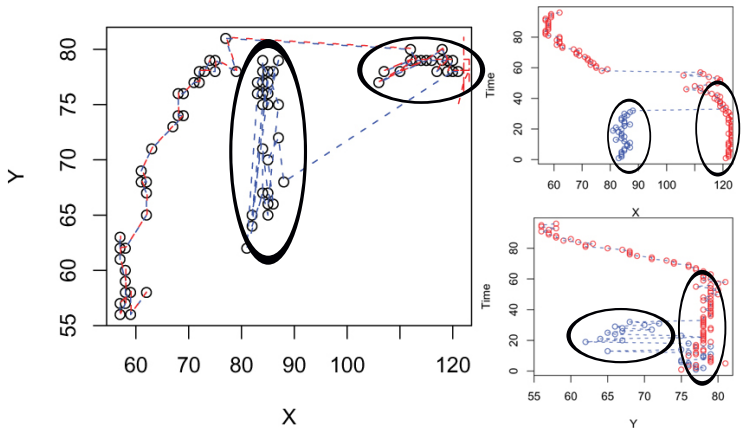

Figure 4.9: illustrating the movement of fire objects in a 2d plane

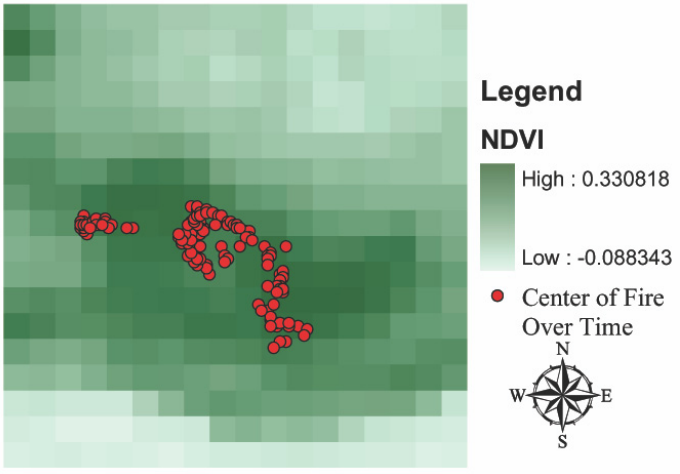

Figure 4.10: Illustration of movement of fire over a NDVI map

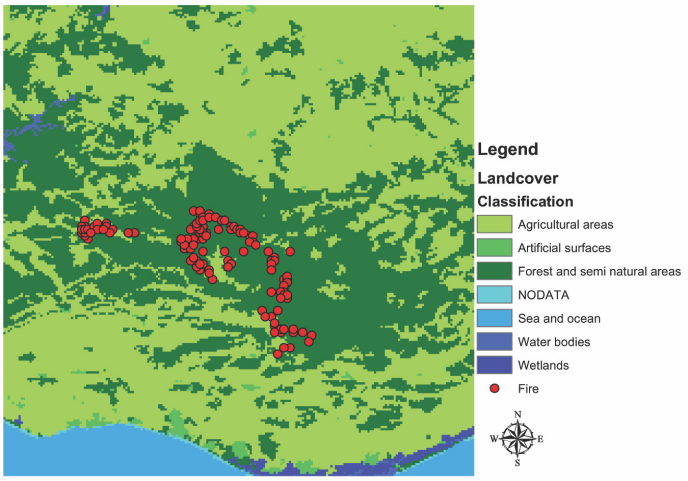

Figure 4.11: Illustration of movement of fire over a landcover map

Table 4.3: Results obtained for the prediction model

	X	Y
Location	0.6956	-0.6812
σ	0.3200	0.9180
I	69.4610	175.2544
WD	-0.1419	-1.2701
WS	-5.7402	-8.5277
NDVI	47156.1000	54026.9602
e	-245305.4462	1755344.4891
R^2	0.9745	0.9425

4.6 Summary

This chapter has detailed the results obtained by four main steps involved in the process of image mining and analysis. In short, the kernel convolution method was able to characterize the patterns over space effectively. Using the FFT to convolve added efficiency while processing a total of 96 images. The patterns obtained were clear and distinct in relation to the background. The extracting algorithm successfully extracted high intensity areas. Each area extracted was characterized as a Gaussian bivariate function. A total of 960 functions were extracted from 96 images. These functions were then tracked in time based on their neighborhood relationships. A total of 185 objects were tracked. A spatio-temporal analysis was carried out for these tracked objects for their relation to their structure and vegetation characteristics. The objects of fire exhibited a swirling pattern. This might be due to the effect of wind, vegetation or topography on fire. Further, there were two objects with a temporal continuity. These objects of fire further merged into one single fire at approximately 18:00 hrs. All objects of fire occurred over regions with relatively high NDVI. During a further analysis it was found that 90.8% of the fire objects occured /moved over landcover that was classified as forests and semi-natural areas. The knowledge extracted was further used to develop a model for predicting movement of the fire. Data on wind speed and direction were also included in this model. R^2 of the developed model was greater than 0.94 which proved to be satisfactory and promising for further detailed modeling.

The next chapter discusses the various stages that were involved in the development of this mining model. It also explains in brief the rationale behind selection of the methods and the algorithms.

Chapter 5

Discussions

5.1 Introduction

Use of several methods in the previous sections are the result of selection of appropriate method among sets of various other methods. This chapter discusses in some detail the selection process and the rationale behind their selection. This chapter is divided into six main sections. Section 5.2 discusses the exploration of characteristics of various bands before arriving at the most suitable band (band 4). Section 5.3 discusses the selection of smoothing function and its extensibility. It discusses the results of various other methods experimented with before arriving at the kernel convolution method. Section 5.4 discusses the extracting algorithms and other methods explored to address the process of extraction. Section 5.5 discusses about the tracking algorithms and some of their pitfalls. Section 5.6 discusses the space-time analysis which was used in this study and its extensibility. Finally section 5.7 discusses image mining model as a whole.

5.2 Exploratory Data Analysis

The spectral profiles of bands 1, 2, 3, 4 and 9 were analyzed in detail before arriving at the suitable band for modeling. Bands 1, 2 and 3 were considered because they were visible bands. The fire and the smoke were clearly identifiable in these bands. The bands 4 and 9 were considered because they were thermal bands of suitable wavelengths for the detection of fire. The rationale at this stage was to see whether any large differences in their radiance were observed so that the classification of the fire and non fire pixels could be made easier. Based on this hypothesis, a pixel having relatively high thermal radiance from the subset was selected. The change within the spectral characteristics of this pixel was analyzed for a 24 hour period. The results were then compared to the spectral profile of the same pixel on a non-fire day (23rd July). This analysis was done to all five bands (1, 2, 3, 4 and 9). The results obtained are displayed in figures 5.1 and 5.2.

Figure 5.1(a) is the subset of the area under study shown as color composite (band 3 displayed in red, band 2 in green and band 1 in blue). This was an image

