RESEARCH SUPPORT OF THE WETNET PROGRAM

Final Technical Report for Research Grant # NAG8-929

Submitted to:

MICHAEL GOODMAN

George C. Marshall Space Flight Center
Marshall Space Flight Center, AL 35812

Principal Investigators:

JOHN E. ESTES
KENNETH C. McGWIRE
JOSEPH SCEPAN

Authors:

SY HENDERSON
MICHAEL LAWLESS

Remote Sensing Research Unit
Department of Geography
University of California
Santa Barbara, CA 93106-4060 U.S.A.

1 October 1995

RESEARCH SUPPORT OF THE WETNET PROGRAM

Final Technical Report for Research Grant # NAG8-929

Section 1

AN EXAMINATION OF THE MICROWAVE VEGETATION INDEX

Author:

SY HENDERSON

1 October 1995

Remote Sensing Research Unit
Department of Geography
University of California
Santa Barbara, CA 93106-4060 U.S.A.

RSRU

TABLE OF CONTENTS

1. Introduction

This study examines various aspects of the Microwave Vegetation Index (MVI). MVI is a derived signal created by differencing the spectral response of the 37 GHz horizontally and vertically polarized passive microwave signals. The microwave signal employed to derive this index is thought to be primarily influenced by vegetation structure, vegetation growth, standing water, and precipitation. The State of California is the study site for this research. Imagery from the Special Sensor Microwave/Imager (SSM/I) is used for the creation of MVI datasets analyzed in this research.

The object of this research is to determine *whether MVI corresponds with some quantifiable vegetation parameter (such as vegetation density) or whether the index is more affected by known biogeophysical parameters such as antecedent precipitation.* A secondary question associated with the above is *whether the vegetation attributes that MVI is employed to determine can be more easily and accurately evaluated by other remote sensing means.* An important associated question to be addressed in the study is the effect of different multi-temporal compositing techniques on the derived MVI dataset.

Fundamental research to create an accurate baseline dataset of global land coverage characteristics is a necessity if society is to study resource exploitation (IGBP, 1992). To produce a more robust dataset a wide range of spectral responses should be studied. Included in this baseline dataset should be passive microwave data.

Vegetation is important because it is one of the primary factors affecting the exchange of energy and moisture between the earth's surface and the atmosphere. Anthropogenic change of the earth's surface such as deforestation in the Amazon has the potential to affect the biosphere (Henderson-Sellers, 1987). Studying MVI is important as a scientific objective due to the possible ability of the index to record and provide data with which to catalogue vegetation change on the earth's surface. Passive microwave is not as affected by weather conditions as are traditional remote sensing systems using visible and near infrared wavelengths. By using the advantages of passive microwave to record vegetation in areas of the world that are commonly cloud covered through some or all of the year MVI may give a greater understanding of the effects of both manmade and natural processes on the spatial distribution and temporal variability of vegetation cover.

Within the last seven years a growing amount of work has been accomplished on the uses of passive microwave imagery to study the earth's surface (Choudhury and Tucker, 1987). The primary reason for the increase in the

amount of work in this portion of the electromagnetic (EM) spectrum has been due to the launching of satellite platforms that image at these wavelengths. In 1978 the Scanning Multichannel Microwave Radiometer (SMMR) was launched. Analysis of the data received from the SMMR allowed designers to develop a more effective and efficient satellite platform for recording passive microwave data. Since 1987 the Department of Defense has launched several meteorological satellites that carry passive microwave instruments. These satellites are called Defense Meteorological Satellite Program (DMSP) and contain the SSM/I sensor.

The physical theory underpinning MVI is well developed (Becker and Choudhury, 1988). A conceptual framework created by Choudhury et al. in 1988-89 within which to assess the efficacy of MVI has been partially developed. However, the people who developed the framework suffered from not understanding the bias created by the nature of the pioneer study site areas, which were seasonal grasslands. The theoretical framework upon which the derivation of MVI is based is of relatively recent origin (Choudhury, 1989), and is evolving to better fit empirical realities. MVI works by bare soils emitting highly polarized energy. Areas with little vegetation cover produce large MVI values. Vegetation acts to depolarize the microwave signal. Heavily vegetated areas produce signals with low MVI values.

This work advances our understanding of the fundamental nature of MVI by studying vegetation as a mixture of structural types, such as forest and grassland. The study further advances our understanding by creating multi-temporal precipitation datasets to compare the affects of precipitation upon MVI. This work will help to lay the groundwork for the use of passive microwave spectral information either as an adjunct to visible and near infrared imagery in areas where that is feasible or for the use of passive microwave alone in areas of moderate cloud coverage.

Three sites in Central and Southern California were studied. The first site located near Fresno, was chosen as an example of an agricultural region. The second site located northeast of Santa Ynez, was chosen as an example of Coastal Range vegetation. The third site located near Rogers Dry lake, was chosen as an example of desert vegetation. These study sites have been designated for multi-temporal studies of MVI, see figure 1.1. At each of these sites MVI, precipitation, and Normalized Difference Vegetation Index (NDVI) multi-temporal data was collected. These sites were chosen to evaluate MVI in diverse conditions to provide researchers with empirical data. These sites were used to see if the MVI multi-temporal trajectory corresponds with expected vegetation growth trajectories.

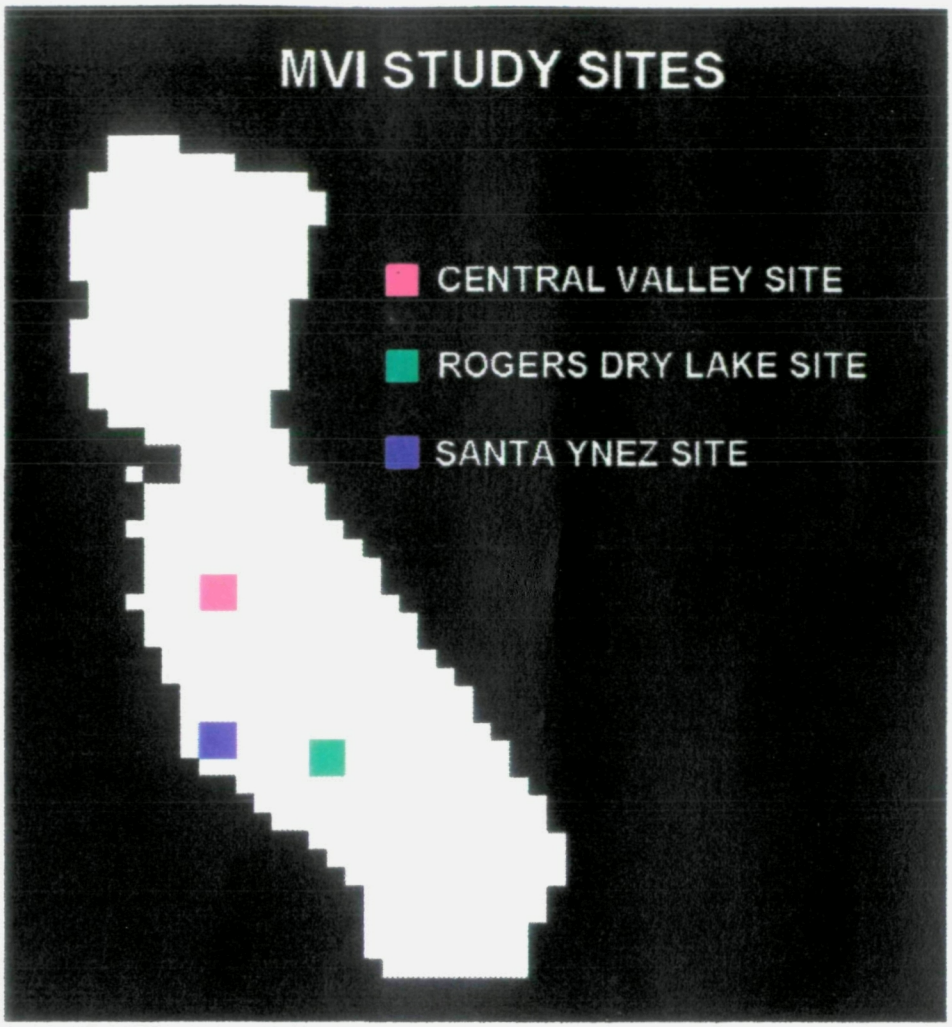


Figure 1.1: The three study sites in California.

Imagery of California as a whole is used as a dataset for questions of a statistical nature that require a large number of samples. Values from the statewide MVI multi-temporal datasets are regressed against 4 biogeophysical parameters; vegetation structure, vegetation primary productivity (using NDVI as a surrogate), water bodies and precipitation in order to gain an understanding of the effects of the parameters on the derived MVI signal.

In this research an MVI dataset spanning the period February 15, 1989 through April 25, 1990 has been created using National Aeronautic and Space Administration (NASA) supplied brightness temperature data. Information from the DMSP satellite 37 GHz wavelength SSM/I sensor in both horizontal and vertical polarization has been processed using the MVI algorithm. In conjunction with the MVI algorithm a multi-temporal compositing technique was used to create datasets that correspond to 14 day periods.

Results of the work should show that each parameter's ability to affect MVI is based upon physical processes inherent at passive microwave frequencies. These processes in general are directly related to the emissivity of a substance, which in turn is related to the dielectric constant of the substance. The primary influence on the dielectric constant of a substance is whether it contains water. The dielectric constant of any substance varies as a function of wavelength and temperature (Ulaby et al., 1986). The second order of influence on the passive microwave signal is surface roughness. Analysis of tertiary order effects on passive microwave physics created by environmental features, conditions and processes are beyond the scale of this paper, but would include such environmental conditions as seasonal temperature changes of the atmospheric column (Choudhury et al., 1992).

Passive microwave remote sensing is predominantly affected by the dielectric constant of the imaged area. The dielectric constant of an area in turn is most affected by the amount of water in or on the ground. Physically the condition that MVI should be most sensitive to is the amount of water in the scene. The first parameter to be compared to the derived MVI value is a data layer containing large surface water bodies, e.g. lakes.

The second parameter employed to attempt to quantify MVI sensitivity to biogeophysical parameters is precipitation. Precipitation should be the second most important contributor to the MVI signal (Fung and Ulaby, 1983, Heymsfield and Fulton, 1992). Large rainfall events should produce a spike in the MVI signal. Precipitation data is composited over 14 day periods to be comparable with the MVI data.

The third parameter to be compared to MVI, vegetation structure, should have a second order effect on MVI. Vegetation structure should affect the MVI signal both because it is an inherent component of surface roughness and because it contains water. It is hypothesized that the denser the vegetation structure the lower the MVI value should be (Justice et al., 1989). For the purposes of this paper vegetation structure is divided into 5 categories based largely on gross morphology. Use of such gross categories is deemed appropriate owing to the coarse resolution of the sensor system. Categories include forest, grass, scrub, shrub and woodland. Spatial locations of categories was derived using the Munz (1968) classification system.

The fourth parameter to be compared to MVI is vegetation growth. Vegetation growth should produce a second order affect on the MVI signal. Theory states that as vegetation grows there should be more vegetation structure, which would increase surface roughness and retained water (Choudhury 1989). The result

should be a decrease in MVI. Vegetation growth will be approximated in this research by NDVI from which information on primary productivity is derived.

As previously stated another aspect of this research involves the study of different techniques for compositing MVI data. The objective is to determine whether the different techniques produce significantly different end products. The first compositing technique to be used for paper work is the one commonly seen in the literature. It is best described in Choudhury (1989). In this method, over a given time span, the second lowest value for a designated pixel is chosen. Theory states that MVI will decrease with increasing cloud coverage due to the depolarizing aspects of the water droplets (Choudhury, 1989). The Choudhury method is used in hopes that cloud contamination of a given scene will be minimized.

The second method, which will be called the Marshall Space Flight Center (MSFC) method, takes the average value for a given pixel over a given time period. The rationale behind this technique postulates that it's use produces the most accurate representation of the ground surface over time (Wetnet Users Manual, 1992).

The necessary datasets were produced to study the different compositing techniques. The compositing period covered 14 days in order to be comparable to the NDVI compositing period. The minimum number of images used for this time period was 2. The maximum number of images for the 14 day time period was 8. The minimum and maximum number of images are doubled for the MSFC approach which use both ascending and descending overpass data.

Three datasets were composited. These consist of a MSFC variant and two Choudhury variants. The two Choudhury data sets were completed in order to study the difference between the morning ascending passes and the evening descending passes. It seems reasonable to assume that the morning and evening passes could be biased by transient weather conditions, such as morning dew or coastal fog or evening thunderclouds.

Follow up work on MVI is logical on four fronts. First, the multi-temporal composites show statistically significant differences. Research to calibrate the absolute bias of these composited datasets particularly with regard to using solely ascending and descending passes is necessary to increase the understanding of MVI.

Second, more research should be attempted on California during a year with heavy precipitation. This research may have been biased by the fact that the 1989-90 period was a time of drought. In particular, more site specific research

should be attempted with MVI. Previous studies have been restricted to continental scale research. While the size of the SSM/I footprint does make site specific research more difficult, the possible benefits of determining whether MVI can follow the expected phenological pattern for a given type of surface cover is worthwhile.

Third, in the literature dealing with MVI the wavelength of interest is 37 GHz. One of the problems the Choudhury technique is designed to minimize is cloud contamination. The 37 GHz data uses only two channels of the SSM/I data. It would be interesting in the future to use data in the other channels such as those enumerated by Neale et al. (1990) to "flag" pixels that are being affected by cloud coverage and standing water.

In the above we have observed that MVI is important because it evaluates vegetation at wavelengths not commonly employed. MVI has been defined as the difference between the horizontal and vertical polarizations of passive microwave at 37 GHz. The study area is the state of California for statistical purposes and three study sites in central and southern California for more in depth study. MVI data was created for California for the time frame February 15, 1989 until April 25, 1990 using three compositing techniques. Finally three possible follow-on topics for the next logical step for study have been suggested. These are: 1) determine the absolute bias created by using different multi-temporal compositing techniques; 2) study California during a year with heavy precipitation; 3) use other wavelengths available with the SSM/I sensor to screen pixels for heavy cloud coverage and standing water.

In this technical report, Section Two contains background information on the State of California and the three MVI study sites. Section Tree describes the methods used to create the MVI and independent variables datasets. Section Four presents the results of the experiment. Section Five summarizes and concludes the work.

2. Background

2.1 Introduction

This chapter presents the basic information necessary to understand the paper work. The chapter starts with a description of the imaging platform and the sensor system used for the creation of the passive microwave imagery analyzed in this study. The second section of this chapter provides the reader with important information on the physical setting of the study area. The final section summarizes the information presented in this chapter.

2.2 Special Sensor Microwave/Imager

The first Special Sensor Microwave/Imager (SSM/I) was launched on June 19, 1987 on the Defense Meteorological Satellite Program (DMSP) Block 5D-2 Spacecraft F8. SSM/I is a seven channel four frequency, linearly polarized, passive microwave radiometric system. SSM/I measures atmospheric, oceanic, and terrestrial microwave brightness temperatures (T_B) at 19.3, 22.2, 37, and 85.5 GHz. All except the 22.2 GHz channel receive signals in both the horizontal (H) and vertical (V) polarizations. The 22.2 GHz channel receives signals only in the vertical polarization (Hollinger, 1990).

SSM/I in a circular sun-synchronous near polar orbit (DMSP, 1989) which results in 14.1 full orbits per day.

2.2.1 Instruments

The instrument consists of an offset parabolic reflector 24 x 26 inches in size located on a spinning mount. A small mirror and a hot reference absorber are positioned off axis to occult the feed once each scan. The mirror reflects cold sky radiation to serve as an absolute calibration reference in conjunction with the hot reference absorber (DMSP, 1989).

Scanning is conducted looking aft of the satellite from left to right with the active scene measurements $\pm$ 51.2° from the aft direction. The result is a swath width of 1,400 km.

2.2.2 Geolocation Error

The operational requirement of geolocating SSM/I data was defined by half the 3 dB beam diameter of the highest resolution channel, 85.5 GHz. The 85.5 GHz channel has a footprint of ~14 km. Errors of 20-30 km. were routinely observed for land/water interfaces during early use of SSM/I data (Poe & Conway, 1990). These geolocation errors create difficulties for ground truthing the data. Prior to July 1989 the Fleet Numerical Oceanography Center (FNOC) used an ephemeris prediction program created by NASA. Spacecraft ephemeris used in geolocation processing of the SSM/I data included subsatellite geodetic latitude, longitude, and altitude over various time frames. Latitude and longitude data was found to be skewed by a time shift of -1.2 seconds. Correcting this error significantly corrected geolocation error (the amount of actual correction depends on each governmental agency receiving SSM/I data as several agencies have, for reasons of their own, used slightly different programs).

A second process used to reduce geolocation error was a trial and error procedure changing spacecraft attitude. After numerous iterations a series of attitude corrections was found to reduce geolocation error to 5-7 km (Poe & Conway, 1990).

2.3 Physical Setting

The first half of this section is devoted to a short description of the physical parameters for the State of California as a whole. The physical conditions represented within the state form the basis for the statistical analysis of MVI. The second portion of this section will describe the three regional study sites which are located in the Fresno, the Santa Ynez and the Rogers Dry Lake areas of California.

Topographically the State of California is quite varied. Elevations range from -276 feet to 14,495 feet above sea level. Weather shows comparable extremes with recorded temperatures ranging from -45° to 134° F. Annual precipitation at selected weather stations across the state has ranged from 161 inches in a year to over a year without any measurable rain (NOAA, 1974).

2.3.1 Geomorphology

The State of California is located at the junction of two tectonic plates; the Pacific plate and the North American plate. California's rocks range from ancient Precambrian to currently forming sediments (Norris and Webb, 1990).

Mountain ranges cover more than one-half of the State's surface (Durrenberger, 1972). Most ranges trend northwest-southeast. The most important range is the Sierra Nevada, lying like a backbone along the eastern margin of the State. It covers around one-fifth of the total area of California. The Central Valley spans about 500 miles in length and has an average width of 40 miles. It is sandwiched between the Sierra Nevada and the Coastal Ranges. The Coast Ranges are to the west of the Great Valley. They span approximately 550 miles and are cleaved by the San Francisco Bay system. The region drained by the Klamath river in northwest California is collectively known as the Klamath mountains. This region covers approximately 30,300 square kilometers. The Modoc plateau in northeastern California covers about 26,000 square kilometers. The Cascade Range extends from southern British Columbia to Lassen peak in California. The Cascades in California cover the north central to north eastern part of the State. The Cascades grade east into the Modoc plateau.

In southern California two important mountain ranges are the Peninsular and Transverse Ranges. The Transverse Range lies just north of Los Angeles basin and extends about 520 kilometers from Point Arguello and San Miguel island to the mountains of the Joshua Tree National Monument. The Peninsular Range extends from the Transverse Range south to the tip of Baja California, Mexico. In California they extend about 200 kilometers and are from 48 to 160 kilometers in width.

The Mojave Desert occupies about 65,000 square kilometers in southeastern California. It roughly corresponds to the "high" desert. The Colorado Desert roughly corresponds to the "low" desert. The region is bounded in the north by the eastern Transverse Ranges, in the east by the Colorado river, in the south by the Mexican border, and in the west by the Peninsular ranges.

2.3.2 Climate

The climate of California is varied as well with weather and precipitation regimes being strongly affected by topography. Isotherms tend to run northwest-southeast parallel to the major mountain ranges of the State. The coastal zone, under the influence of the mid-latitude westerly winds, is primarily affected by the Pacific Ocean. The maritime influence moderates the temperature extremes found in other portions of the State. The coastal zone tends to have warm winters, cool summers and high relative humidity (NOAA, 1974).

As one moves inland from the coast the weather becomes more continental. The temperature extremes increase and there is a larger fluctuation in daily

temperature ranges. Relative humidity decreases. A transition zone is present between the coastal and continental regimes.

There is a pronounced seasonal component to the State's climate and California has basically only two seasons; wet and dry. Summer is generally a dry period over the whole State, although occasionally moist air moves northward from the Gulf of California to produce local showers in the desert and mountain portions of the State. In winter storms progress across the State bringing rain at lower elevations and snow at higher elevations.

The rainy season runs from the end of September through April, with the heaviest precipitation occurring during December, January and February. There is a general north-south gradient for precipitation but, in general, topography produces a more important precipitation gradient (Donley et al. 1979).

The maritime regions of the State are affected by fog banks which extend inshore a comparatively short distance. The fog tends to move inland during the night and evaporates over the land during the day.

2.3.3 Vegetation

Vegetation in California is strongly interconnected with precipitation patterns. Regions of lower elevation, even near the coast, are relatively dry. Along the coast prairie-scrub and sage formations dominate. The Central Valley has been heavily modified by agricultural processes. Before anthropogenic modification, however, most of the area was a treeless bunchgrass prairie.

Along the coast from Northern California to the Monterey region redwoods predominate. At these latitudes as you move inland the vegetation turns to mixed evergreen forest.

In the western Cascades and in the Sierras, as well as on the south facing slopes of the Transverse mountain ranges, there is a succession of plant communities related to the increasing precipitation at higher altitudes. The succession starts at lower altitudes with drought tolerant species such as digger-pine and blue oak. Succession proceeds through ponderosa and mixed pine-fir forests to alpine communities. The species mix is shifted somewhat in southern California at lower altitudes. Chaparral formations are the predominate communities at lower altitudes and instead of a digger-pine or blue oak layer a live oak layer will exist.

In the Modoc plateau region the dominant vegetation is basin sagebrush and wheatgrass. In the Mojave desert the vegetation is mainly creosote and bursage. In this region vegetation is sparsely distributed (Hornbeck, 1983).

2.3.4 Study Sites

Three study sites are used in an attempt to determine if MVI corresponds with the phenological pattern of the vegetation at each of these sites. Each site is 2 x 2 SSM/I pixels in size: 50 x 50 km (2,500 km^2). The three sites are representative of three ecosystems in California.

The first site is representative of the Coastal Range mountains. It is located northeast of Santa Ynez region at approximately 119.5° W, 34.7° N.

Geomorphologically the site predominantly is part of the San Rafael mountains which generally follow a northwest-southeast trend. The southern part of the site consists of a portion of the Santa Ynez valley, which bounded by two mountain systems; the Santa Ynez mountains which follow an east/west trend, and the San Rafael mountains. The San Rafeal mountains are part of the Coastal range. The Santa Ynez mountains are part of the Transverse range. They are an anticline. The southern aspect is so much more apparent that the mountains are sometimes described as a south-dipping homocline with beds strongly overturned to the north, (Norris and Webb, 1990). The region is drained by the Santa Ynez and Santa Maria rivers.

The mountains are the predominant feature of the site and generally range from 700 to 1,400 meters above sea level. The geologic parent material is sedimentary, predominantly made up of Upper Cretaceous marine and Eocene marine material.

Precipitation within the site is quite variable depending on elevation. Using the data from two weather stations, the Santa Maria WSO and the Los Prietos Ranger Station, which respectively approximately border the northwest sector and the southeast sector of the site, the average total rainfall of the two stations is 17.4 inches a year.

Vegetation is predominantly a southern oak forest (*Quercus*) in the valley bottom and lower mountains with the mountainsides predominantly covered with chaparral communities such as *Adenostoma, Arctostaphylos,* and *Ceonothus.* At upper elevations some coniferous vegetation such as juniper-pinyon woodlands (*Juniperus-Pinus*) are found. For research on this and the other two

sites the sites are broken up into 6 vegetation categories. The reason six categories were chosen will be more fully explained in Chapter four. Essentially these categories were chosen to correspond with Munz' (1968) California vegetation categories. The categories are forest, grassland, scrub, shrub, woodland, and other (predominantly agriculture). (For a more explicit representation of land surface coverage at all three sites also see figure 4.8 in Chapter 4). These figures were derived from a Remote Sensing Research Unit (RSRU) database created from a CalVeg map.

