

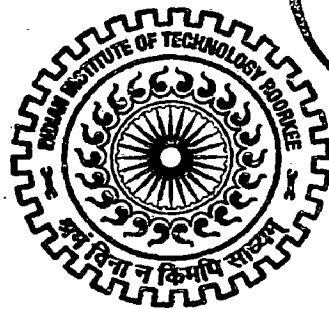
CLOUD DETECTION IN SATELLITE IMAGERY (AVHRR)

A DISSERTATION

Submitted in partial fulfilment of the
requirements for the award of the degree
of
MASTER OF TECHNOLOGY
in
HYDROLOGY

By

PRADNYESH RAMESH WAGHMARE



DEPARTMENT OF HYDROLOGY
INDIAN INSTITUTE OF TECHNOLOGY ROORKEE
ROORKEE-247 667 (INDIA)

JUNE, 2006

CANDIDATE'S DECLARATION

I hereby certify that the work which is being presented in this dissertation entitled "**CLOUD DETECTION IN SATELLITE IMAGERY (AVHRR)**" in partial fulfillment of the requirement for the award of the degree of **Master of Technology in Hydrology**, submitted in the Department of Hydrology of Indian Institute of Technology Roorkee is an authentic record of my work carried out during the period from July, 2005 to June, 2006 under the supervision of **Dr. Dharmendra Singh, Asstt. Professor**, Department of Electronics & Computer, and **Dr.B.S.Mathur, Professor**, Department of Hydrology Indian Institute of Technology, Roorkee.

The matter embodied in this dissertation has not been submitted by me for the award of any other degree or diploma.



(Pradnyesh Ramesh Waghmare)

Candidate's Signature

Dated: 30-06-06

This is to certify that the above statement made by the candidate is correct to the best of our knowledge.



(Dr.Dharmendra Singh)

Asstt. Professor,

Department of Electronics & computer
Indian Institute of Technology,
Roorkee



(Dr.B.S.Mathur)

Professor,

Department of Hydrology
Indian Institute of Technology,
Roorkee

ACKNOWLEDGEMENT

With great pleasure, I express my sincere and profound gratitude to my supervisors, **Dr. B.S.Mathur**, Professor, Hydrology Deptt. and **Dr. Dharmendra singh**, Asst. professor, Electronic and Computer Deptt. Indian Institute of Technology, Roorkee. for their erudite and invaluable guidance throughout this dissertation. I am highly obliged to Dr. B.S.Mathur and Dr. Dharmendra Singh for sparing their valuable time in giving me necessary training, suggestions and providing me the necessary facilities for the experiment.

I am also thankful to all the faculty members of Hydrology department for polishing my knowledge and guiding me like their own child.

Sincere thanks are extended to the staff of the M.tech. Lab. Of Hydrology Deptt. Indian Institute of Technology, Roorkee for the assistance rendered throughout this work. **Mr. Rohit singh** had been extremely helpful for making this programming a complete success.

I wish to extend my special thanks to my parents, my brother, and my sister for help, co-operation and my source of inspiration at every stage of my dissertation work.

I am also grateful to the library staffs especially **Mr. Sharma** for their help in compiling all the literatures required in my report.

I would like to thank MHRD for providing me scholarship for my M.Tech. Program. I would like to thank also **IIT Roorkee** for providing all necessary financial and academic support without which this study would have been a distant dream.

I cannot conclude my acknowledgements without expressing my indebtedness to my friends and Classmates Sanjay, Dhruv, Nehal, Kailash, Prabhakar, S.K.Gangwar, M Rijal, V. Nair, S.N. Yadav, R. Sinha, and many others for their affectionate cooperation and support while staying at Roorkee.

Dated: 30/06/06

Place: Roorkee



Pradnyesh Ramesh Waghmare

M.Tech. (Hydrology)

IIT, Roorkee

ABSTRACT

Jharia is one of the main coal field regions of India where the coal fire burning is the major problem. The present dissertation tries to attempt to detect and monitor cloud and coal fire regions of Jharia by using various image processing technique and application of thresholding cloud detection algorithm with operational Satellite Data (i.e. AVHRR). A Thresholding and Multithresholding algorithm based on Digital Image Processing is developed which is capable to detect cloudy pixels in cloud the hot/fire spots of the study area. The observed suspected regions were highlighted and gave quite good agreement with the results obtained by other researchers. The patterns of different cloud of last 10 years has been observed which provides information of cloudy pixels and tells about any change in or near the cloudy pixels in cloud. The Thresholding technique is applied to reduce the false area location of cloudy surface. Various computer codes (i.e. Calibration, Thresholding, Multi-Threshold, Segmentation, etc) for satellite image processing have been developed. Finally, the study will be quite helpful to develop automated cloudy pixels and hotspot detection software of study areas for AVHRR data.

TABLE OF CONTENTS

CERTIFICATE

ACKNOWLEDGEMENT

ABSTRACT

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

Chapter 1	Introduction	
	1.1 Introduction	1
	1.2 Problem definition	5
	1.3 Aim and Objective	5
	1.4 Types of Satellite data	6
	1.4.1 MODIS	6
	1.4.1.1. MODIS Bands	7
	1.4.2. LANDSAT	9
	1.4.2.1. Automatic cloud cover assessment	10
	1.4.3 MISR	10
	1.4.4. CERES	11
	1.4.5. ASTER	12
	1.4.5.1. Converting Level-1A DNs to Scaled DNs	13
	1.4.6. AVHRR	13

1.4.6.1. Data Formats	14
1.4.6.2. Downloading the data	16
1.4.6.3. Spectral characteristics AVHRR's 5 channels	17

Chapter 2 LITERATURE REVIEW

2.1 Cloud détection techniques	18
--------------------------------	----

Chapter 3 METHODOLOGY

3.1 Algorithm description	28
3.1.1. Introduction	28
3.2 Preprocessing	28
3.2.1. Reprojection	28
3.2.2. Calibration	28
3.3. Change Detection using Image analysis	29
3.4. Multi-thresholding using Image processing	29
3.4.1. Thresholding	29
3.5 algorithm	30
3.5.1. The AVHRR Split-and Merge Clustering algorithm	30
3.5.1.1 Split-and-Merge Iterative Clustering	31
3.5.1.2. Labelling	37
3.5.1.3. The Uncertainty criteria	40
3.5.2. Daytime algorithm	42
3.5.2.1. Detection of cloud-free pixels	42
3.5.3. Contextual algorithm	47

3.5.3.1.	Contextual approach	47
Chapter 4.	RESULT AND DISCUSSION	
4.1	Introduction	48
4.2	Change Detection in Image Analysis	53
4.3	Multi-Thresholding in Image Processing	61
Chapter 5	CONCLUSION AND FUTURE SCOPE	62
References		
Appendix A - Source code		

LIST OF FIGURES

Figure number and name

Figure 3.1 AVHRR Split-and-merge algorithm

Figure 3.2 definition of decision plane

Figure 3.3 Definition of the ambiguous domain used in the labeling

Figure 3.4 Day time Algorithm

Figure 4.1 Processing Stages of Cloud detection

Figure 4.2 band 1(Sept 95)

Figure 4.3 band 2 (Sept 95)

Figure 4.4 band 4 (Sept 95)

Figure 4.5 Color mapping of cloud

Figure 4.6 band 4 (Feb95)

Figure 4.7 band 4 (Sept 97)

Histogram 4.1. June data of Split and Merge Clustering and Daytime Algorithm

Line graph 4.1. June data of Split and Merge Clustering and Daytime Algorithm.

Histogram 4.2. July data of Split and merge Clustering and Daytime algorithm

Line graph 4.2. July data of Split and Merge Clustering and Daytime Algorithm

Histogram 4.3. Sept data of Split and merge Clustering and Daytime algorithm

Line graph 4.3. Sept data of Split and Merge Clustering and Daytime Algorithm

Histogram 4.4. June data of Justice and Flasse-ceccato (contextual) Algorithm.

Line graph 4.4. June data of Justice and Flasse-ceccato (contextual) Algorithm.

Histogram 4.5. July data of Justice and Flasse-ceccato (contextual) Algorithm.

Line graph 4.5. July data of Justice and Flasse-ceccato (contextual) Algorithm

Histogram 4.6. Sept data of Justice and Flasse-ceccato (contextual) Algorithm.

Line graph 4.6. Sept data of Justice and Flasse-ceccato (contextual) Algorithm

LIST OF TABLES

Table Number and Name

Table 1.1 MODIS band

Table 1.2 Data Formats of AVHRR

Table 1.3 spectral characteristics of AVHRR 5 channels

Table 4.1. June data of Split and Merge Clustering and Daytime Algorithm.

Table 4.2. July data of Split and Merge Clustering and Daytime Algorithm.

Table 4.3. Sept data of Split and Merge Clustering and Daytime Algorithm.

Table 4.4. June data of Justice and Flasse-ceccato (contextual) Algorithm.

Table 4.5. July data of Justice and Flasse-ceccato (contextual) Algorithm.

Table 4.6. Sept data of Justice and Flasse-ceccato (contextual) Algorithm.

LIST OF ABBREVIATIONS

- **NOAA/AVHRR** National Oceanic and Atmospheric Administration/
Advanced Very High Resolution Radiometer
- **MODIS** Moderate Resolution Imaging Spectroradiometer
- **MISR** Multi-angle Imaging Spectroradiometer
- **ASTER** Advance spaceborne thermal Emmission and Reflectance
Radiometer.
- **HRPT** High Resolution Picture Transmission
- **HDF** Hierarchical Data Format
- **ANN** Artificial Neural Network
- **VI** Vegetation Indices
- **LST** Land Surface Temperature

1.1. Introduction

Hydrology means science of water. Every drop of water comes from clouds. Cloud is primary source of water. Clouds are masses of condensed water vapor which are visible signs of atmospheric processes at work. Clouds help regulate the earth's energy balance by reflecting and scattering solar radiation and by absorbing the earth's infrared radiation. In addition, clouds help redistribute surplus heat from the equator toward the poles and return water (in the form of precipitation) to the oceans and land masses across the globe. Clouds are essential to the earth-atmosphere system.

Clouds complete the following functions:

1. Clouds help regulate Earth's energy balance by reflecting and scattering solar radiation and by absorbing Earth's infrared energy.
2. Clouds are required for precipitation to occur and, hence are an essential part of the Hydrological cycle.
3. Clouds indicate what type of atmospheric processes are occurring (e.g., cumulus clouds indicate surface heating and atmospheric turbulence).
4. Clouds help redistribute extra heat from the equator toward the poles. Clouds are masses of condensed water vapor.
5. Clouds are formed when water vapor is condensed into liquid water (cloud droplets).

There are three basic requirements for clouds to occur:

1. Water vapor must be present in sufficient amounts so that saturation can be reached by some means.
2. Cloud condensation nuclei (CCN) must be present to provide a surface on which water will condense. (Examples of CCN include dust in the air from

the earth's surface, salt particles from the sea, combustion products, and volcanic or meteorite dust.)

3. Cooling mechanism is required to cool the air temperature to the dew point temperature.

The requirements for sufficient water vapor and CCN are rarely the limiting factors for cloud development. Typically, the limiting factor is a cooling mechanism. Thus, the air temperature needs to be lowered to the dew point temperature for a cloud to form. If a cooling mechanism is not present, clouds will not form. In order to predict the climate several decades into the future; we need to understand many aspects of the climate system, one being the role of cloud in determining climate sensitivity. Clouds affect climate, and are in turn affected by changes in the climate. The relationship is a complicated system of feedbacks, in which clouds modulate Earth's radiation balance.

Everything, from an individual person to Earth as a whole, emits energy. Scientists refer to this energy as radiation. As Earth absorbs incoming sunlight, it warms up. The planet must emit some of this warmth into space or increase in temperature. Two components make up the Earth's outgoing energy: heat (or thermal radiation) that the Earth's surface and atmosphere emit; and sunlight (or solar radiation) that the land, ocean, clouds and aerosols reflect back to space. The balance between incoming sunlight and outgoing energy determines the planet's temperature and, ultimately, climate. Both natural and human-induced changes affect this balance, called the Earth's radiation budget.

Earth's radiation budget is a balance between incoming and outgoing radiation. Clouds affect the radiation budget directly by reflecting sunlight into space (cooling the Earth) or absorbing sunlight and heat emitted by the Earth. When clouds absorb sunlight and heat, less energy escapes to space and the planet warms. To understand how clouds impact the energy budget, scientists need to know the composition of cloud particles, the altitude of clouds and the extent to which clouds at different altitudes overlap each other. Therefore, it is an important aspect to know the percentage of cloud cover over land surfaces

in particular area for a lot of climate modeler. In this direction, satellite images can help a lot.

In recent studies, the image analysis techniques were successfully used for detecting cloud, environmental hazards and extracting changes in the features on the earth surfaces. Environmental hazards like drought (Rarnesh et al 2003), forest fire (Domenikiotis et al 2003), coal fires (Vekerdy et al 1999, Gangopadhyay et al 2005), general fires (Louis et al 2003, Pereira 1999), landslides, floods (Domenikiotis et al 2003), and Automated cloud detection (Gallaudet, T.C., and Simpson, J.J. 1991) etc. were predicted using different image analysis methodologies like thresholding, neural networks, change detection, subpixel analysis etc. using different satellites like NOAA/AVHRR (National Oceanic and Atmospheric Administration! Advanced Very High Resolution Radiometer) MODIS (Moderate Resolution Imaging Spectroradiometer), Landsat TM (Thematic Mapper), and TIROS (Television Infrared Observation Satellite) etc. In similar way, changes in features like vegetation cover (Weiss et al 2001), soil moisture, evapotranspiration (Fabien et al 2000), water properties (Rokade et al 2004) etc. was also analyzed and the information were extracted from the satellite data.

Image analysis or image processing is the area of recognition of individual regions or cloud in an image. The different processes are applied to extract the attributes from the image and recognition of individual cloudy pixels. The area of image processing are so varied that it is needed to categorize images. Based on their source their source these are categorized as imaging using —X-rays Bands, Microwave Bands, Radio Bands, Visible Bands etc. The Bands used in the present dissertation lies in Infrared, Near Infrared, visible and Mid wave Infra-Red range.

Some of the fundamental steps involved in the processing of the image includes: *Image acquisition, Image enhancement, Image restoration, Color image processing, Wavelets, Segmentation, Representation and Description, Recognition etc.* Image acquisition is the basic step which involves preprocessing of the image such as scaling, calibration etc. Next step is the enhancement of an image which is done to highlight some important features such as contrast, brightness etc. Color processing is done to segment more than one feature in an image at the same time. The images of various degrees of

resolution are the key foundation of the wavelet theory. Segmentation is done to identify objects in an image individually. Description is the process of feature selection which deals with the extraction of attributes from a segmented image. Lastly, Recognition deals with assigning of Labels to object based descriptors. These processing techniques are used together and separately to trace the meaningful patterns from the image.

(S.E. Marsh et al 1992) came with his concept of processing technique of vectors and mapping which he employed on the image to have a study on soil property. (L. Olsson et al 1994) worked on Fourier series and its derivatives (the base of image processing in frequency domain) for study of temporal sequence of satellite imagery. Similarly a lot of research works are being carried out in study of movements of cloud detection, thunder storm, forest fires, bush fires, coal fires detection etc.

Cloud detection and its monitoring, which is an existing challenge, can also be detected and its features can be extracted from the satellite image using image processing techniques. Features and patterns of cloud, which are not visible can also be detected using these image analysis techniques.

Clouds have occurred in almost all parts of the world like India, the US, Indonesia, South Australia, China, Germany and many other countries. However, the nature and magnitude of the clouds differs from country to country. Because monsoon season is totally dependant on the geographical location (latitude and longitude), as well as wind pattern. In India, monsoon season possibly four months (June to September) and some parts of India occurs clouds during month of December (in Tamilnadu).

In present project, we have taken the Jharia region for observing the cloud cover over various seasons with satellite images. Jharia coalfield in Bihar is an exclusive storehouse of prime coke coal in the country, consisting of 23 large underground and nine large open cast mines. The mining activities in coalfields started in 1894 and had really intensified in 1925. The Jharia Coalfields is located in the Dhanbad district of Bihar state, and is named after the main mining Jharia. It is situated at the heart of the Damodar valley and is about 250km NW of and about 1150km SE of Delhi. It is confined between the latitudes $23^{\circ}38'$ N and $23^{\circ}50'$ N and longitudes $86^{\circ}07'$ E and $86^{\circ}30'$ E. The maximum length is about 38 km from E to W and 19 km from N to S. The area covered by the coal belt is about 450 km². Many clouds were detected until now and it is very difficult to

monitor or observe the proper change cloud. The field measurements are very cumbersome and expensive. Thus there is an urgent need to use satellite data and image analysis techniques to observe clouds over this region .

The operational satellite data has been used for the image' analysis is from NOAA/AVHRR (National Oceanic and Atmospheric Administration/ Advanced Very High Resolution Radiometer) .The data are in level 1 B format and are of 1km Resolution. These satellite data from earth observation sensor systems is widely used in a range of oceanographic, terrestrial, and atmospheric applications, such as land cover mapping, environmental modeling and monitoring, predictions plenty of work has been done to extract the features from the image in the field of image processing using different image analysis techniques. Some of them includes change detection, Thresholding, segmentation, color mapping .

1.2. Problem definition

The difficulty is to continuously monitor the area due to its tediousness and cost factor. Nature of the cloud is uneven and not continuous. Thus problem that is focused in present dissertation is to remove this difficulty and develop an economical and efficient technique for observing of cloud in Jharia region. For this purpose various image analysis techniques has to be applied on operational satellite data from NOAA/AVHRR satellite for features and recognizing the pattern.