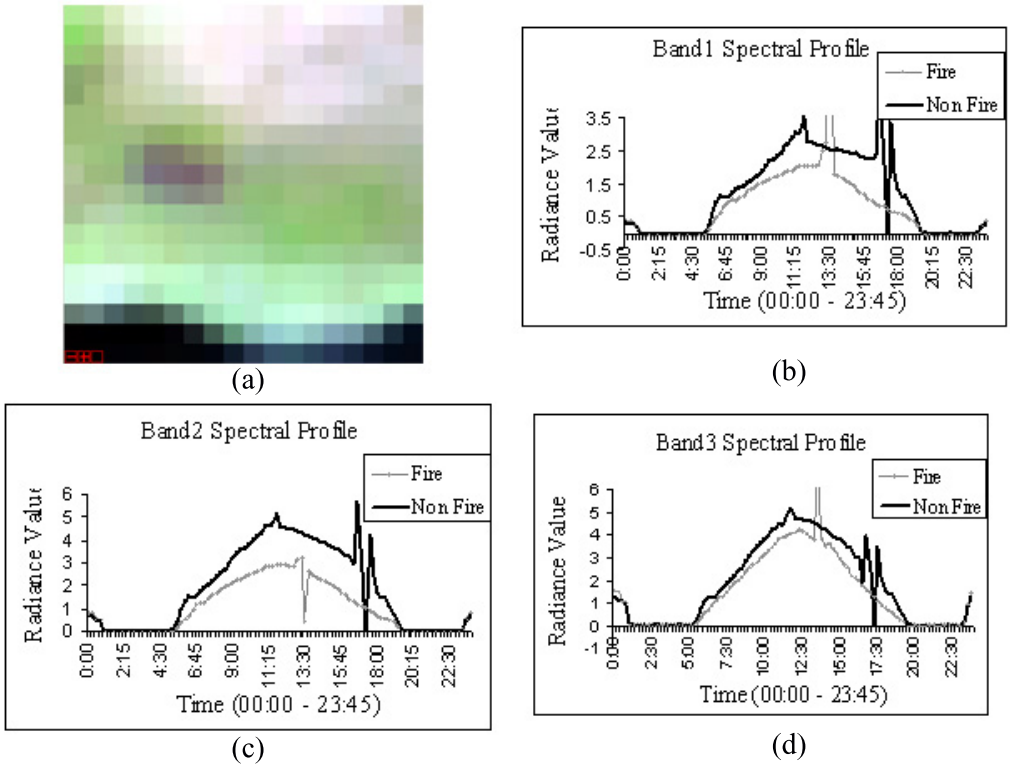

Figure 5.1: False color composite of Meteosat band (3,2,1) and the spectral profile of a fire pixel for a diurnal period

acquired by Meteosat on 28th July, 2004 at noon 12:15 hrs. The grayish black region in the middle of the image denotes the approximate area that was under fire. Figures 5.1(b), (c) and (d) denote the spectral profile of a single pixel in the bands 1, 2 and 3 respectively. One can see from the figures that in the visible region there are no considerable radiance values observed until 06:00 hrs and after 19:00 hrs. This is due to the absence of daylight within that region.

Figures 5.2(a) and (c) shows the image subset of the area under study for bands 4 and 9 respectively. These images were acquired on 28th July, 2004 at 12:15 hours. The white region in the middle of the image denotes the areas under fire. Figure 5.2(b) and (d) show the spectral profile of a single pixel in the bands 4 & 9 respectively. From these figures a variations in every band can be seen. These characteristic could be used further for fire detection and classification. The relative differences between fire and non fire pixels in thermal bands 4 and 9 were considerable, but those observed in the visible bands 1, 2 and 3 were not. Moreover lack of values during day and night time in the visible bands makes these bands unsuitable for diurnal modeling of fire. They could, however, possibly be useful for fire detection during the day.

The variation in infrared channels with respect to their background plays an important role in characterizing fire over space. The variation within pixels that were analyzed previously helped to reduce the number of bands from five

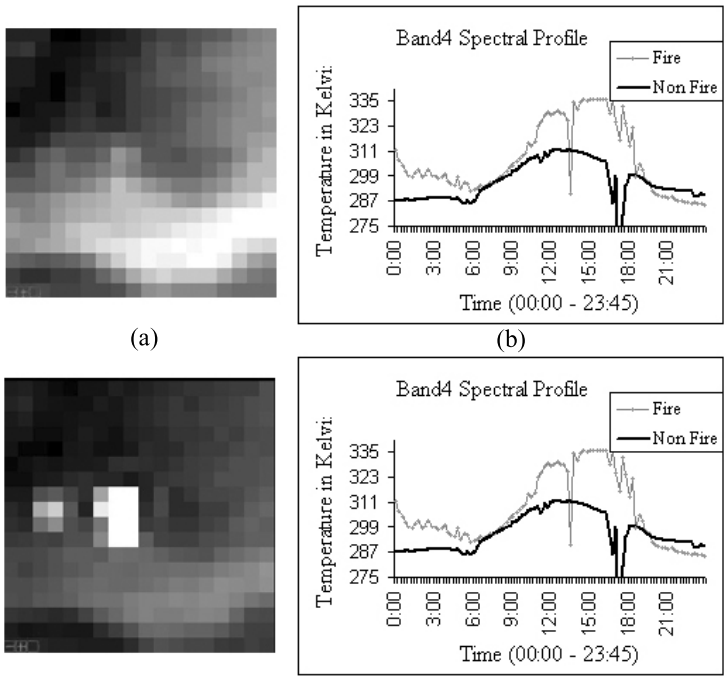

Figure 5.2: Illustration of the band 4 and band 9 with the spectral profile of a fire pixel for a diurnal period

(1,2,3,4 and 9) to two (4 and 9). Subsequent analysis on variation with respect to background was carried out for the bands 4 and 9. The subsets of images during morning, noon and at night were analyzed to detect changes in pattern. The result of this analysis are displayed in figure 5.3

Images 5.3(a), (b) and (c) correspond to band 4 and images 5.3(d), (e) and (f) correspond to band 9. Images 5.3(a) and (d) are of time t=1 (00:00 hours), images 5.3(b) and (e) are of time t=49 (12:15 hours) and images 5.3(c) and (f) are of time t=96 (23:45 hours). The white areas in images (a), (b) and (c) indicate possible areas of fires and are clearly distinct from their respective background during the diurnal cycle. Whereas in images (d), (e) and (f) the fire pixels are not clearly distinct with respect to their background.

Based upon the above mentioned analysis it was concluded that characterization of thermal activity over space can be performed well using band 4, which is in a window of $3.9\mu m$ wavelength. The key 'active fire signal' is, therefore, an increase of the observed radiance in the $3-5\mu m$ region, relative to the surrounding areas. For increasingly smaller or cooler fires, this contrast is progressively attenuated and becomes difficult to discriminate from natural spatial variability of the temperature field [24]. However, the case of fire that was under study was identifiable throughout its diurnal cycle in band 4. A further assumption at this stage was that, if the behavior of fire could be modeled using one band then the same mining model could later be extended for using additional bands.

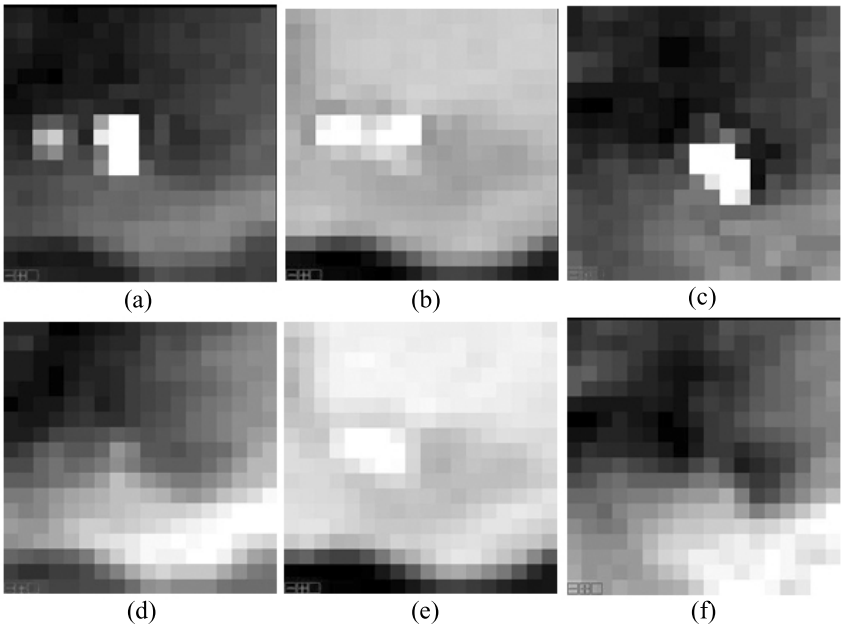

Figure 5.3: Transition of the fire pixels with respect to their background during the diurnal period. (a), (b) and (c) correspond to band 4 & (d), (e) and (f) correspond to band 9. Images(a) and (d) are of time t=1 (00:00 hours), (b) and (e) are of time t=49 (12:15 hours) & (c) and (f) are of time t=96 (23:45 hours)

5.3 Characterizing Patterns over Space

5.3.1 Function

The selection of an appropriate function to describe the phenomenon, as in this case fire was very important. In this research the Gaussian bivariate function was selected based on the nature of fire and behavior of heat in vast area of land. These characteristics made the selected function most appropriate. This image mining model can also be extended to represent other phenomena. In that case the function could be changed to bivariate uniform, spherical uniform, negative binomial or any other function depending upon the users description of the process under study. One of the most important consideration to make while selecting the function is its nature and the extent to which, it could represent the process under study. Further, the complexity of the function is also directly proportional to the processing time. Therefore, it would be advisable to choose simple functions for mining a large amount of remote sensing images.