The second site is representative of the California desert. It is located in the Rogers Dry Lake region at approximately 118.1° W, 35° N. Geomorphologically it is part of the Mojave Desert region of California. Topographically the site is located at approximately 700-900 meters above sea level. It is bounded to the north and west by the southern extension of the Sierra mountains, the Tehachapis, while to the south it is bounded by the Transverse mountain range. The area is predominantly an outflow sink for these ranges. The surficial sediments are predominantly Quaternary alluvium. Underlying these sediments are occasional outcroppings of Mesozoic granitic rock.

Precipitation at the site is measured at the town of Mojave. The average yearly precipitation is an arid 5.62 inches.

Vegetation is extremely sparse in the Rogers Dry Lake test site. The area around the dry lakes contains desert saltbrush *(Atriplex, Grayia, Sarcobatus)* while the surrounding land surface is covered with desert arid-adapted vegetation such as mojave creosote bush (*Larrea*) and Joshua tree scrub (*Salazaria, Tetradymia, Yucca*).

The third site is representative of agriculture in the San Joaquin valley. It is located in the Fresno region at approximately 120.3° W, 36.4° N. Geomorphologically it is part of the Central Valley of California. Topographically it is located approximately 50-70 meters above sea level. It is bounded to the west by the eastern flank of the Coastal Range. Geologically the parent material is sedimentary, predominantly consisting of Upper Cretaceous marine and Pleistocene non-marine material. Precipitation at the site is low. The average yearly precipitation for the city of Fresno is 10.52 inches. Natural vegetation consists primarily of California prairie (*Stipa* spp.) and the majority of the site is covered with a wide variety of agricultural crops.

Figure 2.1 shows average monthly precipitation for the three test sites. Table 2.1 shows vegetation cover mixes at the three test sites.

Figure 2.1 Average Monthly Precipitation (Inches) at MVI Test Sites

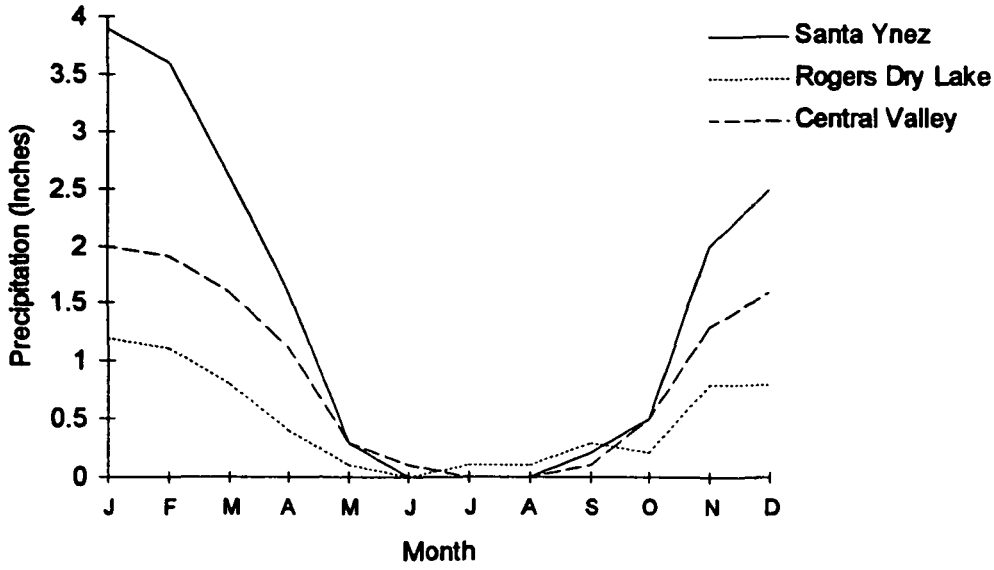


Table 2.1. Test Sites Vegetation cover mixes (%). "Other" is predominantly agricultural landcover.

	Forest	Grassland	Scrub	Shrub	Woodland	Other
Santa Ynez	0.02	0.17	0.08	0.45	0.15	0.13
Rogers Dry Lake	0.19	0.00	0.34	0.00	0.37	0.10
Central Valley	0.00	0.02	0.00	0.00	0.05	0.93

3. Methodology

3.1 Introduction

In this section the creation of the Microwave Vegetation Index (MVI) is detailed. In Section 3.2 a description of the original input dataset and the transformation process used to convert it into a file readable on Earth Resources Data Analysis System (ERDAS) is recorded. Section 3.3 describes the mathematical procedures used on ERDAS, and the creation of a multitemporal dataset. This section also includes a brief description of multitemporal dataset techniques for passive microwave datasets. Section 3.4 details the creation of the four ancillary datasets; precipitation, vegetation, surface water, and Normalized Difference Vegetation Index (NDVI). Section 3.5 covers the creation of the three study sites datasets.

3.2 The 37 GHz Passive Microwave Dataset

The 37 GHz passive microwave dataset was created from Special Sensor Microwave/Imager (SSM/I) data on the Defense Meteorological Satellite Program (DMSP) Block 5D-2 spacecraft. The data was stored at Marshall Space Flight Center (MSFC) in Huntsville, Alabama on 6250 bpi tape. Two primary forms of preprocessing datasets are geometric and radiometric correction. Radiometric corrections were performed at the University of Wisconsin. The brightness temperature, T_B, incident upon the antenna was received and spatially filtered to produce the effective input signal or antenna temperature, T_A, at the input of the feedhorn antenna (DMSP I., 1989). The antenna temperature was calibrated each scan by passing the feedhorn beneath a hot-load black-body radiator at a designed temperature of 250^0 K, and a small calibration reflector which reflects the cosmic background radiation of 3^0 K into the feedhorn field-of-view.

Corrections were applied to increase the accuracy of the scene brightness temperature incident on the reflector antenna pattern. These corrections remove the effects of; feedhorn spillover loss, cross-polarization coupling of the effective far field antenna power pattern, and the sidelobes contribution of the effective far field antenna power pattern (Hollinger et al. 1987; Hollinger, 1990).

In remote sensing there are two basic forms of geometric errors; systematic and nonsystematic. An example of a systematic error would be the distortion caused by the rotation of the earth while the platform images the scene. An example of a nonsystematic error would be a change in the attitude of the platform.

Systematic error of the SSM/I scenes was corrected using satellite ephemeris. As mentioned in Chapter Three, the nonsystematic errors were corrected by using the land/ocean interface for testing procedures.

The data was read onto a MicroVAX 3 running the VMS operating system. The data format was satellite ID (integer * 2), revolution number (integer * 4), y time (integer * 4), five channels of brightness temperature (integer * 2 for each) and latitude/longitude (integer * 2).

The temperature was in decilKelvins. The latitude and longitude were in tenths of a degree. The data was converted to an ASCII text file format using a C program routine.

After the data was converted to an ASCII text file a printout of one of the files containing 37 GHz values and their accompanying latitude/longitude coordinates was created. The brightness temperature values and the latitude/longitude coordinates for the file were verified by a technician from MSFC. Having verified the C routine and it's output all other data files were converted and placed on the Remote Sensing Research Unit's (RSRU) network. A total of 204 data files were read onto the RSRU system. These files correspond to February 15, 1989 to April 25, 1990. The files included both ascending and descending pass information.

The data was then ready to be placed into the image analysis software package, ERDAS. The file was converted from an ASCII text file to a digitized ERDAS file (.dig) using an RSRU piece of C code. These sort of files are commonly used to convert data from a point format to a pixel format.

Using the ERDAS function "GRIDPOL" the 37 GHz data was transformed from a point format to a raster format. The on screen pixel data was compared to the hard copy printout to make sure that the appropriate brightness temperature values were given the appropriate latitude and longitude coordinates. The pixels were then randomly checked to make sure that the pixels were in the correct grid for their latitude and longitude. Using the latitude values on a given pass the data was then split into ascending and descending datasets.

3.3 Passive Microwave Data Set Transformations

Having placed the passive microwave data successfully into ERDAS the data was screened twice to remove any corrupted data. All pixels with a T_B greater than 350^0 K or less than 150^0 K were set to a null value. These cutoff values were chosen because normal surface temperatures range from approximately

250^0 to 300^0 K. The data was then transformed into the Microwave Vegetation Index format using a simple differencing routine to subtract the horizontal 37 GHz temperature from the 37 GHz vertical temperature. All derived MVI values that were less than zero or greater than 90 were set to a null value in order to delete information that was problematic (Choudhury, 1989).

A disconcerting problem arose in that the produced MVI values were lower than expected, according to the literature, for given earth surfaces. The original data was checked with MSFC to verify accuracy. The before and after values created by differencing the horizontal and vertical polarizations were also compared. The end product values were confirmed. The assumption thereafter became that while the values were somewhat different from what was expected from the literature, they were the actual values to the best of our knowledge.

The data was now ready to be multitemporally composited. Multitemporal compositing is a technique used to minimize the effects of environmental processes, such as cloud coverage or standing water, associated with a rain burst event. Cloud coverage would lower the MVI value, while standing water would increase the MVI value. Multitemporal compositing consists of taking a series of datasets from a given period of time and, through the use of an algorithm, determining what the value for a given pixel is over the time span in question. Various multitemporal compositing algorithms have been created to manipulate MVI.

Due to the problems created by cloud contamination an algorithm using the second lowest value for a given pixel over a one month period became widely used (Choudhury, 1989; Choudhury et al., 1990; Justice et al., 1989). Choudhury is now contemplating using a two tier approach (Choudhury, personal communication, 1994). The first cull would take the lowest value for a one week span. This procedure would be used for three consecutive weeks. In the second step the three values would then be listed and the second lowest value would be chosen to represent the given pixel for the three week period. This approach seems to have all the same problems associated with cloud coverage and surficial water, and it is relatively cumbrous.

The National Aeronautics and Space Administration (NASA) uses a different approach for compositing procedures. NASA uses a two week average value for a given pixel. It combines all values for a given pixel in *n* number of passes and divides the total value by *n*. This procedure has the advantage of simplicity. The problem is that MVI values still could vary based on average cloud cover over a given area. The NASA technique uses both the ascending and descending pass values.

For this experiment, three multitemporal composites were created. The first composite used the NASA technique of averaging all values. The second and third datasets used a modified Choudhury approach. A two week compositing period was used instead of a one month period so that the MVI data would coincide with Normalized Difference Vegetation Index (NDVI) data. Within this two week span the second lowest value for a pixel was chosen. The second multitemporal compositing dataset used only ascending values while the third dataset used only descending values. C language code was written to perform the three different multitemporal composites.

The original 204 37 GHz files were translated into 32 two week period multitemporal datasets. The number of files used to create an individual multitemporal period ranged from 2 to 8 depending on the period in question. The number of files used to create an individual multitemporal period was 4 to 16 for the MSFC approach which used both ascending and descending passes. Three of these 32 period datasets were created in order to correspond to the three different multitemporal compositing techniques.

The datasets were then converted to a Lambert azimuthal equal area projection using the "CCVRT" command. The datasets were converted to this projection

because the NDVI data the RSRU receives is in the Lambert azimuthal equal area format. The CalVeg datasets that the RSRU has are also in the Lambert azimuthal equal area projection. In the transformation the longitude of central meridian was 100^0 W, the latitude of origin was 45^0 N, and the easting and northing were both set to zero.

Creating a mask for the State of California was the next step. The RSRU has a file created from the CalVeg map that is used to mask out the state of California during image processing. It is co-registered to NDVI datasets received from EROS Data Center. Since the CalVeg map is at a scale of 1 km per pixel and the MVI datasets were at a scale of 25 km per pixel it was necessary to coarsen the mask to a 25 km scale. To do this a 5 x 5 kernel in the "DCONVLV" function (ERDAS cannot create a 25 x 25 kernel) was used. Each pixel in the 5 x 5 window was multiplied by 1. The total value of the 25 pixels were then divided by 25 to produce a value for the central pixel in the matrix. This was followed by use of the "LRECTIFY" command to rescale the resulting image by a factor of five. These two commands were used twice to coarsen the California image to a 25 km per pixel scale.

The outlined portion of the mask was coded to have a value of one and the areas outside of the state were coded to zero. The mask was applied to the MVI

datasets using the "MASK" command. An MVI image before and after the mask was applied can be seen in figure 3.1.

The next problem was to resolve the question of snow coverage. As mentioned in chapter three, snow coverage affects the MVI signal. It was decided that attempting to determine the average snow depth for any given area for a given time period was beyond the scope of this project. Instead, snowfall over the state of California during the 1989-90 period was determined using data gathered and archived at the National Climatic Data Center, (Asheville, N.C.), and dispersed by Earth-Info Inc., Boulder, CO, on a CD-ROM titled "Weather of the Western United States. The records of 376 weather stations in the State of California were included on the CD.

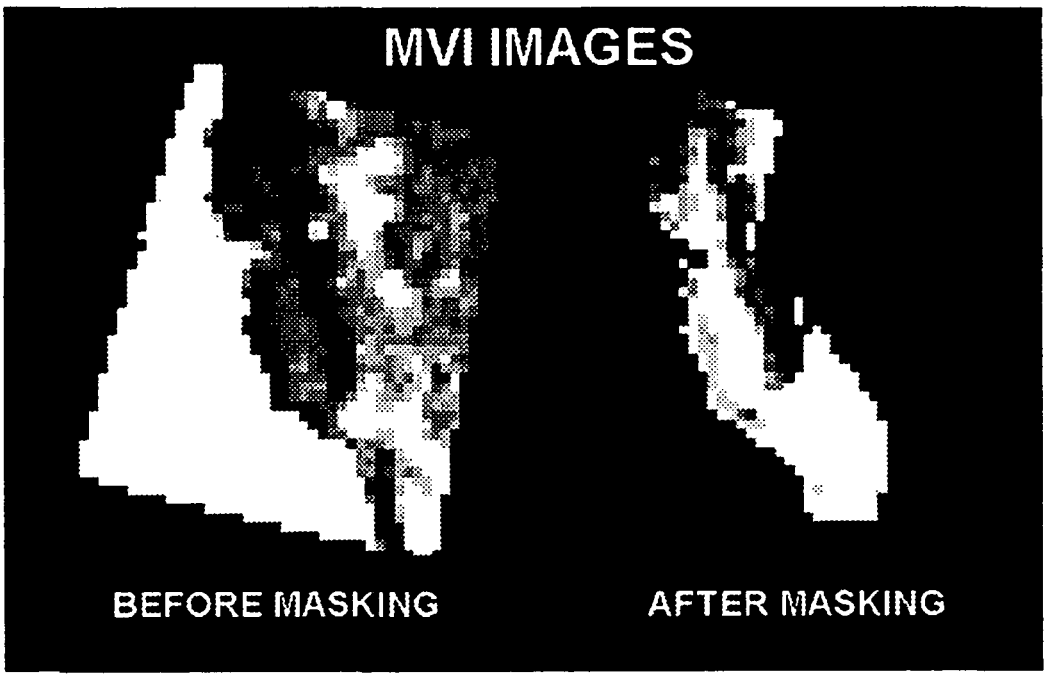


Figure 3.1. MVI Image data before and after masking.

Snowfall patterns were analyzed and, as expected, there were marked altitudinal and latitudinal gradients. Four snowfall masks for the State of California were created in order to minimize the effects of snowfall contamination on the derived MVI signal. The four masks were based on the average altitude of the pixels in question. The four altitude cutoffs were 6,500, 6,000, 4,000, and 2,000 feet above mean sea level. All pixels with an average altitude above 6,500 feet were considered snow covered all year round. Essentially, the 6,500 foot pixels correspond to high summer, the 6,000 foot high pixels to spring and fall, the 4,000 foot high pixels to winter, and the 2,000 foot high pixels to February 1990, see Table 3.1.

Each pixels altitude was determined using Digital Elevation Model (DEM) data. The data was part of the RSRU database for the State of California and was registered to the same Lambert azimuthal equal area coordinates as the RSRU CalVeg dataset. The same coarsening process as mentioned above for the State mask was used to create DEM pixels at 625 km^2 per pixel scale. The effect of the coarsening procedure on the DEM pixels was to create an average height value for the individual pixels. It is of interest to note that the average height of the pixel encompassing Death Valley, the lowest spot in North America, was on the order of 4,000 feet above sea level. This average elevation gives an idea of the variability of terrain entailed in a 625 km^2 pixel, with all the possible effects that may entail upon MVI. Employing the coarsened DEM data the 4 snow masks were created using the average height for each pixel, see Figures 3.2 and 3.3.

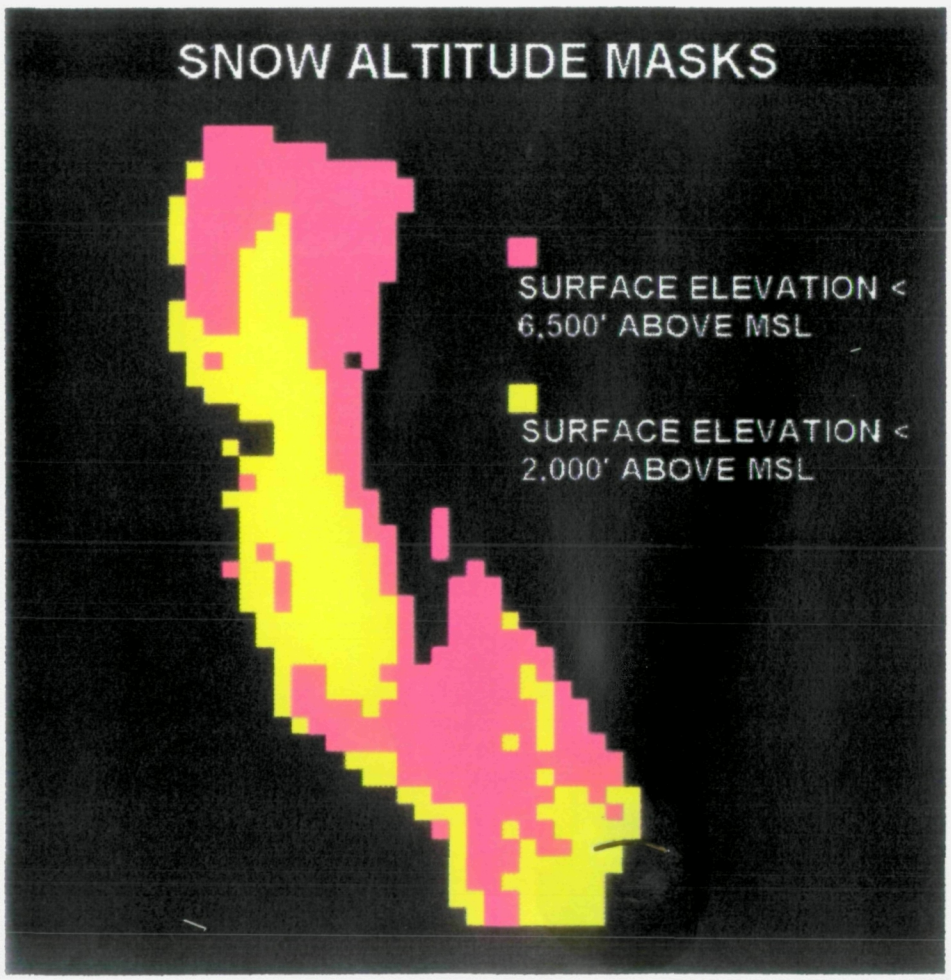


Figure 3.2. MVI Image Data Snow Altitude Masks

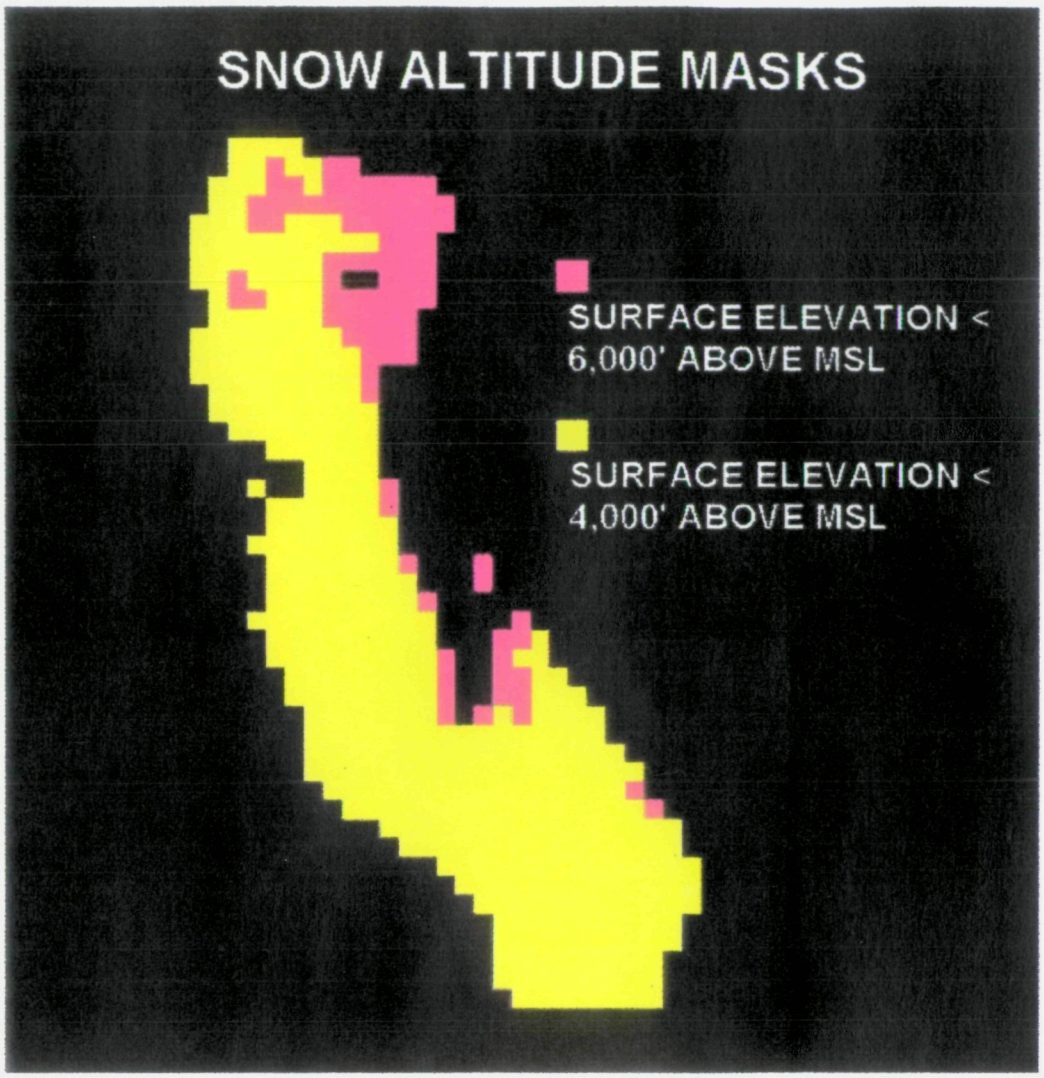


Figure 3.3. MVI Image Data Snow Altitude Masks

Table 3.1: Altitude mask cut-offs. Column one shows the multitemporal period, column two shows the beginning Julian date of the temporal period, column three shows the altitude mask used for the given date, and column four shows whether there was coincident NDVI coverage for the given period

Table 3.1

Period	Julian Date	Altitude Mask	NDVI Coverage
1	46	6000	
2	60	6000	
3	74	6000	
4	88	6000	YES
5	102	6000	
6	116	6000	YES
7	130	6000	YES
8	144	6000	YES
9	158	6500	YES
10	172	6500	YES
11	186	6500	YES
12	200	6500	YES
13	214	6500	YES
14	228	6500	YES
15	242	6500	YES
16	256	6500	YES
17	270	6000	YES
18	284	6000	YES
19	298	6000	
20	312	6000	
21	326	6000	
22	340	6000	
23	354	6000	
24	3	4000	
25	17	4000	
26	31	4000	
27	45	2000	
28	59	2000	YES
29	73	4000	YES
30	87	6000	YES
31	101	6000	YES
32	115	6000	YES

3.4 Creation of Independent Variable Datasets

The next step was to create the four independent variable datasets. The four datasets were vegetation coverage, water bodies, precipitation, and NDVI.