1.3 Aim and objective

Aim of this dissertation is to develop an approach for observe clouds in Jharia region with operational satellite data (NOAA/AVHRR) using image analysis technique.

To achieve this aim, the objectives are defined as

- 1 Study of NOAA/AVHRR which has 1 Km.Sq. resolution.
- 2 Impact of various bands on the clouds.
- 3 Algorithm to be developed for detecting the clouds.

4. Assessment of percentage cloudy pixel in whole images.
5. Specification for clouds detection from satellite image to the future users.

1.4 Types of Satellite data

1.4.1 MODIS (Moderate Resolution Imaging Spectroradiometer)

Modis is a key instrument aboard the Terra (EOS AM) and Aqua (EOS PM) satellites. Terra's orbit around the Earth is timed so that it passes from north to south across the equator in the morning, while Aqua passes south to north over the equator in the afternoon. Terra MODIS and Aqua MODIS are viewing the entire Earth's surface every 1 to 2 days, acquiring data in 36 spectral bands, or groups of wavelengths (see MODIS Technical Specifications). These data will improve our understanding of global dynamics and processes occurring on the land, in the oceans, and in the lower atmosphere. MODIS is playing a vital role in the development of validated, global, interactive Earth system models able to predict global change accurately enough to assist policy makers in making sound decisions concerning the protection of our environment.

The MODIS instrument is operating on both the Terra and Aqua spacecraft. It has a viewing swath width of 2,330 km and views the entire surface of the Earth every one to two days. Its detectors measure 36 spectral bands between 0.405 and 14.385 μm , and it acquires data at three spatial resolutions -- 250m, 500m, and 1,000m.

The many data products derived from MODIS observations describe features of the land, oceans and the atmosphere that can be used for studies of processes and trends on local to global scales. As just noted, MODIS products are available from several sources. MODIS Level 1 and atmosphere products are available through the Goddard DAAC at the Goddard Space Flight Center (GSFC) in Greenbelt, MD. Land Products are available through the Land Processes DAAC at the U. S. Geological Survey EROS Data Center (EDC). Cryosphere data products (snow and sea ice cover) are available from the National Snow and Ice Data Center (NSIDC) in Boulder, Colorado. Ocean color products and sea surface temperature products along with information about these products are obtainable at the OCDPS at GSFC. Users with an appropriate x-band

receiving system may capture regional data directly from the spacecraft using the MODIS Direct Broadcast signal.

Table 1.1 MODIS band

Orbit:	705 km, 10:30 a.m. descending node (Terra) or 1:30 p.m. ascending node (Aqua), sun-synchronous, near-polar, circular
Scan Rate:	20.3 rpm, cross track
Swath	2330 km (cross track) by 10 km (along track at nadir)
Dimensions:	
Telescope:	17.78 cm diam. off-axis, afocal (collimated), with intermediate field stop
Size:	1.0 x 1.6 x 1.0 m
Weight:	228.7 kg
Power:	162.5 W (single orbit average)
Data Rate:	10.6 Mbps (peak daytime); 6.1 Mbps (orbital average)
Quantization:	12 bits
Spatial	250 m (bands 1-2)
Resolution:	500 m (bands 3-7) 1000 m (bands 8-36)
Design Life:	6 years

Primary Use	Band	Bandwidth1	Spectral Radiance2	Required SNR3
Land/Cloud/Aerosols	1	620 - 670	21.8	128
Boundaries	2	841 - 876	24.7	201
Land/Cloud/Aerosols	3	459 - 479	35.3	243
Properties	4	545 - 565	29.0	228
	5	1230 - 1250	5.4	74

	6	1628 - 1652	7.3	275
	7	2105 - 2155	1.0	110
Ocean Color/ Phytoplankton/ Biogeochemistry	8	405 - 420	44.9	880
	9	438 - 448	41.9	838
	10	483 - 493	32.1	802
	11	526 - 536	27.9	754
	12	546 - 556	21.0	750
	13	662 - 672	9.5	910
	14	673 - 683	8.7	1087
	15	743 - 753	10.2	586
	16	862 - 877	6.2	516
Atmospheric Water Vapor	17	890 - 920	10.0	167
	18	931 - 941	3.6	57
	19	915 - 965	15.0	250

Primary Use	Band	Bandwidth1	Spectral Radiance2	Required NE[delta]T(K)4
Surface/Cloud Temperature	20	3.660 - 3.840	0.45(300K)	0.05
	21	3.929 - 3.989	2.38(335K)	2.00
	22	3.929 - 3.989	0.67(300K)	0.07
	23	4.020 - 4.080	0.79(300K)	0.07
Atmospheric Temperature	24	4.433 - 4.498	0.17(250K)	0.25
	25	4.482 - 4.549	0.59(275K)	0.25
Cirrus Clouds Water Vapor	26	1.360 - 1.390	6.00	150(SNR)
	27	6.535 - 6.895	1.16(240K)	0.25
	28	7.175 - 7.475	2.18(250K)	0.25
Cloud Properties	29	8.400 - 8.700	9.58(300K)	0.05
Ozone	30	9.580 - 9.880	3.69(250K)	0.25
Surface/Cloud	31	10.780 - 11.280	9.55(300K)	0.05

	32	11.770 - 12.270	8.94(300K)	0.05
Cloud Top	33	13.185 - 13.485	4.52(260K)	0.25
Altitude	34	13.485 - 13.785	3.76(250K)	0.25
	35	13.785 - 14.085	3.11(240K)	0.25
	36	14.085 - 14.385	2.08(220K)	0.35

¹ Bands 1 to 19 are in nm; Bands 20 to 36 are in μm

² Spectral Radiance values are $(\text{W}/\text{m}^2 \cdot \mu\text{m}\cdot\text{sr})$

³ SNR = Signal-to-noise ratio

⁴ $\text{NE}(\Delta T)$ = Noise-equivalent temperature difference

Note: Performance goal is 30-40% better than required

1.4.2. LANDSAT

The Landsat Mission, initiated in 1972 with the launch of the Landsat 1 satellite, began as an experiment to determine the usefulness of Earth orbiting spacecraft in providing land usage data as well as future land and ocean development data. The first Landsat satellites proved to be extremely useful in collecting data relating to climatic change. Overtime, the instruments onboard these satellites have been improved, and have enhanced the quality and quantity of the data the satellites have collected. The newest Landsat satellite, Landsat 7 which was launched in April of 1999, is continuing the Landsat mission to collect data in order for scientists to increase knowledge on the causes of the changes of the Earth.

Regardless of the source of satellite images, clouds often obstruct such images. Since clouds are prominent features of an image, the proper identification of clouds is important for two reasons. Larger images of the Earth's surface are formed from smaller images taken by satellites with a process known as Image Registration. In this process, prominent features from multiple satellite images are matched to these features in other images. The coinciding features are then use to align one image to another, resulting in the creation of a larger image. Unfortunately, clouds can mask prominent features and can also be mistaken with prominent features from other images. As a result,

inaccurately assessed features such as clouds may then be used to orient one image with another resulting in misaligned images. In order to align images of the same area more accurately, clouds need to be removed so that only valid land features are used to match images.

1.4.2.1. Automatic cloud cover assessment

Landsat 5 was the first of the Landsat satellites generations to use a cloud detection system, known as Automatic Cloud Cover Assessment (ACCA) for image processing. It was inaccurate in assessing cloud cover because the algorithms used identified some landmasses as clouds because of the similar characteristics. A second cloud detection system was created for image processing for Landsat 7, which included improved algorithms to detect clouds more accurately. As shown in **Error! Reference source not found.**, the improved ACCA algorithm used for the processing of the Landsat 7 images discriminates between clouds and difficult terrain features better than the algorithm used for the Landsat 5 satellite.

1.4.3. Multi-angle Imaging SpectroRadiometer (MISR) imagery.

With the launch of the Multi-angle Imaging SpectroRadiometer (MISR) onboard the National Aeronautics and Space Administration (NASA) Terra satellite in 1999 novel electromagnetic radiation measurements made at nine angles became available for scientific study. The MISR sensor consists of nine cameras (Figure 1), with each camera viewing Earth scenes at a different angle in four spectral bands (blue, green, red, and near-infrared). The view zenith angles of the nine cameras are 70.5° (Df), 60° (Cf), 45.6° (Bf), and 26.1° (Af) in the forward direction, 0.0° (An) in the nadir direction and 26.1° (Aa), 45.6° (Ba), 60° (Ca) and 70.5° (Da) in the aft direction. The “f” in the letter designation of the cameras represents the “forward” direction and the “a” represents the “aft” direction. The Da camera collects data from a location seven minutes after the Df camera. The nominal resolution of the MISR radiances is 275 m by 275 m at the Earth’s surface, but the blue, green and near-infrared radiances are aggregated onboard to a 1.1 km by 1.1 km resolution to reduce data transmission from the Terra satellite.

The MISR cameras cover a swath at the Earth's surface that is approximately 360 km wide and extends across the daylight side of the Earth from the Arctic down to Antarctica in about 40-45 minutes. There are 233 geographically distinct, but overlapping, MISR swaths, which are also called paths. MISR collects data from all paths on a repeat cycle of 16 days; that is, it covers the exact same path every 16 days. In the MISR data products each path is subdivided into 180 blocks, with the block numbers increasing from the north to south pole. Each complete trip of MISR around the earth is given its own orbit number.

However, the MISR operational cloud detection algorithms were designed before MISR was launched and they were not particularly targeted at detecting clouds over bright surfaces in polar regions. As a result, MISR operational algorithms do not work well over polar regions. Moreover, one orbit of data needs to be processed before the next orbit comes, so an operational cloud detection algorithm is necessary. The massive MISR data size (MISR collects 3.3 megabits per second on average and 9.0 megabits per second at peak) poses a hurdle for operational processing.

Especially of cloud coverage as clouds play an important role in modulating the sensitivity of the Arctic to increasing surface air temperatures. Ascertaining the properties of clouds in the Arctic is a challenging problem because liquid and ice-water cloud particles often have similar scattering properties to the particles that compose ice- and snow-covered surfaces. As a result, the amount of visible and infrared electromagnetic radiation emanating from clouds and snow- and ice-covered surfaces is often similar, which leads to problems in the detection of clouds over these surface types.

1.4.4. CERES

The Clouds and the Earth's Radiant Energy System (CERES) experiment is one of the highest priority scientific satellite instruments developed for EOS. CERES products include both solar-reflected and Earth-emitted radiation from the top of the atmosphere to the Earth's surface. Cloud properties are determined using simultaneous measurements by other EOS instruments such as the Moderate Resolution Imaging Spectroradiometer (MODIS). Analyses of the CERES data, which build upon the foundation laid by previous missions such as the Earth Radiation Budget Experiment

(ERBE), will lead to a better understanding of the role of clouds and the energy cycle in global climate change.

CERES instruments were launched aboard the Tropical Rainfall Measuring Mission (TRMM) in November 1997 and on the EOS Terra satellite in December 1999. Two additional instruments will fly on the EOS Aqua spacecraft in 2002. Multiple satellites are needed to provide adequate temporal sampling since clouds and radiative fluxes vary throughout the day. The first 24 months of CERES data collected on both TRMM and Terra demonstrate that the CERES instruments are substantially improved over the ERBE instruments. The CERES data show lower noise, improved ties to the ground calibration in absolute terms, and smaller fields of view. CERES instrument calibration stability on TRMM and Terra is typically better than 0.2%, and calibration consistency from ground to space is better than 0.25%. Onboard calibration sources provide traceability of the measurements to the International Temperature Scale of 1990 at the 0.2% level. Such levels of accuracy have never before been achieved for radiation budget instruments .

1.4.5. ASTER(Advance spaceborne thermal Emmission and Reflectance Radiometer).

The ASTER instrument aboard the Terra satellite has a complement of three different telescopes with varying pointing capabilities. ASTER data are primarily acquired and processed in support of a global mapping mission, and based on Data Acquisition Requests (DARs) from authorized users. ASTER Level-1 data are produced at the Ground Data System (GDS) facility of the Earth Remote Sensing Data Analysis Center (ERSDAC) in Tokyo, Japan, and subsequently sent to the Land Processes (LP) Distributed Active Archive Center (DAAC) for archiving, distribution, and higher-level product generation. The ASTER Level-1A data set contains reconstructed, instrument digital numbers (DNs) derived from the telemetry streams of the 3 telescopes: Visible Near Infrared (VNIR), Shortwave Infrared (SWIR), and Thermal Infrared (TIR). The Level-1A data contains depacketized, demultiplexed, and realigned instrument image data with their geometric correction coefficients and radiometric calibration coefficients calculated and appended but not applied. It also includes corrections for the SWIR

parallax, and intra- and inter-telescope registration information. The EOS Data Gateway (EDG) provides GDS-produced Level-1A browse images for each of the three sensors, and allows users to view and/or request these images.

1.4.5.1 Converting Level-1A DNs to Scaled DNs

An ASTER Level-1A data set contains raw digital numbers (DNs) quantized as 8-bit unsigned integers. To convert these to scaled or calibrated DNs, the ASTER Level-1A DNs are converted on a detector-by-detector basis using the Slope/Inclination (A), Gain (G) and Offset (D) values from the Radiometric Conversion Coefficients (RCC) table that is appended with the Level-1A data set in the HDF-EOS file. The RCC information determines how the Level-1A DNs are converted into Level-1B calibrated DNs. The gain and offset information in the RCC table are used in that conversion

1.4.6 . AVHRR (Advance Very High Resolution Radiometer)

The National oceanic and Atmospheric Agency(NOAA) operates two types of satellites intended for studying meteorology. These satellites collect electromagnetic radiation in the visible light and infrared portions of the spectrum. Geostationary meteorological satellites stay in one place above the Earth, by orbiting with the Earth as it rotates, and provide continuous information for the area they can see. Polar-orbiting satellites are designed to provide complementary information to the geostationary meteorology satellites. They have a sun – synchronous, near-polar orbit, which means that a satellite travel from the North Pole to the South Pole as the earth rotates below it. The significance of a sun-synchronous orbit is that the satellite passes over the same part of the earth at approximately the same local time each day, ensuring comparable daylight condition over time.

The primary sensor on board the NOAA polar-orbiting satellite is the AVHRR (Advance Very High Resolution Radiometer). The AVHRR instrument consists of an array of small sensors that record (as digital number) the amount of visible and infrared radiation reflected and emitted from the earth surface. This provides images of the earth's surface showing elements that cannot normally be viewed with the human eye.

Initially, the NOAA/AVHRR satellites were designed to observe the earth's weather in the form of cloud patterns. However, further research on the sensors clearly demonstrated that they could be used for more than just monitoring weather phenomena. There are 5 sensors or "channels"(bands) on board each AVHRR satellite, each designed to record information from a different part of the earth

The AVHRR sensor provides for global (pole to pole) on board collection of data from all spectral channels. Each pass of the satellite provides a 2399 km (1491 mi) wide swath. The satellite orbits the Earth 14 times each day from 833 km (517 mi) above its surface. It provides four to six band multispectral data from the NOAA polar-orbiting satellite series. It is providing the global coverage since June 1979, with morning and afternoon acquisitions available. The resolution is 1.1 kilometer at nadir.

1.4.6.1. Data Formats

The AVHRR data is saved in the common HRPT (high resolution picture transmission) format. There are three types of AVHRR data: High Resolution Picture Transmission (HRPT), Local Area Coverage (LAC) and Global Area Coverage (GAC). HRPT data are full resolution (1 km) image data that are transmitted to a ground station as they are collected (in real time). LAC are also full resolution data, but are recorded with an on-board tape recorder over selected areas for subsequent transmission to a ground station during the next overpass. GAC data are low resolution data (4 km) that provide sub sampled global coverage recorded on the satellite's tape recorders which are then transmitted to a ground station.

These instruments are carried on various satellites such as TIROS-N and NOAA-11 satellite. The AVHRR 5 channel scanning radiometer with 1.1 Km resolution is sensitive in the visible 7 near infra red and infra red "window" region. This instrument will be carried through NOAA-J(14), NOAA-K, L, M (15, 16, 17) and will have similar instrument with six channels and other improvements.

Record	Byte	Bytes	contents
1	1-2	2	Scan the line
	3-8	6	Time code
	9-12	4	Quality indicator
	13-52	40	Calibration coefficient
	53	1	Number of meaningful zenith angles and earth location points appended to scan
	54-104	51	Solar zenith angles.
	105-308	204	Earth location.
	309-448	140	Telemetry header
	449-7400	6952	LAC/HRPT video data
2	1-6704	6704	LAC/HRPT video data
	6705-6724	20	Additional decimal portion of 51 solar zenith angles.
	6725-6726	2	Clock drift delta in millisec X 2 + indicator 0 → no time adjustment 1 → time adjustments
	6727-7400	674	Space

1.4.6.2. Downloading the data

NOAA/AVHRR data is easily available on: <http://www.saa.noaa.gov>

Step1 Log on to the above website and sign in as a registered user.

Step2 Specify the AVHRR data from the list of product given.

Step3 Specify the dates you wish in the “Choose a date/time” box.

Step4 Select the correct spatial range of the latitude and longitude i.e.

Latitude: West - 85 East - 87

Longitude North -24.5 South -22.5 :

Step5 Click on the “search” button.

Step6 Select the image and place it in the CART specifying the FTP/pull to download data option.

Step7 Place the order by giving some basic information.

The following is the Band details of NOAA AVHRR data whose sensing is done in the visible near-infrared, and thermal infrared portions of the electromagnetic spectrum.