5.3.2 Method

In this research an exploratory analysis of various methods was done to find the best class of methods to describe the characteristics of the patterns which are within a single image. Explored methods included loess, splines, kernels and kernel convolution. The results obtained form these methods are shown in figure 5.4. Figure 5.4(a) represents the actual image on which the methods were

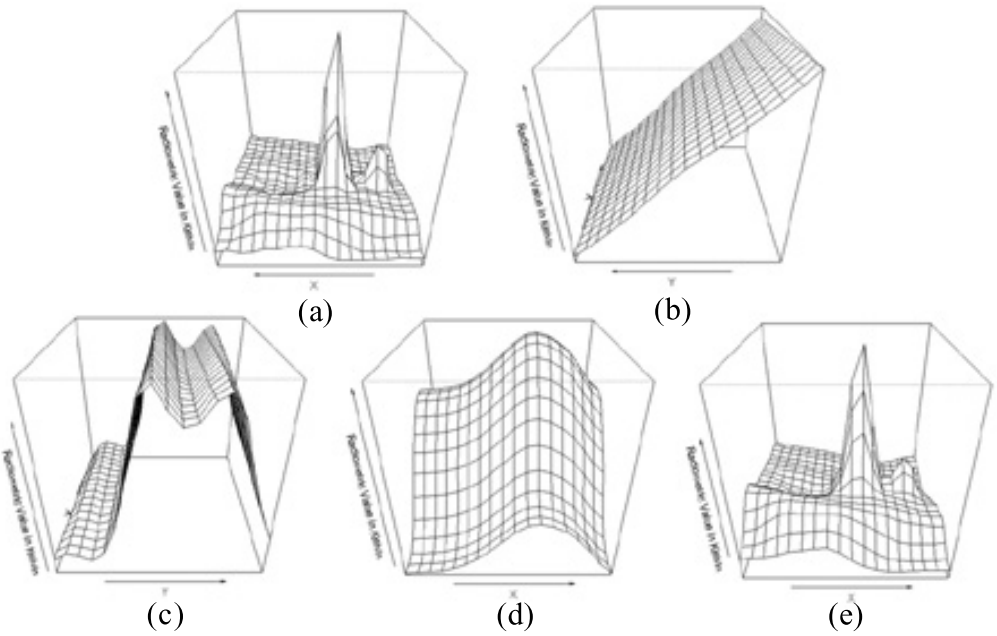

Figure 5.4: Actual image (a) along with the results of methods such as loess (b), splines (c), kernel (d) and kernel convolution (e)

applied. Figure 5.4(b), (c), (d) and (e) represent results obtained by implementing loess function, spline, kernel and kernel convolution methods respectively.

Based on results and characteristic features, the kernel convolution method was selected for characterizing the patterns over space. This method, based on convolution does support extensibility in terms of dimensions and dimension reduction (see equation 3.6 and 3.7 on page 27). This factor can be very helpful in extending this approach to define the patterns of variations for other phenomena that requires more than one band to classify it over space.

The method of kernel convolution was also applied to the image with normalized background. This image was generated by subtracting the thermal radiance of the image acquired on 28^{th} with the mean thermal radiance of images acquired on days 23^{rd}, 24^{th} and 25^{th}) for the same time. The results of the image with normalized background were similar to that of the actual image except for the fact that the background was exaggerated for early and late hours of the day and much smoothed for images of late morning and mid-day. The comparison of results of kernel convolution applied to actual image and normalized image is illustrated in the figure 5.5. Since this differences in radiance did not improve the model significantly, the results from the actual image were considered for further modeling.

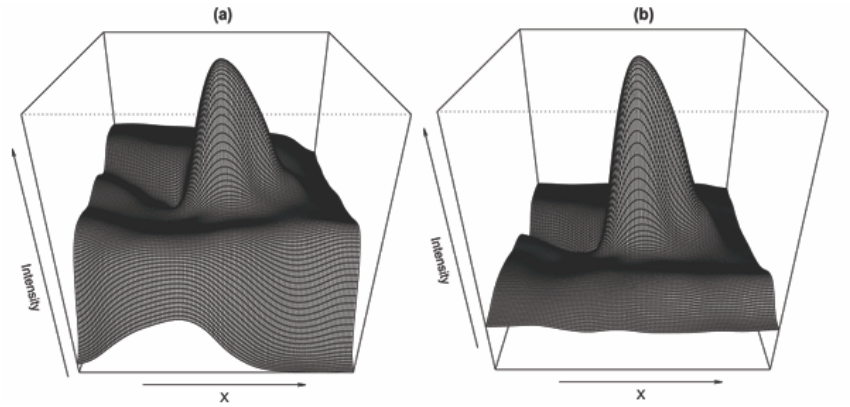

Figure 5.5: Showing the result of kernel convolution applied to actual image and image with normalized background. (a) shows the result of the actual image, (b) shows the result of the normalized image

5.4 Extraction of Fire Objects

Extracting the fire must lead to identification of all fires over space s at any time t. Two methods have been explored to automate this process. The first method characterizes the regions of high thermal intensity under the condition that they are of a rectangular nature. This led to the extraction of rectangular regions from the image containing the possible fire object. Figure 5.6 explains the result of this process for an image. This method generalizes the fire objects to be of rectangular nature. This also led to miscalculation of volume of fire objects.

To overcome this inefficiency of the previous algorithm, another algorithm was developed. This algorithm redefined fire objects to be composed of a Gaussian bivariate distribution instead of a rectangular region. The Gaussian function was selected based upon the same rationale as that of, selection of function under subsection 3.3.1 for characterizing (refer page 24). This algorithm thus developed facilitated in decomposing the image into a series of functions that define the fire objects and a background. Figure 5.7 illustrates the results of this algorithm using Gaussian bivariate function (also refer figure 4.4, 4.5 and 4.6 on page 39).

The results of the extracting algorithm 2 also did over fit and under fit the fire objects. Figure 5.8(a) and (b) illustrate the amount of over fitting and under fitting respectively. This amount of exaggeration was considerably less and its success rate was also higher as compared to the previous method. Based on the results, this algorithm was selected for the final image mining model to extract the possible fire objects from the imagery.

5.5 Tracking of Fire Objects

The algorithm for tracking fire objects over time was developed based on the location (x and y), spread (σ_x and σ_y) and their neighborhood relation in space, i.e.