The vegetation coverage and water body datasets were part of the CalVeg dataset of the RSRU. The two variables were coarsened in the same manner as above to a 625 km^2 per pixel scale. For the vegetation data set, 5 types of ground cover were used: forest, grassland, scrub, shrub and woodland, see figures 3.4 and 3.5. A listing of the various subcategories that went into each category can be found in Table 3.2. Reviewing this table gives one an idea of the heterogeneity of ground cover in any given SSM/I pixel due to the size of the pixel.

Figure 3.4: Forest, scrub, and water body ground cover in California.

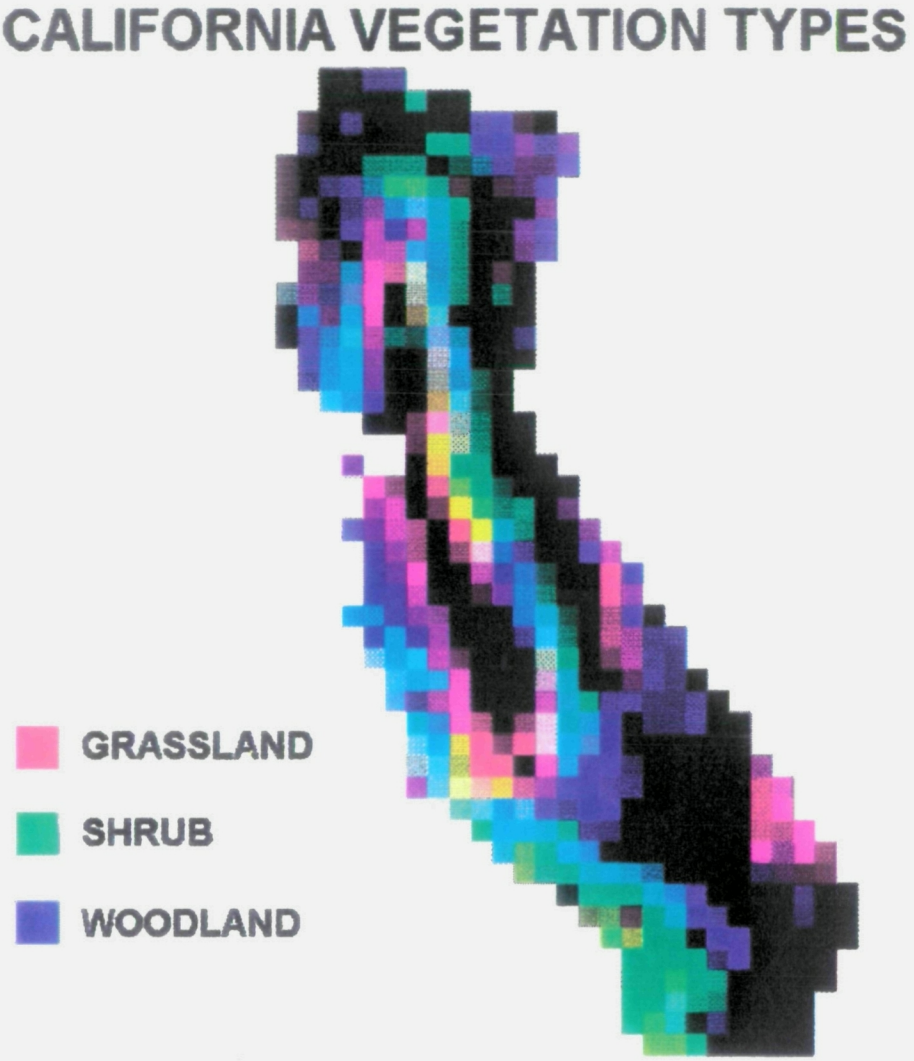


Figure 3.5: Grassland, shrub, and woodland ground cover in California.

Table 3.2

Major Plant Vegetation Types and Plant Communities of California

I. Strand	1. Coastal Strand
II. Salt Marsh	2. Coastal Salt Marsh
III. Freshwater Marsh	3. Freshwater Marsh
IV. Scrub	4. Northern Coastal Scrub
	5. Coastal Sage Scrub
	6. Sagebrush Scrub
	7. Shadscale Scrub
	8. Creosote Bush Scrub
	9. Alkali Sink
V. Coniferous Forest	10. North Coastal Coniferous Forest
	11. Closed-Cone Pine Forest
	12. Redwood Forest
	13. Douglas-Fir Forest
	14. Yellow Pine Forest
	15. Red Fir Forest
	16. Lodgepole Forest
	17. Subalpine Forest
	18. Bristle-cone Pine Forest
VI. Mixed Evergreen Forest	19.Mixed Evergreen Forest
VII. Woodland-Savanna	20. Northern Oak Woodland
	21. Southern Oak Woodland
	22. Foothill Woodland
VII. Chaparral	23. Chaparral
IX. Grassland	24. Coastal Prarie
	25. Valley Grassland
X. Alpine Fell-fields	26. Alpine Fell Fields
XI. Desert Woodland	27. Northern Juniper Woodland
	28. Pinon Juniper Woodland
	29. Joshua Tree Woodland

Table 3.2: The plant communities used to produce the five vegetation categories. The forest category is created by combining Coniferous Forest (V) and Mixed Evergreen Forest (VI), the grassland category is vegetation type IX, the scrub category is vegetation type IV, the shrub category is vegetation type VIII, and the woodland category is created by combining Woodland-Savanna (VII) and Desert Woodland (XI). (From Munz, 1968).

It was necessary to choose the category in which to place some vegetation types. The most difficult was the woodland category. The Desert Woodland category includes needle bearing species such as Pinyon Pine. The Woodland -Savanna category includes deciduous communities such as the Northern Oak woodland. While there are questions whether it is appropriate to include needle bearing species with leaf bearing species, for simplicities sake, and to remain as faithful to Munz (1968) as possible, the Woodland category created for MVI research included both the Desert Woodland and Woodland-Savanna vegetation types.

A sixth vegetation category called "other" is the default category for any ground cover that is not one of the 5 vegetation categories. Predominantly this category includes non-natural vegetation coverage, mostly agricultural.

The NDVI dataset came from EROS Data Center. Two different sources of data were used. The 1989 data coverage was given to the RSRU on 6250 bpi tape. The receipt of the 1989 data was the driving force behind acquiring 1989 SSM/I data. In hindsight, this was unfortunate in that the 1989 data only covered the period mid-March through mid-October, it had registration problems for 2 of the datasets, and the 1989-90 rainy season was the driest rainy season in several decades.

The 1989 data was in Albers equal area projection. It was converted to Lambert azimuthal equal area projection using the "CCVRT" command. Of the 15 data periods for the 1989 NDVI dataset the second period contained bad data and the thirteenth period had data that was badly misregistered. The data for the second period was not used, and the thirteenth period data was only used for a 3 period moving average after seeing the results of coarsening the pixel scale from 1 km per pixel to 25 km per pixel.

NDVI data for the 1990 period was on a CD-ROM. It was registered to a a Lambert azimuthal equal area projection with the same central meridian and latitude of origin as the MVI datasets so no projection transformation was necessary. The 1989 and 1990 NDVI datasets were linked together into a single set. In table 3.1 one can see the NDVI periods and the corresponding MVI periods, along with the snow altitude mask used for the given period. The NDVI dataset was coarsened to a 625 km^2 per pixel scale using the "DCONVLV" and "LRECTIFY" procedures outlined above. See figures 3.6 and 3.7 for a before and after look at NDVI image coarsening to 625 km^2 per pixel scale.

Figure 3.6: NDVI image of California.

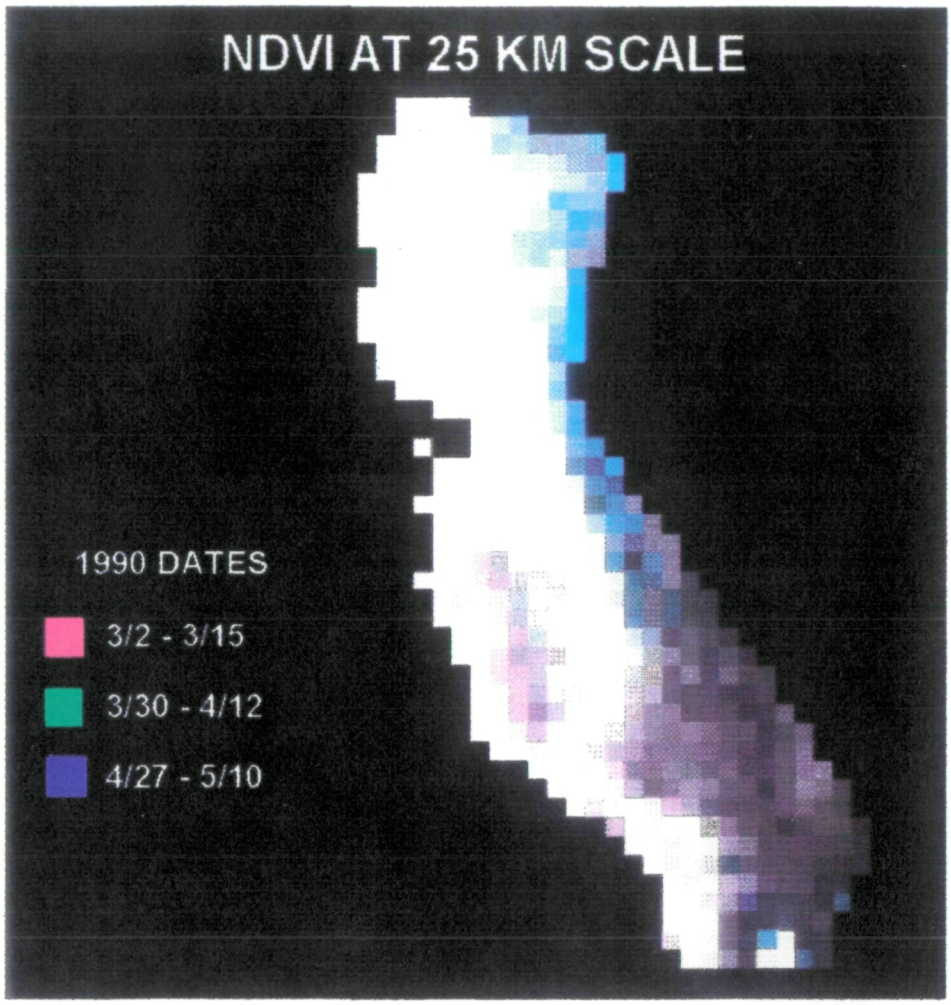


Figure 3.7: **NDVI image of California after resampling spatial resolution to 625 km^2 (25x25 km) per pixel.**

Precipitation data was transferred from the Earth-Info western United States weather CD-ROM. The data was originally gathered and archived by the National Climatic Data Center, (Asheville, N.C.). There are national standards about gauge type, height, exposure, etc., which the Weather Service attempts to verify despite staff shortages. The major concern is unrecorded location changes which can cause fairly serious discontinuities. Three hundred and seventy six weather stations covering the state of California were included in the database. The data was read onto an IBM PS/2 computer and precipitation data

were copied onto floppy disks for transfer to the RSRU Sun/UNIX workstation network.

A C program was used to transform the data into a format readable in ARC/INFO. Transferring the data into ARC/INFO was a two stage process. First, the points designating the weather stations latitude and longitude coordinates were created. Second, the point identification numbers were then linked with the precipitation records corresponding to the ID's.

Using the ARC/INFO command "KRIGING" the precipitation data were interpolated from point data to a two dimensional surface. Kriging is a form of linear regression whereby an unknown value Z_0 is estimated by a linear combination of *n* known values, Z_a, a = 1, ..., *n*. If the n known values correspond to the same attribute as Z_0 the process can be called kriging.

The precipitation file created by the above process was in a latitude/longitude projection. The file was changed to a Lambert azimuthal equal area projection and gridded to a 25 km pixel size using the "PROJECT" command. The resulting file was exported to ERDAS for viewing as an image.

The individual pixel values for the MVI datasets and the four independent variables were then stripped out into a tab delineated table format using a C program. The program created two output datasets. The first set of values were used for statistical analysis. The second set were used to test the statistical hypothesis derived from the first data set.

The two datasets were exported from the Sun/UNIX platform to an "EXCEL" (Microsoft Corp.) spreadsheet on a PC. The data sets were further transferred from the Excel spreadsheet to the "SYSTAT" (SPSS Corp.) statistics package for statistical analysis.

3.5 Creation of the Three Study Site Datasets

Our original intent was to examine how MVI responded to a dark target invariant site and a light target invariant site. Only 1 pixel per site was to be used. It was hoped that a spectral space could be identified for MVI and in future work the area between the two end points would be filled in. The light target invariant site decided upon was the Rogers Dry Lake area north of the Tehachapis and the dark target invariant site was the Pacific ocean. The oceanic response was so obviously different from the land response that it was felt that any research on land/ocean differences for MVI was not worthwhile. Instead, three sites in

California were decided upon; the Santa Ynez region, the Fresno region, and the Rogers Dry Lake area.

Multi-date registration of a single pixel was believed to have inherent problems. The most notable of these problems was the accuracy level of geolocation being only 7 km while the pixel as a whole was only 25 km. A two tiered approach to studying the State of California became the preferred approach. The first tier was to use the whole State of California for statistical analysis of the interactions of MVI with four independent variables. The second tier was to use the three study sites previously mentioned to compare a multitemporal MVI trajectory with multitemporal trajectories for NDVI and precipitation.

In order to lessen the effects of the geolocation error created by using only single pixels for the three study sites, larger study sites were contrived. Instead of one pixel per site 4 pixels per site were used. The 4 pixels were in a 2 x 2 matrix. The result was a study area of 50 km on a side, or 2,500 km^2.

As mentioned in Chapter Two, California is an extremely heterogeneous State. It is unrealistic to hypothesize that an area as large as 2,500 km^2 would be homogeneous, except perhaps for some areas in the Mojave desert.

In order to minimize inhomogeneity at the sites, United States Geological Survey (USGS) land coverage maps at 1:250,000 scale were studied. Fifty kilometer squares with optimal single vegetation type land coverage were determined. The optimal 4 pixel matrix in Lambert azimuthal equal area coordinates was then compared to the actual pixel matrix of the image. The four pixel image matrix that most closely approached the optimal configuration was chosen.

In all cases this led to the inclusion of vegetation types that would not commonly be considered part of the study sites general vegetation type. For example, the Rogers Dry Lake site would normally be considered desert. However, the pixel that actually covers the lakebed also covers part of the Tehachapi mountains. The result is that 55% of the lakebed pixel is covered by Forest vegetation, see figure 3.8. Again, this illustrates the problems associated with such a large ground surface footprint when dealing with vegetation coverage in a rapidly varying land surface.

	CENTRAL VALLEY		ROGERS DRY LAKE		SANTA YNEZ	
FOREST	0	0	55	4	0	0
	0	0	16	0	6	0
GRASSLAND	0	0	0	0	6	17
	9	0	0	0	9	37
SCRUB	0	0	2	38	0	29
	0	0	58	36	0	1
SHRUB	0	0	0	0	59	13
	0	0	0	0	59	48
WOODLAND	0	0	32	46	34	0
	21	0	7	61	23	3

Figure 3.8: By pixel percent ground cover of the three study sites by vegetation category:

Central Valley site = 0% forest, 2% grassland, 0% scrub, 0% shrub, 5% woodland, 93% other.

Rogers Dry Lake site = 19% forest, 0% grassland, 34 % scrub, 0% shrub, 37 woodland, 10% other.

Santa Ynez site = 2% forest, 17% grassland, 8% scrub, 45% shrub, 15% woodland, 13% other.

The three study sites were individually masked out from the California MVI, NDVI, and precipitation images. Using the "BSTATS" command in ERDAS the mean values and standard deviation for each site over the 32 multitemporal periods were determined. These values were entered into an "EXCEL" spreadsheet where they were used to create multitemporal trajectories.

4. Results

4.1 Introduction

In this chapter the results of my paper research are reported. Section 4.2 examines the relationship between different compositing techniques on the derived Microwave Vegetation Index (MVI) signal. Section 4.3 covers the relationship between MVI and four independent variables; precipitation, standing water, vegetation groundcover, and Normalized Difference Vegetation Index (NDVI). Section 4.4 develops an experimental model to predict MVI. The final section, 4.5, examines the relationship between the multi-temporal trajectories of MVI, NDVI, and precipitation at the three study sites.

4.2 The Relationship Between MVI Multi-temporal Compositing Techniques

In this section the relationship between the Choudhury multi-temporal compositing technique and the Marshall Space Flight Center (MSFC) technique is detailed. In the thirty-two time periods studied, the three compositing techniques were compared using two methods. The first method was to compare the correlation between the different compositing techniques. The second method was to use an ANOVA test to contrast the p values for comparing both the variance and the mean.

Four sets of tests were conducted. These sets of tests were based upon the four altitude masks used to screen the data sets. The four masked data sets correspond to the data stripped from four altitudes above mean sea level, 2,000 feet, 4,000 feet, 6,000 feet, and 6,500 feet. The higher the altitude mask employed the more data points collected, e.g. for the data stripped from the image using a 2,000 foot mask there were 122 sample points, while for the data set stripped from the image using the 6,500 foot mask there were 264 sample points. The reason there are more data points for the higher elevation mask is that there is more of the State to examine the higher the elevation cut-off.

If enough data points were invalid for any given period, $n < 30$, pairwise deletion was used. In other words if column A and column B had values in their respective cell 1's, but column C did not then the relationship between A and B would be determined, but not A and C or B and C. This only occured for the statistical tests using the second period of the 6,500 foot mask, e.g. those in the following tables that include 65AV2, 65NA2 or 65ND2. All other combinations used a listwise technique, where all values in a row were used.

4.2.1 Correlation Between the Three Compositing Techniques

In this section the correlation between the various compositing techniques is compared. The objective is to determine whether there is a high degree of correlation between the techniques. A high degree of correlation would tend to confirm a relative similarity between the data sets. In all of the following cases a Pearson correlation was used. The statistics package used for all statistical analysis in this paper is Systat for Windows, Version 5.02, run on a PC.

The naming conventions used for the various data sets are as follows: The first number designates the altitude mask used, for example 2 corresponds to the 2,000 foot elevation mask. The next two letters show what compositing method was used, e.g. AV stands for averaging method, NA stands for ascending passes using the Choudhury method, and ND represents descending passes using the Choudhury method. The final number represents the time period for that elevation mask, see Table 4.1. For example, the seventh time period of the 6,000 foot elevation mask using solely the descending passes would be represented as 6ND7.

The first data set to be studied was the one created by using the 2,000 foot elevation mask. There are two compositing periods in this data set. As can be seen in table 4.1 the correlation between the averaging method and the data sets using the Choudhury method, consisting solely of the ascending or descending passes, was very high. The highest correlations in both time periods were between the averaging method and the ascending passes. The lowest correlations in both periods were between the ascending passes and the descending passes.

Table 4.1: Correlation between the Choudhury and MSFC derived MVI data sets using the 2,000 feet above MSL mask. AV represents the MSFC averaging method, NA and ND respectively represent the ascending and descending Choudhury data sets.

	2AV1	2AV2	2NA1	2NA2	2ND1	2ND2
2AV1	1.000					
2AV2	0.976	1.000				
2NA1	0.969	0.943	1.000			
2NA2	0.954	0.973	0.930	1.000		
2ND1	0.945	0.931	0.887	0.898	1.000	
2ND2	0.925	0.934	0.896	0.900	0.895	1.000

The second data set studied used the 4,000 foot elevation mask. For ease of presentation the four time periods were split up into two groups of two time periods each. The first group consists of the first two time periods and the second group consists of the last two time periods. It should be noted, see Table 4.1, that between the third and fourth time periods of the 4,000 foot elevation mask there were two time periods using another elevation mask.

The highest correlation in all four time periods was between the averaged data set and the ascending data set. The lowest correlation in all four cases was between the ascending and descending data sets, see Table 4.2.

Table 4.2: Correlation between the Choudhury and MSFC derived MVI data sets using the 4,000 feet above MSL mask.

	4AV1	4AV2	4NA1	4NA2	4ND1	4ND2
4AV1	1.000					
4AV2	0.965	1.000				
4NA1	0.956	0.936	1.000			
4NA2	0.929	0.957	0.920	1.000		
4ND1	0.932	0.920	0.879	0.888	1.000	
4ND2	0.905	0.943	0.878	0.877	0.884	1.000

	4AV3	4AV4	4NA3	4NA4	4ND3	4ND4
4AV3	1.000					
4AV4	0.955	1.000				
4NA3	0.970	0.926	1.000			
4NA4	0.945	0.972	0.914	1.000		
4ND3	0.960	0.913	0.916	0.914	1.000	
4ND4	0.889	0.926	0.881	0.890	0.861	1.000

For ease of interpretation the 6,000 foot above mean sea level (MSL) was split into 6 subgroups, each containing three time periods 6(1-3), 6(4-6), 6(7-9), 6(10-12), 6(13-15), and 6(16-18). For example, the first subgroup contains information on the first three time periods. Note that there are temporal gaps between periods 8 and 17, and periods 23 and 30 when other altitude masks were used, see Table 4.1.

In this data set thirteen out of eighteen times the highest correlation was between the average data set and the ascending data set. Five times the

highest correlation was between the average data set and the descending data set. The lowest correlation in all eighteen cases was between the ascending and descending data set, see Tables 4.3-4.8.

Table 4.3: Correlation between the Choudhury and MSFC derived MVI data sets using the 6,000 feet above MSL mask.

	6AV1	6AV2	6AV3	6NA1	6NA2	6NA3	6ND1	6ND2	6ND3
6AV1	1.000								
6AV2	0.972	1.000							
6AV3	0.969	0.981	1.000						
6NA1	0.974	0.948	0.944	1.000					
6NA2	0.952	0.977	0.956	0.940	1.000				
6NA3	0.934	0.944	0.964	0.917	0.930	1.000			
6ND1	0.931	0.900	0.896	0.894	0.889	0.876	1.000		
6ND2	0.916	0.954	0.930	0.891	0.921	0.898	0.886	1.000	
6ND3	0.925	0.941	0.957	0.906	0.925	0.907	0.878	0.918	1.000

Table 4.4: Correlation between the Choudhury and MSFC derived MVI data sets using the 6,000 feet above MSL mask.