1.4.6.3 The spectral characteristics of these 5 channels are

Band Number	NOAA Satellites: 6,8,10	NOAA Satellites: 7,9,11,12,14	NOAA Satellites: 15,16,17	Location in spectrum	Primary use
1	0.58 – 0.68	0.58 – 0.68	0.58 – 0.68	Visible Red	Daytime cloud / surface mapping
2	0.725 – 1.10	0.725 – 1.10	0.725 – 1.10	Near Infra red	Surface water delineation, ice & snow melt
3 (A)			1.58 – 1.64	Thermal infrared	Snow / ice discrimination
3 (B)	3.55 - 3.93	3.55 - 3.93	3.55 - 3.93	Thermal infrared	Sea Surface temperature, night time cloud mapping
4	10.50 – 11.50	10.50 – 11.50	10.50 – 11.50	Thermal infrared	Sea Surface temperature, night time cloud mapping
5	Band 4 repeated	11.50 – 12.50	11.50 – 12.50		Sea Surface temperature, night time cloud mapping
	(In micrometers)	(In micrometers)	(In micrometers)		

2.1 CLOUD DETECTION TECHNIQUES

Various cloud detection methods have been proposed, namely for AVHRR, GOES, and Meteosat images. They are usually based on the contrast (visible as well as thermal) between cloud (usually bright, cold) and surface (usually darker and warmer). The thermal band, which allows the detection of clouds through temperature differences with the surface, is also a critical tool for cloud identification, particularly for thin clouds that do not reflect much sunlight

The SPOT VGT is a relatively recent sensor with four optical bands in the visible, near- infrared and short-wave infrared. It has a footprint resolution of about 1 km with one morning pass (Equator crossing time: 10h30 AM local time). Linear threshold or clustering cloud detection methods that rely on thermal infrared bands have been developed for cloud detection of AVHRR, GOES and Meteosat images

Detection of clouds in satellite imagery is straightforward in most circumstances. In most situations, clouds are brighter than the background surface at visible wavelengths and colder than the background surface at thermal infrared wavelengths. Infrared imagery is usually viewed as a negative image so that clouds appear white several problem areas for cloud detection exist, and cloud detection is difficult, if the image pixel size is significantly larger than the size of individual cloud elements. This is particularly problematic for thin cirrus, low stratiform cloud decks, polar regions and multilayer cloud systems. Cloud detection is difficult at night in areas where the thermal contrast between the surface and the cloud is low, for example, in the case of marine stratocumulus regions and areas of fog. [Arthur P Cracknell, Cracknell P Cracknell, 1994]

In regions of thin cirrus, the observed radiance is a combination of the radiance from cloud and the ground. In the Polar regions and other areas of snow cover, the contrast is low at visible wavelength and the thermal gradient in the atmosphere is

often weak. In the polar winter, where a strong surface inversion forms, the thermal contrast is inverted. Detection of clouds in problem areas can be enhanced by the use of multispectral imagery to detect the signatures of cloud types. Predominately, cloud detection relies on optical and thermal infrared observations, but wavelengths regions in the near infrared have been used to distinguish between clouds and snow ($1.6 \mu\text{m}$) and to detect the presence of fog ($3.7 \mu\text{m}$). Such multispectral analysis offers the opportunity to develop cloud type signatures for improved classification of clouds. Pattern recognition techniques have also been applied to problem areas such as wintertime polar clouds. [Arthur P Cracknell, Cracknell P Cracknell, 1994]

Cloud masking is essentially the same process as cloud detection, except that the focus is on the elimination of cloud-affected pixels from further analysis. When determining surface parameters such as surface temperature and vegetation indices from satellite radiance data, the presence of clouds can seriously contaminate the result.

A number of techniques have been used for cloud masking. A simple threshold brightness temperature (e.g. AVHRR channel 5) can be used, to indicate cloud - affected pixels. Alternatively, a ratio $Q = R_1 / R_2$, where R_i is optical - band reflectance, can be used to distinguish between ocean (Q generally less than 1), land (Q greater than 1) and cloud ($Q \approx 1$). Some techniques have been the difference in brightness temperature between AVHRR channels 4 and 5 as indicators of presence of cloud. The Spatial coherence of cloud - affected areas is generally larger than that of cloud free areas, and this can also be used for cloud masking although the technique is not useful where the background variance is large, e.g. in coastal regions. Most cloud masking schemes involve the derivation of empirical threshold and their success depend on the geographical area of application. [Arthur P Cracknell, Cracknell P Cracknell]

They have used vegetation indices, surface temperature and emissivity to determine a regression model with the soil color. This model is an image processing technique which is a part of color image processing. Model along with the VIs and the emissivity were used to prepare soil color grids and in turn used to detect the degradation areas.

People also have come up with image analysis techniques like thresholding at different range of values rather than using indices for extracting features from the image.

(Donna et al 1995) used many image processing techniques for determining the thickness of the sea. They made use of Region thresholding, Gaussian curve approximation, threshold clustering, histogram computation etc for removing noise and clearing image to draw out features of sea and water from the image. (M Craig Doboson et al 1996) used image segmentation to get the knowledge about the Land Cover classification. (Giglio et al 1999) took the combination of band 3 and band 4 of NQAA/AVI-IRR satellite and the mean of these bands were calculated.

Data from AVHRR which is a satellite from NOAA was also used for image analysis and features extraction. T11, T12 bands important for image processing. They employed these for cloud and water masking. Multi-thresholding technique involving T4 and T5 were used for detecting the cloudy pixels.

Apart from these methods other optimization techniques like GA (Genetic Algorithm), Neural Networks, SVM (Support Vector Machine) etc. have been used along with image processing to remove false points. Neural Networks with Statistical methods for classification of Multi-resource Remote Sensing Data. They found the results for Neural Network to be more accurate and better than the statistical approach. Neural Networks have been used for classification of land cover, detection of cloud, etc. has identified cloud from AVHRR images. Both ANN and multi-threshold techniques were used for differentiating cloud and land surface. The study involves two major steps — identifying potential areas covered by cloud using the neural networks or threshold classifier, then removing false-classified pixels by applying additional tests, texture analysis and spatial filtrations. The multi-threshold approach is based on differences in the reflectance of AVHRR band 1 (R1) and band 2 (R2) and in the brightness temperature of band 4 (BT4) and band 5 (BT5). The 2 steps involved were marking potential cloud pixels and removing false pixels. 3 X3 and 9X9 median filtering was also done to remove false and noisy cloud pixels in the output of neural network.

Cloud detection by ANN has been proposed by Yhann and Simpson (1995) for AVHRR images and by Lewis *et al.* (1997) for Meteosat images. Faure *et al.* (2001) proposed ANN retrieval of cloud parameters as a feasibility study using synthetic radiometric data for new sensors such as the moderate resolution imaging spectroradiometer (MODIS) and global imager (GLI).

A detailed description of MLP implementation with BP (Back Propagation) training is given in the following six steps (Lippmann, 1987):

Step 1. Initialize weights and biases

Set all weights w and biases θ to small random values between 0 and 1.

Step 2. Present input and desired outputs

Present a continuous valued input vector x_0, x_1, \dots, x_{N-1} and specify the desired output vector d_0, d_1, \dots, d_{M-1} .

Step 3. Compute actual outputs, y_j

$$y_j = \frac{1}{1 + e^{-(x_j - \theta_j)}} \quad x_j = \sum_i y_i w_{ij} \quad (1)$$

Step 4. Update weights and biases

$$w_{ij}(t+1) = w_{ij}(t) + \eta \delta_j y_i \quad (2)$$

where $w_{ij}(t)$ is the weight from node i at the former layer to node j at the latter layer at time t , η is a learning rate and δ_j is an error term for node j . If node k is a output node, then

$$\delta_k = y_k(1 - y_k)(d_k - y_k) \quad (3)$$

where d_k is the desired output at node k and y_k is the actual output. If node k is an internal hidden node, then

$$\delta_j = y_j(1 - y_j) \sum_k \delta_k w_{jk} \quad (4)$$

Convergence is sometimes faster and weight changes are smoothed if a momentum term (μ) is added by

$$w_{ij}(t+1) = w_{ij}(t) + \eta \delta_j x_j + \mu (w_{ij}(t) - w_{ij}(t-1)) \quad (5)$$

$$w_j(t+1) = w_j(t) + \eta \delta_k + \mu (w_j(t) - w_j(t-1)) \quad (6)$$

where in general $0 < \eta < 1$ and $0 < \mu < 1$. In this study, those two parameters were set to $\eta=0.5$ and $\mu=0.2$, respectively.

Step 5. Compute mean square error (MSE)

$$MSE = \frac{1}{2} \sum_j (d_j - y_j)^2 \quad (7)$$

Step 6. Repeat by going to 2 until MSE becomes 0.005 or 1000 iterations.

Various method using for Cloud detection by earlier researcher.

Simpson, J.J. and Gobat, J. I.(1995) have done Improved cloud detection in GOES scenes over Oceans, in which accurate cloud detection in GOES data over the ocean is a difficult task complicated by poor spatial resolution (4 Km) in the GOES IR data , relatively coarse , quantization (6 bits) for GOES VIS data , a visible sensing region of the spectrum not ideally suited for cloud versus ocean segmentation, and relative small oceanic signal dynamic range compared to that of either cloud or land structures found in a typical GOES scene. The GOES adapted LDTNLR (Local Dynamic Threshold Non – Linear Rayleigh) ocean cloud mask (GALOCM) algorithm for cloud detection in GOES scenes.

Over the oceans provides a computationally efficient , scene – specific way to circumvent these difficulties. The algorithm consists of four steps

6. generate a cloud mask using the LDTNLR algorithm of Simpson and Humphrey
7. generate a second cloud mask using an adaptive threshold.
8. Divide the pixels in the scene into three groups (both methods agree that pixel is ocean, pixel is cloud, or the pixel is in contention.) and
9. Iteratively apply an adaptive threshold to the contested pixels. Convergence occurs when pixels are no longer in contention based on statistical criteria. Results show that the GALOCM method produces accurate cloud masks over the oceans which are neither regionally dependent nor temporally specific. GOES scenes containing ocean, cloud and land are best cloudscreened using a combination of the GOES Split – and – merge clustering and the GALOCM algorithm.
10. Simpson, J. J. And Keller, R.H. (1995) have done, An improved fuzzy logic segmentation of sea ice, cloud and ocean in remotely sensed arctic imagery, in which the accurate segmentation of sea ice from cloud and from cloud – free ocean in polar AVHRR imagery is important for many scientific application (e.g. sea ice – albedo feedback mechanism, heat exchange between ocean and atmosphere in polar regions). Unfortunately, it is a difficult task complicated by the common visible reflectance characteristics of sea ice and cloud. Moreover, AVHRR channels 3 data historically have been contaminated by highly variable sensor noise which generally has hampered their use in the classification of polar scenes. Likewise, polar scenes often contain pixels with mixed classes (e. g. Sea ice and cloud). This article uses a combination of fuzzy logic classification methods , noise reduction in AVHRR channel 3 data using Wiener filtering methods (Simpson and Yhann , 1994) and a physically motivated rule base which makes effective use of the Wiener filter channel 3 data to more accurately segment polar imagery. The new methods improved classification skill compared to more traditional methods as well as its regional independence is demonstrated. The algorithm is computationally efficient and hence is suitable for analyzing the large volumes of polar imagery needed in many global change studies.

Saunders, R.W. and Kriebel, K.T. have done, An improved method for detecting clear sky and cloudy radiances from AVHRR data. To obtain accurate estimates of surface and cloud parameters from satellite radiance data a scheme has to be devised which identifies cloud - free and cloud - filled pixels (i.e. fields of view). Such a scheme has been developed for application to high resolution (1.1 Km pixel) images recorded over western Europe and the North Atlantic by the AVHRR on the TIROS - N / NOAA polar orbiters. The consists of five daytime or five night time tests applied to each individual pixel to determine whether that pixels is cloud - free, partly cloudy or cloud - filled. The pixel is only identified as cloud - free or cloud - filled if it is passes all yhe tests to identify that condition ; otherwise it assumed to be partly cloudy . Surface parameters (e.g. Skin temperature, reflectance, vegetation index, snow cover) can then be inferred from cloud - free radiances, and cloud parameters (e. g. cloud top temperature, optical depth and liquid water content) from the cloud - filled radiances. Only 8fractional cloud cover is derived from the partly cloudy pixels which, together with the number of cloud - filled pixels, gives total cloud cover over a given area. The schemes has been successfully applied to data for all seasons, including images with unusually cold or warm surface temperatures.

A simple method for cloud Detection over land using Daytime AVHRR Data (MyoungSeok Suh, Kwangmi Jang, KyoungYoon park, 1997) in which, 5 steps threshold method using different combination on channels was developed to detect the cloud-contaminated pixels from NOAA-14/AVHRR daytime imagery. The first two thresholds were applied to the infra - red (channel 4) and visible (channel 1) imagery to detect the low temperature pixels (high or middle - level clouds) and the high reflectance pixels (low-level clouds). To detect the cloud edge and small cumulus cloud, spatial coherency test using local standard deviation (LSD) was applied to near infra - red images. And split - window threshold (the brightness temperature difference between channel 4 and 5) was applied to detect the optically thin cirrus. Finally, inverse relation between visible and infra-red imagery was tested the other contaminated pixels.

NOAA/AVHRR Imagery detected automatic cloud, by real -time processing.(M.Derrien, B.Fanki, L.Harang,H.LeGleau, A.Noyalet, D.Pochic)Polar

orbiting satellites allows a description of cloud cover, oceanic, and continental surfaces that is used by MétéoFrance for nowcasting activities and as input for numerical weather prediction models (NWP). A real-time processing scheme has been designed at the Centre de Météorologie Spatialement (CMS) in Lannion to extract cloud cover and surface parameters from NOAA-11 AVHRR imagery received at CMS. The key step of this scheme is cloud detection. It is based upon threshold tests applied to different combinations of channels. Its main originality is its complete automation by the computation of the $11 \mu\text{m}$ infrared threshold from a monthly sea surface temperature (SST) climatology over the oceans and from air temperature (near the surface) forecast by NWP over land. A special test has been implemented to detect cloud edges and sub pixel clouds over continental surfaces during daytime. It is applied daily in deferred time only to compute normalized difference vegetation index (NDVI). This scheme has been used operationally since February 1990, and its quality has been checked. It has enabled the routine production of various products. A nighttime cloud classification is sent to all French forecasters; NDVI values are computed daily and used to map the vegetation cover; and SST and thermal fronts are derived operationally from nighttime imagery.

Cirrus cloud detection from Airborne Imaging Spectrometer Data shows that a malfunction of the water vapor channel at $1.38 \mu\text{m}$ to detect cirrus clouds, (E. Ben-Dor). A combination of elevation, vegetation coverage, water vapor content, and albedo characteristics (mostly governed by the terrain) are the major factors affecting cirrus cloud detection. Using the criteria of relative low radiance and high signal-to-noise ratio amongst several targets and across the $1.84\text{--}1.92 \mu\text{m}$ spectral region, the $1.84\text{--}1.92 \mu\text{m}$ channel was found to more effectively mask ground signals than the $1.382 \mu\text{m}$ channel. Over targets having moderate elevation, dry conditions, minimal vegetation, and high albedos, both spectra regions present significant ground signals that can mistakenly be attributed to cirrus cloud particles. It is strongly recommended that for accurate cirrus cloud detection, both spectral regions around $1.38 \mu\text{m}$ and $1.88 \mu\text{m}$ be examined along with the above-mentioned factors.

Accurate cloud detection in Advanced Very High Resolution Radiometer (AVHRR) data over land is a difficult task complicated by spatially and temporally varying land surface reflectances and emissivities. The AVHRR Split-and-Merge Clustering (ASMC) algorithm for cloud detection in AVHRR scenes over land provides a computationally efficient, scene-specific, objective way to circumvent these difficulties, (J.J.Simpson and J.I.Gobat, 1996). The algorithm consists of two steps:

- 1) a lit-and-merge clustering of the input data (calibrate channel 2 albedo, calibrated channel 4 temperature, and a channel 3 - channel 4 temperature difference), which segments the scene into its natural groupings; and 2) a cluster-labelling procedure that uses scene-specific, joint three-dimensional adaptive labelling thresholds (as opposed to constant static thresholds) to label the clusters as either cloud, cloud-free land, or uncertain. The uncertain class is used for those pixels whose signature is not clearly cloud-free land or cloud (e.g., pixels at cloud boundaries that often contain sub pixel cloud and land information that has been averaged together by the integrating aperture function of the AVHRR instrument). Results show that the ASMC algorithm is neither regionally nor temporally specific and can be used over a large range of solar altitudes. Sensitivity of the segmentation and labelling steps to the choice of input variables also was studied. Results obtained with the ASMC algorithm also compare favorably with those obtained from a wide range of currently used algorithms to detect cloud over land in AVHRR data. Moreover, the ASMC algorithm can be adapted for use with data to be taken by the Moderate Resolution Imaging Spectrometer-Nadir (MODIS-N).