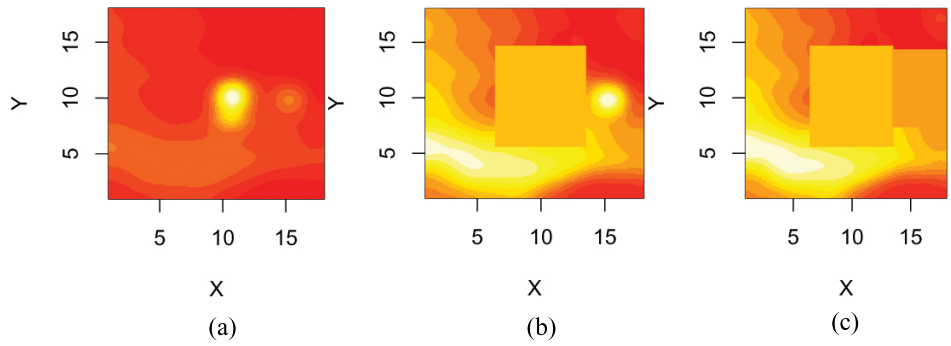

Figure 5.6: Illustrating the process involved in the extraction of fire objects (algorithm1). (a) show the actual image, (b) and (c) show the result of steps followed in extracting two fire objects. The rectangular areas correspond to the areas of fire objects

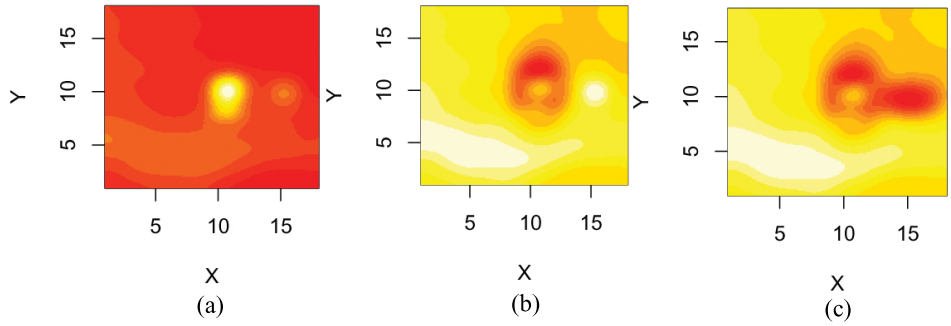

Figure 5.7: Illustrating the process involved in the extraction of fire objects (algorithm2). (a) show the actual image, (b) and (c) show the result of steps followed in extracting two fire objects. The color red shows regions of low intensity and white shows the regions of high intensity.

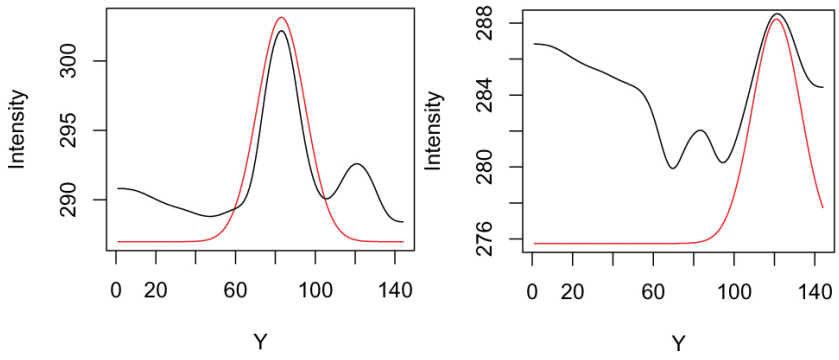

Figure 5.8: Illustration showing the amount of over fitting (a) and under fitting (b) of the function to define the fire objects. The red line represents the profile of the function defined by the algorithm and the black line represents the actual profile of the object

with objects of time t_{n+1}. There could be other attributes of the fire objects such as volume, shape, angle, etc., that could have been incorporated into the algorithm. But since the algorithm was performing sufficiently well based on the location and spread over time further attention to these other attributes was not given. Nevertheless other problems were faced by the tracking algorithm, such as splitting of objects in time. The aspect of merging was well captured by the algorithm but the aspect of splitting of fire objects needed further improvement. The initial starting point of the fire needed to be given as an input to the algorithm. This property complicated the process of tracking objects that were splitting in time. In the algorithm that was used for the final model, the aspect of splitting is to be guided by the user to an extent. This led to the algorithm in terminating several times. Further thought is to be given to make this algorithm more efficient.

5.6 Space-Time Analysis

The space-time analysis did throw some light on utilization of the extracted knowledge. In this research the extracted knowledge has been utilized for understanding the influence of wind, NDVI and landcover. Nevertheless the knowledge could also be used to study the influence of other features such as topography, wetness index, fuel index, etc. From this study it was found that fire moved over the regions having high NDVI. This fact was contradicting to the statements made by Lambin et al [36] in their study (see, page 16). This might be due to the reason that fire hazard is site specific or the effect of NDVI is dependent upon the sensor resolution. This aspect needs to be further analyzed before arriving at any conclusion. Furthermore this research attempted to model the movement of the fire using the data of wind, vegetation, spread of fire, etc.,. The rational at this stage for developing such a model was not to make a prediction for any fire detected from Meteosat, but to test the hypothesis whether such prediction models could be developed from the acquired knowledge. Based upon the results of the model ($R^2 > 0.94$) we can conclude that, further research along this direction could lead to the development of a more refined global models for predicting the movement of fire from Meteosat imagery. Such models could also be used in conjunction with other micro level models that are already existing (ref chapter 2.3.3) to make the predictions even stronger.

5.7 Extension of the Architecture

The architecture of the image mining model is illustrated in figure 5.9. Separate functions and methods have been developed in 'R' script to automate the processes. The details related to the functions and their corresponding source codes are presented in the annex. The extensibility of the components to address other phenomena from Meteosat imagery is as follows:

- Selection of the bands: In this research band 4 of Meteosat was selected for modeling the fire. The fire was well differentiable in this band and

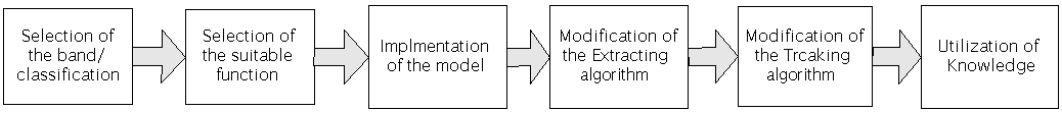

Figure 5.9: The architecture of the Meteosat image mining model

was modeled successfully. But a combination of bands or a classification technique could also be incorporated for addressing the same phenomenon or any other phenomena. Scalability of the kernel convolution method in terms of dimension could further facilitate in this process.

- Selection of the function: In this research a Gaussian bivariate function was selected. Other functions could be used to address other phenomena. The selection of the suitable function can be based on experience of the miner or on exploratory analysis of the data.

- Method: The methods of kernel convolution was found to be very effective. This study applied this method for characterizing one variable over two-dimensional plane. But the kernel methods could be extended for multi variable over three or more dimensional planes. This extensibility of the method would be promising for addressing complicated phenomena that are characterized by multidimensional variables.

- Extraction algorithm: In this research the peaks having relatively maximum intensities were extracted since fire pixels have relatively higher intensity values. But the same algorithm could also be used for characterizing minima's. The choice of characterizing the extraction algorithm to be a maximization or minimization problem would depend upon the type of classification needed for differentiating the event from its background. This algorithm could be extended for both cases by incorporating minor modifications.

- Tracking algorithm: The performance of this algorithm was satisfactory for the case of fire which was under study. Only the relationship with respect to the center of the objects and their spread was considered in this research. Further parameters such as volume can be included to increase the effectiveness of this algorithm for tracking other dynamic phenomena.

- Utilization of the knowledge: The utilization of the knowledge extracted was demonstrated using simple techniques such as overlay and linear regression models. But the same knowledge could also be used for developing complex model using advanced techniques. Further only the characteristics related to vegetation were considered, but there are other factors such as topography which play a vital role in the movement and spread of fire. This information, combined with other techniques, is being explored to address further issues.

5.8 Summary

This chapter has detailed the various steps which were followed before arriving at the final model for monitoring of forest fire from Meteosat imagery. Several of the bands were suitable to characterize the phenomenon but band 4 was found to be the most suitable. From the various classes of methods that were available the kernel convolution technique fitted the data better than other methods such as; splines, loess and other kernel methods. The kernel convolution method was then experimented with the images having their background normalized. But this process did not give much improvement in the modeling. The selected method was then applied for 96 images and the fire objects from the resulting image were then extracted and tracked. There are few problems which are related to the tracking algorithm but its performance was found to be satisfactory for the case at hand. The results of the tracked objects were analyzed using space-time cube. Finally, this chapter summarized the components of the image mining model that was developed and briefly discusses the extensions that could be made possible within them.