	6AV4	6AV5	6AV6	6NA4	6NA5	6NA6	6ND4	6ND5	6ND6
6AV4	1.000								
6AV5	0.920	1.000							
6AV6	0.894	0.950	1.000						
6NA4	0.945	0.878	0.847	1.000					
6NA5	0.850	0.939	0.905	0.794	1.000				
6NA6	0.820	0.893	0.934	0.816	0.874	1.000			
6ND4	0.966	0.887	0.865	0.867	0.824	0.784	1.000		
6ND5	0.876	0.904	0.867	0.828	0.780	0.773	0.837	1.000	
6ND6	0.814	0.838	0.885	0.743	0.762	0.772	0.787	0.853	1.000

Table 4.5: Correlation between the Choudhury and MSFC derived MVI data sets using the 6,000 feet above MSL mask.

	6AV7	6AV8	6AV9	6NA7	6NA8	6NA9	6ND7	6ND8	6ND9
6AV7	1.000								
6AV8	0.968	1.000							
6AV9	0.871	0.878	1.000						
6NA7	0.925	0.922	0.792	1.000					
6NA8	0.944	0.966	0.837	0.872	1.000				
6NA9	0.839	0.826	0.959	0.735	0.814	1.000			
6ND7	0.941	0.909	0.824	0.806	0.885	0.795	1.000		
6ND8	0.944	0.975	0.881	0.887	0.921	0.849	0.912	1.000	
6ND9	0.774	0.792	0.937	0.724	0.734	0.893	0.691	0.794	1.000

Table 4.6: Correlation between the Choudhury and MSFC derived MVI data sets using the 6,000 feet above MSL mask.

	6AV10	6AV11	6AV12	6NA10	6NA11	6NA12	6ND10	6ND11	6ND12
6AV10	1.000								
6AV11	0.980	1.000							
6AV12	0.979	0.988	1.000						
6NA10	0.976	0.955	0.958	1.000					
6NA11	0.954	0.962	0.952	0.942	1.000				
6NA12	0.944	0.948	0.953	0.946	0.945	1.000			
6ND10	0.961	0.930	0.928	0.919	0.915	0.903	1.000		
6ND11	0.948	0.960	0.955	0.910	0.916	0.897	0.908	1.000	
6ND12	0.933	0.928	0.942	0.892	0.898	0.851	0.902	0.951	1.000

Table 4.7: Correlation between the Choudhury and MSFC derived MVI data sets using the 6,000 feet above MSL mask.

	6AV13	6AV14	6AV15	6NA13	6NA14
	6NA15	6ND13	6ND14	6ND15	
6AV13	1.000				
6AV14	0.977	1.000			
6AV15	0.979	0.987	1.000		
6NA13	0.972	0.949	0.950	1.000	
6NA14	0.953	0.970	0.959	0.946	1.000
6NA15	0.949	0.950	0.960	0.934	0.943
6ND13	0.965	0.939	0.943	0.938	0.920
6ND14	0.947	0.970	0.961	0.915	0.918
6ND15	0.929	0.939	0.954	0.895	0.906

	1.000				
	0.905	1.000			
	0.910	0.922	1.000		
	0.877	0.928	0.933	1.000	

Table 4.8: Correlation between the Choudhury and MSFC derived MVI data sets using the 6,000 feet above MSL mask.

	6AV16	6AV17	6AV18	6NA16	6NA17
	6NA18	6ND16	6ND17	6ND18	
6AV16	1.000				
6AV17	0.973	1.000			
6AV18	0.943	0.963	1.000		
6NA16	0.966	0.941	0.918	1.000	
6NA17	0.947	0.976	0.930	0.933	1.000
6NA18	0.940	0.954	0.960	0.926	0.931
6ND16	0.949	0.922	0.896	0.880	0.877
6ND17	0.918	0.937	0.912	0.887	0.890
6ND18	0.900	0.915	0.965	0.870	0.880

	1.000				
	0.895	1.000			
	0.886	0.899	1.000		
	0.893	0.867	0.907	1.000	

The 6,500 foot above MSL data set was split into 3 subgroups for ease of viewing. The first and second subgroups each contained 3 time periods, 65(1-3), and 65(4-6). The third subgroup only contained 2 time periods, 65(7-8). The highest correlation five out of eight times was between the average data set and the descending data set. Three times the highest correlation was between the average and ascending data sets. One time the lowest correlation was between the average and ascending data sets (65AV2 and 65NA2), the other seven times the lowest correlation was between the ascending and descending data sets, see Tables 4.9 -4.11.

Table 4.9: Correlation between the Choudhury and MSFC derived MVI data sets using the 6,500 feet above MSL mask.

	65AV1	65AV2	65AV3	65NA1	65NA2
	65NA3	65ND1	65ND2	65ND3	
65AV1	1.000				
65AV2	0.872	1.000			
65AV3	0.941	0.851	1.000		
65NA1	0.969	0.830	0.925	1.000	
65NA2	0.920	0.907	0.906	0.909	1.000
65NA3	0.893	0.808	0.947	0.839	0.864
65ND1	0.975	0.770	0.919	0.918	0.846
65ND2	0.905	0.958	0.887	0.871	0.958
65ND3	0.932	0.818	0.962	0.899	0.794

	1.000								
	0.833	1.000							
	0.827	0.706	1.000						
	0.843	0.913	0.834	1.000					

Table 4.10: Correlation between the Choudhury and MSFC derived MVI data sets using the 6,500 feet above MSL mask.

	65AV4	65AV5	65AV6	65NA4	65NA5
	65NA6	65ND4	65ND5	65ND6	
65AV4	1.000				
65AV5	0.974	1.000			
65AV6	0.933	0.963	1.000		
65NA4	0.969	0.945	0.914	1.000	
65NA5	0.913	0.953	0.933	0.931	1.000
65NA6	0.883	0.922	0.958	0.862	0.879
65ND4	0.973	0.958	0.922	0.910	0.884
65ND5	0.944	0.961	0.923	0.901	0.871
65ND6	0.920	0.932	0.957	0.912	0.912

	1.000								
	0.913	1.000							
	0.888	1.000							
	0.930	1.000							
	0.892	0.891	0.923	1.000					

Table 4.11: Correlation between the Choudhury and MSFC derived MVI data sets using the 6,500 feet above MSL mask.

	65AV7	65AV8	65NA7	65NA8	65ND7
	65ND8				
65AV7	1.000				
65AV8	0.967	1.000			
65NA7	0.979	0.948	1.000		
65NA8	0.942	0.974	0.930	1.000	
65ND7	0.961	0.924	0.927	0.892	1.000
65ND8	0.904	0.954	0.890	0.907	0.872

	1.000

One aspect of the correlation analysis of MVI values across compositing techniques is the relatively high correlation values, on the order of 0.85 to 0.95, across both compositing techniques and temporal periods. These high values would tend to suggest a structural basis for MVI values. At 37 GHz the most likely structural form affecting MVI would be vegetation. The interaction between MVI and vegetation is discussed more fully in section 4.3.3.

Examining the correlations from a temporal viewpoint reveals a definite pattern, see Table 4.12. During the 32 periods studied the highest correlation occured between the MSFC averaging technique data and the Choudhury technique data using the ascending passes 22 times. Ten times out of the 32 possible periods the descending data and the averaged data have the highest correlation.

Table 4.12: This table shows the high correlation and the low correlation between the different compositing techniques used for deriving MVI. Note that the high correlation tends to be between the averaging technique and the ascending passes, while the low correlation is almost entirely between the ascending and descending techniques.

PERIOD	JULIAN DATE	ALTITUDE MASK	HIGH CORRELATION	LOW CORRELATION
1	46	6000	AVERAGE/ASCENDING	ASCENDING/DESCENDING
2	60	6000	AVERAGE/ASCENDING	ASCENDING/DESCENDING
3	74	6000	AVERAGE/ASCENDING	ASCENDING/DESCENDING
4	88	6000	AVERAGE/DESCENDING	ASCENDING/DESCENDING
5	102	6000	AVERAGE/ASCENDING	ASCENDING/DESCENDING
6	116	6000	AVERAGE/ASCENDING	ASCENDING/DESCENDING
7	130	6000	AVERAGE/DESCENDING	ASCENDING/DESCENDING
8	144	6000	AVERAGE/DESCENDING	ASCENDING/DESCENDING
9	158	6500	AVERAGE/DESCENDING	ASCENDING/DESCENDING
10	172	6500	AVERAGE/DESCENDING	AVERAGE/ASCENDING
11	186	6500	AVERAGE/DESCENDING	ASCENDING/DESCENDING
12	200	6500	AVERAGE/DESCENDING	ASCENDING/DESCENDING
13	214	6500	AVERAGE/DESCENDING	ASCENDING/DESCENDING
14	228	6500	AVERAGE/ASCENDING	ASCENDING/DESCENDING
15	242	6500	AVERAGE/ASCENDING	ASCENDING/DESCENDING
16	256	6500	AVERAGE/ASCENDING	ASCENDING/DESCENDING
17	270	6000	AVERAGE/ASCENDING	ASCENDING/DESCENDING
18	284	6000	AVERAGE/ASCENDING	ASCENDING/DESCENDING
19	298	6000	AVERAGE/ASCENDING	ASCENDING/DESCENDING
20	312	6000	AVERAGE/ASCENDING	ASCENDING/DESCENDING
21	326	6000	AVERAGE/ASCENDING	ASCENDING/DESCENDING
22	340	6000	AVERAGE/DESCENDING	ASCENDING/DESCENDING
23	354	6000	AVERAGE/DESCENDING	ASCENDING/DESCENDING
24	3	4000	AVERAGE/ASCENDING	ASCENDING/DESCENDING
25	17	4000	AVERAGE/ASCENDING	ASCENDING/DESCENDING
26	31	4000	AVERAGE/ASCENDING	ASCENDING/DESCENDING
27	45	2000	AVERAGE/ASCENDING	ASCENDING/DESCENDING
28	59	2000	AVERAGE/ASCENDING	ASCENDING/DESCENDING
29	73	4000	AVERAGE/ASCENDING	ASCENDING/DESCENDING
30	87	6000	AVERAGE/ASCENDING	ASCENDING/DESCENDING
31	101	6000	AVERAGE/ASCENDING	ASCENDING/DESCENDING
32	115	6000	AVERAGE/DESCENDING	ASCENDING/DESCENDING

The lowest correlation between the three compositing techniques occurs almost exclusively between the ascending and descending passes. Thirty one times

out of the 32 periods the ascending and descending passes showed the lowest correlation. The only exception occured during the second period of the 6,500 foot data set. During that period the lowest correlation was between the average and ascending data sets. This 10th period data set also had the largest amount of data values set to null values due to questions about their validity because of extreme values.

Examining the correlations it is not surprising that the lowest correlations were mainly between the ascending and descending passes. The averaging technique by definition included half of it's data from either the ascending or descending passes. Therefore, the MSFC data set would be more likely to have a higher correlation with either of those data sets than they would have with each other.

The generally higher correlation, 22 out of 32 periods, between the MSFC technique and the Choudhury technique using solely ascending passes is more difficult to explain. Originally it was thought that one could explain the correlation based on physical properties and the Choudhury algorithm. As the ascending passes are morning passes it seemed reasonable to expect the imaged areas to be generally cooler than they would be during a late afternoon pass. The cooler temperatures would lead to a small but noticeable difference between the MVI values in the morning versus the evening, creating lower MVI values overall. Since the Choudhury technique is based on lower MVI values this could explain the higher overall correlation between the MSFC technique and the Choudhury technique using ascending passes only. Unfortunately this turned out to be a case of circular logic in that it was essentially comparing the Choudhury technique to itself.

It was expected that the evening passes would have a higher MVI value based on warmer ground temperatures. This was the case. By definition, selecting the second lowest MVI values for the descending passes should push the MVI value lower moving it closer to the average value, while selecting the second lowest ascending value should push the ascending MVI values lower and thus away from the average value. One would therefore expect a higher correlation between the data set produced by the average technique and the descending data set. Another possibility for the relatively high correlation between the average and ascending data sets could have been the relative spread of the variance. A lower variance would tend to suggest a smoother temporal profile, such as that produced by an averaging technique. Again, this turned out not to be the case as the descending data set had a lower overall variance. We are currently unsure as to what is causing the higher correlation between the ascending and average data sets. This is an area for future research.

4.2.2 ANOVA Comparisons of the Three Different Multi-temporal Compositing Techniques

Having examined the correlation between the different multi-temporal techniques, ANOVA tests were then employed to determine whether the different data sets were producing comparable data.

ANOVA techniques generally involve separating the total variance in a series of measurements into various components or sources. The tests of equality operate by simultaneously considering both differences in mean and in variances (Davis, 1986). The hypothesis and alternative can be described as:

$$H_0: m_1 = m_2 = m_3 = \ldots m_n$$

$$H_1: \text{at least one mean is different}$$

In effect, the observer is attempting to determine whether a number of populations are identical.

From the high correlations of the data sets it would seem reasonable to expect that the different compositing methods were comparable. In particular, the high correlations between the averaging method and the Choudhury method using ascending passes would tend to indicate equivalent data sets. The results of the ANOVA test suggest however, that the different compositing procedures produce statistically significant differences.

In the analysis that follows there are three possible combinations to compare for each time period. The possible combinations are: average versus ascending, average versus descending, and ascending versus descending.

For the 2,000 foot above mean sea level (MSL) data set the probability of the ascending and average data sets producing equivalent values was 0.000. The highest likelihood at 0.294 of an equivalent data set was between the average method and the descending data set for the first period, 2AV1 and 2ND1. Surprisingly the second highest possibility of an equivalent data set at 0.195 was between the ascending and descending first period data set 2NA1 and 2ND1, see Table 4.13.

Table 4.13: Probability of the different compositing techniques producing equivalent results. Comparisons are only given for comparable time periods.

	2AV1	2AV2	2NA1	2NA2
2NA1	0			
2NA2		0		
2ND1	0.294		0.195	
2ND2		0.11		0.071

In general the 4,000 foot above MSL data set produced low probabilities of comparable values between the three compositing methods. Only two of the 12 possible combinations produced a probability above 5%. The probability of comparable values between the average method and descending method for period four, 4AV4 and 4ND4, was 0.573. The value for 4NA1 and 4ND1 was 0.089, Table 4.14.

Table 4.14: Probability of the different compositing techniques producing equivalent results. Comparisons are only given for comparable time periods.

	4AV1	4AV2	4NA1	4NA2
4NA1	0			
4NA2		0		
4ND1	0		0.089	
4ND2		0		0.003

	4AV3	4AV4	4NA3	4NA4
4NA3	0			
4NA4		0		
4ND3	0		0	
4ND4		0.573		0.009

For the 54 possible combinations for the 6,000 foot above MSL data set only 10 showed a probability above 5% that the data sets were comparable. The highest probability at 0.767 of a comparable data set was between 6NA4 and 6ND4. Only three other combinations, 6AV10/6ND10, 6NA14/6ND14, and

6NA16/6ND16, showed probabilities above 50% that the data sets were comparable.

Thirty six of the 54 possible combinations showed .000 probability that the data sets were comparable. A further 5 combinations had probabilities below 0.007. In general the lowest probability of analogous data sets was between the average and both the ascending and descending data sets. Generally there were higher likelihoods of the ascending and descending passes being comparable. No other specific patterns were evident as shown in Table 4.15.

Table 4 - 15	6AV1	6AV2	6NA1	6NA2
6NA1	0			
6NA2		0		
6ND1	0		0.001	
6ND2		0		0
	6AV3	6AV4	6NA3	6NA4
6NA3	0			
6NA4		0		
6ND3	0		0.007	
6ND4		0		0.767
	6AV5	6AV6	6NA5	6NA6
6NA5	0.001			
6NA6		0.003		
6ND5	0		0.049	
6ND6		0		0
	6AV7	6AV8	6NA7	6NA8
6NA7	0.156			
6NA8		0		
6ND7	0		0	
6ND8		0		0
	6AV9	6AV10	6NA9	6NA10
6NA9	0			
6NA10		0.064		
6ND9	0.038		0	
6ND10		0.688		0.476
	6AV11	6AV12	6NA11	6NA12
6NA11	0			
6NA12		0		
6ND11	0		0	
6ND12		0		0
	6AV13	6AV14	6NA13	6NA14
6NA13	0			
6NA14		0		
6ND13	0		0.486	
6ND14		0		0.580
	6AV15	6AV16	6NA15	6NA16
6NA15	0			
6NA16		0		
6ND15	0		0	
6ND16		0.001		0.717
	6AV17	6AV18	6NA17	6NA18
6NA17	0			
6NA18		0.013		
6ND17	0.470		0	
6ND18		0		0.083

For the 6,500 foot above MSL data set, which corresponded to high summer, there were 24 possible combinations. Of these 24 combinations only 10 had a probability above 5% that the data values were comparable. The largest probability at 0.976 was between 65NA3 and 65ND3. Only four other combinations, 65AV5/65ND5, 65NA6/65ND6, and 65AV8/65ND8, had a probability above 20% for comparable data sets. Seven of the combinations had a probability of .000 that their values were comparable, see Table 4.16.

Table 4.16: Probability of the different compositing techniques producing equivalent results. Comparisons are only given for comparable time periods.

	65AV1	65AV2	65NA1	65NA2
65NA1	0			
65NA2		0.051		
65ND1	0		0	
65ND2		0.047		0.087
	65AV3	65AV4	65NA3	65NA4
65NA3	0.014			
65NA4		0		
65ND3	0		0.976	
65ND4		0.003		0.120
	65AV5	65AV6	65NA5	65NA6
65NA5	0.003			
65NA6		0		
65ND5	0.497		0.032	
65ND6		0.036		0.205
	65AV7	65AV8	65NA7	65NA8
65NA7	0.065			
65NA8		0		
65ND7	0.148		0.580	
65ND8		0.570		0.002

These results were unexpected. With the high correlations between the various combinations it was assumed that the groups would show analogously high probabilities for comparable data sets. It appears that the ascending and descending data sets are the most comparable of the three data sets, but still not very comparable. It seems reasonable to believe that the cause for the relatively similar response of these two data sets is the algorithm used, e.g. creating data sets with a lower mean value.

ANOVA tests compare both mean and standard deviation. While it is not too surprising that the mean values between the Choudhury method and the MSFC method are not comparable, it was assumed that the standard deviation would be comparable. Perhaps what is causing the discrepancy between expectation and reality is that if you are taking the second lowest value for a given datapoint you are throwing out data with a wider spread of available values.

Maybe the most interesting aspect of the difference between MVI data sets using different compositing techniques is the possibility that by using either morning or evening passes there is a bias introduced into the equation. This question seems like a logical area for follow-on studies.

Another area for follow on studies is the actual compositing methods themselves. The MSFC and Choudhury compositing techniques were created to ameliorate contamination of the given pixel by cloud coverage or standing surface water. Neither the MSFC or Choudhury approach scientifically attacks the problem of pixel contamination. I suggest that it would be far more useful to use some of the algorithms created by Neale et al. (1990), such as 85V - 37V (used as a precipitation flag) or 22V - 19V (used to determine land that has recently been affected by heavy precipitation), as screening mechanisms to throw out pixels affected by environmental contamination. Using these algorithms first the researcher could then average all remaining values.

4.3 The Relationship Between MVI and Water Bodies, Precipitation, Primary Productivity, and Ground Cover

In this section the statistical relationship between MVI and precipitation, NDVI, water bodies, and vegetation coverage is detailed. In this first section the transformation of the data sets is covered.

In statistical regression analysis five major assumptions are made (Montgomery and Peck, 1982):

1) The relationship between x and y is closely approximated by a straight line.

2) The error term has a zero mean.

3) The error term has a constant variance.

4) The errors are uncorrelated.

5) The errors are normally distributed.

Using all 8 independent variables (five vegetation categories, water bodies, NDVI, and precipitation) as independent regressor variables a variety of tests were employed to check on the above assumptions before any work was done on the data set.

Using the residuals from the regression and various diagnostic techniques within the Systat statistical package it was apparent that the errors were not normally distributed and the error term did not have a constant variance (Kirby, 1993). Therefore, a variety of transformations were attempted. The most stable result occured when a natural log transformation was applied to the MVI dependent variable. All future writing in sections 4.3 and 4.4 will be based upon the log transformed MVI variable.

4.3.1 The Relationship Between MVI and Water Bodies

Theoretically, the strongest affect on MVI should be lakes and other water bodies(Choudhury, 1989). A correlation between the water bodies data layer and the MVI values showed very little correlation, see Table 4.17. The largest correlation achieved was 0.252 during the fourth MVI compositing period.

The number of observation points for each period is 122. California at the 25 km scale of an SSM/I pixel has relatively few water bodies. The result of having relatively few water bodies is that of the 122 observation points in the above sample only 15 contained greater than 1% water coverage. In effect the correlation equation between water bodies and MVI was driven by 107 observation points that contained no substantial water bodies. In order to resolve whether water bodies by themselves affected MVI a second correlation attempt was tried using only those observation points that contained water bodies. The 15 observation points containing at least 1% water body coverage were correlated and as can be seen in Table 4.17 the correlation between MVI

and water bodies jumps dramatically, with the highest correlation reaching .618 in period four.

Table 4.17: Correlation between MVI values and water bodies. In column one is the MVI period. In column two is the correlation between MVI values and water bodies using all 122 observation points. In column three is the correlation between MVI and the 15 observation points that contained at least 1% surface coverage by water bodies (WATERNZ stands for water no zeros).

	WATER	WATERNZ
MVI1	0.237	0.565
MVI2	0.199	0.496
MVI3	0.218	0.529
MVI4	0.252	0.618
MVI5	0.208	0.522
MVI6	0.191	0.496
MVI7	0.183	0.473
MVI8	0.162	0.440
MVI9	0.173	0.447
MVI11	0.195	0.431
MVI12	0.164	0.434
MVI13	0.191	0.453
MVI14	0.218	0.509
MVI15	0.201	0.481
MVI16	0.231	0.562
MVI17	0.230	0.521
MVI18	0.211	0.521
MVI19	0.225	0.538
MVI20	0.223	0.530
MVI21	0.213	0.505
MVI22	0.221	0.511
MVI23	0.220	0.500
MVI24	0.167	0.387
MVI25	0.184	0.487
MVI26	0.184	0.483
MVI27	0.199	0.504
MVI28	0.204	0.550
MVI29	0.204	0.558
MVI30	0.186	0.507
MVI31	0.175	0.466
MVI32	0.160	0.460

A graphical representation of the relationship between water bodies and MVI can be seen in figure 5.1. Note that relatively few observation points with high leverage drive the equation.

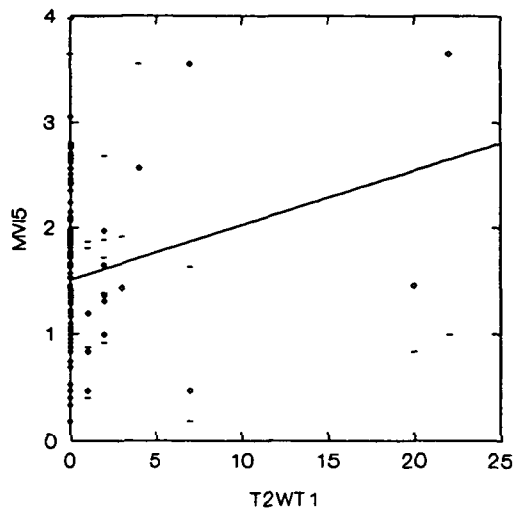


Figure 4.1: Relationship between water bodies and MVI values. On the X axis is percentage surface area covered by water. On the Y axis is the logged MVI value.

The observed result of MVI's relationship to water is that in an area of relatively minimal surface coverage by water bodies other conditions are more important factors on the derived MVI signal.

4.3.2 The Relationship Between MVI and Precipitation

The second most important variable for affecting MVI should be precipitation (Choudhury, 1989). The correlation between MVI and precipitation can be seen in Table 5.18.