Valid estimates of sea surface temperature (SST) from satellite data [e.g., the Along Track Scanning Radiometer (ATSR)] are critically dependent upon the identification and removal of cloud from the data, but few cloud screening algorithms for ATSR data have appeared in the literature,(J.J.Simpson, Andrew Schmist,and Andrew harris.,1998). A new algorithm, the ATSR Split-and-Merge Clustering (ATSR/SMC) algorithm, for cloud masking ATSR data is presented which evaluates every pixel in the image, is statistically reproducible, computationally efficient, and requires no knowledge of cloud type. Moreover, it is effective in detecting multilayer cloud structures in a scene, which is a difficult task because such systems generally have bimodal statistical

distributions. It also accurately detects glint radiance, which is quite common in at least one of the $1.6 \mu\text{m}$ views, sub pixel cloud contamination near cloud boundaries and low-lying marine stratiform cloud. Historically, these issues have interfered with ATSR-based SST retrieval [see the work of Jones et al., (1996a, b) and the references cited therein]. The SSTs derived from the cloud-free ocean pixels were validated with 96 buoy observations and the mean difference (buoy—SST) was $+0.24^{\circ}\text{C} \pm 0.51^{\circ}\text{C}$. For the 103 pairs of images (forward/nadir views) tested, the mean $11 \mu\text{m}$ BTs that result from SADIST (standard ATSR processing) vs. ATSR/SMC cloud detection are 0.4°C (daytime) and 0.6°C (nighttime) colder for SADIST than for ATSR/SMC, even though the SADIST cloud masks generally over detect clouds relative to ATSR/SMC cloud masks. The conclusion that the new procedure produces cloud masks which are superior to the standard ATSR operational cloud mask product and it retains substantially more valid pixels. The algorithm can be used in tropical and midlatitude regions; it is not designed to detect sea ice, and consequently should not be used in Polar Regions. Finally, the approach can easily be adapted to ATSR—2 data and to other data to be taken from soon to be launched sensors.

3. ALGORITHM DESCRIPTION

3.1. Introduction

In order to detect the fire spots, cloudy pixels more than one algorithm have been proposed in this dissertation which are based on change detection, multi-thresholding and thresholding. In change detection, the pixel values of different years of data were compared. Multi-thresholding algorithm is based on differences in the values of AVHRR bands 3, 4 and 5 in the land surface temperature (LST_Night). Thresholding algorithm is based on all the bands of AVHRR.

3.2. Preprocessing

The preprocessing of AVHRR is done in order to make the data more meaningful and for easier processing at later stages. In case of AVHRR the preprocessing involves two basic steps — reprojection and calibration

3.2.1. Reprojection

The downloaded data has the information of geographic coordinates but they are either in meters or need to be projected. The data used here are gridded products of AVHRR which is reprojected to its geographic Lat/Long using Everest 1956 datum.

3.2.2. Calibration

Absolute calibration is essential for a variety of scientific studies and image analysis applications. Calibration is done in order to convert the DN (Digital Number) values of the bands to meaningful values. In this dissertation, we used LST Day, LST_Night which actually represents temperature in K.

Three major steps are involved in the method used here in detection of cloudy pixels.

- 1 Change Detection using Image analysis.

2. Multi-thresholding using Image processing.

3.3. Change Detection using Image analysis

Change detection is a technique based on the comparison of temporal development curve for successive years of remotely sensed indicators derived from high temporal resolution data such as those from NOAA/AVHRR or MODIS etc. When the time trajectory of a particular pixel or Lat/Long deviates from the expected value for that pixel or Lat/Long, a process of change can be detected. 10 years data using for image analysis.

This method involves separately classifying images for two different dates. The classified images are then compared pixel by pixel to detect change in its values. Result is then displayed in a change matrix or change map or by plotting graph in XY plane.

3.4 Multi-thresholding using Image processing

3.4.1. Thresholding

Consider an image of gray scale where each pixel $f(x, y)$ represents some pixel in the image. Each pixel has some gray scale value in the image. If the image is separated based on the fact if $f(x, y) > T$ (some pre-decided threshold value) are assigned black or 0 and $f(x, y) < T$ are assigned white or 1. This makes the image in zeroes and ones. The technique discussed is called Thresholding.

Thermal anomalies and their background can be separated through a threshold, which is defined as the first histogram turning point after the mean plus the standard deviation. A proper threshold can minimize the number of false alarms. When a threshold is set, the result map can be sliced to a bit map representing cloudy induced thermal anomalies and the non-cloudy area. Furthermore, some false alarms, such as fire, can be removed by taking appropriate threshold values.

3.6. ALGORITHM

In order to detect the cloud and fire pixels, four algorithms have been used in this dissertation which is based on multi-thresholding. Two algorithms used for to detect cloudy pixels.

3.6.1 THE AVHRR SPLIT-AND-MERGE CLUSTERING CLOUD-DETECTION ALGORITHM

The algorithm, called the AVHRR Split-and-Merge Clustering(ASMC) cloud-detection algorithm, uses an iterative , nested , partitional , hierarchical clustering procedure to segment the AVHRR scene into its natural divisions and then joint , three dimensional , adaptive thresholding procedure to label the segments as cloudy , cloud – free , or uncertain. [Ref]

The AVHRR Split-and-Merge Clustering algorithm uses high spatial resolution channel-2 visible albedo, infrared temperature derived from channel-4 data, and a channel 3 minus channel 4 temperature difference. Although channel-1 data may give better contrast between land and cloud than channel-2 data, channel-2 data are preferred over channel-1 data for this application because channel-1 data are much more affected by time-and- space varying atmospheric processes (e.g., Rayleigh and aerosol scattering) than are channel-2 data. A detailed discussion of this point, including results from radiative transfer model computations, is given by Simpson and Gobat (1995b). Prefiltering the channel-3 data with a Wiener filter to remove noise (Simpson and Yhann, 1994) generally has not proved necessary in this application based on tests that used both raw and filtered channel-3 data to form the required channel 3 minus 4 temperature difference. Extremely noisy channel-3 data, however, may require preconditioning with the Wiener filter methods. The algorithm has two steps 1) a split-and- merges clustering procedure, and 2) an adaptive-labelling procedure.

3.6.1.1. Split-and-Merge Iterative Clustering

Clustering refers to any type of method that attempts to automatically partition a given data set by identifying the natural groupings of the data within a specified feature space by optimizing a chosen clustering criterion (Pairman and Kittler, 1986). Hence, this approach is dependent on the global properties of the data. Clustering attempts to make all the elements of a given cluster as similar as possible (i.e., minimize within-cluster variance) and simultaneously make individual clusters as distinct as possible (i.e., maximize between-cluster variance). Scatter matrices used here provide a computationally convenient way in which to formulate this minimization problem. Moreover, Gallaudet and Simpson (1991) have shown that minimizing the within-cluster scatter matrix simultaneously maximizes the between-cluster scatter matrix if total variance is to be conserved.

We define n_d as the number of bands in the input image. Then we can describe the input data as $N n_d$ no-dimensional column vectors where N is the number of valid pixels in the input image. Each column vector corresponds to a given pixel location within the input image and each component of the column vector corresponds to a single spectral or derived value at that location. For the ASMC algorithm, $n_d = 3$; each vector is defined as

$$Y = \begin{bmatrix} a_2 \\ T_4 \\ \delta \end{bmatrix} \quad (1)$$

where a_2 is the channel-2 albedo, T_4 is the surface temperature derived from channel 4, and δ is the temperature difference between channels 3 and 4, subject to the constraint that negative differences are mapped to 0.0.

The clustering operation proceeds as follows:

1. Distribute the initial K , cluster centers, m_k , equidistant throughout the range of the input data

**DEPARTMENT OF HYDROLOGY
INDIAN INSTITUTE OF TECHNOLOGY
ROORKEE**

No.HYD/Conf/134

Dated July 17, 2006

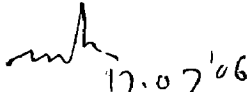
Asstt.Registrar (PGS&R)

Please find enclosed herewith in a separately sealed envelope the marks to **Mr. Waghmare Pradnyesh Ramesh**, Trainee Officer M.Tech. Hydrology (33rd batch), Enroll No. 47903 for his M.Tech. Thesis Viva-Voce Examination held on July 13, 2006.

One copy of his M.Tech. Thesis as per title given below is also sent herewith. One copy of the thesis meant for guide has been retained by me.


Title of Thesis : "Cloud Detection in Satellite Imagery (AVHRR)"

Encl: As above


17.07'06
(D.K.Srivastava)
Chairman, DRC

cc:

1. O.C. Library, Department of Hydrology, with a copy of the above thesis for library of Department of Hydrology.
2. Prof. & Head, Department of Hydrology for information please.


(D.K.Srivastava)
Chairman, DRC

$$m_k = y'_{\min} + \left(\frac{k-1}{K-1} \right) (y'_{\max} - y'_{\min}) \dots \quad (2)$$

where the maximum and minimum vectors of the entire data set y'_{\max} and y'_{\min} respectively, are defined as the component-wise maximum and minimum of the n_d spectral or derived inputs.

2. Assign each input vector, y , to the nearest cluster center by computing the Euclidean distance from the vector to each cluster center

$$d_k = (y - m_k)^T (y - m_k) \quad k=1 \dots K \quad (3)$$

where K is the total number of clusters) and assign the vector to the cluster that corresponds to the minimum d_k . In Eq. (3) and elsewhere superscript T denotes the transpose operation.

3. Compute the between-cluster scatter matrix, S_b

$$S_b = \sum_{k=1}^K n_k (m_k - m)(m_k - m)^T \quad (4)$$

where n_k the number of vectors in cluster k ,
and m is the mean vector of the entire input data

$$m = \frac{1}{N} \sum_{k=1}^K n_k m_k \quad (5)$$

4. Check for convergence based on the trace of the between-cluster scatter matrix, $T_r(S_b)$. $T_r(S_b)$ is a measure of between-cluster variance. Convergence there is negligible change in this measure from one iteration to the next, that is, if

$$\left| \frac{T_r(S_b) - T_r(S'_b)}{T_r(S_b)} \right| < \text{tol} \quad (6)$$

then we stop the clustering process and move on to the labelling stage. In the convergence check above, T_r (S'_b) represents the between-cluster from the previous iteration and tol is a predefined numerical tolerance.

5. For each cluster k , compute the distance between the component-wise cluster maximum and minimum vectors,

$$d_m = (y_{\max} - y_{\min})^T (y_{\max} - y_{\min}) \quad (7)$$

where the cluster minimum and maximum vectors are defined by a procedure analogous to that used for the image minimum and maximum vectors in (2); they are the component-wise cluster minimum and maximum vectors, If d_m is greater than the splitting threshold (T_s) split the cluster by reassigning each vector to one of two new clusters, based on whether it is closer to the maximum or minimum.

6. Repeat step 5 until nothing is split during that step.

7. For every pair of clusters, compute the distance, d_b , between the cluster mean vectors

$$d_b = (m_i - m_j)^T (m_i - m_j) \quad (8)$$

If d_b is less than the merging threshold (T_m), combine all the vectors in clusters i and j into single cluster.

8. Repeat step 7 until nothing is merged during that step.

9. Based on the current K cluster centers, repeat the process beginning at step 2.

The process outlined above uses four predetermined parameters: the number of initial clusters the splitting threshold (T_s), the merging threshold (T_m), and the convergence tolerance (tol). All our are reatively easy to determine because, as Gallaudet and Simpson (1991) have shown, the clustering procedure is largely insensitive to the values chosen. Specifically, Gallaudet and Simpson (1991) show the sensitivity contours of the T_s and T_m thresholds to: a) within- cluster variance, b) number of split and merge operations in the first iteration of the clustering procedure, c) the percentage of pixels rejected as cloud, d) the total number of iterations required for convergence to a stable cluster population, and e) the final number of clusters at convergence.

Moreover, a detailed discussion of the sensitivity of the clustering procedure to the initial number of (K .) and to the convergence criteria (tol) is given. Because the ASMC clustering procedure is a slight adaptation of that used by the original PCTSMC algorithm of Gallaudet and Simpson (1991). The same rules apply for selection of input parameters to the clustering procedure of the ASMC algorithm. Thus, for the ASMC algorithm, these parameters are specified as follows. K , the initial number of clusters, is chosen as 30 because this number is small enough that the algorithm still performs acceptably fast and large enough that the clustering result is generally stable for any number larger than this. The splitting and merging thresholds are defined by Gallaudet and Simpson (1991)

$$\beta (y'_{\max} - y'_{\min})^T (y'_{\max} - y'_{\min}) \quad (9)$$

That is, as some percentage of the distance between the input data set maximum and minimum vectors. β was chosen as 0.01. The convergence tolerance was taken as 0.05. This means that if the traces of the between-cluster scatter matrix changes by less than 5% from one iteration to the next, then the algorithm will accept the current set of

sters and proceeds to the labelling stage.

e method of clustering just described combines both partitional and hierarchica
roaches. It consists of a partitional clustering algorithm augmented by a splitting-and
rging step at each iteration, Combining a partitional with a hierarchical method ha
eral advantages over the use of either method 'alone (see Gallaudet and Simpson
)1 and the references contained therein).

K, T_s, T_m, tol

$$Y = \begin{bmatrix} a_2 \\ T_4 \\ \delta \end{bmatrix}$$

Cluster center $m_k = y'_{\min} + \left(\frac{k-1}{K-1} \right) (y'_{\max} - y'_{\min})$

Euclidean distance $d_k = (y - m_k)^T (y - m_k)$

Mean vector of the entire input data

$$m = \frac{1}{N} \sum_{k=1}^K n_k m_k$$

Between - cluster scatter matrix

$$S_b = \sum_{k=1}^K n_k (m_k - m)(m_k - m)^T$$

Check convergence based on

$$\left| \frac{T_r(S_b) - T_r(S'_b)}{T_r(S_b)} \right| < \text{tol}$$

Each cluster distance between max and min component -wise vector

$$d_m = (y_{\max} - y_{\min})^T (y_{\max} - y_{\min})$$

$$d_m > T_s$$

Every pair cluster (d_{ij}) cluster mean vectors

$$d_{ij} = (m_i - m_j)^T (m_i - m_j)$$

$$d_{ij} < T_m$$

Labelling (T_s)

Split and Merge Clustering algorithm

3.4.1.2. Labelling

Once the data have been segmented into their natural classes, a labelling algorithm is used to label each cluster (and thus all the pixels in that cluster) as either cloudy or clear. The labelling criteria for ASMC are based on an adaptively determined decision plane that split the three-dimensional input space into clear and cloudy regions, Figure 1 illustrates three-dimensional vector space of the problem (albedo, temperature, $T_3 - T_4$ temperature difference) as the dashed outline; the plane that splits this space into clear and cloudy regions is drawn as the solid triangle.

Given n final clusters from step 1 of the ASMC algorithm, we define a_{\min} as the minimum albedo component of the n cluster mean vectors, Similarly, δ_{\min} is the minimum $T_3 - T_4$ temperature difference and T_{\max} is the maximum temperature component. The scene-specific adaptive labelling threshold values, a_{th} , δ_{th} , and T_{th} , are determined from an adaptive threshold of the appropriate component of the input data of the entire image using the following procedure:

- 1) for a given input component, select the image mean as an initial threshold
 - 2) use this value to divide the image into two groups having new means m_1 and m_2
 - 3) define a new threshold as the average of m_1 and m_2 ; and
 - 4) iterate until the adaptively computed threshold, remains constant.
- The complementary roles of the split and merge thresholds used to segment the image and the adaptive labelling thresholds used to label the segments (cloudy versus cloud-free).

Given a cluster mean vector with components $(\alpha_0, \delta_0, T_0)$, the cluster is labeled as cloud if it falls outside the polyhedron defined by the adaptive labelling thresholds and the extrema of the cluster means (Fig. 1a). Algebraically, a cluster will be labeled as cloud if the inner product of the cluster mean vector and the adaptively determined decision

plane is positive (i.e., if the normal vector from the decision plane to the cluster center points outward from the polyhedron). If we write the equation of the plane as

$$ma + nT + \delta - (\delta_{\min} + ma_{\min} + nT_{th}) = 0 \dots\dots\dots (10a)$$

where m and a are defined as

$$m = \left(\frac{\delta_{th} - \delta_{\min}}{a_{th} - a_{\min}} \right) \dots\dots\dots (10b)$$

and

$$n = \left(\frac{\delta_{\min} - \delta_{th}}{T_{\max} - T_{th}} \right) \dots\dots\dots (10c)$$

Then the signed distance, d_s , from the plane to the cluster cent

$$d_s = \frac{ma_0 + nT_0 + \delta_0 - \delta_{\min} - ma_{\min} - nT_{th}}{\sqrt{(m^2 + n^2 + 1)}} \dots\dots\dots (10d)$$

it is important to emphasize that eqs10a and 10d provide a cluster-specific labelling rule for determining if a given cluster falls outside the decision polyhedron ($d_s > 0$) and hence is a cloud (Figure 1a). Eqs. 10a and 10d do not compute a single static threshold for the entire scene.

Complementary Roles of Clustering and Labelling Procedures

The clustering procedure (step 1 of the ASMC algorithm) uses cluster split – (T_s) and merge (T_m)-thresholds to partition the original spectral and derived input space (a_2, T_4, δ) into a set of homogeneous regions in cluster space. The adaptive-labelling thresholds ($a_{th}, \delta_{th},$ and T_{th}) partition the cluster space into the geophysical domains of interest (cloudy, cloud-free). The cluster thresholds are determined solely from the spectral and derived properties of the specific image under study using the dynamic computational procedures. The adaptive-labelling thresholds are computed as defined in the section entitled Labelling.

If the adaptive-labelling thresholds were used singularly, then they would behave very much like the traditionally used static thresholds. Use of the decision polyhedron, however, allows us to apply the individual adaptive thresholds in a novel simultaneous sense (novel and simultaneous as opposed to the more traditional simple “and” or “or” logical conditions). Moreover, the joint adaptive-labelling thresholds are applied to the clusters and not to individual pixels as generally has been done. The variance associated with even small subsets of adjacent pixels can give rise to noise in a statistical sense, which often causes single and /or logically linked sequences of static thresholds to break down. This problem is resolved by the initial clustering step of the ASMC algorithm because the joint-labelling criteria are only applied to statistically homogeneous structures (i.e., the clusters) whose statistical signatures are more clearly indicative of cloud or cloud-free conditions than are the signatures of individual pixels.

The difference between traditional thresholding of individual pixels and the use of clustering [with splitting (T_s) - and merging (T_m).threshold] to segment the scene into homogenous natural groupings, followed by cluster labelling (with adaptive labelling thresholds ($a_{th}, \delta_{th}, T_{th}$)) is clearly shown in Figure 1a.