Chapter 6

Conclusion

To summarize, the problem of fire still exists and is of serious concern. Even though many algorithms have been developed for fire, the lack of timely monitoring using remote sensing is still a concern for European and African regions. In order to address that issue, this research aimed at developing a model using Meteosat Second Generation (MSG) for monitoring and analyzing the behavior of fire in space and time. The temporal resolution of Meteosat is high (an image every 15 minutes leading to 96 images a day). This makes it humanly impossible to analyze them manually. At the parallel front there are new tools and techniques which are being developed to address the problem of handling large databases. Through this research, a synergic merger model for fire pattern extraction and analysis from MSG using techniques of image mining was conceptualized and developed.

Various properties of the Meteosat Image data were analyzed visually with respect to the phenomenon under study. Based on this analysis the band 4 was selected since it was the most suitable for monitoring the event of fire as compared to the rest of the bands. The assumption was that if the fire could be modeled using one band then the inclusion of further band and data could be built upon it. The selected band was characterized over space using kernel convolution. FFT was used to increase the efficiency of the processing during convolution. The patterns developed were then fed into as an input for extracting the fire objects. The fire objects were first defined based on the Gaussian bivariate function. Then the possible fire pixels were extracted from the image using image decomposition, where the image is separated into a series of function (objects) and a background. The function was composed of the location of the fire, its intensity and spread. These extracted objects were then taken as an input for tracking them over time. The tracking algorithm was based on identifying the least spatial distance within adjoining time frame between all the objects. This led to a series of objects in time. The patterns of these objects over space and time represented the knowledge gained from the system. This knowledge was then used for further understanding the characteristic behavior of the fire. The process of the movement of the fire with respect to the change in its structure based on time, vegetation and wind were analyzed. The task of converting the image into matrices, characterizing, extracting and tracking

were automated. Separate functions were developed and implemented in the environment R to ease the processes.

Based on the results, the kernel convolution method was able to characterize the patterns over space effectively. The patterns obtained were distinct in relation to the background. The extracting algorithm defined and extracted the high intensity areas as a Gaussian bivariate function. A total of 960 functions were extracted from 96 images. These functions were then tracked based on their neighborhood relationships in time. A total of 185 objects were tracked. A spatio-temporal analysis was carried out for these tracked objects for analyzing relation to their structure and vegetation characteristics. The objects of fire exhibited a swirling pattern. This might be due to the effect of wind or topography on fire. Further, there were two objects with a temporal continuity. These objects of fire further merged into one single fire at approximately 18:00 hrs. All objects of fire occurred in high NDVI regions. During a further analysis it was found that 90.8% of the fire objects spread over landcover that was classified as forests and semi-natural areas. The knowledge extracted was also further utilized by developing a model for predicting the movement of the fire. Data on wind speed and direction were included in this model. The R^2 of the developed model was greater than 0.94 which proved to be satisfactory and promising for further detailed modeling for predicting the movement of the fire.

The developed mining model is very promising for monitoring and analyzing the behavior of fire. The Meteosat satellite has lot of potential for giving rapid response, since the data are disseminated within $12\ mins$ of reception. Furthermore the meteorological products which are generated from it could be used for further understanding of behavior of the phenomenon under study. This model could therefore be improved and extended for spatio-temporal analysis and monitoring of other dynamic phenomena from Meteosat imagery or other high temporal resolution imagery.