Precipitation did not show a high correlation with the MVI value. The highest correlation between MVI and precipitation occured during period 21 with a value of -0.691. The mean correlation between precipitation and MVI was -0.375.

Since at any given time over a 1+ year cycle large areas of California have no precipitation, a second correlation run was tried using only those observation points having some precipitation at the time of observance. The results for using only observation points with some precipitation did not significantly change the outcome, see column three in Table 5.18. The highest correlation was 0.933

during period ten. However, period ten contained only 11 observation points, so this high correlation may be somewhat spurious. In general, the correlation results between MVI and precipitation was actually worse for the data set using only observation points that had some precipitation compared to the data set using all observation points. The average correlation between MVI and precipitation using only observation points that had precipitation was -.296.

Table 4.18: Correlation between precipitation and MVI. The second column contains correlation values for all observation points. The third column contains correlation values only for observation points that contained precipitation. Blank cells represent data sets that did not contain enough coincident observation points to produce a valid correlation. Due to the nature of precipitation the correlation procedure used for this data set was pairwise correlation, not the more common listwise procedure).

PERIOD	MVI CORRELATION	MVI CORRELATION (w/o 0 Precipitation)
1	-0.396	-0.547
2	-0.436	-0.326
3	-0.391	-0.326
4	-0.354	-0.185
5	-0.608	-0.541
6	-0.332	-0.273
7	-0.632	-0.591
8	-0.349	-0.345
9	-0.317	-0.413
10	-0.488	0.933
11	0.217	
12	0.380	0.751
13	0.264	-0.074
14	-0.546	-0.831
15		
16	-0.451	-0.346
17	-0.584	-0.312
18	-0.609	-0.562
19	-0.556	-0.519
20		
21	-0.691	-0.597
22	-0.058	-0.390
23	-0.468	-0.334
24	-0.665	-0.628
25	-0.619	-0.558
26	-0.686	-0.628
27	-0.378	-0.256
28	-0.680	-0.603
29	-0.325	0.048
30	0.085	0.071
31	-0.583	-0.473
32	0.018	0.267

Since vegetation growth often lags precipitation by some factor of time the correlation between MVI and precipitation from both one and two periods before the MVI reading were checked. Comparing MVI to precipitation in the previous 14 day period the mean correlation increased from -0.375 to -0.394. The highest individual correlation was -0.706 during an early spring period. The mean correlation increased to -0.405 with a high individual correlation of -0.762 when comparing MVI to precipitation two periods in advance of the MVI period. A graphical representation of the relationship between MVI and antecedent precipitation can be found in Figures 5.2 and 5.3. Figure 4.2 depicts the rainy period, and Figure 5.3 depicts the dry summer. Note in Figure 4.3 that if the zero precipitation values were taken out the relationship would be essentially flat.

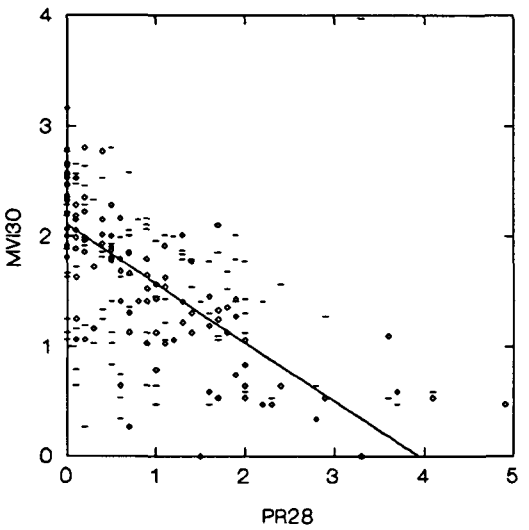


Figure 4.2: The relationship between MVI and antecedent precipitation during the rainy season. The X axis represents precipitation in inches during period 28 (beginning Julian date 59). The Y axis represents logged MVI values during period 30 (beginning Julian date 87).

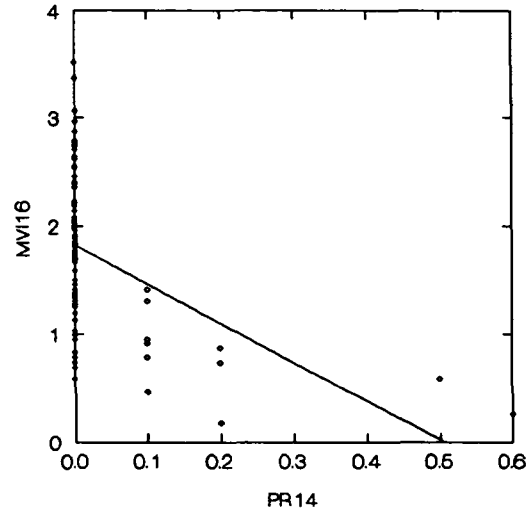


Figure 4.3: The relationship between MVI and antecedent precipitation during the dry season. The X axis represents precipitation in inches during period 14 (beginning Julian date 228). The Y axis represents logged MVI values during period 16 (beginning Julian date 256).

The negative correlation between MVI and precipitation during comparable time periods was a surprise. MVI values should increase over wetted areas, therefore one would expect a positive correlation between MVI and precipitation. Perhaps the negative correlation between MVI and comparable time periods precipitation can be attributed to the gross phenology of California where a large percent of the State is covered by grasses or desert adapted vegetation both of which tend to be rapidly responsive to changes in water availability, which would increase the surface roughness and could thus, in turn, decrease MVI.

After a period of time necessary for the onset of vegetation growth coincident with increased precipitation, MVI should actually decrease and thus lead to a negative correlation between precipitation and MVI. The increase in the negative correlation between MVI and precipitation in the preceding two periods therefore makes sense.

Precipitation, and in particular precedent precipitation, does show some affect to MVI. However, the average correlation (using the precedent precipitation) of -0.405 is not as large as had originally been expected. While there was no set numerical correlation value that had been expected, there are a wide variety of studies that show that precipitation dramatically affects the emissivity of soils and thus the TB, (Choudhury et al., 1987; Neale et al., 1990; Sen et al., 1990).

4.3.3 The Relationship Between MVI and Ground Cover

Ground cover should affect MVI both due to the increase in surface roughness and the increase in water coverage over the soil surface (Choudhury, 1989). In general the thicker the vegetation coverage in terms of optical depth the lower MVI should be (Van de Griend and Owe, 1993). As discussed in Section 3, ground vegetation was split into five types; forest, grassland, scrub, shrub and woodland.

A sixth category called other was also used for some statistical techniques. The "other" category in the CalVeg base map is mainly agricultural land. The "other" category was also used for any pixel that did not contain at least 30% ground cover in any of the main five categories. This was done to avoid designating a pixel a given class coverage type if it wasn't covered by at least 1/3 of the vegetation category.

One result not previously discussed in Section 4.3 is the high correlation between each of the MVI data sets. When checking multi-temporal MVI versus various other variables one result became clear. The MVI data sets generally show a correlation of between 0.900 and 0.950, which infers a strong structural component, see Table 4.19. The most logical structural component at 37 GHz would be vegetation.

Table 4.19: The correlation between various MVI periods, roughly corresponding to 3 month intervals.

	MVI1	MVI7	MVI13	MVI19	MVI25	MVI31
MVI1	1.000					
MVI7	0.895	1.000				
MVI13	0.869	0.829	1.000			
MVI19	0.951	0.903	0.903	1.000		
MVI25	0.936	0.913	0.868	0.947	1.000	
MVI31	0.945	0.906	0.919	0.957	0.949	1.000

As vegetation should increase surface roughness, and thus decrease MVI, it is expected that the largest decrease in MVI would occur in forested regions. Statistically this would appear as a large negative correlation. The least vegetated regions, scrub, should have the highest MVI values and thus produce

a large positive correlation with MVI. In between these two endpoint categories it is expected that the next to lowest MVI values would be in the woodland category, followed by the shrub category, and finally the next to highest MVI values should come from the grassland category.

Repetitive runs of each of the MVI temporal periods were compared to the five vegetation categories. For ease of discussion four time periods corresponding to early winter, early spring, early summer, and early fall have been broken out, see Table 4.20.

Table 4.20: Correlation between MVI and Groundcover. The columns represent different time periods. The beginning Julian date of the MVI time period in question is shown at the top of the columns. The dates correspond to winter, spring, summer and fall.

	MVI BEGINNING JULIAN DATE			
GROUNDCOVER	3	101	186	284
FOREST	-0.574	-0.512	-0.590	-0.453
GRASSLAND	-0.158	-0.191	-0.153	-0.169
SCRUB	0.492	0.591	0.566	0.630
SHRUB	-0.202	-0.140	-0.077	-0.212
WOODLAND	-0.438	-0.446	-0.322	-0.451

In the above case theoretical considerations correlate well with the observed data. Forested areas show the largest negative correlation, followed by woodland. Grassland and shrub have values more closely interlinked than was expected. Perhaps what is happening between these two categories is that at a scale of 1 km2 (which was the original pixel size basis before expanding the pixel scale to 625 km2) shrub and grass are often intermixed in California.

While it is not part of this paper to determine an algorithm to discriminate different vegetation types it is useful to show a plot of MVI versus different category types, see figures 4.4 - 4.7.

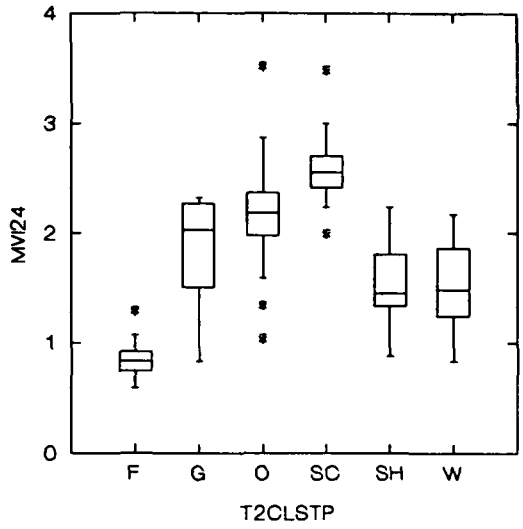


Figure 4.4: Box plot of logged MVI values versus different vegetation categories during the winter. On the Y axis MVI24 is the 24th MVI period, with starting Julian date of 3. On the X axis F stands for Forest, G is Grassland, O is Other, SC is Scrub, SH is Shrub, W is Woodland. The median value is signified by the horizontal line in the middle of the box. The upper and lower end of the box show the interquartile range, (Wilkinson, 1989a).

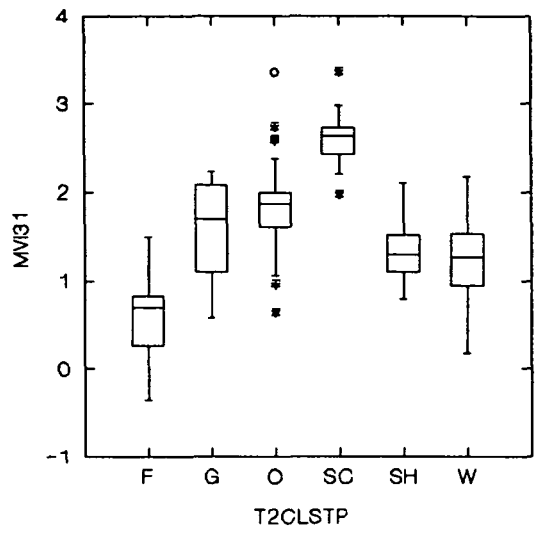


Figure 4.5: Box plot of logged MVI values versus different vegetation categories during the spring. On the Y axis MVI31 is the 31st MVI period, with starting Julian date of 101. On the X axis F stands for Forest, G is Grassland, O is Other, SC is Scrub, SH is Shrub, W is Woodland. The median value is signified by the horizontal line in the middle of the box. The upper and lower end of the box show the interquartile range.

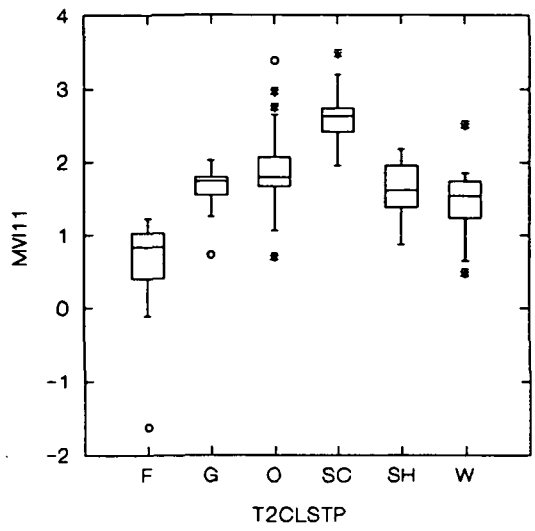


Figure 4.6: Box plot of logged MVI values versus different vegetation categories during the summer. On the Y axis MVI11 is the 11th MVI period, with starting Julian date of 186. On the X axis F stands for Forest, G is Grassland, O is Other, SC is Scrub, SH is Shrub, W is Woodland. The median value is signified by the horizontal line in the middle of the box. The upper and lower end of the box show the interquartile range.

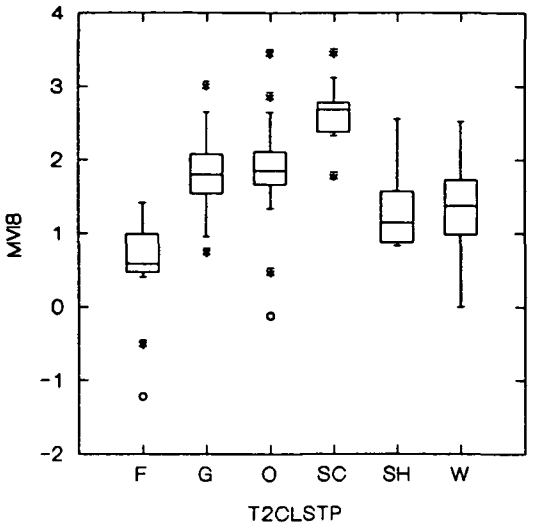


Figure 4.7: Box plot of logged MVI values versus different vegetation categories during the fall. On the Y axis MVI18 is the 18th MVI period, with starting Julian date of 284. On the X axis F stands for Forest, G is Grassland, O is Other, SC is Scrub, SH is Shrub, W is Woodland. The median value is signified by the horizontal line in the middle of the box. The upper and lower end of the box show the interquartile range.

As can be seen, from Figures 4.4 through 4.7, it is relatively easy to discriminate forested areas from scrub covered areas. The other 3 known categories show a more mixed response. The catch-all category Other shows quite a wide spread, as might be expected.

By looking at the above graphs it seems reasonable to hypothesize that MVI can be used to differentiate gross vegetation differences. The primary difficulty will be to differentiate intermediate conditions resulting from either ground cover that appears relatively comparable at a 625 km2 scale, such as grassland and shrub, or a heterogeneous pixel having no single overriding category.

Somewhat disconcertingly, given the large correlation between MVI across dates, the structural component of vegetation categories did not appear to have a large correlation coefficient with MVI. The strongest groundcover correlation is scrub, yet even scrub only shows a correlation of 0.630. At 37 GHz, or 8.1 mm, it is difficult to believe that either gross surface morphology or soil structure would create the structural components necessary to produce comparable MVI

across many dates. It is difficult to believe either gross surface morphology or soil structure could produce the structural components necessary to affect MVI because by definition gross surface morphology has to do with large surface features (Norris and Webb, 1990), while soil structure tends to be dominated by features less than 5 mm in size (Brady, 1984).

4.3.2 The Relationship Between MVI and NDVI

MVI should decrease as vegetation density increases. Aside from gross vegetational differences such as forest and grassland, vegetation ground coverage varies throughout the phenological cycle of the vegetation in question. Deciduous vegetation in particular should show an increase in surface roughness as leaves develop, which would lead to a decrease in MVI (Choudhury, 1989). A secondary aspect of vegetation growth which affects MVI is that an increasing amount of water resides in the plant tissues leading to a decrease in MVI values (Prince and Choudhury, 1989).

One way to determine an increase in vegetation primary productivity is by NDVI (Box et al., 1989). Using NDVI as a surrogate for vegetation growth a correlation analysis was run between MVI and NDVI. Correlation values ranged from -0.600 to -0.868, with a mean correlation value of -0.692. The -0.868 value was substantially different from other correlation values. It corresponded to period 10 which contained 18 useable observation points. All the other correlations used at least 110 observation points. The next highest correlation after -0.868 was -0.742, see Table 4.21.

Table 4.21: Correlation between MVI and NDVI. The mean correlation = -0.692.

TIME PERIOD	CORRELATION WITH NDVI
MVI4	-0.696
MVI6	-0.723
MVI7	-0.696
MVI8	-0.686
MVI9	-0.696
MVI10	-0.868
MVI11	-0.635
MVI12	-0.6
MVI13	-0.617
MVI14	-0.626
MVI15	-0.625
MVI17	-0.716
MVI18	-0.732
MVI28	-0.653
MVI29	-0.715
MVI30	-0.742
MVI31	-0.729
MVI32	-0.702

The relationship between MVI and NDVI appears to be the strongest of the four independent variables. This result was somewhat of a surprise because theoretically NDVI tends to explain primary productivity while MVI tends to show surface roughness features. This relationship may be biased because the majority of California's ecosystems are dominated by grass or scrub which would tend to show the largest percentage change in vegetation surface area over a growing season. A visual representation of the relationship between MVI and NDVI can be seen in Figures 5.8-5.11. Of these four figures Figure 5.8 is shifted about 2 months from figure 5.4 above, but the latter three compare temporally with the vegetation ground coverage figures shown above.

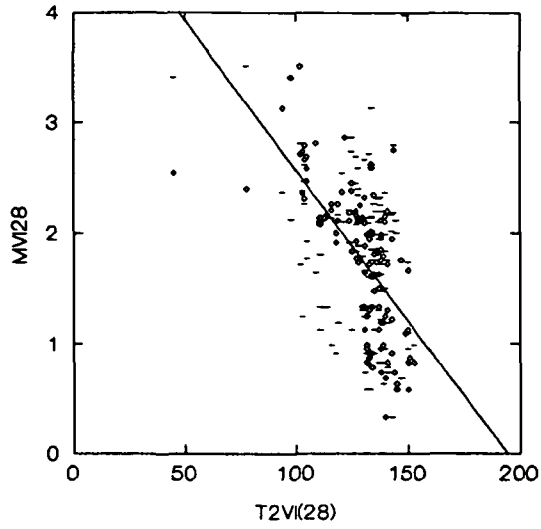


Figure 4.8: Relationship between MVI (X-axis) and NDVI (Y-axis) during period 28 (beginning Julian date 59).

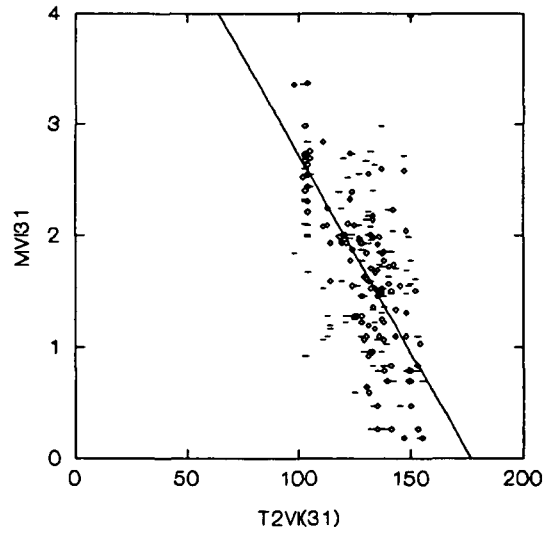


Figure 4.9: Relationship between MVI (X-axis) and NDVI (Y-axis) during period 31 (beginning Julian date 101).

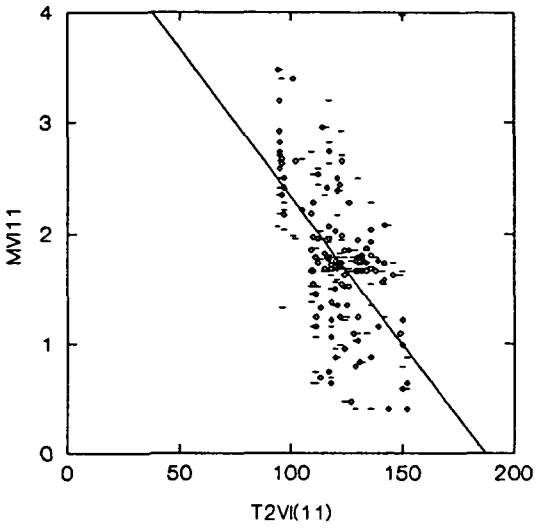


Figure 4.10: Relationship between MVI (X-axis) and NDVI (Y-axis) during period 11 (beginning Julian date 186).

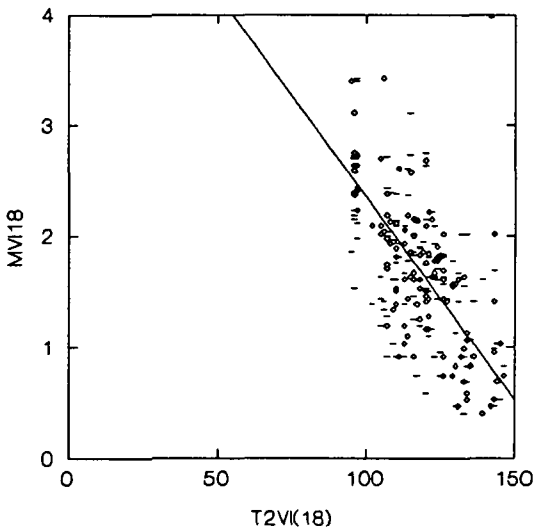


Figure 4.11: Relationship between MVI (X-axis) and NDVI (Y-axis) during period 18 (beginning Julian date 284).

4.4 Development of an Experimental Model to Predict MVI

The following section describes the development of an operational procedure to predict MVI values in California from known data. It starts with a discussion of the development of predictor variables. The use of the predictor variables to evolve a statistically valid model follows. A discussion of results concludes this section.

The first step in developing an experimental model was to create a 78 x 78 correlation matrix using all independent and dependent variables. The matrix consisted of 32 MVI variables, 32 precipitation variables, 18 NDVI variables, 5 vegetation type variables, and 1 water body variable. Some of these results have been discussed in Section 4.3, e.g. MVI versus precipitation, NDVI, vegetation coverage, and water body coverage.

The next step was to try and consolidate the multi-temporal data sets, e.g. MVI, NDVI and precipitation, into one individual variable for each multi-temporal data set. This was done by running a principal components analysis on these variables.

Principal components are the eigenvectors of a variance-covariance or a correlation matrix (Davis, 1986). The eigenvectors of the matrix represent the orientations of the principle axes of an n-dimensional ellipsoid, while the eigenvalues depict the lengths of the principal semi-axes.

It was hoped that by doing a principal components analysis on the precipitation data sets one of the first three variables produced by the analysis would represent the seasonal aspects of precipitation. A principal components analysis based on a correlation matrix was used. Generally a correlation matrix transformation normalizes the different data sets first so that each set has comparable variance. Due to several of the 32 precipitation periods showing essentially no rain at any observation points (which means that the statistical program would view them as constants), only 21 of the periods were used for the analysis. The periods used were; 1-8, 16-19, 21, 23-28, 30 and 31.

The first principal component explained approximately 62.6% of the variance while the second and third components explained approximately 8.0% and 6.6% respectively. Generally the first 3 principal components will explain between 85 to 95% of the variance, so these three principal components contained somewhat less information than was expected.