3.4.1.3. The Uncertainty Criteria

In addition to the binary clear / cloud classification defined by the sign of the distance from the decision plane to cluster center (Eq. 10d), the labeling scheme also can make use of the actual distance from the decision plane and use it to attach a reliability estimate to the resulting label. Thus, a cluster with large negative distance is interpreted as clear land with a high degree of certainty; a large positive is interpreted as cloud with a high degree of certainty. The smaller the distance (either positive or negative), the less certain is the final label (either cloud or clear land). Because of the uncertain labels associated with cluster that lie very close to the decision plane, additional planes (parallel to the original) can be added to the decision rules to define zones of ambiguity or zones where very uncertain labels are likely to occur. The definition of these additional planes depends on the application (i.e., recovering as much clear land as possible, ensuring that clear land labels are assigned with a high degree of certainty, or identifying regions of possible subpixel cloud contamination).

For the purposes of this article, two additional planes were defined. The first, just outside of the original decision plane effectively increases the size of the polyhedron in which clear land labels should be assigned.

The second additional surface is even further beyond the original plane and the area between this plane and the first additional plane defines a region where ambiguous labels are assigned. The actual placements of these planes are defined as 5% and 12% of the maximum corner-to-corner distance beyond the original decision plane (where the corner-to-corner distance refers to the dashed cube in Fig. 1a). With the addition of these two planes, the labelling rules can be summarized as follows:

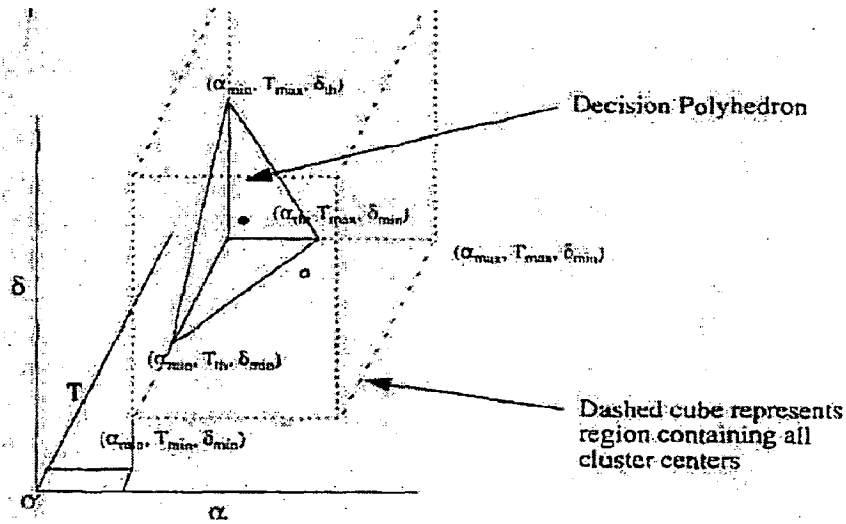
$$D = \sqrt{((a_{\max} - a_{\min})^2 + (T_{\max} - T_{\min})^2 + (\delta_{\max} - \delta_{\min})^2)}$$

$$ds < 0.05D \rightarrow \text{clear}$$

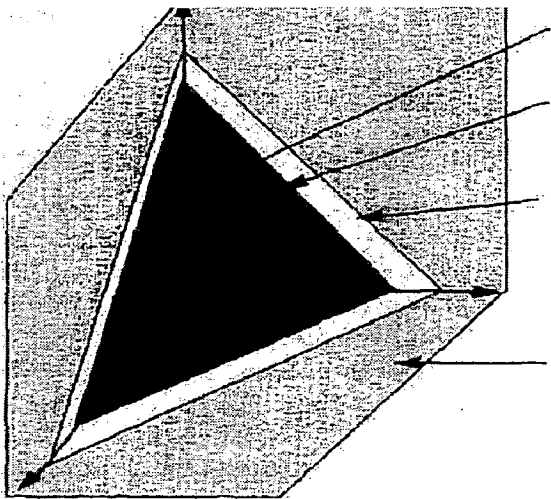
$$0.05D \leq ds \leq 0.21 \rightarrow \text{ambiguous}$$

$$ds > 0.12D \rightarrow \text{cloud} \dots\dots\dots (11)$$

where ds , is the signed distance defined by Eq10d, and D is the corner-to-corner distance of the cube that completely encloses all of the cluster-mean vectors. Thus, if ds is negative or a small positive number less than $0.05 \times D$, the cluster is labeled cloud-free land



(Figure1a)



Decision polyhedron
 First additional plane (labeled as clear)
 Second additional plane (ambiguous) outside f
 other regions labeled as cloud

(figure1b)

Figure 1a) definition of decision plane used to separate cloudy from cloud-free land pixels in the labeling step of the ASMC algorithm.

Figure 1b) Definition of the ambiguous domain used in the labeling step of the ASMC algorithm.

3.6.2. DAYTIME ALGORITHM

3.6.2.1. Detection of cloud-free pixels

The overall philosophy of this cloud detection scheme is to apply up to five tests to detect cloud and then to identify a pixel as cloud-free only *if all* the tests prove negative. This does lead to the possibility that some tests will incorrectly identify some cloud-free pixels as cloud-contaminated but this is the safest way to ensure no cloud-contaminated pixels escape detection. These tests can vary depending on whether it is night or day and on the underlying surface type which is divided into three classes, sea, land and coast (i.e. mixed). The scheme uses all available channels of the AVHRR, which are in the following wavelength ranges: 0.58—0.68 μm , 0.72—1.10 μm , 3.55—3.93 μm , 10.3—11.3 μm and (for AVHRR/2) 11.5—12.5 μm for channels 1 to 5 respectively. During the day bi-directional reflectance from channels 1 and 2 and infrared brightness temperatures (i.e. equivalent black body temperatures) from channels 4 and 5 are required for input to the scheme whereas at night the infrared brightness temperatures from channels 3, 4 and 5 are required. [ref. Kidwell, 1985 and Lauritson et al., 1979.] If channel 5 is not available, as is the case for TIROS-N and NOAA-6, -8 and -10, the scheme is simplified by omitting all tests using channel 4 and 5 differences and wherever possible substituting channel 4 for channel 5. More details of the AVHRR instrument and calibration procedure can be found in Kidwell (1985) and Lauritson et al. (1979). Some of the cloud detection tests have already been described elsewhere (Saunders 1986 a) and so are only mentioned here briefly for the sake of completeness. However, a few improvements have been made and additional tests included.

The first test applied to both daytime and night-time data is an infrared threshold test using the measured AVHRR/2 12pm brightness temperature (or 11 μ m if no channel 5 is available) as a check on cloud contamination. The 12 μ m brightness temperature is used because clouds have a greater optical depth at these wavelengths (Olesen and Grassl 1985). If the measured brightness temperature is below a certain threshold temperature the pixel is rejected as cloud-contaminated. A problem arises in defining an appropriate threshold temperature. Over the sea it is straightforward as the sea surface temperature (SST) varies only slowly during the year. An operational scheme could use the last 5-day mean SST for each 1 degree latitude/longitude grid point corrected for the maximum likely atmospheric absorption effects to give a brightness temperature at the top of the atmosphere. Over the land, however, the large day-to-day variability in surface temperature due to different meteorological conditions makes defining a threshold temperature more difficult. Operationally a forecast surface skin temperature could be used from a mesoscale model to define a threshold top-of-the-atmosphere brightness temperature. During the development of the scheme the thresholds were determined interactively from the data. An 11 μ m brightness temperature image was displayed and the user identified cloud-free land and sea areas which were likely to be the coldest in the image. The brightness temperature over these areas were then determined and temperatures 2 deg K less than the measured values were used as threshold values. This approach is only suitable for images of small sections of one pass, as thresholds for Scotland for instance will be inappropriate over Spain. Over the coast the temperature threshold for the coldest surface (i.e. land or sea) is used, which is normally the sea during the day and the land at night.

The second test is a local uniformity or spatial coherence test applied on 3 x 3 pixel array of 11 μ m brightness temperatures. During the day it is only applied over the sea (with a standard deviation threshold of 0.2deg K), as the horizontal temperature variations over cloud-free land can be considerable. We have found that the surface temperature variability over land is less at night, making a spatial coherence test possible, and so it is used with an increased standard deviation threshold of 1deg K. The test is never applied over coastal areas where there are usually large variations in surface temperature.

The third test applied during the day (i.e. when the solar elevation is greater than 10 degrees) is a dynamic reflectance threshold test. Over land and sea a cloud-free peak can be identified in the reflectance histograms of about 50 x 50 pixel arrays, allowing a reflectance threshold to be set at a slightly higher reflectance. All pixels with reflectances above this threshold are assumed to be cloud-contaminated. Identifying a cloud-free reflectance peak and then setting a threshold value removes uncertainties due to variations in calibration and changes in surface reflectance with solar zenith/azimuth angles, etc. If a fixed reflectance threshold is applied these variations will mean that it is often too high or too low. Over coastal areas, however, identification of the cloud-free peak is more difficult so a fixed reflectance threshold of 15 per cent has to be used. Over sea, channel 2 reflectances are used, as they are less sensitive to aerosol and molecular scattering effects. Over the land channel 1 reflectances are used since the reflectance of land surfaces in channel 1 is much less than in channel 2, which increases the contrast between land and cloud. More details of all these tests described above can be found in Saunders (1986 a).

The fourth test used during the day makes use of the ratio of near-infrared bidirectional reflectance's (AVHRR channel 2) to visible bi-directional reflectance's (AVHRR channel 1). Bi-directional reflectance R in this paper is defined as

$$R_n = \left(\frac{G_n C + Y_n}{\cos \theta_0} \right) \dots\dots\dots (1)$$

where R_n is in units of percentage reflectance, the gain G_n and intercept Y_n are given in the appendices of Lauritson et al. (1979), C is the raw count value received from the satellite for channel n and θ_0 is the appropriate solar zenith angle. The ratio used in the test is then simply defined as

$$Q = \frac{R_2}{R_1} \dots\dots\dots (2)$$

This ratio Q is close to unity over clouds, as the reflectance of clouds only decreases slightly at near-infrared wavelengths and anisotropy effects are similar in both channels and hence cancel. Over cloud-free water, however, enhanced backscattering at the shorter wavelengths due to molecular and aerosol scattering causes the visible reflectance to be often twice that in the near-infrared (outside sun glint), giving values of Q of around 0.5. Over land with growing vegetation the reflectance increases markedly at near-infrared wavelengths compared to shorter visible wavelengths (Swain and Davis 1978). Even over desert or during the winter when the vegetation is dormant the reflectance is higher at the longer wavelengths (except over snow and ice), ensuring that Q is always greater than unity. The land and sea peaks are well defined at the 'dark' end with cloudy reflectances producing a broad higher reflectance tail. The cloud-free sea and land radiances are well separated with a well-defined cloudy peak close to unity. To identify cloud-free pixels a cloud-free sea or land peak is identified from the histogram of Q and then only pixels with values of Q closer to the cloud-free peak(s) than pre-defined values (0.06 over sea and 0.2 over Land) identified as cloud-free. This works well over sea but over land, as there is often no well-defined peak due to the large variability of the ratio overland. In this case a default threshold of 1.6 is set where all pixels with a value of Q less than this are assumed to be cloud-contaminated. If a cloud-free peak over sea cannot be identified, all pixels with values of Q greater than 0.75 are assumed to be cloud-contaminated. One problem is in areas of sun glint where Q can approach unity, having the same value as over cloud, and so the test is not applied over these areas.

The final test applied to both daytime and night-time data examines the difference between the $11 \mu\text{m}$ (channel 4) and $12 \mu\text{m}$ (channel 5) brightness temperatures. Figure shows the overall flow diagram of the scheme to detect cloud-contaminated pixels during the day. A pixel must pass all of the tests described above to be assigned cloud-free. More details of the first three tests are given in Saunders (1986 a).

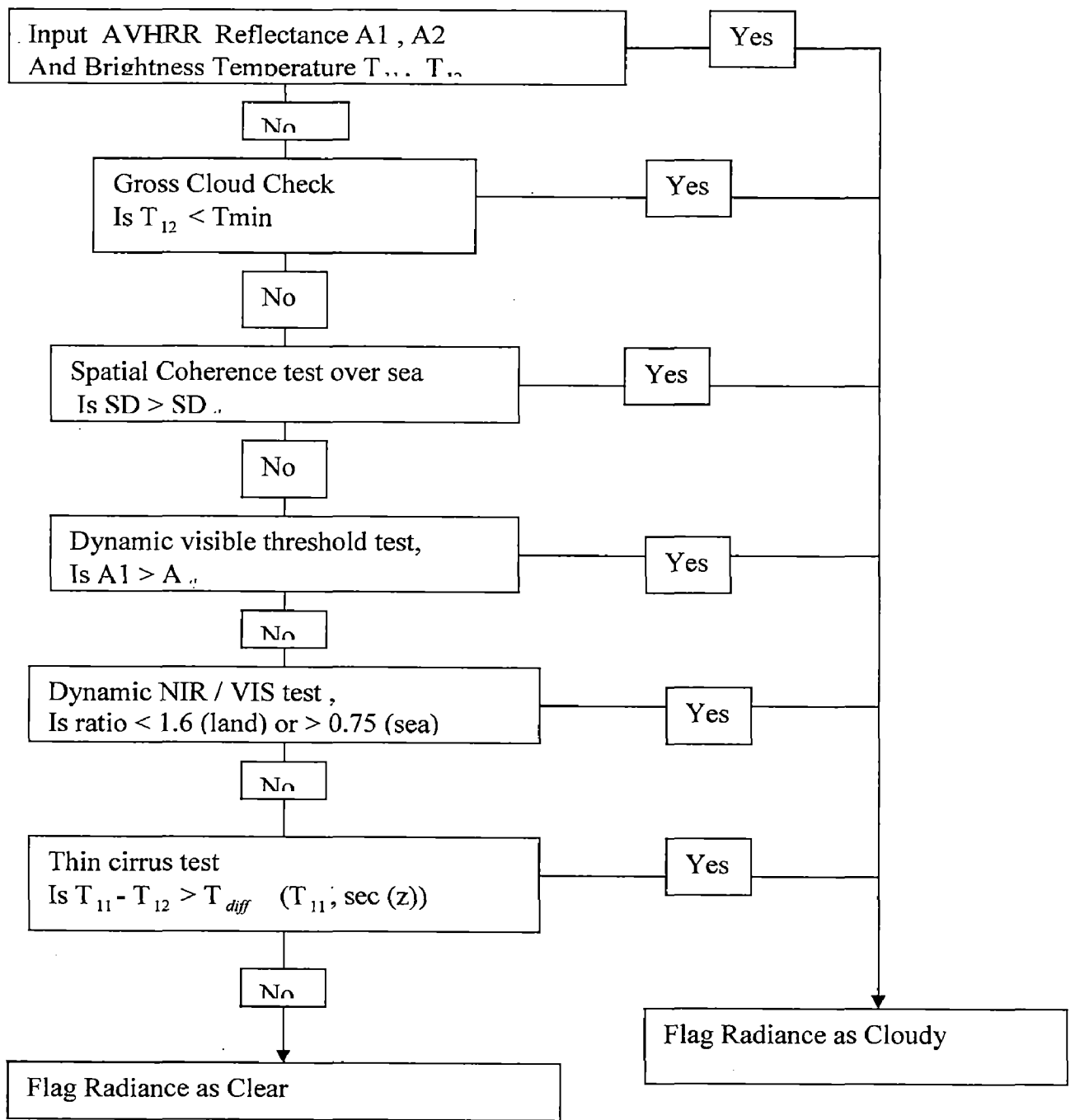


Figure: 2 Day time Algorithm

3.6.3. CONTEXTUAL ALGORITHM

Regular information about fire events at local to national scales is a necessary prerequisite to understanding and documenting the extent of their occurrence, in space and time.

Under an ongoing programmed of activities for improving direct access to environmental information where most needed , the local applications of Remote sensing Techniques group (LARST) at the Natural Resources Institute (NRI) uses national Oceanic and Atmospheric Administration (NOAA) , Advanced Very High Resolution radiometer (AVHRR) data as a primary source of information for detecting vegetation fires.

3.6.3.1. Contextual approach

The principles of this contextual approach were first found in a fire detection algorithm review by Justice and Dowty (1994). When interpreting an image visually, the human eye usually spots a fire because of the fire itself and its surroundings. This is exactly the way the contextual algorithm works: a decision about whether a pixel is a fire is made by comparing the values of a possible fire with those of its immediate neighbors. If the contrast between the two is high enough, the pixel is identified as a fire. The main difference from 'traditional threshold algorithm 'is that a decision is made on a relative basis rather than an absolute one. The algorithm is self – adaptive and automatically detects fire under different conditions.

Test) since the band width of AVHRR channel 3 covers parts of both the solar and thermal ranges of the electromagnetic spectrum , it is important to reject those pixels whose value in channel 3 would saturate or would be too high due to high reflection rather than high temperature (e. g. from bright soils , clouds or sun glinths).

This is assessed by looking at the reflectance in channel 2.

PF is a not a fire if

$$\rho_2 \geq 20 \%$$

Where ρ_2 is the top – of – atmosphere bidirectional reflectance factor for AVHRR channel 2.

4.1 Introduction

In this dissertation report, an attempt has been made to detect the cloudy pixels in NOAA/AVHRR images using different image analysis techniques. To accomplish this change detection, multi-thresholding techniques were developed and applied in order to extract features from the image. The images of AVHRR satellite were first calibrated, georeferenced, and then image processing techniques were applied for realizing the objectives. The data was reprojected to geographic Lat/Long using Everest-1956 datum for obtaining processed image. Figure 4.1 shows the procedure followed in processing an image for detection of cloudy pixels from cloud using image analysis technique.