Appendix A

Source Code

```
# The source code which was developed in R to execute & automate the processes
# This is the source of only the functions developed in R. This doesn't include
#  the source of the interfaces developed in Visual Basic
# need to load the following libraries before doing the process
# rgdal, lattice, rgl, fields
############################################################
####### CODE FOR CHANGE OF FORMAT FOR THE DATA BASE#######
############################################################

fire_day<- "f1_25"
for(band_no in 1:2)
{
        #band_no<-2
        file_var1<-paste("pictures/",fire_day,"/",band_no,"/",sep="")
        for(time_stamp in 1:96)
        {
                file_var2<-paste(fire_day,"_",band_no,"_",time_stamp,
                        ".img",sep="")

                file_var<-paste(file_var1,file_var2,sep="")
                logo <- system.file(file_var, package="rgdal")[1]
                x <- new("GDALReadOnlyDataset", logo)
                h <- getRasterTable(x, band = NULL, offset = c(0, 0),
                        region.dim = dim(x))
                hrow<-h[,1]
                hrow<-round((round(round(hrow)-3666493))/3000+1)
                hcol<-h[,2]
                hcol<-round(-((hcol)+657088.3)/3000.4+1)
                d <- data.frame(cbind(time_stamp, hrow, hcol, h[,3]))
                GDAL.close(x)
                file_var3<-paste(band_no,".csv",sep="")
                write.table(d, file = file_var3, append = TRUE, quote = TRUE,
                 sep = ",", eol = "\n", na = "NA", dec = ".", row.names = FALSE,
                 col.names = FALSE, qmethod = c("escape", "double"))
        }
}
############################################################
######open the file and converting it into one table######
###########made complex but effecient####################
############################################################
start_band<-1
end_band<-2
```

```
for(band_no in start_band:end_band)
{
file_var<-paste(band_no,".csv",sep="")
l<-read.table(file_var, header = FALSE, sep =",")
if(band_no >= start_band+1)
        {
                k<-data.frame(cbind(k, l[,4]))
        }
        if(band_no <= start_band)
        {
                k<-l
        }
}
write.table(k, file = "f1_new.csv", append = FALSE, quote = TRUE,
 sep =  ",", eol = "\n", na = "NA", dec = ".", row.names = FALSE,
 col.names = FALSE, qmethod= c("escape", "double"))

############################################################
########openeing the files and converting into one#########
#################table made simple##########################
############################################################
for(band_no in 2:3)
{
    file_var<-paste(,band_no,".csv",sep="")
    l<-read.table(file_var, header = FALSE, sep =",")
    l<-data.frame(cbind(l,band_no))
write.table(l, file = "f1_new.csv", append = TRUE, quote = TRUE,
 sep = ",", eol = "\n", na = "NA", dec = ".", row.names = FALSE,
 col.names = FALSE, qmethod= c("escape", "double"))
}

############################################################
#######To get the data and make a subset with it###########
############################################################
p<-read.csv("f1_28.csv", header = FALSE, sep = ",",
 quote="\"",    dec=".")

colnames(p) <- c("time","x","y","1","2","3","4","5",
 "6","7","8","9","10","11")

p1<-subset(p,time<2, select = c(time,x,y,7))
colnames(p1) <- c("time","x","y","band4")
xy<-data.matrix(subset(p1,select = c(x,y)))

############################################################
#######to make the average of the dates####################
############################################################
##where p1 p2 and p3 are the subset of three days of fire files
##do the above procedure for all the dates i.e 23,24 and 25
##name them as p2,p3,p4 respectively
##then the averaging is then simple

p2<-(p4+p2+p3)/3
write.table(p_mean, file = "f1_mean.csv", append = FALSE,
 quote  = TRUE, sep = ",", eol = "\n", na = "NA", dec = ".",
 row.names = FALSE, col.names = FALSE,qmethod= c("escape", "double"))

############################################################
```

```
####################mean data creation####################
############################################################
p<-list(NULL)
p_mean<-list(NULL)

p<-read.csv("f1_28.csv", header = FALSE, sep = ",",
 quote="\"",    dec=".")

colnames(p) <- c("time","x","y","1","2","3","4","5",
 "6","7","8","9","10","11")

p_mean<-read.csv("f1_mean.csv", header = FALSE, sep = ",",
 quote="\"", dec=".")

colnames(p_mean) <- c("time","x","y","1","2","3","4","5",
 "6","7","8","9","10","11")

pded<-p
for(i in 4:14)
pded[,i]<-p[,i]-p_mean[,i]
write.table(pded, file = "f1_deducted.csv", append = FALSE,
 quote = TRUE, sep =",", eol = "\n", na = "NA", dec = ".",
 row.names = FALSE, col.names = FALSE,qmethod= c("escape", "double"))

############################################################
###############bivariate normal distribution###############
############################################################
f<-function(x,y,mu1,mu2,sigma1,sigma2,rho,h)
{
 exp((-h/(2*(1-rho^2)))*(((x-mu1)/sigma1)^2-2*rho*(x-mu1)*(y-mu2)/
  (sigma1*sigma2)+((y-mu2)/sigma2)^2))/
    (2*pi*sigma1*sigma2*sqrt(1-rho^2))
}

############################################################
#################2D process convoution######################
############################################################
ca<-function(l1,h)
{
        df1<-list(NULL)
        a1<-list(NULL)
        a1p<-list(NULL)
        w<-list(NULL)
        s<-list(NULL)

        y<-l1[,4]
        n<-length(y)
        s<-cbind(l1[,2],l1[,3]) # of data points
        m<-n # number of support sites for x(s)
        w<-s

        #w<-spatial(h)
        #m<-(length(w)/2)

        # create the matrix K
        K<-matrix(NA,ncol=m,nrow=n)

 ## the model is not dependent on the rho
```

```
 #it is only dependent on the sigma i.e spread
 #the approximate value for the minimum temperature helps in the
 #regression to read better values of min and max this must be mainly
 #due to the fact that a good start point helps in sppedy process

        for(i in 1:m) K[,i]<-f(s[,1],s[,2],w[i,1],
         w[i,2],3,3,0.1,280)

        df1<-data.frame(y=y,K=K,sub=1)
        df1$sub<-as.factor(df1$sub)

        # now a fit a mixed model using lme
        a1<-lme(fixed= y ~ 1,random= list(sub=pdIdent(~K-1)),
         data=df1, na.action=na.omit)

        # obtain and plot the fitted values
        a1p<-as.vector(predict(a1,df1))
        a1p<-matrix(a1p,nr=18)

        return(a1p)
}

############################################################
###############generating spatial weights##################
############################################################
spatial<-function(value)
{
        w1<-seq(18,0.1,-value)
        w2<-w1
        l<-length(w1)*length(w1)
        w<-matrix(NA,ncol=2,nrow=l)
        len<-1
        count2<-1
        for(count1 in 1:l)
        {
                w[count1,1]<-w1[count2]
                w[count1,2]<-w2[len]
                if(count2 == sqrt(l))
                {
                        count2<-0
                        len<-len+1
                }
        count2=count2+1
        }
        return(w)
}

############################################################
###################inversing a matrix#######################
############################################################
inv_matrix<-function(x)
{
        d<-dim(x)
        #row matrix
        i<-d[1]
        #colomn matrix
        j<-d[2]
        l<-matrix(NA,ncol=i,nrow=j)
```

```
        for(a in 1:i)
        {
                for(b in 1:j)
                {
                        l[b,a]<-x[a,b]
                }
        }
        return(l)
}

#############################################################
##################making singular matrix#####################
#############################################################
sin_matrix<-function(x)
{
        d<-dim(x)
        #row matrix
        i<-d[1]
        #colomn matrix
        j<-d[2]
        l<-0
        count=1
        for(a in 1:i)
        {
                for(b in 1:j)
                {
                        l[count]<-x[a,b]
                        count<-count+1
                }
        }
        return(l)
}
#############################################################
#####extracting images of particular time sequence########
#############################################################
p<-list(NULL)
p<-read.csv("f1_28.csv", header = FALSE, sep = ",",
 quote="\"",    dec=".")

colnames(p) <- c("time","x","y","1","2","3","4","5",
 "6","7","8","9","10","11")

p1<-list(NULL)
s2d<-list(NULL)
sreg<-list(NULL)
con2d<-list(NULL)

#############################################################
#######executing the command and storing the data#########
#############################################################
for(im in 1:96)
{
        y1<-list(NULL)
        y2<-list(NULL)
        h<-1

        p1[[im]]<-subset(p,time==im, select = c(time,x,y,7))
        colnames(p1[[im]]) <- c("time","x","y","band4")
```

```
        y1<-subset(p1[[im]],select = (band4))
        xy<-subset(p1[[im]], select = cbind(x,y))

        k<-p1[[im]]
        xy2<-cbind(k[,2],k[,3])
        y2<-k[,4]
        #zxy<-cbind(p1[,2],p1[,3])

        ##applying smooth.2d##
        #need to look into the cov. function
        #and understand the technical details

        s2d[[im]]<-smooth.2d (y1, ind = xy, weight.obj = NULL,
         setup = FALSE, nrow = 144, ncol = 144, surface = TRUE,
          cov.function = gauss.cov, Mwidth = 18, Nwidth = 18)

        ##applyting sm.regression
        sreg[[im]]<-sm.regression(xy2, y2, h=c(.25,.25),
         design.mat = NA, model = "none", display="none")

        ## applying 2d convolution
        ## before applying this function got to load
        ## the bivariate normal distribution function and
        ## 2d convolution function

        l1<-p1[[im]]

#this is to run the program without running the nlme
#take care "nlme" model does'nt use the FFT
#so it will takes a lot of time this is not necessary
#for the further modeling but if someone is interested
#in comparing results then this would be interesting

#use this bracket to complete the code else it will run full

#}

###convolution############

        df1<-list(NULL)
        a1<-list(NULL)
        a1p<-list(NULL)
        w<-list(NULL)
        s<-list(NULL)

        y<-l1[,4]
        n<-length(y)
        s<-cbind(l1[,2],l1[,3]) # of data points
        m<-n # number of support sites for x(s)
        w<-s

        K<-matrix(NA,ncol=m,nrow=n)
        for(i in 1:m) K[,i]<-f(s[,1],s[,2],w[i,1],w[i,2],
         3,3,0.1,280)

        df1<-data.frame(y=y,K=K,sub=1)
        df1$sub<-as.factor(df1$sub)
```

```
        a1<-lme(fixed= y ~ 1,random= list(sub=pdIdent(~K-1)),
         data=df1, na.action=na.omit)