Looking at the component loadings, the elements of the eigen vectors that are used to compute the scores of observations, the first set of component loadings

showed a strong positive coefficient ranging from approximately 0.660 to 0.940. My surmise is that this component is a structural component, possibly related to orographic lifting, or some other structural effect of California geography on precipitation. What I was looking for in the second or third component was some sort of a variation between positive and negative coefficients that would be interpretable as seasonal changes in precipitation. While these variations did occur, the relative component loadings were generally of a small magnitude.

The first three principal component loadings were applied to the 21 precipitation data sets to produce three new/derived variables. These variables were to be used in conjunction with the principle component variables produced from the two other multi-temporal data sets; MVI and NDVI. The expectation was that the multi-temporal data could be distilled to one set of data.

Principal components analysis was applied to the NDVI data. All 18 temporal periods were used. A correlation matrix was used. The first principal component explained 88.1% of the variance, while the second and third explained 7.3% and 2.0% respectively. The first principal component showed high positive loadings ranging from approximately 0.85 to 0.95. The second principal component appeared to show a possible seasonal component. The periods corresponding to high summer had a negative coefficient while the other time periods had positive coefficients.

The first three principal component loadings were applied to the 18 NDVI data sets to produce three new/derived variables.

Finally a principal component transformation was applied to the 32 MVI data sets. This transformation was done in order to compare like objects to like, in this case to compare the transformed data of the NDVI and precipitation data sets to transformed MVI data.

A correlation matrix was used for the transformation. The first principal component explained 96.6% of the variance of the logged MVI data while the second and third variables explained respectively 1.7% and 0.5% of the variance. The loadings on the first principal component showed a very strong structural element. Essentially all coefficients ranged between 0.980 and 0.995. The second principal component showed some possibility of having a temporal component as did the third principal component.

The first three principal component loadings were applied to the 32 MVI data sets to produce three new/derived variables.

A correlation matrix to compare the first three principal components of the precipitation, NDVI and MVI data sets was created, see Table 4.22. The only variable that appears to have a fairly strong relation with MVI is the first principal component of precipitation versus the first principal component of MVI, with a correlation of -0.516. It seems logical to hypothesize that the structural basis of MVI would indeed be related to the structural basis of precipitation. The structural basis of precipitation in California has two main components. The two structural components are orographic lifting and the north/south gradient for precipitation, e.g. generally speaking the higher the altitude or the higher the latitude the more precipitation one would expect.

Table 4.22: Correlation between the first three principal components of MVI, precipitation, and NDVI.

	MVI1	MVI2	MVI3	PRCP1	PRCP2	PRCP3	NDVI1	NDVI2
MVI1	1.000							
MVI2	0.331	1.000						
MVI3	-0.267	-0.131	1.000					
PRCP1	-0.516	-0.042	0.376	1.000				
PRCP2	0.198	0.064	0.225	-0.234	1.000			
PRCP3	0.371	0.019	-0.301	-0.856	0.097	1.000		
NDVI1	-0.406	0.226	0.104	0.747	-0.452	-0.615	1.000	
NDVI2	-0.201	-0.246	-0.275	0.061	-0.275	-0.113	0.058	1.000
NDVI3	-0.020	-0.212	0.218	0.129	0.305	-0.122	-0.350	-0.089

The principal components analysis was an attempt to determine what sort of temporal processes might be affecting MVI. During the period in question, 1989-90, it appears that the temporal effect was not great. This period was also a period of significant drought across the State of California, consequently it would be worthwhile to perform this type of analysis on a year with significant precipitation. If successful one would expect one of the first three principal components for each of MVI, NDVI, and precipitation to show strong seasonal changes.

The 3 MVI principal components were then correlated with ground cover to see if there was any relationship between them. The results, see Table 5.23, generally did not show a strong relationship between any of the ground cover variables and the principal component variables.

Table 4.23: The correlation between the first three principal components of MVI versus various ground cover types.

	MVI1	MVI2	MVI3
FOREST	-0.423	0.082	0.455
GRASSLAND	-0.098	-0.033	-0.068
SCRUB	0.233	-0.226	0.187
SHRUB	-0.104	-0.140	-0.211
WOODLAND	-0.320	-0.213	-0.058
WATER	0.205	-0.055	0.156

As no significant insights into MVI were derived from principal components analysis no further analyses were conducted.

4.4.1 Variable selection

In developing a regression model for the MVI data it was first necessary to determine which variables would be included in the model. While using all 8 possible independent variables; forest, grassland, scrub, shrub, woodland, water bodies, NDVI, and precipitation, would have the maximum possible "information content" it would also produce a problem. The problem is that the variance of the prediction increases as the number of regressors increases (Montgomery and Peck, 1982).

One way to evaluate any given subset of regressor variables is to check the coefficient of multiple determination (R^2). Generally, the higher the R^2 the more effective the model is in predicting the actual value of the dependent variable. Other ways to determine the optimal subset of independent variables includes using the adjusted R^2 or the residual mean square. In the case at hand, the adjusted coefficient of multiple determination, which takes the number of parameters in the model into account through the degrees of freedom (Neter et al., 1990), was used to help develop the model. The adjusted coefficient of multiple determination is defined by:

$$\overline{R}_p^2 = 1 - (\frac{n-1}{n-p})(1 - \overline{R}_p^2)$$

Where:

R_p^2 is the coefficient of multiple determination for a subset regression model with p terms.

p is the number of parameters.

n is the number of observations.

The following description of variable selection is somewhat simplified due to the nature of adding and subtracting multiple variables to reach an optimal model. The following describes the basic process used, as well as the ultimate combination of independent variables used for the descriptive model. The write-up does not describe the full iterative combination of variables for any given time period.

If all variables were used, with precipitation being lagged two periods, and using the water bodies data set that only includes observation points with at least 1% water body coverage the regular R^2 would be 0.774. However, using the adjusted R^2, this value changes to 0.412 because n was only fifteen.

If all variables were used, including the full water body data set which contains 122 observation points, the regular R^2 is 0.675. The adjusted R^2 is to 0.649. All future discussions of adjusted R^2, unless otherwise noted, will employ the full water body data layer, containing 122 observation points, for the State of California.

Interestingly, if the precipitation data set is not used, even though it shows a correlation of -0.482, the adjusted R^2 changes to 0.668. Clearly the spatial distribution of a two week series of storms, during the first half of March 1989, causes a conflict with the underlying spatial distribution of vegetation. This effect was noticeable to a greater or lesser amount in other temporal periods. Even if the first principal component for precipitation is used, which should contain the structural basis for precipitation, the adjusted R^2 still is only 0.666. For this reason precipitation was deleted as an independent variable, and the adjusted R2 that all following adjusted R^2 are compared against is 0.668.

Having a base adjusted R^2 of 0.668 means that approximately 2/3 of the output value of the MVI could be accounted for by the independent variables. This means that approximately 1/3 of the data content of the MVI data sets used in this experiment is based on other factors. Clearly this leaves a large gap in the development of an empirically based model.

As can be seen in Table 4.20 both the grassland and shrub categories appear to add little to MVI. If the grassland category is deleted the adjusted R^2 changes to

0.665. If the shrub category is deleted the adjusted R^2 changes to 0.645. Deleting both categories leads to an adjusted R^2 of 0.644. By deleting both categories the adjusted R^2 has changed by only 3.2%, (e.g. 0.645/0.668 = 96.6).

The full water body data set also appears to have a low correlation, approximately 0.210, with MVI. However, it became obvious that although some independent variables might have low correlations with MVI they actually produced higher adjusted R^2 values than other independent variables with much higher correlations. While a higher adjusted R^2 does not necessarily lead to a better fitting model (Montgomery and Peck, 1982), (and as was found out by this researcher by trial and error) it is a reasonable starting point. By deleting the water body variable the adjusted R^2 changes to 0.617. However, by deleting the scrub variable with it's much higher correlation of approximately 0.570 the adjusted R^2 only dropped to 0.629. At this point four of the original eight variables have been dropped and the adjusted R^2 has only changed 5.8%.

The rest of the variables; forest, woodland, water bodies, and NDVI were iteratively deleted. Only the water body data set could be deleted without a major change in the adjusted R^2. When the water body data set was deleted the adjusted R^2 dropped to 0.609. The end result was that 91.2% of the information content of the eight independent variables was accounted for by three of the variables.

Having determined an optimal subset of independent variables it is necessary to resolve whether there is a problem with multicollinearity. When there is multicollinearity the variance of the regression coefficients will be degraded, significance tests will not be as meaningful, and the estimated regression coefficients will be highly unstable.

Two methods were used to determine whether multicollinearity of the independent variables could cause a problem with the predicted MVI values. The first method was to check on whether there was a high degree of correlation between the three optimal independent variables; forest, woodland, and NDVI. The result can be seen in Table 4.24. A correlation of 0.0 would be optimal. The high correlation of 0.384 should not be large enough to cause multicollinearity problems according to Kirby (1993).

Table 4.24: Correlation between the three independent variables.

	FOREST	WOODLAND	NDVI
FOREST	1		
WOODLAND	-0.1	1	
NDVI	0.305	0.384	1

A second method to determine whether multicollinearity will cause problems within a regression equation is called tolerance. Tolerance is automatically computed by Systat when computing a regression. Tolerance is equal to 1 minus the proportion of variance in each predictor accounted for by all other predictors (Wilkinson, 1989b; Kirby 1993). If the value of tolerance for a given independent variable is close to zero then that variable is close to being a linear combination of other predictors. The tolerance values for the forest, woodland and NDVI variables were 0.854, 0.819, and 0.756 respectively. The net result of the correlation tests and the tolerance tests was to conclude that the three variables could be used as independent variables to create a regression model for MVI.

Applying forest, woodland, and NDVI independent variables to a regression equation with MVI being the dependent variable led to the following model equation being developed:

$$MVI = 4.912 + (-0.009)*F + (-0.011)*W + (-0.023)*NDVI$$

Where:

F is forest and W is woodland.

This equation was based on a period with a beginning Julian date of 88, early spring.

As mentioned in Section 3, when the information was stripped out of the various data layers a little more than half of the data were funneled into one set, while the rest of the data were funneled into a second set. The first subset of data were used for the statistical tests enumerated in this chapter and to develop the model. The above equation was applied to the second subset of data to create a set of expected MVI values.

These expected MVI values were then correlated and plotted against the actual MVI values of the second subset. The results, as can be seen in Table 4.25, showed a mean correlation of 0.694. If the questionable results from period 10 are not used the mean correlation increases to 0.712. For a model to be truly useful a correlation of over 0.900 is desired.

Table 4.25: Correlation between expected MVI value and actual MVI value. Only 18 MVI periods are used as there were no NDVI values for the 14 other periods. All periods except for period 10 used 104 observation points. Period 10 only used 13 observation points.

TIME PERIOD	CORRELATION
MVI4	0.728
MVI6	0.695
MVI7	0.644
MVI8	0.649
MVI9	0.714
MVI10	0.402
MVI11	0.687
MVI12	0.723
MVI13	0.739
MVI14	0.757
MVI15	0.725
MVI17	0.735
MVI18	0.743
MVI28	0.726
MVI29	0.74
MVI30	0.727
MVI31	0.714
MVI32	0.652

The average correlation is 15% greater than would be expected due to the adjusted R^2, e.g. the adjusted R^2 was 0.609 while the average correlation was 0.712. Since the equation used was based on an early spring time period, the time frame when there normally would be rainfall in California, a second model equation was tried using a dry summertime period. The period used was 15, with a beginning Julian date of 242. Using the same three independent variables the adjusted R^2 was actually better for this period than for the spring period. The adjusted R^2 for this period was 0.691. The resulting model equation was:

$$MVI = 4.327 + (-0.012)*F + (-0.013)*W + (-0.018)*NDVI$$

The expected results when correlated with the actual results were actually not as good as the first model. The average correlation was 0.676 when all periods were taken into account and 0.691 when period 10 was deleted. One possible reason why this produced results with a somewhat lower correlation with the expected MVI value than the first equation is that structural data such as forest or woodland would be more important during a summer period, due to increasing senescence of grasslands etc., and thus might be overweighted.

Plots of the the expected versus actual logged MVI value (using the spring model equation values) can be seen in figures 5.12-5.15.

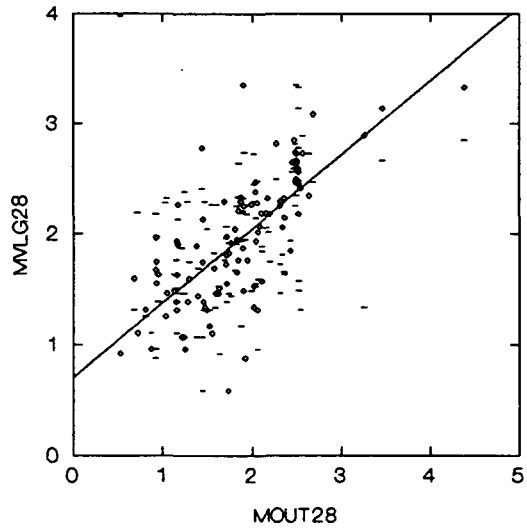


Figure 4.12: Plot of expected MVI value (x-axis) versus actual MVI value (y-axis). Beginning Julian date for the data set is 59. The correlation between the two is 0.726.

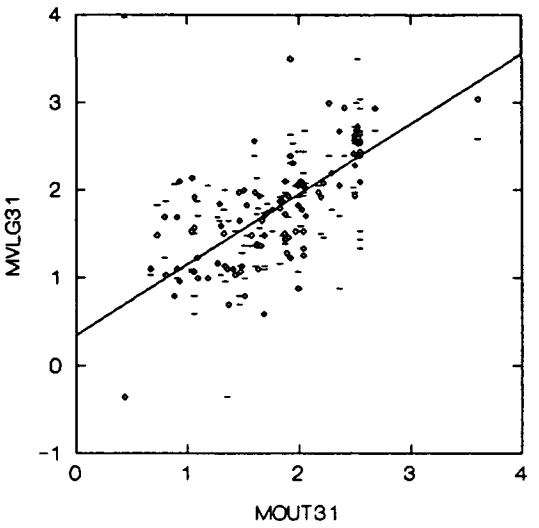


Figure 4.13: Plot of expected MVI value (x-axis) versus actual MVI value (y-axis). Beginning Julian date for the data set is 101. The correlation between the two is 0.714.

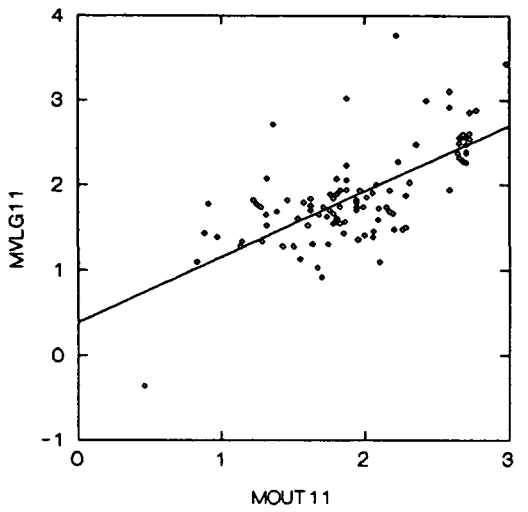


Figure 4.14: Plot of expected MVI value (x-axis) versus actual MVI value (y-axis). Beginning Julian date for the data set is 186. The correlation between the two is 0.687.

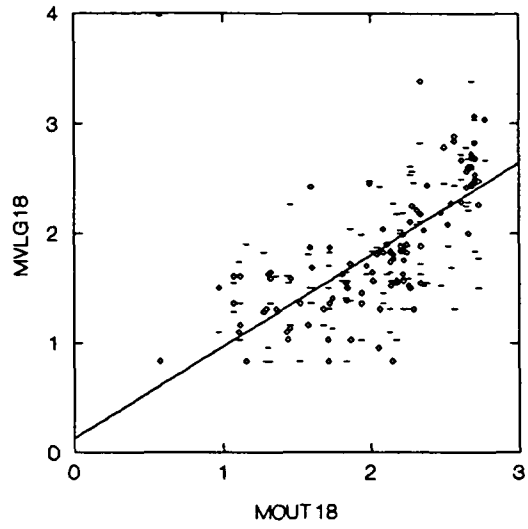


Figure 4.15: Plot of expected MVI value (x-axis) versus actual MVI value (y-axis). Beginning Julian date for the data set is 284. The correlation between the two is 0.743.

4.4.2 Discussion of Results

The results of estimating MVI using the sundry independent variables were not as robust as expected. The general results were correlations in the 0.70 to 0.75 range. Of particular surprise was precipitation's lack of importance to the model. It was expected that precipitation would be the second most important variable.

Precipitations effect on the passive microwave vegetation index signal in any given period can vary. Precipitation could increase MVI if there is standing water, saturated soil, or well soaked ground. Precipitation could decrease MVI if the precipitation occured during the early part of the period, stimulating plant growth over the rest of the period. However, precipitation, over any sort of a time lag, in this case 4 weeks, should lead to a definite lowering of MVI due to the increase in vegetation density. This lowering of MVI did not occur as strongly as expected.

The period 1989-90 in California was a period of drought. It is likely that precipitation did not prove more important to affecting the passive microwave signal because there was not enough precipitation to create large increases in vegetation biomass.

It was hypothesized that MVI would be most affected by water bodies. While water bodies showed some promise of use as an indicator of MVI values in any

given pixel, there are not enough large water bodies in the State of California to make a statistically significant statement in this area. In the tests used, water bodies clearly affected the R^2 values of the linear regression. However, due to the paucity of data points the adjusted R^2 values were lower than expected.

These results are indicative of two aspects of using passive microwave with a ground footprint the size of SSM/I's footprint for remote sensing of vegetative coverage. With a pixel size of 625 km2 the pixel will, in all likelihood, be extremely heterogeneous. Unless the water body makes up a fairly significant portion of the pixel the likelihood is that the pixel's values will be comparable to the values of the surrounding pixels. A corollary of this is that predicting MVI's value by water surface area would best be accomplished in regions with significant water body coverage (if the researcher is interested in vegetation coverage, this is a spurious argument). If, on the other hand, what the researcher is interested in is water coverage, then MVI becomes a more appropriate tool.

Vegetation coverage was expected to be the third most important indicator of MVI value. In particular, regions with dense vegetation coverage, such as forests and woods, were expected to produce low MVI values. These two variables do appear to be useful in estimating MVI values. The paucity of vegetation in desert areas did not prove as useful a tool for designing a model. Scrub covered areas are clearly distinguishable from other areas, yet they do not in themselves appear to be good estimators of MVI.

NDVI was expected to show another aspect of vegetation coverage; primary productivity. A high NDVI would tend to presage a low MVI. This appears to be the case. NDVI turns out to be one of the three most robust predictor variables.

Using NDVI to predict MVI somewhat begs the question of a passive microwave vegetation index. NDVI is a well known and well used vegetation indicator. However, by using it one is in essence abdicating the usage of passive microwave and returning to the well worn path of visible and near infrared remote sensing. While this may be entirely appropriate in this case, we still do not have enough data to make recommendations in other areas with different landscape characteristics.

4.5 MVI Multi-temporal Trajectories at the Three Study Sites

The three study sites consist of the Santa Ynez region, the Rogers Dry Lake region, and the Fresno region (figure 1.1). Each sites MVI, NDVI and

precipitation multi-temporal trajectories were plotted to see if there appeared to be some sort of an interconnection between them.

4.5.1 Rogers Dry Lake Site

The three sites each represent a different type of ground cover. The Rogers Dry Lake site represents desert vegetation. In such a regime one would expect a high MVI value. Vegetation growth would tend to be conditional on the occasional rain burst. Figures 4.16 to 4.19 show the multi-temporal trajectories of MVI, precipitation and NDVI. Two forms of multi-temporal trajectories are used. The first sort uses the mean value for the given site for the given time period. The second type creates a 3 period running average for the site. For example if period 15 were the designated period the values for periods 14, 15, and 16 would be added together and divided by three. This is an attempt to remove ambient noise from the data set.

There does not appear to be any pattern evident in the curves depicted in these figures. The correlation between MVI and precipitation at the site is -0.318.

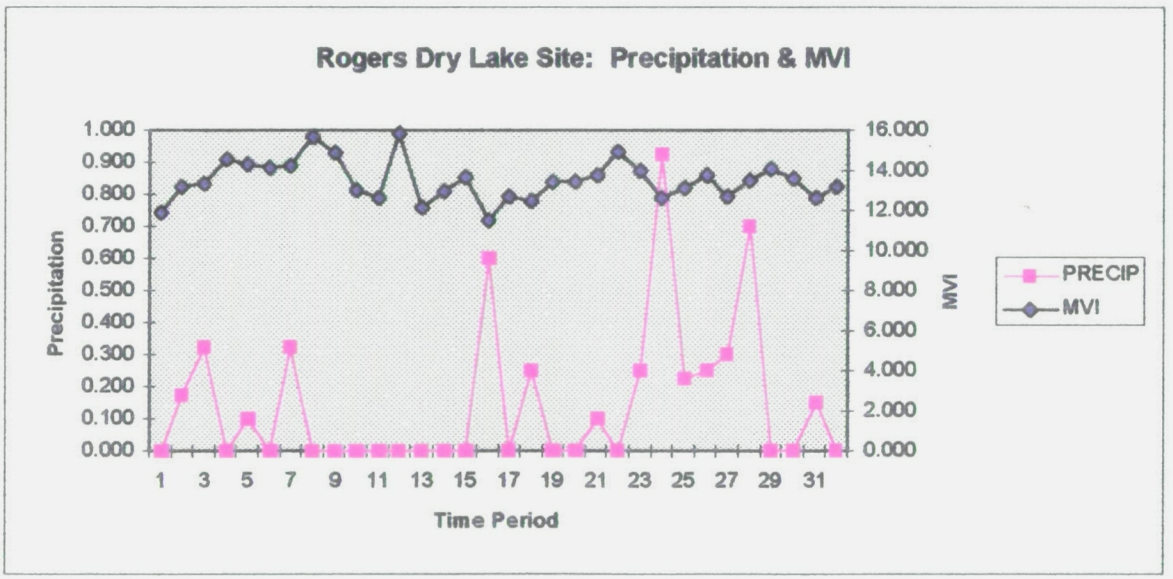


Figure 4.16: Multi-temporal trajectories of MVI and precipitation at the Rogers Dry Lake site. Precipitation, in inches, is on the left axis, MVI value is on the right axis. Time periods are on the X-axis. Each time period is equal to two weeks. Beginning Julian date for time period 1 is 46, for period 16 is 256, for period twenty four is 3 and for period 32 is 115.