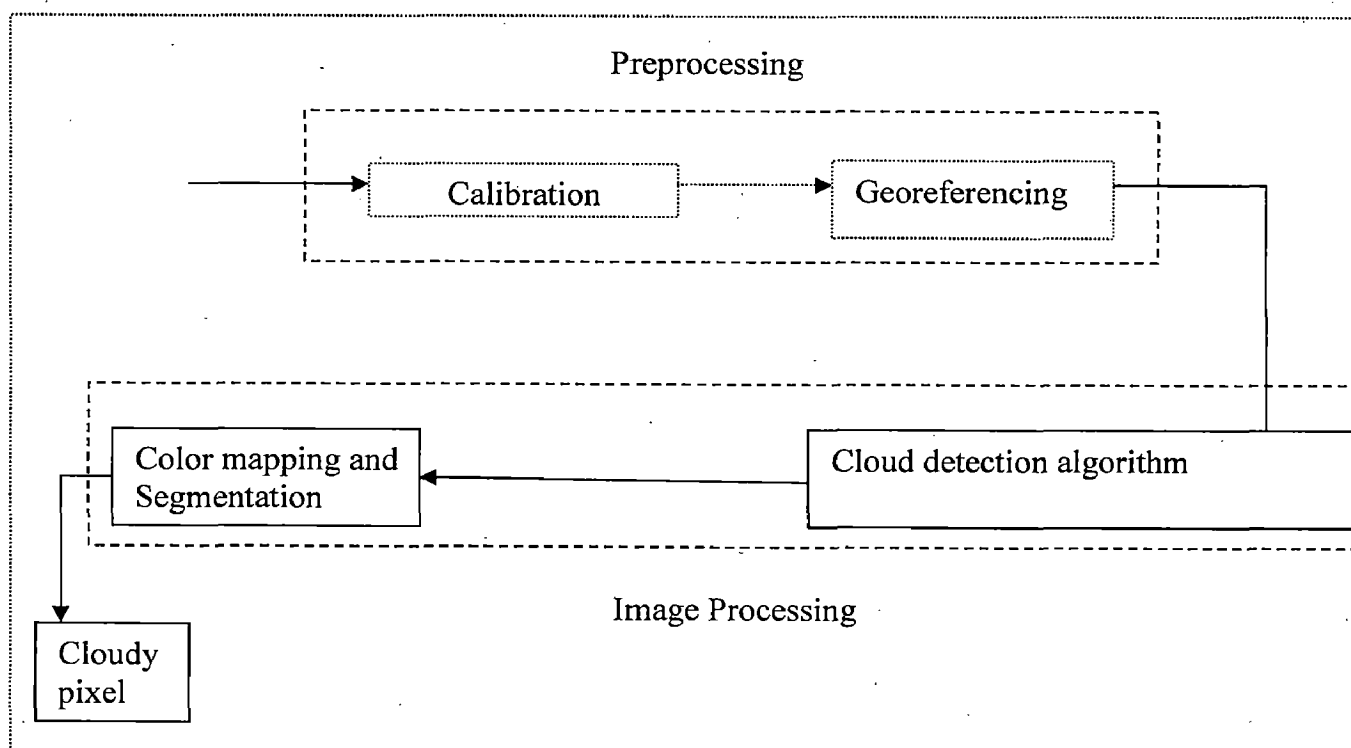


Figure 4.1: Processing Stages of Cloud detection

Multi-thresholding technique was combined along with image processing techniques for observing cloud in Jharia (Jharkhand), India, using operational satellite (NOAA/AVHRR) images. Cloudy pixels of Jharia region are obtained from given image data that shown approximately similar trends during year 1995 to 2004. The multi-thresholding techniques discriminated cloud, and land surfaces successfully. The observed results encourage further application of these techniques which includes the calculation and assessment of actual cloudy surface in cloud image using image processing. With the use of temporal satellite data of various automated detection techniques can be developed and can be compared by using different optimization techniques.

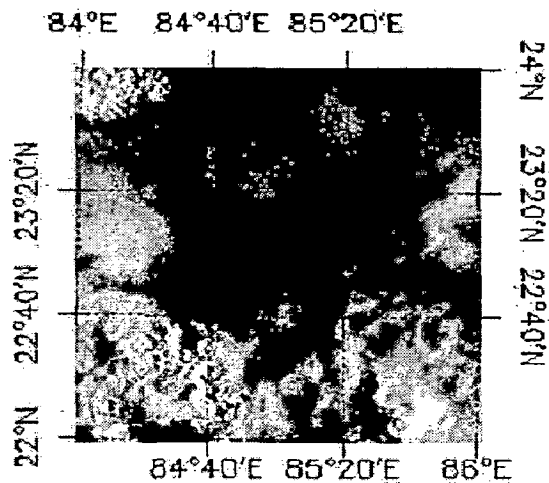


Figure 4.2 Band 1(Sept. 1995)

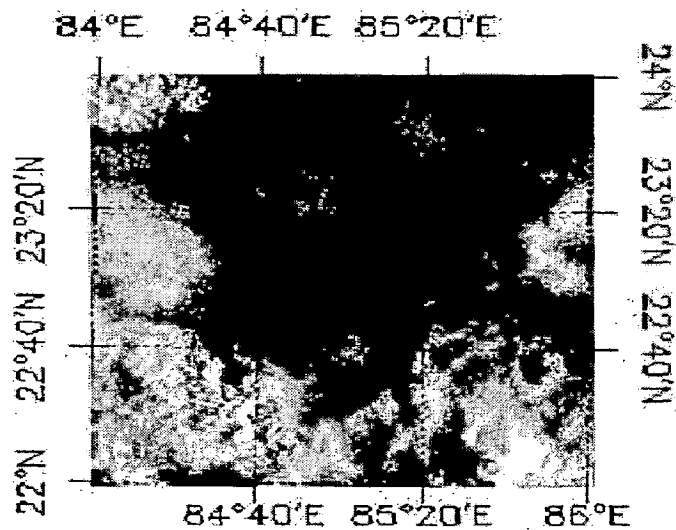


Figure 4.3 Band 2 (Sept. 1995)

Figure 4.2 band1 (Sept 1995), figure 4.3 band 2 (Sept 1995) and figure 4.3 band 4 (Sept 1995) are showing cloudy area in the sky. Band 1 and band 2 are shown the reflectance of the cloud and land surface. Band 4 is shown the temperature of the cloud and land surface.

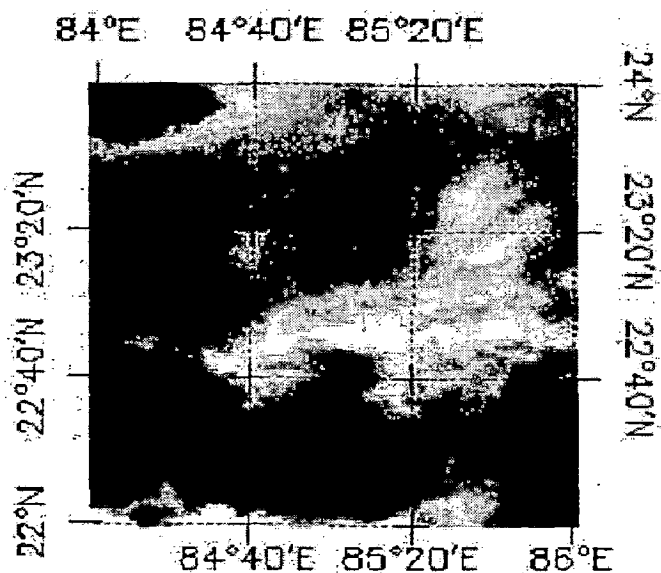


Figure 4.4 band 4 (Sept. 1995)



Different bands of AVHRR images are shown cloudy pixels in the cloud image. Band 1, 2, and 4 images represent reflectivity and temperature.

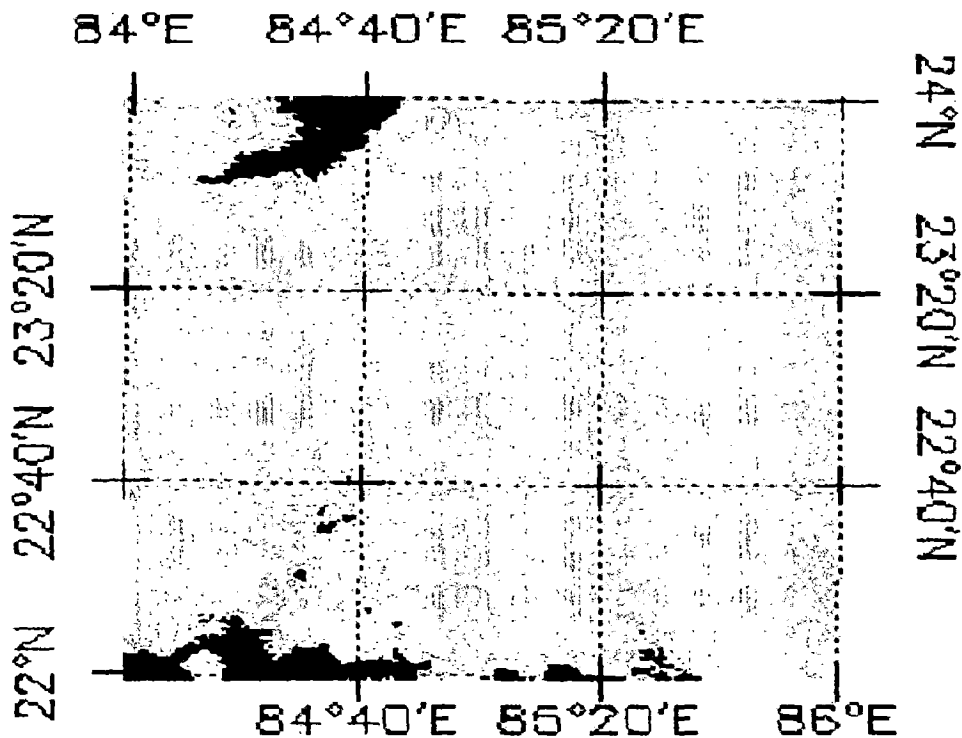


Figure 4.5 Color mapping of cloud

Above image (Fig: 4.5) is showing a different color, i.e. red. This color represents land pixels and pink color corresponds to cloudy pixels. This is the easy method to differentiate the cloudy pixels and land surfaces.

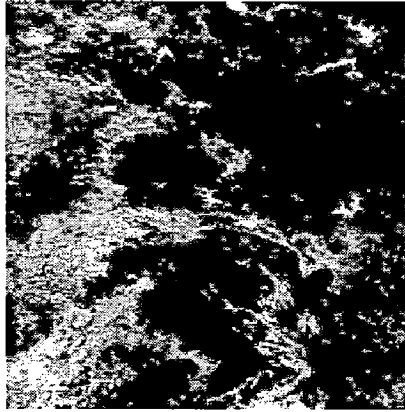


Figure 4.6 Band 4 (Feb. 1995)

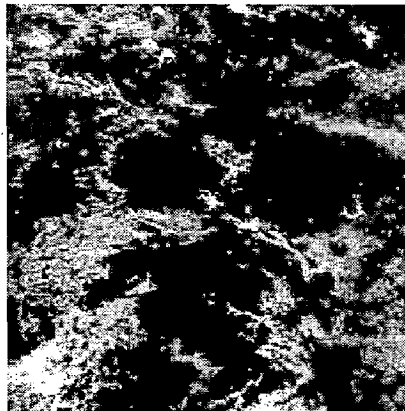


Figure 4.7. Band 4 (Sept 97)

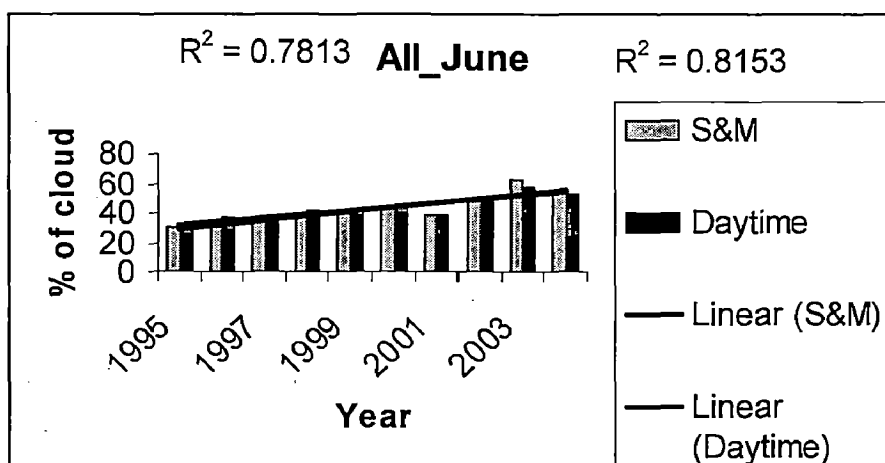
Observations of data (cloud image data) of the same month (i.e. Sept) but for different years (1995 and 1997), shows changes in the cloudy pixels, due to different Meteorological conditions. Due to this, we get different image data for the same month of different years, which shows changes in value of the cloudy pixels.

4.2 Change Detection in Image Analysis

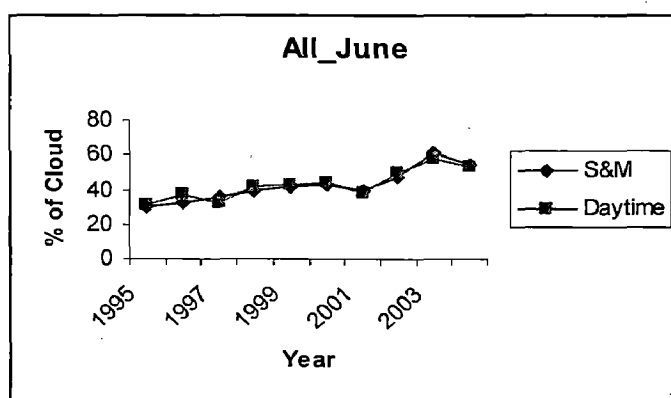
Further processing of these band images gave more idea about features of the region. Graphs for different years that demonstrate the change in the suspected region were plotted (below) and changes in the cloud images of Jharia region (Long north-24.5° to south-22.5° and Lat west-85° to east 87°) were observed. This implies the changing cloud properties of the region.

No.	Method / Year	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004
1	S&M	30	33	36	39	42	43	39	48	62	54
2	Daytime	31	37	33	42	43	44	38	50	58	53

Table 4.1. June data of Split and Merge Clustering and Daytime Algorithm.



Histogram 4.1. June data of Split and Merge Clustering and Daytime Algorithm.



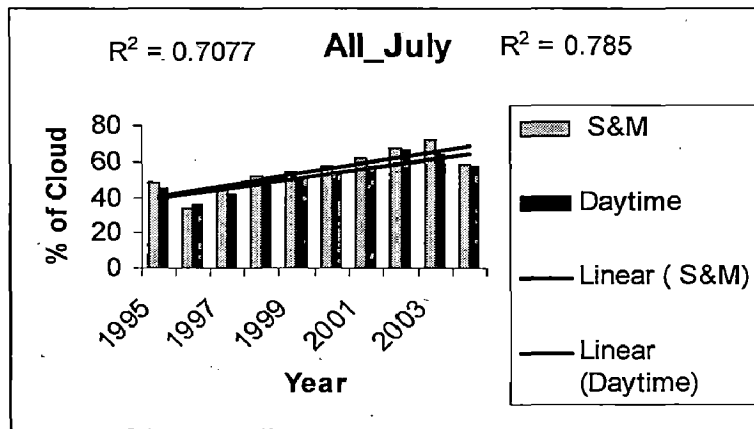
Line graph 4.1. June data of Split and Merge Clustering and Daytime Algorithm.

‘Split and Merge Clustering’ and ‘Daytime’ algorithms are used for checking and detecting percentage of cloudy pixels in the obtained image data for the month of June. The variations in the value of cloudy pixels of the same month (June) of the different year (from 1995 to 2004) are shown in the Histogram 4.1 and Line graph 4.1.

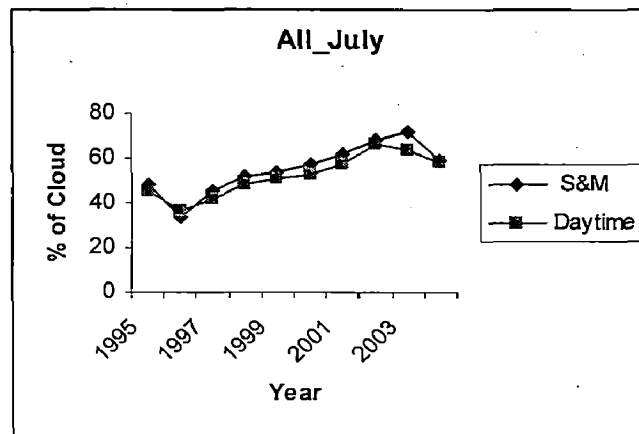
Trend line of Split and Merge Clustering and Daytime algorithms of June data is continuously ascending upwards along with an increment in the year. Further the cloudy pixels are varying year by year.

No.	Method / Year	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004
1	S&M	48	34	45	52	54	57	62	68	72	59
2	Daytime	45	36	42	48	51	53	57	66	64	58

Table 4.2. July data of Split and Merge Clustering and Daytime Algorithm.