        a1p<-as.vector(predict(a1,df1))
        con2d[[im]]<-matrix(a1p,nr=18)
}

##the result of the function generation are stored in the RAM
##from the next stage the data is not extracted from the file
##they are used from the RAM

############################################################
####################distance extarction#####################
############################################################
peak_ext<-function(v,grid)
{
        ####################################
        ###to get the counts#################
        ####################################
        #grid<-144
        numb1<-which.max(v)
        col_numb<-list(NULL)
        row_cumb<-list(NULL)
        col_numb<-round((numb1/grid)+0.4999)
        row_numb<-numb1-((col_numb-1)*grid)

        j<-0
        count=list(NULL)
        count<-matrix(1,ncol=1,nrow=4)
        while (j==0)
        {
                if(col_numb-(count[1]) > 0)
                {
                        if(v[row_numb,col_numb-(count[1]-1)] >
                         v[row_numb,(col_numb-count[1])])

                        count[1]=count[1]+1
                        if(col_numb-(count[1]) < 1) break
                        if(v[row_numb,col_numb-(count[1]-1)] <=
                         v[row_numb,(col_numb-count[1])]) j<-1
                }
                else j<-1
        }

        j<-0
        while (j==0)
        {
                if(col_numb+(count[2]) < 145)
                {
                        if(v[row_numb,col_numb+(count[2]-1)] >
                         v[row_numb,(col_numb+count[2])])

                        count[2]=count[2]+1
                        if(col_numb+(count[2]) > grid) break
                        if(v[row_numb,col_numb+(count[2]-1)] <=
                         v[row_numb,(col_numb+count[2])]) j<-1
                }
                else break
```

```
        }

        j<-0
        while (j==0)
        {
                if(row_numb-(count[3]) > 0)
                {
                        if(v[row_numb-(count[3]-1),col_numb] >
                         v[(row_numb-count[3]),col_numb])

                        count[3]=count[3]+1
                        if(row_numb-(count[3])< 1) break
                        if(v[row_numb-(count[3]-1),col_numb] <=
                         v[(row_numb-count[3]),col_numb]) j<-1
                }
        else j<-1
        }

        j<-0
        while (j==0)
        {
                if(row_numb+(count[4])< 145)
                {
                        if(v[row_numb+(count[4]-1),col_numb] >
                         v[(row_numb+count[4]),col_numb])

                        count[4]=count[4]+1
                        if(row_numb+(count[4])> grid) break
                        if(v[row_numb+(count[4]-1),col_numb] <=
                         v[(row_numb+count[4]),col_numb]) j<-1
                }
                if(row_numb+(count[4])> grid) break
                }
return(count)
}

############################################################
###############function used for extarction################
############################################################

f<-function(x,y,mu1,mu2,sigma1,sigma2,rho)
{
 exp((-1/(2*(1-rho^2)))*(((x-mu1)/sigma1)^2-2*rho*(x-mu1)*(y-mu2)/
   (sigma1*sigma2)+((y-mu2)/sigma2)^2))/(2*pi*sigma1*sigma2*sqrt(1-rho^2))
}

############################################################
###############extracting peaks#############################
############################################################
Sys.time()
for(count1 in 1:96)
{

        im<-s2d[[count1]]$z
        storage<-matrix(NA,nrow=10,ncol=6)
        count2<-1
        jj<-0
        while(jj==0)
```

```
{
        ##########################################
        #############finding spread###############
        ##########################################
        col_numb<-list(NULL)
        row_cumb<-list(NULL)
        h<-which.max(im)
        col_numb<-round((h/grid)+0.4999)
        row_numb<-h-((col_numb-1)*grid)
        mu1<-row_numb
        mu2<-col_numb
        sigma<-peak_ext(im,grid=144)
        sigma_min1<-round((min(sigma[3],sigma[4])-1)/2)
        if(sigma_min1<8) sigma_min1<-8
        sigma_min2<-round((min(sigma[1],sigma[2])-1)/2)
        if(sigma_min2<8) sigma_min2<-8

        ##########################################
        ############executing the function########
        ##########################################
        l<-matrix(h,ncol=144,nrow=144)
        for(i in 1:144)
        {
        for(j in 1:144)
        {
        l[i,j]<- f(x=i,y=j,mu1=mu1,mu2=mu2,
         sigma1=sigma_min1,sigma2=sigma_min2,rho=0)
        }
        }
        #persp(x,x,l)
        ##########################################
        ####rescaling the function with respect###
        ###############to the originl height#######
        ##########################################
        h1<-(max(im)-min(im))
        h2<-max(l)-min(l)
        h3<-min(im)
        h4<-min(l)

        for(i in 1:144)
        {
        for(j in 1:144)
        {
        l[i,j]<-(l[i,j]-h4)*(h1/h2)+h3
        }
        }
        ##########################################
        ########storing the function##############
        ##########################################
        if(sigma_min1>8 && sigma_min2>8)
        {
        storage[count2,1]<-count1
        storage[count2,2]<-mu1
        storage[count2,3]<-mu2
        storage[count2,4]<-sigma_min1
        storage[count2,5]<-sigma_min2
        storage[count2,6]<-max(l)
        count2<-count2+1
```

```
                }
                ########################################
                ##deducting the function from the image###
                ########################################
                img<-(im-l)+283
                if(count2>1) jj<-1 else im<-img

        }
write.table(storage, file = "peaks.csv", append = TRUE,
 quote = TRUE, sep = ",", eol = "\n", na = "NA", dec = ".",
 row.names=TRUE,col.names =FALSE,qmethod= c("escape","double"))
}
Sys.time()
###################

#############################################################
###################connecting centers#######################
#############################################################
connecting_centers<- function(tn, tn1)
{
        st<-1
        storage3<-matrix(0,ncol=11,nrow=10)
        storage3[,11]<-tn[1,2]
        count<-dim(tn)
        countj<-dim(tn1)
        jump<-1
        for(i in 1:10)
        {
                for(j in 1:10)
                {
                        #20 threshold

                        k1<-sqrt((tn[i,3]-tn1[j,3])^2+
                         (tn[i,4]-tn1[j,4])^2)

                        if(k1< 20)
                        {
                                storage3[i,j]<-1
                        }
                }
        }
write.table(storage3, file = "connection.csv", append = TRUE,
 quote = TRUE, sep = ",", eol = "\n", na = "NA", dec =  ".",
 row.names = TRUE, col.names =FALSE, qmethod=c("escape","double"))
}

#############################################################
#############connecting centers for 96 images##############
#############################################################
sto<-list(NULL)
sto1<-list(NULL)
sto2<-list(NULL)

sto<-read.csv("peaks.csv", header = FALSE, sep = ",",
 quote="\"", dec=".")

colnames(sto) <- c("ID","No","time","x","y","sigma1",
 "sigma2","temperature")
```

```
for(count in 1:95)
{
sto1<-subset(sto,time==count,select=c(No,time,x,y))
sto2<-subset(sto,time==(count+1),select=c(No,time,x,y))
connecting_centers(sto1,sto2)
}

############################################################
#####################exracting fire#########################
############################################################
fires1<-read.csv("connection.csv", header = FALSE, sep = ",",
 quote="\"", dec=".")
colnames(fires1) <- c("ID","1","2","3","4","5",
 "6","7","8","9","10","time")

storage<-matrix(0,ncol=3,nrow=960)
storage<-as.data.frame(storage)
colnames(storage) <- c("time","order","no")

storage[1,1]<-10000
storage[1,2]<-1
storage[1,3]<-1
store_count<-2
j<-0

for(count in 1:95)
{
        if(count==1) j<-10000 else j<-count-1

        f1<-subset(fires1,time==count,select=c(1,2,3,4,5,
         6,7,8,9,10,11,time))

        st<-subset(storage,time==j,select=c(3))
        colnames(st) <- c("no")
        k1<-dim(st)
        if(k1[1]>1)
        {
        for(sort1 in 1:(k1[1]-1))
        {
                for(sort2 in (sort1+1):k1[1])
                {
                        if(st[sort1,1]==st[sort2,1])
                         st[sort2,1]<-0
                }
        }
        st<-subset(st,no > 0,select=c(1))
        }

        k<-dim(st)
        for(count2 in 1:k[1])
        {
                #if(st[count2,1]<6)
                #{
                f2<-subset(f1,ID==st[count2,1],select=c(1,2,3,4,5,
                 6,7,8,9,10,11,time))

                for(i in 2:11)
```

```
                {
                if(f2[1,i]> 0)
                        {
                        storage[store_count,1]<-count
                        storage[store_count,2]<-st[count2,1]
                        storage[store_count,3]<-(i-1)
                        store_count<-store_count+1
                        }
                }
                #}
        }
}

write.table(storage, file = "fire1.csv", append = TRUE,
 quote =TRUE, sep = ",", eol = "\n", na = "NA", dec = ".",
 row.names=TRUE,col.names =FALSE, qmethod=c("escape","double"))

##do the same for the number of fires one wants to extract
##then same the aggregation of the results in the file fires.csv
##it should have the following colomns
##x, y, spreadx,spready, intensity, time

############################################################
######################exracting fire########################
############################################################
cor<-read.csv("coordinates.csv", header = FALSE, sep = ",",
 quote="\"",    dec=".")

f1<-read.csv("fires.csv", header = FALSE, sep = ",",
 quote="\"", dec=".")

colnames(f1) <- c("lat","long","x","y","temp","time")
for(i in 1:96)
{
f1[i,1]<-cor[(145-f1[i,3])+(146-f1[i,4])*144,3]
f1[i,2]<-cor[(145-f1[i,3])+(146-f1[i,4])*144,4]
}
write.table(f1, file = "fire_corr3.csv", append = FALSE,quote = TRUE,
 sep = ",", eol = "\n", na= "NA",dec = ".", row.names = FALSE,
 col.names = FALSE, qmethod =   c("escape", "double"))

############################################################
###############additional commands#########################
############################################################
#can use persp(x,y,z) for displaying in perspective
#can use image(x,y,z) for displaying the image
#try using wireframe to provide light and shade to the model
#try out using rgl for interactive 3d environment

#once the required database is  generated one can either use
#rgl for the display or any other software for better display
```