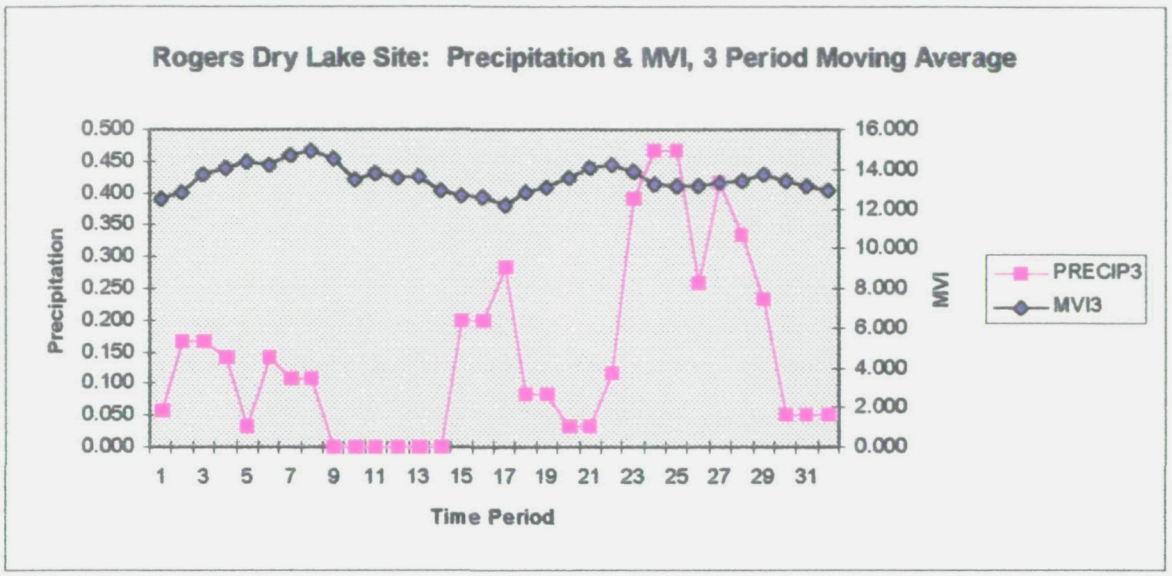


Figure 4.17: Multi-temporal trajectories of MVI and precipitation at the Rogers Dry Lake site using a 3 period moving average. Precipitation, in inches, is on the left axis, MVI value is on the right axis. Time periods are on the X-axis. Each time period is equal to two weeks. Beginning Julian date for time period 1 is 46, for period 16 is 256, for period 24 is 3 and for period 32 is 115.

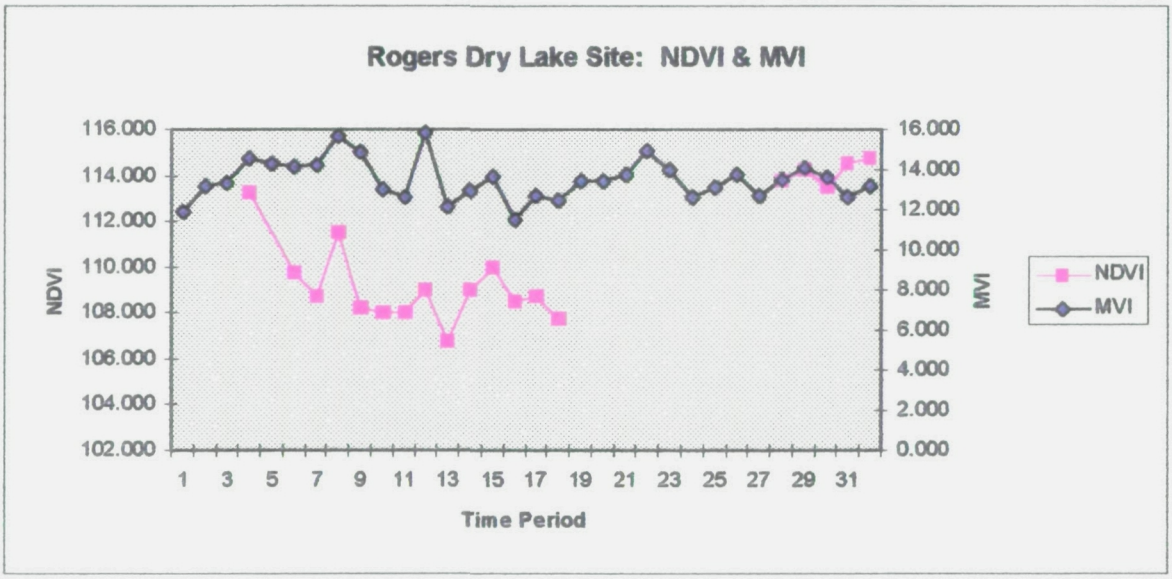


Figure 4.18: Multi-temporal trajectory of MVI and NDVI at the Rogers Dry Lake site. NDVI is on the left axis, MVI is on the right axis. Time periods are on the X-axis. Each time period is equal to two weeks. Beginning Julian date for time period 1 is 46, for period 16 is 256, for period 24 is 3 and for period 32 is 115.

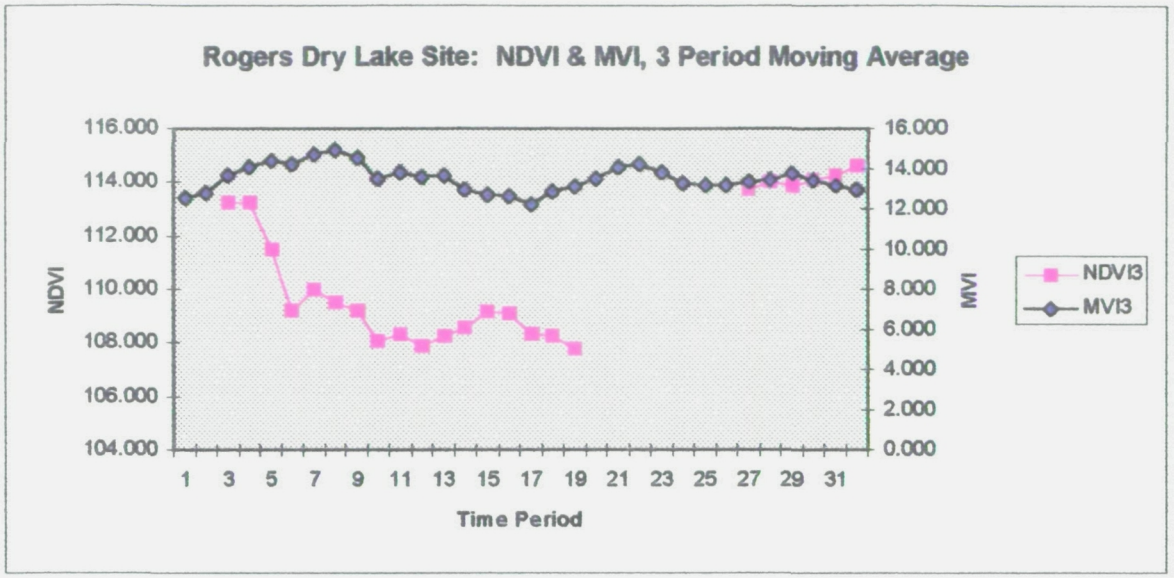


Figure 4.19: Multi-temporal trajectory of MVI and NDVI at the Rogers Dry Lake site using a 3 period moving average. NDVI is on the left axis, MVI value is on the right axis. Time periods are on the X-axis. Each time period is equal to two weeks. Beginning Julian date for time period 1 is 46, for period 16 is 256, for period 24 is 3 and for period 32 is 115.

The only possible pattern to find, using the three period moving average (Figure 4.17) is the decrease in MVI starting around period 22 and continuing through period 26 in conjunction with an increase with rainfall during the same periods. However, if the whole time frame is looked at the relationship between precipitation and MVI seems insignificant. Considering the relatively minimal rainfall this is not surprising.

NDVI and MVI do not appear to show a pattern. The correlation between NDVI and MVI using the individual values is 0.211, while using the 3 week moving average is 0.051. The almost uncorrelated response appears to be based more on the MVI response than the NDVI response, due to the uneven nature of the MVI values. While the MVI value appears to be bouncing around, in absolute values it is only showing a spread of 40. Considering the lack of rain during the 1989-90 period it is possible that the values shown over the desert are more a function of water in the atmospheric column than in any actual change of vegetation on the ground.

4.5.2 Central Valley Site

The Central Valley site is predominantly composed of agricultural vegetation. Unlike the rest of the State of California, which tends to have a spring bloom, the Fresno region often exhibits a summer growing season due to irrigation. Figures 5.20 - 5.23 show the relationship between MVI, precipitation and NDVI.

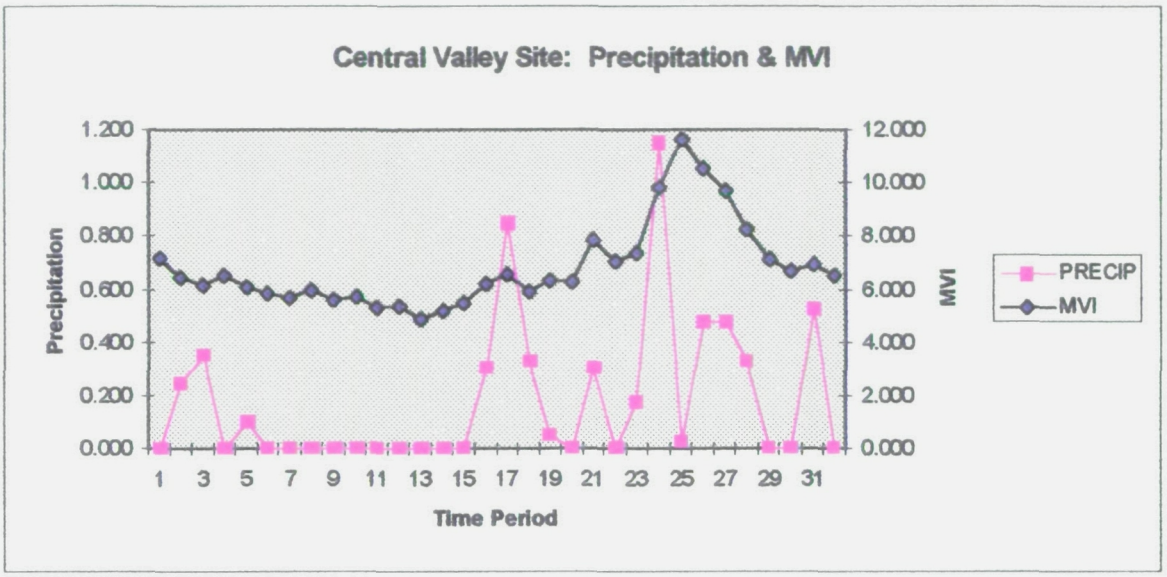


Figure 4.20: Multi-temporal trajectories of MVI and precipitation at the Central Valley site. Precipitation, in inches, is on the left axis, MVI value is on the right axis. Time periods are on the X-axis. Each time period is equal to two weeks. Beginning Julian date for time period 1 is 46, for period 16 is 256, for period 24 is 3 and for period 32 is 115.

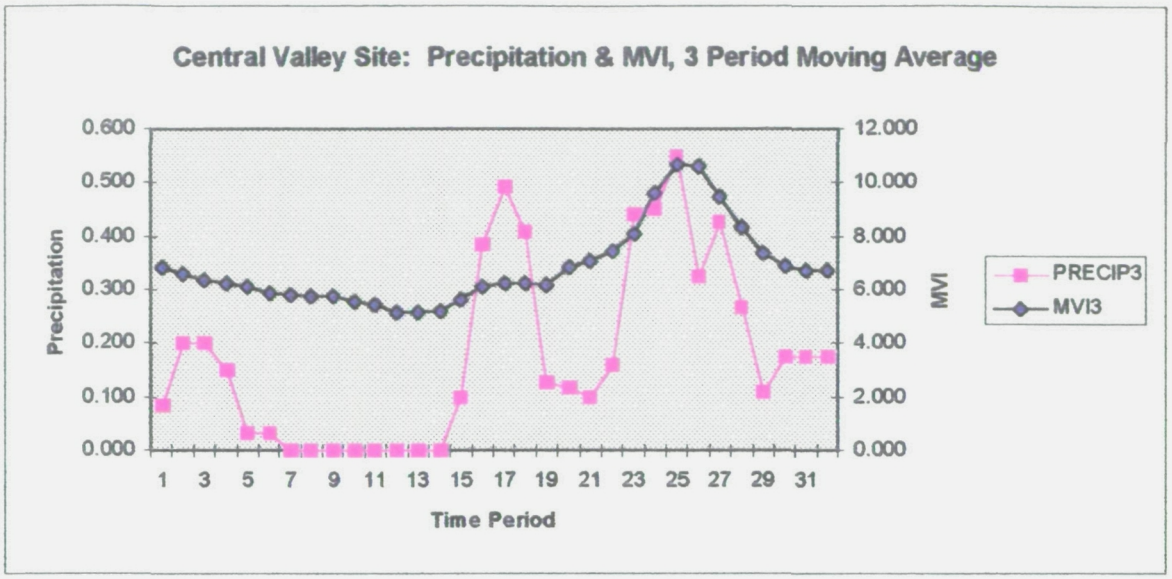


Figure 4.21: Multi-temporal trajectories of MVI and precipitation at the Central Valley site using a 3 period moving average. Precipitation, in inches, is on the left axis, MVI value is on the right axis. Time periods are on the X-axis. Each time period is equal to two weeks. Beginning Julian date for time period 1 is 46, for period 16 is 256, for period 24 is 3 and for period 32 is 115.

The Central Valley site follows the expected multi-temporal trajectory. As the spring progresses into the summer MVI values decline. Various of the crops would be harvested in August and September, starting approximately at period 14, which would lead to an increase in MVI values. The relationship between MVI and precipitation should be less important for an irrigated region than for natural vegetation. It appears, using the three period moving average, that there is some relationship between precipitation and an increase of MVI. Of interest is the fact that MVI shows a positive correlation with precipitation. For the individual periods approach, the correlation is 0.474 while for the 3 week moving average the correlation is 0.715.

Due to precipitation wetting the soil and thus increasing polarization differences, it was expected that precipitation and MVI would show a positive correlation. Over the State as a whole this was not the case. It is possible that the correlation shown here is spurious; a function of the anthropogenic changes of irrigation. Irrigation may cause higher than average soil moisture. The spike in MVI values cresting in period 25, starting Julian date 17, may be a function of agricultural practices, such as turning over the soil. Finally, the topography may have something to due with the positive correlation. As a whole, the rest of the State is fairly mountainous, while the Central Valley is quite flat. Water would

tend to pond on a flat site and thus lead to an increase in MVI. Figures 4.22 and 4.23 show the relationship between the multi-temporal trajectories of MVI and NDVI.

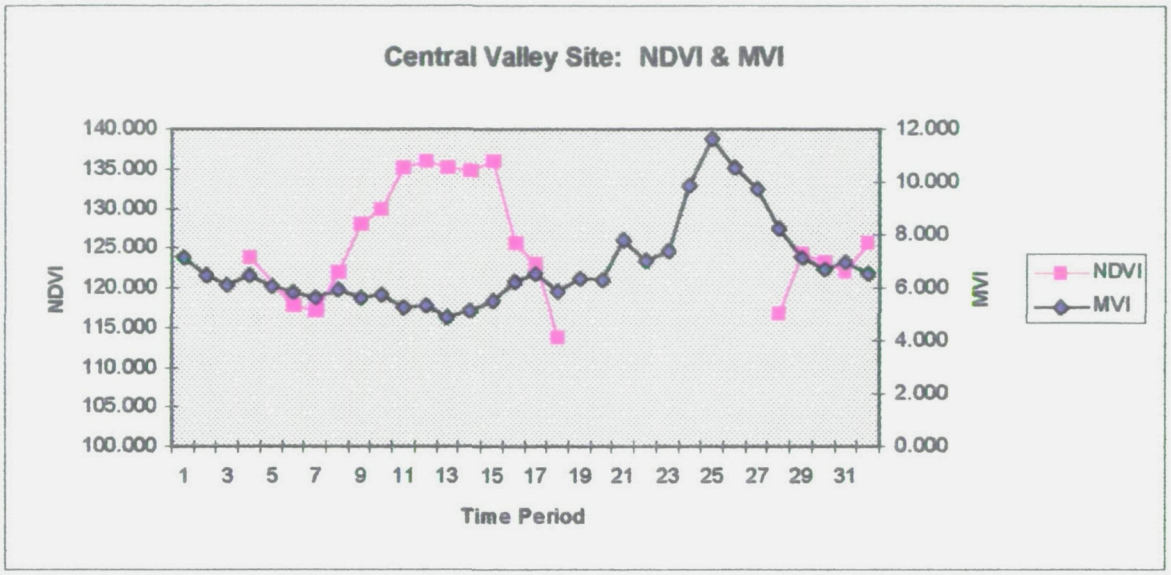


Figure 4.22: Multi-temporal trajectory of MVI and NDVI at the Central Valley site. NDVI is on the left axis, MVI value is on the right axis. Time periods are on the X-axis. Each time period is equal to two weeks. Beginning Julian date for time period 1 is 46, for period 16 is 256, for period 24 is 3 and for period 32 is 115.

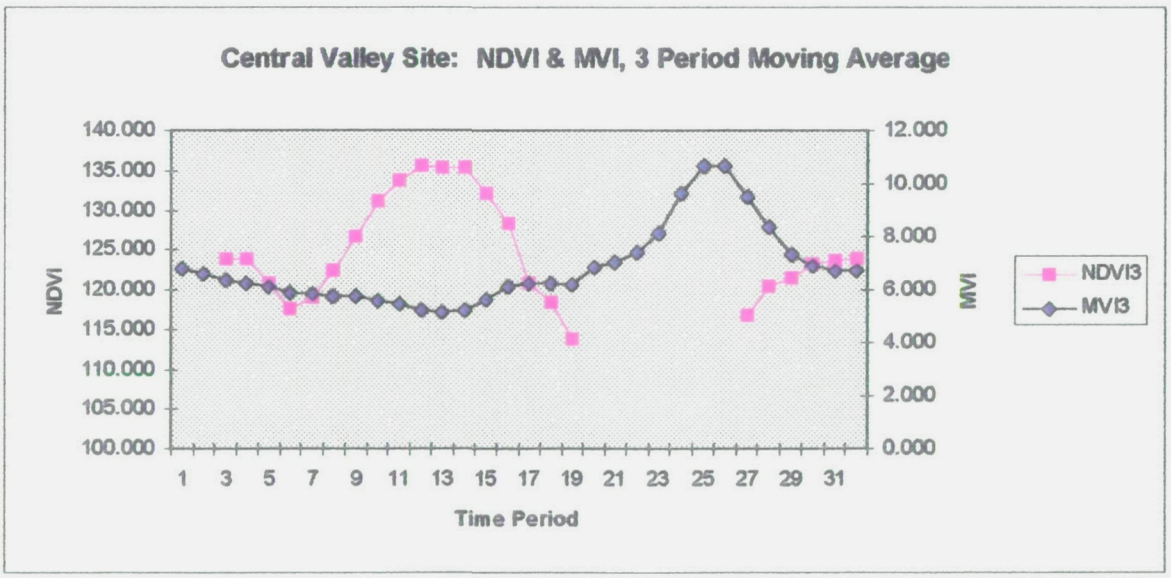


Figure 4.23: Multi-temporal trajectory of MVI and NDVI at the Central Valley site using a 3 period moving average. NDVI is on the left axis, MVI value is on the

right axis. Time periods are on the X-axis. Each time period is equal to two weeks. Beginning Julian date for time period 1 is 46, for period 16 is 256, for period 24 is 3 and for period 32 is 115.

The relationship between MVI and NDVI is about as would be expected. When NDVI increase MVI decreases and vice versa. The correlation between MVI and NDVI is fairly standard at -0.600 for the individual periods and -0.625 for the three week moving average.

4.5.3 Santa Ynez Site

The third site is located in the Santa Ynez region. It's predominant coverage, 45%, is by shrub type vegetation. One would expect a spring low in MVI followed by a summer high. Figures 4.24 - 4.27 show the multi-temporal trajectories of MVI, precipitation, and NDVI at the Santa Ynez site.

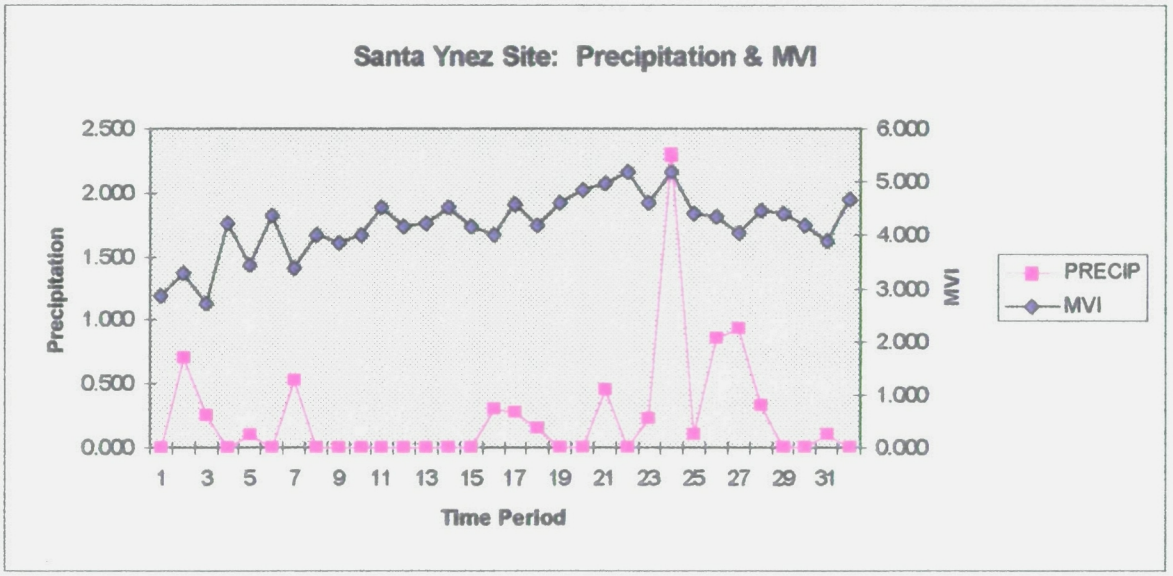


Figure 4.24: Multi-temporal trajectories of MVI and precipitation at the Santa Ynez site. Precipitation, in inches, is on the left axis, MVI value is on the right axis. Time periods are on the X-axis. Each time period is equal to two weeks. Beginning Julian date for time period 1 is 46, for period 16 is 256, for period 24 is 3 and for period 32 is 115.

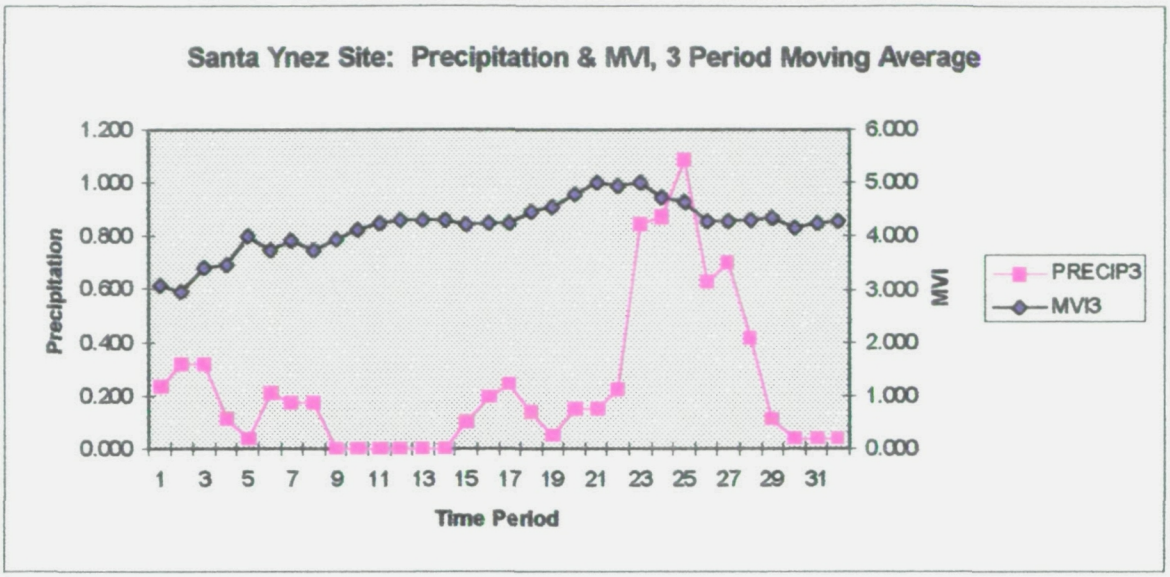


Figure 4.25: Multi-temporal trajectories of MVI and precipitation at the Santa Ynez site using a 3 period moving average. Precipitation, in inches, is on the left axis, MVI value is on the right axis. Time periods are on the X-axis. Each time period is equal to two weeks. Beginning Julian date for time period 1 is 46, for period 16 is 256, for period 24 is 3 and for period 32 is 115.