Histogram 4.2. July data of Split and merge Clustering and Daytime algorithm



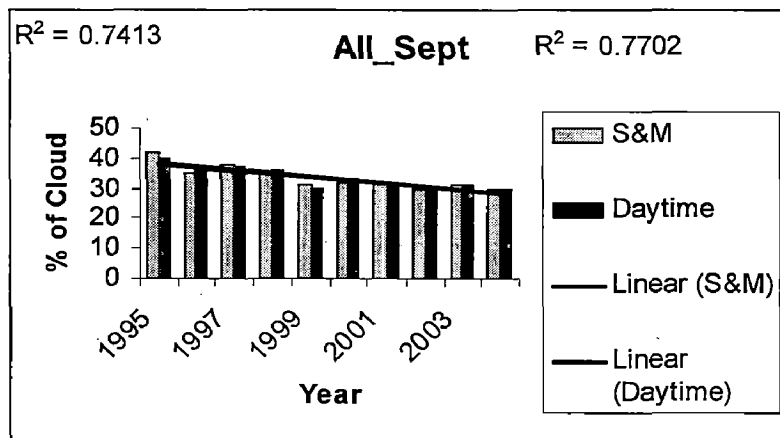
Line graph 4.2. July data of Split and Merge Clustering and Daytime Algorithm

In a similar way the analysis of the cloudy pixel data for the month of July for different years can be carried out. And one can apply the ‘Split and Merge Clustering’ and ‘Daytime’ algorithms and plot the respective histogram and line graph. These are shown in: Histogram 4.2 and Line graph 4.2.

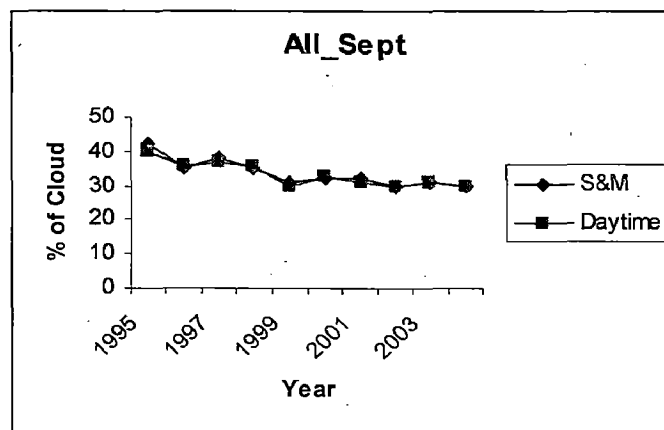
The same inference about the ascending nature of the trend line in the histogram along with the varying cloudy pixel data can be suggested as was done in the previous case for the month of June.

No.	Method / Year	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004
1	S&M	42	35	38	35	31	32	32	30	31	30
2	Daytime	40	36	37	36	30	33	31	30	31	30

Table 4.3. Sept data of Split and Merge Clustering and Daytime Algorithm.



Histogram 4.3. Sept data of Split and merge Clustering and Daytime algorithm



Line graph 4.3. Sept data of Split and Merge Clustering and Daytime Algorithm

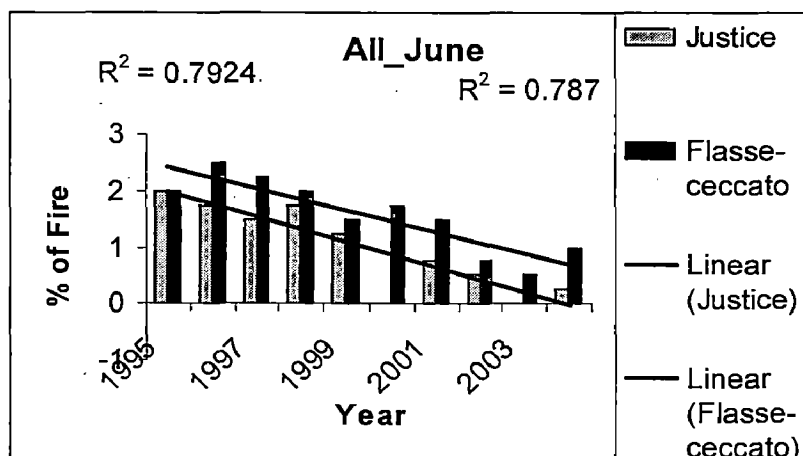
Accordingly, the histogram obtained as a result of the plotting the cloudy image data following the mentioned algorithms are also shown in fig: Histogram 4.3 and Line graph 4.3. The month considered over here is September.

It is seen that the trend line obtained here is showing somewhat different characteristics; i.e. there is a descending nature in the image data, as the year progresses.

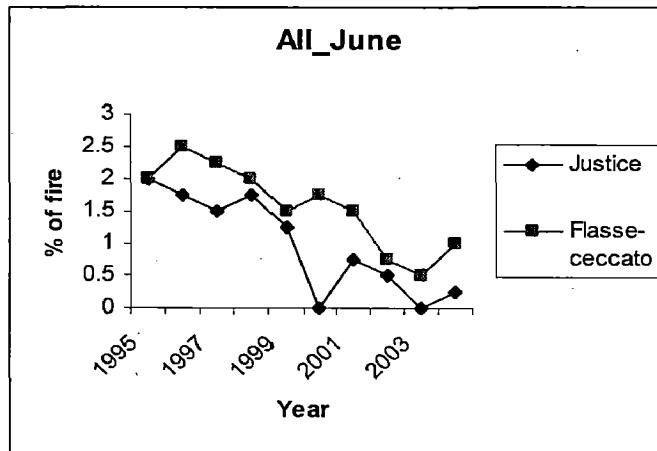
Contextual Algorithm:

No.	Method / Year	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004
1	Justice	2	1.75	1.5	1.75	1.25	0	0.75	0.5	0	0.25
2	Flasse-ceccato	2	2.5	2.25	2	1.5	1.75	1.5	0.75	0.5	1

Table 4.4. June data of Justice and Flasse-ceccato (contextual) Algorithm.



Histogram 4.4. June data of Justice and Flasse-ceccato (contextual) Algorithm.

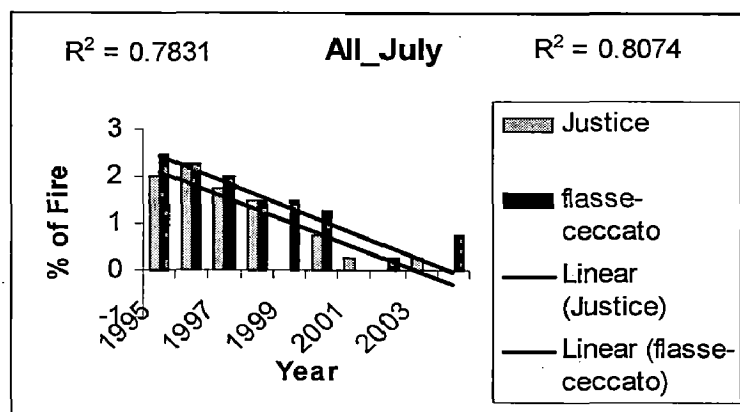


Line graph 4.4. June data of Justice and Flasse-ceccato (contextual) Algorithm.

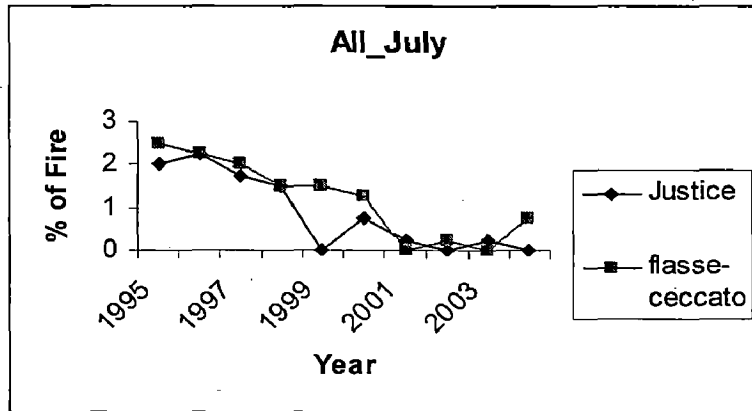
A contextual algorithm can be used to detect the fire pixels. Above shown figures is the result of applying this contextual algorithm. The trend line obtained shows a sharp decline in the contextual data. The month under consideration is June.

No.	Method / Year	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004
1	Justice	2	2.25	1.75	1.5	0	0.75	0.25	0	0.25	0
2	flasse-ceccato	2.5	2.25	2	1.5	1.5	1.25	0	0.25	0	0.75

Table 4.5. July data of Justice and Flasse-ceccato (contextual) Algorithm.



Histogram 4.5. July data of Justice and Flasse-ceccato (contextual) Algorithm.

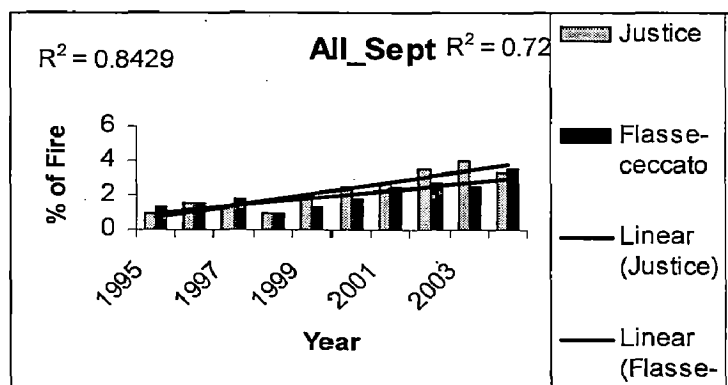


Line graph 4.5. July data of Justice and Flasse-ceccato (contextual) Algorithm

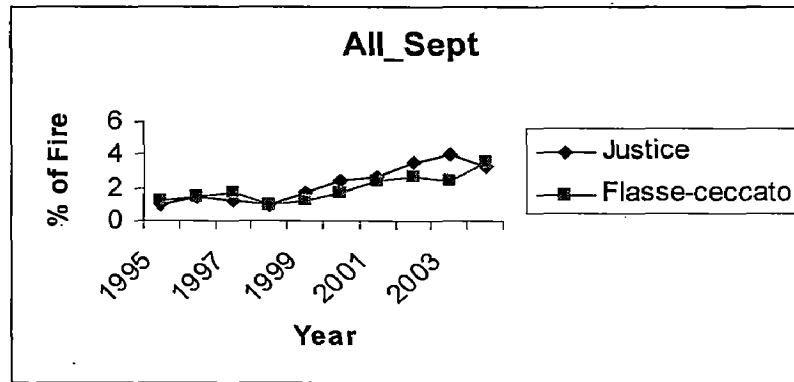
The fire pixel data for the month of July are plotted for different years (from 1995 to 2004), using the Contextual algorithm. These are shown in table 4.5 and its histogram 4.5. There is a similar trend found as was the case with the month of June.

No.	Method / Year	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004
1	Justice	1	1.5	1.25	1	1.75	2.5	2.75	3.5	4	3.25
2	Flasse-ceccato	1.25	1.5	1.75	1	1.25	1.75	2.5	2.75	2.5	3.5

Table 4.6. Sept data of Justice and Flasse-ceccato (contextual) Algorithm.



Histogram 4.6. Sept data of Justice and Flasse-ceccato (contextual) Algorithm.



Line graph 4.6. Sept data of Justice and Flasse-ceccato (contextual) Algorithm

However, for the month of September, there is a different pattern found in the histogram obtained as a result of Contextual algorithm. The trend line is showing some ascending nature. This can be dedicated to the changes caused in season and meteorological aspects.

In this dissertation work, the cloudy image data for the months of June, July and September are shown along with their histograms and line charts.

The histograms plotted are based on the application of various algorithms available such as; Cloud detection algorithm and Contextual algorithm. The data used in this procedure is for ten different years (from 1995 to 2004). A detailed study shows that there is a drastic change in the data obtained and the nature of the respective trend lines for the months of June and September. This difference can be devoted to the various changes brought in the image data because of seasonal changes and meteorological parameters.

Individual month data for 10 year (from 1995 to 2004) are shown in tables (from Table 4.1 to 4.6) and the respective histogram (Histogram4.1 to 4.6) and Line graph (Line graph 4.1 to 4.6) can be referred for the analysis.

Line graphs obtained for different months considered (June, July and September) for a data ranging between ten years (from 1995 to 2004) are showing fluctuations in the values of the cloudy pixels, i.e. year by year, for the same month, cloudy pixels varied.

The algorithms that are used in this work are Daytime algorithm and Split and Merge Clustering algorithm. The histograms thus obtained for the same month of different years as a result of application of these two algorithms shows minute variations in the cloud image data. However the nature of the trend lines is similar for both types of algorithms. In the month of June and July, cloudy pixels are easily achieved because of monsoon period. Contextual algorithms that detect the fire pixel data (Justice and Flasse-ceccato) are used in this work. These two algorithms again give minute changes in the data obtained that are comparable. Further the nature of trend lines for different months considered, illustrate same characteristics. i.e. either constantly descending or ascending in nature.

Observing the changes in the cloudy pixel data over the regions of Jharia for the year 1995 as compared to other years (1996 to 2004), it is concluded that the changes in the image data may be devoted to the variations in the meteorological parameters, such as, wind direction, temperature, evapotranspiration, etc. The images of Sept 1995 and Sept 1997 are showing variations in the value of cloudy pixels. Since AVHRR data is available from the year 1995 to 2004, large temporal coverage can be made for application of change detection algorithm.

4.3. Multi-Thresholding in Image Processing

Thresholding was applied at more than one threshold value which made the cloud and fog coverage pixels segmented in the image. The choice of temperature to detect the cloud pixels was decided by taking the minimum threshold 300 K. In figure 4.5, pink color represents the cloud whereas the red color corresponds to land.

The work was more emphasized on the use of AVHRR and image processing for the detection of cloudy pixels in the image data. This dissertation work show good results that are near to perfection. AVHRR Split-and-Merge clustering and Daytime algorithm, both are showing good results. Both the algorithms are showing a minimum of 30% cloudy pixels. Coming to maximum of the cloudy pixels, the split and merge clustering algorithm is showing a maximum of 72% cloudy pixels. In case of Daytime algorithm this maximum is found to be 66%. Hence the intersection of these two algorithms can be given as 30% cloudy pixels of image data.

Conclusions:

All the methods discussed in this report were found to be useful in the detection of cloudy pixels. Both the algorithms (Split and Merge Clustering and Daytime) are representing cloudy pixels in the region (Jharia, Jharkhand) to a minimum of 30%, during the monsoon period. Change detection, threshold and multi-threshold techniques were combined along with image processing techniques for the detection of cloud on Jharia region of Jharkhand, India, using satellite (AVHRR) images. Some changes were seen during the years 1995 to 2004 in cloudy pixels of cloud images over the Jharia region, and are showing approximately changing trends for changing years (1995 to 2004). The multi-threshold technique was found to be accurate due to highlighting of cloud pixels. The technique though successful in discriminating cloudy pixels from other land surfaces, highlights the false points. The application of multi-thresholding along with segmentation technique of image analysis provides precise results in locating the cloudy pixels in the a Jharia region of Jharkhand. The observed results encourage further work into these techniques which includes the calculation and assessment of actual cloudy pixels using image processing and also correlate it with various meteorological parameters. There is a scope of observing the change detection in large time series data provided with the actual ground truth points. With the use of temporal satellite data an automated detection technique can be developed and can be compared by using different optimization techniques.

REFERENCES

1. Kriebel, K.T., G. Gesell, M. Kastner, and H. Mannstein,(2003), The cloud analysis tool APOLLO : improvements and validations, *International Journal of Remote Sensing*, 24, 2389-2408,
2. R.W.Saunders, K.T. Kriebel,(1988), An improved method for detecting clear sky and cloudy radiances from AVHRR data, *International Journal of Remote Sensing*,9, Nr,1
3. N.K.Sakellariou, H.G.Leighton and Z.Li,(1993), Identification of clear and cloudy pixels at high latitudes from AVHRR Radiances, *International Journal of Remote Sensing*,14,2005-2024.
4. Takashi Yamanouchi, Sadao Kawaguchi,Cloud distribution in the Antartic from AVHRR data and radiation measurements at the surface, *Internatuinal Journal of Remote Sensing*,13,111-127.
5. Gallaudet,T.C.,and Simpson,J.J.(1991), Automated cloud screening of AVHRR Imagery using split-and-merge clustering,*Remote Sensing of Environment*,38,77-121.
6. Gallegos,S.C., Hawkins,J.D., and Cheng, C.F.(1993), A new automated method of Cloud masking for Advanced Very High Resolution Radiometer full- resolution data over the ocean,*Journal of Geophysics.Res.* 98,8505-8516.
7. Pairman,D,and Kittler,J.(1986),Cluster algorithms for use with images of clouds, *International Journal of Remote Sensing*, ,7, 855-866.
8. Rossow,W.B., Mosher, F., Kinsella, E. et al.(1985), ISCCP cloud algorithm Intercomparison. *J.Clim.Appl.Meteor.*24, 877-903.
9. R.W.Saunders, K.T. Kriebel,(1988a) , An improved method for detecting clear Sky and cloudy radiances form AVHRR data, *International Journal of Remote Sensing*,9,123-150.
10. Simpson,J.J., and Humphrey,C.(1990),An automated cloud screening algorithm For daytime Advance Very High Resolution Radiometer imagery , *Journal of Geophysics,res.*,95,13459-13481.