Bibliography

[1] M. Kafatos, X. S. Wang, Z. Li, R. Yang, and D. Ziskin, "Information technology implementation for a distributed data system serving earth scientists: seasonal to interannual ESIP," *Proceedings of SSDBM*, 1998.

[2] P.-N. Tan, M. Steinbach, and V. Kumar, "Finding spatio-temporal patterns in earth science data," NASA grant project, University of Minnesota, 2002.

[3] S. Velickov, P. D. Solomatine, X. Yu, and K. R. Price, "Application of data mining techniques for remote sensing image," *Hydroinformatics '2000 - 4th International conference on hydroinformatics*, 2000. Hydroinformatics '2000 - 4th International conference on hydroinformatics, Iowa City, USA.

[4] Q. Ding, Q. Ding, and W. Perrizo, "Association rule mining on remotely sensed images," *Database Systems users & Research Group*, 2003.

[5] S. Openshaw, "Geographical data mining: Key design issues," *Geocomputation*, no. 99, p. 14, 1999.

[6] "University of berkeley, research web site." http://www.sims.berkeley.edu/research/projects/how-much-info/summary.html, accessed on January 06, 2005, December 2004.

[7] J. Schmetz, P. Pili, S. Tjemkes, D. Just, J. Kerkmann, S. Rota, and A. Ratier, "An introduction to meteosat second generation," technical document, American Meteological Society, 14 May,2002 2002.

[8] J.Cihlar, A.Belward, and Y.Govaerts, "Meteosat second generation oppurtunities for land surface research and application," Technical Document:EUMETSAT Scientific Publication EUM SP 01, EUMETSAT, 1999 1999.

[9] A. Muzy, G. Wainer, and E. Innocenti, "Comparing simulation methods for fire spreading across a fuel bed," *Proceedings of AI, Simulation and Planning in High Autonomy Systems*, vol. AIS'2002, pp. 219–224, 2002.

[10] P. Cabena, *Discovering data mining from concept to implementation*. Prentice hall, 1998.

[11] R. .Groth, *Data mining : a hands - on approach for business professionals*. Prentice hall, 1999.

[12] H. J. Miller and J. Han, *Geographic Data Mining and Knowledge Discovery*. London and Newyork: Taylor & Francis, 2001.

[13] D. Hand, H. Mannila, and P. Smyth, *Principles of Data Mining*, vol. 1. A Bradford Book, The MIT Press, 2001.

[14] R. Stern, R. Coe, E. Allan, and I. Dale, *Good statistical practice for natural resources research*. Wallingford etc: CABI, 2004.

[15] K. Koperski, G. B. Marchisio, and K. Meyer, "Inforrmation fusion and mining from satellite imagery and gis data," report on the project, Insightful Corporation, 2002.

[16] W. Shi, P. F. Fisher, and M. F. Goodchild, *Spatial data quality*. Taylor & Francis, 2002.

[17] G. M. Foody, "Uncertainty, knowledge discovery and data mining in gis," *Progress in physical geography*, vol. vol.27, no 1, pp. 113–121(9), 2003.

[18] EUMETSAT, "Metosat second generation system overview," Technical Document EUM TD 07, ESA, 25 May,2001 2001.

[19] I. Galindo, P. López-Pérez, and M. Evangelista-Salazar, "Real-time avhrr forest fire detection in mexico (1998-2000)," *International Journal of Remote Sensing*, vol. 24, no. 2003, pp. 9–22, 2003.

[20] "Envi, environment for visulaization, web site." http://www.rsinc.com/envi/, January 2005.

[21] "S+spatialstats, web site." http://www.insightful.com/products/default.asp, January 2005.

[22] "R an open source software, web site." http://sal.agecon.uiuc.edu/csiss/Rgeo/, January 2005.

[23] "Matrix laboratory, web site." http://www.spatial-statistics.com/software_index.htm, January 2005.

[24] J. M. C.Pereira and Y. Govaerts, "Potential fire applications from msg/ seviri observations," *Technical Memorandum No.7*, 2001.

[25] European-Commission, J. San-Miguel-Ayanz, P. Barbosa, G. Libert, G. Schmuck, and E. S. (Eds), "Forest fires in europe, 2002 fire campaign," Tech. Rep. SPI.03.83.EN, Official Publication of the European Communities, 2003.

[26] "Govermental non profit organisation for research development on global change particularly in mediterranean basin and subtropical africa, web site." http://medias.obs-mip.fr:8000/igac/html/book/chap4/chap4.html, January 2005.

[27] "Scripps institution of oceanography, ucsd, web site." http://meteora.ucsd.edu/s2k/s2k_showcase/iacobellis1.html, January 2005.

[28] "Nasa web site." http://sky.arc.nasa.gov/chatfield/Trace3D_Paper.html, December 2004.

[29] "DLR Germany, web site." http://www.caf.dlr.de/caf/anwendungen/umwelt/feuer/;internal&action=_setlanguage.action?LANGUAGE=en, December 2004.

[30] Y. Rauste, A. Sephton, V. Kelh, T. Vainio, M. Heikinheimo, K. Soini, O. Frauenberger, and J. S. Miguel-Ayanz, "Esa forest fire operational study, requirements and analysis report (rar)," Report Version 2.3/ AO/1-3468/98/I-DC, VTT Automation in collaboration with Finnish Ministry of the Interior, Finnish Meteorological Institute (FMI), Space Systems Finland and EC Joint Research Centre, 24.11.1999 1999.

[31] "Canadian wildfire fire information system, web site." http://fms.nofc.cfs.nrcan.gc.ca/en/background/bi_FM3_Intro_e.php, January 2005.

[32] "Cooperative institute for meteorological satellite studies space science and engineering center, university of wisconsin–madison, web site." http://fms.nofc.cfs.nrcan.gc.ca/en/background/bi_FM3_Intro_e.php, January 2005.

[33] "Advanced fire information system, international strategy for disaster reduction, web site." http://www.fire.uni-freiburg.de/GlobalNetworks/Africa/Afrifirenet_4e.html, January 2005.

[34] J. Ameghino, "Models of complex physical systems using cell–DEVS."

[35] "Fire research, national institute of standards and research." http://www.bfrl.nist.gov/866/fmabbs.html, January 2005.

[36] E. Lambin, K.Goyvaerts, and C.Petit, "Remotely-sensed indicators of burning efficiency of savannah and forest fires," *International Journal of Remote Sensing*, vol. 24, no. 15, pp. 3105–3118, 2003.

[37] L. Giglio and C. O.Justice, "Effect of wavelength selection on characterization of fire size and temperature," *International Journal of Remote Sensing*, vol. 24, no. 17, pp. 3515–3520, 2003.

[38] C. Mbow, K. Goita, and G. B. Benie, "Spectral indices and fire behavior simulation for fire risk assessment in savanna ecosystems," *Remote Sensing of Environment*, vol. 91, no. 2004, pp. 1–13, 2004.

[39] N. Koutsias and M. Karteris, "Classification analyses of vegetation for delineating forest fire fuel complexes in a mediterranean test site using satellite remote sensing and gis," *International Journal of Remote Sensing*, vol. 24, no. 15, pp. 3093–3104, 2003.

[40] "News article, web site." http://www.fire.uni-freiburg.de/GFMCnew/2004/0729/20040729_southern_europe.htm, January 2005.

[41] J. Shawe-Taylor and N. Cristianini, *Kernel Methods for Pattern Analysis*. Cambridge, United Kingdom: Cambridge University Press, 2004.

[42] A. W.Bowman and A. Azzalini, *Applied Smoothing Techniques for Data Analysis*. Oxford Statistical Science Series, New York, USA: Clarendon Press. Oxford, 1997.

[43] "Online dictionary." http://en/wikipedia.org/wiki/Convolution, January 2005.

[44] D. higdon, "Space and space-time modeling using process convolution," *Quantitative methods for current environmental issues*, pp. 37–56, 2002.

[45] M. P. Wand and M.C.Jones, *Kernel Smoothing*. Monographs on Statistics and Applied Probability, London, UK: Chapman & Hall, 1995.

[46] J. Williams, *Geographic Information from space*. Wiley–Praxis series in remote sensing, Canada: John Wiley & Sons in association with Praxis publishing ltd, 1995.