11. Simpson,J.J., and Gobat,J.I.(1995a), Improved cloud detection in GOES scenes Over land, Remote Sensing of Environment ,52,36-54.
12. Simpson,J.J., and Gobat,J.I.(1995a), Improved cloud detection in GOES scenes Over oceans, Remote Sensing of Environment , 52,79-94.
13. Yhann,S.R., and Simpson,J.J.(1995), Application of neural networks to AVHRR Cloud segmentation, IEEE Trans.Geo.Remote Sensing,33,590-604.
14. Di. Girolamo, L and davies R.(1994), A band differenced angular signature technique for cirrus cloud detection, IEEE Trans.Geo.Remote Sensing ,32,890-896.
15. Simpson,J.J. , A.B. Schmitt and A.Harris.(1998a), Improved cloud detection in Along track scanning Radiometer (ATSR) data over the ocean,Remote Esnsing Of Environment,65,1-24.
16. Simpson,J.J. , T. McIntire, Z.Jin, and J.R.Still (1999a),Improved cloud height estimation under arbitrary viewing and illumination conditions using AVHRR data, Remote Sensing of Environment
17. Gurney,C.M.(1982), The use of contextual information to detect cumulous clouds and cloud shadows in Landsat data, International Journal of Remote, Sensing ,3, 51-62.
- 18 Ben-Dor,E. (1994), A precaution regarding cirrus cloud detection from airborne imaging spectrometer data using the 1.38 μ m water vapour band, Remote Sensing of Environment ,50, 346-350.
19. P.M.Barbora(1999), An algorithm for extracting burned areas from time series of AVHRR Gac data applied at a continental scale, Remote Sensing of Environment ,69-71,253.
20. Seze, G. and M. desbois, (1987), cloud cover analysis from satellite imagery using spatial and temporal characteristics of the data ,Journal of applied Meteorology, 26,287-303.

```
1) function [newcluster, newcenter, newtrace] = asmc_part1(b2, b4, del,
clustercenter, K)
```

```
m = clustercenter;
```

```
N = length(b2);
```

```
clusters = zeros(N,100);
```

```
% Assigning each vector to its nearest cluster.
```

```
disp('Assigning each vector to its nearest cluster');
```

```
for i = 1:N
```

```
    y = [b2(i), b4(i), del(i)];
```

```
    for k = 1:K
```

```
        d(k) = (y - m(:,k))*(y - m(:,k));
```

```
    end
```

```
    mind = find(d==min(d));
```

```
    current = clusters(:,mind);
```

```
    location = find(current==0);
```

```
    clusters(location(1),mind) = i;
```

```
end
```

```
% Computing the mean vector of the image
```

```
disp('Computing the mean vector of the image');
```

```
sum = zeros(3,1);
```

```
for k = 1:K
```

```
    sum = sum + length(find(clusters(:,k)~=0))*m(:,k);
```

```
end
```

```
mean = sum/N;
```

```
% Computing the between-cluster scatter matrix.
```

```
disp('Computing the between-cluster scatter matrix');
```

```
sb = zeros(3,3);
```

```
for k = 1:K
```

```
    sb = sb + length(find(clusters(:,k)~=0))*(m(:,k) - mean)*(m(:,k)-mean)';
```

```
end
```

```
newtrace = trace(sb);
```

```
newcluster = clusters;
```

```
newcenter = m;
```

```

2)      function [newclusters, newcenter] = asmc_part2(b2, b4, del, clusters,
clustercenter, Ts, Tm; K)
    N = length(b2);
    % checking whether to split the cluster or not.
    m = clustercenter;
    disp('checking whether to split the cluster or not');
    for k = 1:K
        current = clusters(:,k);
        location = find(current==0);
        current = current(1:location(1)-1);
        tb2 = zeros(length(current),1);
        tb4 = zeros(length(current),1);
        tdel = zeros(length(current),1);
        for i = 1:length(current)
            l = current(i);
            tb2(i) = b2(l);
            tb4(i) = b4(l);
            tdel(i) = del(l);
        end
        mincluster_vector_y = [min(tb2), min(tb4), min(tdel)];
        maxcluster_vector_y = [max(tb2), max(tb4), max(tdel)];

        dm = (maxcluster_vector_y - mincluster_vector_y)'* (maxcluster_vector_y -
mincluster_vector_y);

        if dm>Ts
            %      disp('the cluster number which is getting splitted is');
            %      disp(k);
            newcluster1 = zeros(N,1);
            newcluster2 = zeros(N,1);

            K = K+1;
            for i = 1:length(current)
                ytemp = [tb2(i), tb4(i), tdel(i)];
                d1 = (ytemp - mincluster_vector_y)'* (ytemp - mincluster_vector_y);
                d2 = (ytemp - maxcluster_vector_y)'* (ytemp - maxcluster_vector_y);
                if d1<d2
                    place = find(newcluster1==0);
                    newcluster1(place(1)) = current(i);
                else
                    place = find(newcluster2==0);
                    newcluster2(place(1)) = current(i);
                end
            end
        end
    end
end

```

```

        clusters(:,k) = newcluster1;
        clusters(:,K) = newcluster2;
%       display('The cluster no. which got splitted and the new clusters were formed as
position');
%       disp(k);disp(k);disp(K);
%       display('The length of cluster and the cluster center at position k is');
%       disp(length(newcluster1)); disp(mincluster_vector_y);
%       display('The length of cluster and the cluster center at position K is');
%       disp(length(newcluster2)); disp(maxcluster_vector_y);
        m(:,k) = mincluster_vector_y;
        m(:,K) = maxcluster_vector_y;
        k = 1;
    end
end
% disp('after splitting the clusters, value of K is');
% display(K);
% disp('after splitting the clusters, the new cluster center is');
% disp(m(:,1:K));

% checking whether to merge the clusters or not.
disp('checking whether to merge the clusters or not');
for j = 1:K-1
    for k = j+1:K
        db = (m(:,j)-m(:,k))'*(m(:,j)-m(:,k));
        if db<Tm
%           disp('Now the cluster number j and k are getting merged, where j and k
are');
%           disp(j); disp(k);
            oldcluster1 = clusters(:,j);
            oldcluster2 = clusters(:,k);
            newcluster = zeros(N,1);
            placeold1 = find(oldcluster1==0);
            placeold2 = find(oldcluster2==0);
            finalloc = placeold1(1) + placeold2(1) -2;

            newcluster(1:placeold1(1)-1) = oldcluster1(1:placeold1(1)-1);
            newcluster(placeold1(1): finalloc) = oldcluster2(1:placeold2(1)-1);
%           display('the length of old clusters at position j and k was');
%           disp(placeold1(1)-1); disp(placeold2(1)-1);
%           disp('the length of new cluster at position j is');
%           disp(finalloc);
            clusters(:,j) = newcluster;
            clusters(:,k) = 0;
%           display('the old and new cluster center at position j is');
%           disp(m(:,j));
%           disp('the cluster center at position k was');

```

```

%         disp(m(:,k));
%         m(:,k) = 0;
%         disp('After merging, the cluster center at position k is now');
%         disp(m(:,k));
%         j = 1;
%         k = 1;
%     end
% end
end

```

```

% disp('after merging the clusters, value of K is');
% display(K);
% disp('after merging all the clusters, the new cluster center is');
% disp(m(:,1:K));

```

```

countclust = 0;
for e = 1:K;
    if clusters(:,e)==0
        countclust = countclust+1;
    end
end
% disp('the number of clusters which are having atleast one vector is');
% disp(K-countclust);

```

```

countcen = 0;
for e = 1:K;
    if m(:,e)==0
        countcen = countcen+1;
    end
end
% disp('the number of centers which are having atleast one vector is');
% disp(K-countcen);

```

```

newclusters = zeros(N,K-countclust);
newm = zeros(3,K-countcen);

```

```

count = 1;
for e = 1:K;
    if m(:,e)==0
        count = count;
    else
        newm(:,count) = m(:,e);
        newclusters(:,count) = clusters(:,e);
        count = count+1;
    end
end

```



```

    end
end

% disp('newcluster size is');
% display(size(newclusters));

% disp('new m size is');
% display(size(newm));

newcenter = newm;

```

3) function [final_clusters, final_centers] = cloud_detect(roi)

```

b1 = roi(:,8);
b2 = roi(:,9);
b3 = roi(:,10);
b4 = roi(:,11);
del = b3-b4;
del(find(del<0))=0;
K = 30;
N = length(b1);
b2min = min(b2);
b4min = min(b4);
delmin = min(del);

b2max = max(b2);
b4max = max(b4);
delmax = max(del);

ymax = [b2max, b4max, delmax]';
ymin = [b2min, b4min, delmin]';

Ts = 0.01*(ymax - ymin)'*(ymax - ymin);
Tm = Ts;
Tol = 0.05;

% Computing cluster centers for each cluster
m = zeros(3,100);
for k = 1:K
    m(:,k) = ymin + ((k-1)/(K-1))*(ymax-ymin);
end

```

```

disp('The initial clustercenter is');
display(m(:,1:K));
oldtr = 0;
clustercenter = m;
[newclusters, newcenter, newtrace] = asmc_part1(b2, b4, del, clustercenter, K);
clusters = newclusters;
clustercenter = newcenter;
checkontol = abs((newtrace - oldtr)/newtrace);
% disp('the value of checkontol for the first time is');
% display(checkontol);
while checkontol >= Tol
    oldtr = newtrace;
    [newclusters, newcenter] = asmc_part2(b2, b4, del, clusters, clustercenter, Ts, Tm, K);
%    disp('after calling asmc_part2 first time, now the cluster center is');
%    display(newcenter);
    clusters(:, :) = 0;
    clustercenter(:, :) = 0;
    clustercenter(:, 1:size(newcenter, 2)) = newcenter;
    K = size(newcenter, 2);
    disp('Now K is');
    display(K);
    [newclusters, newcenter, newtrace] = asmc_part1(b2, b4, del, clustercenter, K);
    checkontol = abs((newtrace - oldtr)/newtrace);
    disp('the value of checkontol for the next time is');
    display(checkontol);
    clusters = newclusters;
    clustercenter = newcenter;
end

final_clusters = newclusters(:, 1:K);
final_centers = newcenter(:, 1:K);

```

4) function [cloudarray] = label_clusters(roi, cluster, clustercenter, outputfile)

```
b1 = roi(:,8);  
b2 = roi(:,9);  
b3 = roi(:,10);  
b4 = roi(:,11);  
b5 = roi(:,12);  
xloc = roi(:,2);  
yloc = roi(:,3);  
del = b3-b4;  
del(find(del<0))=0;  
K = size(cluster,2);  
N = length(b2);
```

```
countclust = 0;  
for e = 1:K;  
    if cluster(:,e)==0  
        countclust = countclust+1;  
    end  
end  
countclust
```

```
newcluster = zeros(N,K-countclust);  
count = 1;  
for e = 1:K;  
    if cluster(:,e)==0  
        count = count;  
    else  
        newcluster(:,count) = cluster(:,e);  
        count = count+1;  
    end  
end  
size(newcluster)
```

```
%calculating cluster mean vectors  
meanvector = zeros(3,K-countclust);  
cloudarray = zeros(1,K-countclust);
```

```
for k=1:K-countclust  
    temp = newcluster(:,k);  
    location = find(temp==0);  
    current = temp(1:location(1)-1);  
    tempb2 = zeros(length(current),1);  
    tempb4 = zeros(length(current),1);
```

```

tempdel = zeros(length(current),1);
for i=1:length(current)
    loc = current(i);
    tempb2(i) = b2(loc);
    tempb4(i) = b4(loc);
    tempdel(i) = del(loc);
end
sumvector = [sum(tempb2), sum(tempb4), sum(tempdel)];
meanvector(:,k) = sumvector/length(current);
end

amin = min(meanvector(1,:));
tmax = max(meanvector(2,:));
delmin = min(meanvector(3,:));

aTh = mean(b2);
delTh = mean(del);
tTh = mean(b4);

preaTh = 0;
predelTh = 0;
pretTh = 0;

while aTh~=preaTh
    temp1 = b2.*(b2>aTh);
    temp2 = b2.*(b2<=aTh);

    zerotemp1 = length(find(temp1==0));
    zerotemp2 = length(find(temp2==0));

    preaTh = aTh;

    m1 = sum(temp1)/(length(temp1) - zerotemp1);
    m2 = sum(temp2)/(length(temp2) - zerotemp2);

    aTh = (m1+m2)/2;
end

while delTh~=predelTh
    temp1 = del.*(del>delTh);
    temp2 = del.*(del<=delTh);

    zerotemp1 = length(find(temp1==0));
    zerotemp2 = length(find(temp2==0));

    predelTh = delTh;

```

```

m1 = sum(temp1)/(length(temp1) - zerotemp1);
m2 = sum(temp2)/(length(temp2) - zerotemp2);

delTh = (m1+m2)/2;
end

while tTh~=pretTh
temp1 = b4.*(b4>tTh);
temp2 = b4.*(b4<=tTh);

zerotemp1 = length(find(temp1==0));
zerotemp2 = length(find(temp2==0));

pretTh = tTh;

m1 = sum(temp1)/(length(temp1) - zerotemp1);
m2 = sum(temp2)/(length(temp2) - zerotemp2);

tTh = (m1+m2)/2;
end

m = (delTh - delmin)/(aTh - amin)
n = (delmin - delTh)/(tmax - tTh)

for k=1:K-countclust
a0 = meanvector(1,k)
t0 = meanvector(2,k)
del0 = meanvector(3,k)

ds = ((m*a0) + (n*t0) + del0 - delmin - (m*amin) - (n*tTh))/sqrt(m^2 + n^2 + 1)
if ds>0
cloudarray(k) = 1;
end
end

fid = fopen(outputfile,'w');
count=1;
for k=1:K-countclust
if cloudarray(k)==1
temp = newcluster(:,k);
location = find(temp==0);
current = temp(1:location(1)-1);

tempb1 = zeros(length(current),1);

```

```
tempb2 = zeros(length(current),1);
tempb3 = zeros(length(current),1);
tempb4 = zeros(length(current),1);
tempb5 = zeros(length(current),1);
tempx = zeros(length(current),1);
tempy = zeros(length(current),1);
```

```
for i=1:length(current)
    loc = current(i);
```

```
fprintf(fid, '%u\t%u\t%u\t%u\t%u\t%u\t%u\t%u\n', count, xloc(loc), yloc(loc), b1(loc), b2(lo
c), b3(loc), b4(loc), b5(loc));
    count = count+1;
end
end
end
fclose(fid);
```

```

5)      function cl6=daytime1(im)
b1=im(:,:,1);
b2=im(:,:,2);
b3=im(:,:,3);
b4=im(:,:,4);
b5=im(:,:,5);
delta=(b3-b4);
stdev_4=pstd(b4);
ratio=(b2/b1);
T11_T12=(b4-b5);
Tmin=300;
SDthr=273;
ALBthr=10;
td=T11_T12;
[a,b]=size(ratio);
count1=0;
count2=0;
for r=1:1:a
    for s=1:1:b
        if ((b5(r,s)>Tmin) && (stdev_4>SDthr) && (b1(r,s)>ALBthr) &&
(0.75<ratio(r,s)<1.6) && (td(r,s)> 0.81))
            cloudepix1=r
            cloudepix2=s
            disp('pixel is cloudy');
            count1=count1+1;
            else
            %disp('pixel is non-cloudy');
            count2=count2+1;
        end
    end
end
end
fprintf('total cloudy pixel is%6d\n ',count1);
fprintf('total non-cloudy pixel is %6d\n',count2);
cl6=1;

```

```

6) function cl=fiery(im)
b3=im(:,:,3);
b4=im(:,:,4);
delta=(b3-b4);
mean_4=pmean(b4);
mean_del=pmean(delta);
stdev_del=pstd(delta);
delta_t=(2*stdev_del)+0.1;
value=mean_del+delta_t;
[a,b]=size(delta);
count1=0;
count2=0;
for r=1:1:a
    for s=1:1:b
        if (delta(r,s)>value && b4(r,s)>mean_4)
            firepix1=r
            firepix2=s
            disp('fire pixel');
            count1=count1+1;
        else
            %disp('pixel is non-fire');
            count2=count2+1;
        end
    end
end
end
fprintf('total fire pixel is%6d\n ',count1);
fprintf('total non-fire pixel is %6d\n',count2);
cl=1;

```


7)

```
function c11=fiery1(im,m,n)
b3=im(:,:,3);
b4=im(:,:,4);
delta=(b3-b4);
mean_4=pmean(b4);
mean_del=pmean(delta);
stdev_del=pstd(delta);
delta_t=(2*stdev_del)+0.1;
value=mean_del+delta_t;
if (delta(m,n)>value && b4(m,n)>mean_4)
    disp('fire pixel');
else
    disp('pixel is non-fire');
end
c11=1;
```

```

8)      function cl2=fiery2(im)
b3=im(:,:,3);
b4=im(:,:,4);
delta=(b3-b4);
mean_3=pmean(b3);
mean_del=pmean(delta);
stdev_3=pstd(b3);
stdev_del=pstd(delta);
value1=(b3-(mean_3+2*stdev_3));
value2=(mean_del+2*stdev_del);
[a,b]=size(delta);
count1=0;
count2=0;
for r=1:1:a
    for s=1:1:b
        if (abs(value1(r,s))>3 && (delta(r,s)>value2))
            cloudy pix1=r
            cloudy pix2=s
            disp('fire pixel');
            count1=count1+1;
        else
            %disp('pixel is non-fire');
            count2=count2+1;
        end
    end
end
end
fprintf('total fire pixel is%6d\n ',count1);
fprintf('total non-fire pixel is %6d\n',count2);
cl2=1;

```

```
9) function cl3=fiery3(im,m,n)
b3=im(:,:,3);
b4=im(:,:,4);
delta=(b3-b4);
mean_3=pmean(b3);
stdev_3=pstd(b3);
mean_del=pmean(delta);
stdev_del=pstd(delta);
value1=b3-(mean_3+2*stdev_3);
value2=mean_del+2*stdev_del;
if (value1(m,n)>3 && delta(m,n)>value2)
    disp('fire pixel');
else
    disp('pixel is non-fire');
end
cl3=1;
```