Looking at Figures 4.24 and 4.25 the MVI pattern is appropriate for the ground cover type. There is a late winter/early spring low followed by a summer high. The second springs low appears higher than the first, which is what would be expected of a continuing and worsening drought.

There is no strong correlation between MVI and precipitation at the Santa Ynez site. The correlation value for the individual periods is 0.153, while for the 3 week moving average it is 0.236. In Figure 4.24 period 24 shows a large rainfall spike which appears to result an increase in the MVI value. Otherwise, rainfall does not seem an appropriate predictor of MVI values.

Figures 4.26 and 4.27 show the relationship between the multi-temporal trajectories of MVI and NDVI.

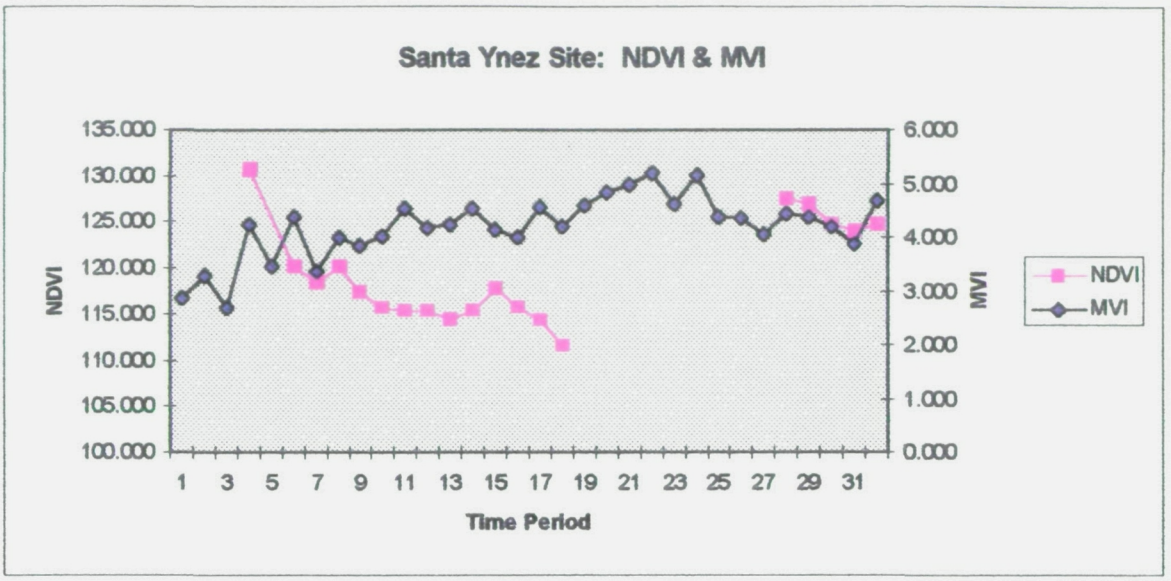


Figure 4.26: Multi-temporal trajectory of MVI and NDVI at the Santa Ynez site. NDVI is on the left axis, MVI value is on the right axis. Time periods are on the X-axis. Each time period is equal to two weeks. Beginning Julian date for time period 1 is 46, for period 16 is 256, for period 24 is 3 and for period 32 is 115.

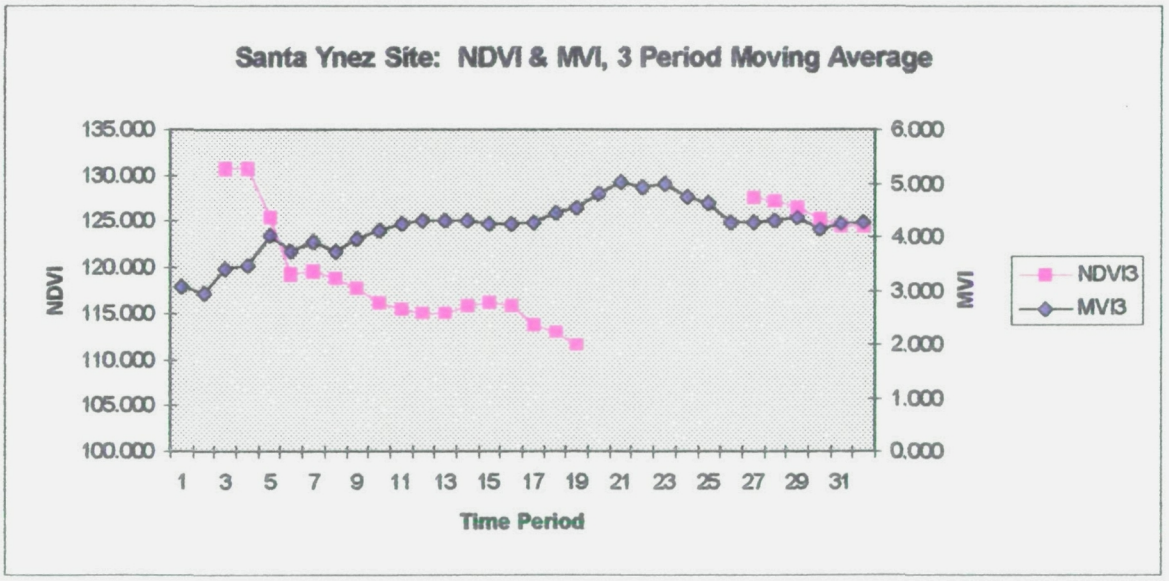


Figure 4.27: Multi-temporal trajectory of MVI and NDVI at the Santa Ynez site using a 3 period moving average. NDVI is on the left axis, MVI value is on the right axis. Time periods are on the X-axis. Each time period is equal to two weeks. Beginning Julian date for time period 1 is 46, for period 16 is 256, for period 24 is 3 and for period 32 is 115.

The NDVI and MVI multi-temporal trajectories fundamentally mirror each other. In general when MVI trends downward NDVI trends upward. Of interest is the time frame around period sixteen when MVI trends down and NDVI trends upward. Some chaparral species have a late season growth spurt. This perhaps could account for the slight change downward around period sixteen.

The correlation between NDVI and MVI at the site was much worse than expected. For the individual periods it was 0.101. For the 3 week average the correlation was -0.366. While looking at Figures 4.26 and 4.27 it appears that MVI and NDVI show a good relationship, statistically this is not the case. Perhaps the driving force behind this fundamental imbalance is the final five periods of the sequence. While the first group of 13 NDVI values seems to mirror the MVI values the last 5 NDVI values appear to show a comparable trend to the MVI values.

In looking at the three study sites the results are inconclusive. The Rogers Dry lake site shows little relationship between the theoretical MVI multi-temporal trajectory, eg. rapid changes in MVI values coincident with precipitation events, and reality. The Central Valley site shows the best relationship between theory and reality. The MVI trend is downward over the growing season followed by an increase after the harvesting period. The Santa Ynez site falls somewhere in between. There is a general upward trend during the summer and a small drop-off during the rainy season.

The period mid-February 1989 to late April 1990 was a period of severe drought. Plants in natural ecosystems do not flourish without water. The general concept of a vegetation index is to determine biomass and plant productivity. A vegetation index such as MVI is more sensitive to biomass than plant productivity. The three sites were chosen with the hope that they would show the highest percentage change in biomass during the phenological cycle. With the extreme drought there was comparatively little vegetation growth. In order to better evaluate MVI it is necessary to overview both individual study sites and the State as a whole during a wetter year, perhaps the winter/spring of 1992-3 or 1994-5. As it now stands, MVI shows a potential value, but not one that is a significant advance on other vegetation indices such as NDVI.

5. Summary and Conclusions

There has been substantial progress in recent years in developing longterm multi-temporal multi-channel passive microwave datasets of the earth's surface. The acquisition and analysis of these datasets has been accelerated by the launching of Defense Meteorological Satellite Program (DMSP) based Special Sensor Microwave/Imager (SSM/I) sensors beginning in 1987. SSM/I is a seven channel, four frequency, linearly polarized, passive microwave radiometric system.

In 1989 the National Aeronautics and Space Administration (NASA) sponsored a prototype earth science and application data system test called Wetnet. One aspect of Wetnet was the distribution of data from SSM/I to multidisciplinary groups of scientists from around the world. The work in this paper is part of the Wetnet project.

An outgrowth of the increased distribution of passive microwave data of the earth's surface has been the development of numerous algorithms to classify these datasets. One of the algorithms created was the Microwave Vegetation Index (MVI), which is based on differencing the 37 GHz horizontally and vertically polarized signals. Vegetation indices are generally used to reduce multi-dimensional data into a single dimension emphasizing a specific aspect of the vegetation coverage within an image.

The objective of this paper was to determine whether MVI corresponds with some quantifiable vegetation parameter, such as vegetation type, or whether the index was more affected by known geophysical parameters such as antecedent precipitation.

Three features of MVI were studied. The first feature studied examined the possibility of different MVI multi-temporal compositing techniques producing different results. The two compositing algorithms used were the Choudhury method, wherein the second lowest value for a pixel for a given time period was used, and the Marshall Space Flight Center (MSFC) method, which averages all values. The second approach for studying MVI was to model MVI values using vegetation ground cover, water bodies, the Normalized Difference Vegetation Index (NDVI) and antecedent precipitation. Finally, the relationship between MVI and NDVI and antecedent precipitation at three study sites representing desert, agricultural and coastal mountain vegetation was detailed.

The study area for this paper was the State of California. California's weather and vegetation are characteristic examples of a Mediterranean regime. Precipitation tends to occur in the winter and spring followed by a drought during

the summer and fall. Vegetation growth is dependent upon the seasonal nature of the precipitation, with most plant species adapted to a spring growing season followed by summer senescence. Chapter Two details the physical aspects of the State of California.

In order to determine vegetation change in California using MVI the physical processes inherent at microwave frequencies must be understood. The spectral, polarization and angular variations of the radiation emitted, absorbed and scattered by a medium are a function of three aspects of the medium. The first aspect is the geometrical structure of the interior and exterior of the medium. The second aspect is the dielectric property of the medium. Finally, the temperature of the medium affects the amount of energy emitted. The energy emitted by a given medium is further modified by the signal from surrounding objects and the atmosphere. Chapter Three itemizes various physical processes that affect passive microwave signals.

In Chapter Four the development of the experimental datasets was detailed. The data were transferred from MSFC to the Remote Sensing Research Unit (RSRU) on tape. The data were radiometrically corrected at the University of Wisconsin and geometrically corrected at RSRU.

After screening the dataset for bad values the data were multi-temporally composited using two week periods. Three derived MVI datasets were created; one using the MSFC approach of averaging all values, and two employing the Choudhury method of using the second lowest value for the given time span. One Choudhury dataset used only the ascending satellite passes data, while the second used only the descending passes data. Thirty-two periods worth of data were created for each method.

The derived MVI values were somewhat different from previous literature values for comparable surface areas. This was an area of concern, as two possibilities were immediately apparent, bad data or operator error. The processed horizontally and vertically polarized 37 GHz data were spot checked with the original data at MSFC and found to be accurate. The only possible operator error would have been in applying the simple mathematical procedure (differencing) to the processed horizontally and vertically polarized datasets, this was double checked and verified to be accurate. The assumption thereafter became that while the values were somewhat different from what was expected from the literature, they were the actual values to the best of our knowledge.

The vegetation and water body independent variable datasets were created by coarsening CalVeg maps from the RSRU database to a 25 km per pixel scale. The NDVI independent variable dataset was also coarsened to 25 km scale.

The precipitation dataset involved transferring data from a CD to ARC/INFO where the point data were kriged into a two dimensional surface. The precipitation data were then changed to the appropriate pixel scale. The data from the dependent and independent variables were then split into two groups; one for creation of the model and one for testing of the model.

Two methods were used to determine the relationship between the three different compositing techniques; correlation and Analysis of Variance (ANOVA). The correlation method showed one distinct trend and one fairly strong trend. Thirty-one out of thirty-two times, the lowest correlation between the three compositing datasets was between ascending and descending datasets created using the Choudhury method. The highest correlation, 22 out of 32 times, was between the ascending dataset and the average dataset.

The lower correlations between the ascending and descending datasets is probably a function of the averaging algorithm used for the MSFC technique. By definition the averaging technique would include half of it's values from either the ascending or descending datasets. The result is that the averaging technique would tend to have a higher correlation with either the ascending or descending datasets than they would have with each other.

The higher correlation between the averaging technique and the Choudhury technique using solely ascending passes is more difficult to explain. Theory would tend to suggest that there should be a higher correlation between the average and descending datasets. Various attempts were made to determine why the average and ascending datasets showed a higher correlation. None were succesful. This is an area for future research.

An ANOVA test was applied to the datasets created by the three compositing techniques. The results proved quite surprising. While the correlations between the various techniques was generally quite high, on the order of 0.85 to 0.95, the probability of having equivalent results among the three techniques according to the ANOVA tests was quite low. Of the 96 possible permutations comparing the three techniques, only 26 permutations showed a probability higher than 5% that the datasets were comparable.

It was expected that the tests would show significantly different means between the different techniques due to the nature of the Choudhury algorithm, which by definition takes the second lowest value for a given pixel. It was hypothesized that the standard deviation would be roughly comparable, but this did not turn out to be the case. The standard deviations were significantly different. One reason for the difference is that by taking only the second lowest value for a given pixel, data with a wider spread of available values was discarded.

One of the most interesting aspects of the difference between MVI datasets created by using different compositing techniques was the evidence that using either morning or evening passes introduces a bias into the analysis. This aspect is a logical area for further studies.

The next approach to studying MVI was to investigate three study sites in California. The three sites were located in the Central Valley near Fresno, in the Coastal Mountains near Santa Ynez, and in the Mojave Desert near Rogers Dry Lake. They represent three forms of vegetation coverage, respectively, agriculture, chaparral/woodland, and desert. Multi-temporal trajectories of MVI, NDVI, and precipitation were constructed for each site.

It was expected that the Central Valley site would show a spring high in MVI followed by a summer low due to agricultural practices. This pattern was apparent. The Santa Ynez site was expected to have a spring low in MVI followed by a summer high due to precipitation induced growth. This pattern was somewhat evident. The Rogers Dry Lake site was expected to show peaks and troughs in MVI values based on individual precipitation events. This pattern was not at all evident. To a large extent the multi-temporal trajectories at the desert site appeared to show no pattern whatsoever.

The final approach to understanding MVI was to detail the relationship between MVI and; water bodies, precipitation, vegetation ground cover and NDVI. Based on the understanding gained by studying these parameters, an effort to predict MVI values given known values for the above four variables was attempted.

Using standard statistical diagnostic techniques on the regression residuals it was apparent that the errors were not normally distributed and the error term did not have a constant variance. Therefore, a log transformation of the MVI dependent variable was used to produce more stable results.

It was expected that water bodies would produce the strongest effect on MVI values. The average correlation between MVI and the water body datasets was approximately 0.20. Realizing that this value may be skewed due to the large amount of observation points containing less than 1% surface water body coverage, a second correlation analysis was tried using only those observation points with more than 1% water body coverage. The correlation increased to circa 0.50. The results suggest that in areas such as the State of California, with relatively small surface coverage by water bodies, other conditions are more important factors in driving the derived MVI signal.

It was expected that the second most important variable affecting MVI values would be precipitation. The mean correlation between precipitation and MVI was -0.375. A second correlation test using only observation points that had precipitation during the time period in question produced a mean value of -0.296.

As vegetation growth lags precipitation by some temporal factor, correlations between MVI and precipitation from both one and two time periods before the MVI reading were checked. Comparing MVI to precipitation in the previous 14 day period the mean correlation increased from -0.375 to -0.394. The mean correlation increased to -0.405 when comparing MVI to precipitation two periods (28 days) in advance of the MVI period.

Precipitation was expected to affect MVI in two ways. MVI values should increase over wetted areas which would show up as a positive correlation between MVI and precipitation in the same time periods. After some period of time the increased water availability produced by precipitation should lead to increased vegetation growth. Therefore, one would expect a negative correlation between MVI and precipitation in previous periods. Conceivably the negative correlation mentioned above between MVI and precipitation in comparable time periods is a function of the gross phenology of California where a large portion of the State is covered by grasses or desert adapted vegetation, both of which tend to be rapidly responsive to changes in water availability. This rapid response would increase surface roughness and thus, in turn, decrease MVI.

It was expected that the relationship between MVI and different types of vegetation coverage would vary depending upon the Leaf Area Index (LAI) and the optical depth properties of the vegetation. The higher the LAI and the greater the optical depth of the vegetation the lower the expected MVI. Five vegetation categories were used; forest, grassland, scrub, shrub and woodland. A sixth catch-all category, predominantly agricultural vegetation, called "other" was also used.

The correlation between MVI values and vegetative ground cover followed the theoretical relationship mentioned above. Forest had an average correlation of approximately -0.53, woodland was next at -0.41, grassland and shrub both showed a correlation of roughly -0.16, and scrub showed a high positive correlation of approximately 0.57. These results support the hypothesis that MVI can be used to discriminate grossly different vegetation types.

NDVI is often used as a surrogate for vegetation growth. The mean correlation between MVI and NDVI was -0.692. Surprisingly, the relationship between MVI

and NDVI appears to be the strongest of the four independent variables. Due to physical theory one would expect NDVI to show primary productivity, while MVI would tend to show surface roughness, which in most cases would be driven by pre-existing vegetative surface structure, e.g. tree trunks etc. The relationship between MVI and NDVI may be biased in this study because so much of California's surface area is covered by grass or scrub which would tend to show a large percentage change in vegetation surface area over a growing season.

An attempt to consolidate the multi-temporal datasets, e.g. MVI, NDVI and precipitation, into one individual variable per dataset was tried. This was done by running a principal components analysis on these three variables.

It was hoped that a principal components analysis on the multi-temporal datasets would help to extract the seasonal aspects of the given datasets. If there was indeed a seasonal signal, one of the three principal components for each of the multi-temporal datasets might help to isolate it.

The results of the principal component analysis did not show a strong temporal affect on MVI. In a normal year the temporal effect would be based on the rainy-winter dry-summer precipitation pattern. The period in question, 1989-90, was a period of significant drought in California. It appears that there was not enough precipitation to make this a worthwhile avenue of inquiry.

Having examined the relationship between MVI and water bodies, precipitation, vegetation surface type, and NDVI the results were used to create a model to predict MVI. The vegetation surface type was broken down into five categories, forest, grassland, scrub, shrub, and woodland. These eight independent variables were used as the starting point for predicting MVI.

The optimal subset of variables to predict MVI was chosen using the adjusted coefficient of multiple determination (adjusted R^2) values of the regression equation. The original adjusted R^2 was 0.668. Iteratively adding and deleting the independent variables led to determining the optimal subset consisted of three variables; forest, woodland and NDVI. These three variables produced an adjusted R^2 of 0.609. In other words more than 91% of the information content of the eight independent variables was contained by these three variables. Several tests to determine whether multicollinearity was a problem with the three variables proved negative.

Two different weightings of the three variables were tried. One was based on an early spring period, while the second was based on a late summer period. The predicted MVI values, using the spring period weightings, produced an average correlation of 0.712 with the actual MVI values. The predicted MVI values, using

the summer weightings, produced a mean correlation of 0.691 with the actual MVI values.

It was hypothesized that water bodies would be the variable that would produce the greatest effect on MVI. While water bodies showed some promise of use as an indicator of MVI values there are not enough large water bodies in California to make a statistically significant statement for this hypothesis. The results do shed light on one aspect of using passive microwave data with a ground footprint the size of SSM/I's footprint for remote sensing of vegetative cover. A 625 km^2 pixel will, in all likelihood, be extremely heterogeneous. Water bodies must make up a fairly significant portion of the pixel if it is to show a difference from surrounding pixels.

Precipitation was expected to be the second most important factor affecting MVI values. One would expect areas with heavy rainfall to originally show a high MVI value, followed by a decreasing MVI as vegetation green-up proceeded. The results of the experiment showed a relatively small effect on MVI values caused by precipitation. Due to lack of significant rainfall, the causal relationship between precipitation and MVI values was degraded, because vegetation growth was constrained.

Vegetation coverage was expected to be the third most important indicator of MVI value. The relationship between MVI and vegetation coverage proved to be stronger than expected. Forest and woodland coverage proved to be two of the three most important predictor variables for determining MVI values.

Plant growth was expected to be the fourth most important indicator of MVI value. NDVI was used as a surrogate for vegetation primary productivity. NDVI turned out to be one of the three most important predictor variables. While NDVI may prove useful in determining MVI this is a somewhat backward approach to developing a passive microwave vegetation index.

The research performed on MVI at the RSRU suggests that the use of MVI for determining vegetation change is a relatively inferior tool, compared to other indices such as NDVI. MVI is able to distinguish between areas of gross vegetation difference, e.g. forested areas versus desert. Due to the size of the sensor footprint, 625 km^2, most pixels will contain a heterogeneous mix of vegetation and, possibly, water coverage.

MVI does appear to show some usefulness for judging absolute variations between vegetation biomass at passive microwave frequencies. The results of this paper point to three main areas of future study.

The first area of future study would look at the possible bias introduced into a study by using solely a morning or an evening pass. The work here suggests that the datasets created solely from either ascending or descending passes are significantly different. Research to calibrate the absolute bias of these composited datasets is necessary to increase the understanding of MVI.

The second area of study would be to use other wavelengths from the SSM/I sensor to aid in the compositing procedure used for MVI. Both the Choudhury method and the MSFC method for compositing MVI values are designed to minimize the effects of cloud coverage on MVI. Using some of the Neale et al. (1990) algorithms that are flags for precipitation or standing water, such as differencing the 85 and 37 GHz vertically polarized signals, would allow the researcher to screen out precursor 37 GHz data that is contaminated by unacceptably high environmental problems such as cloud coverage. This would allow the averaging method to be used with a much higher degree of accuracy in the composited end-product.

The third area of future study would be to determine the relationship between MVI, precipitation and NDVI in California during a year with heavy precipitation, perhaps the winter of 1992-3. By determining the typical response of MVI to California's surface during a rainy year a continuum of MVI values for a Mediterranean regime would be created. This would allow future researchers to determine if future events are producing values out of the expected range. More importantly, by comparing expected multi-temporal trajectories of MVI at various study sites, such as the three used above, with the actual multi-temporal trajectories a future researcher would be able to determine whether MVI does indeed show valid possibilities for long term analysis of the earth's surface. If severe rainfall events, such as occured in California in January 1995, do not significantly affect MVI then the usefulness of the index would have to be questioned.

It might also be quite useful to examine the effectiveness of MVI for discriminating vegetation differences and phenological changes if a new sensor with a ground resolution of, say, 1 km was launched. A serious problem with MVI is the heterogeneity of the pixels studied due to their size. If the sensor's resolution were to dramatically improve it may increase the usefulness of MVI.

MVI currently does not appear to be a useful tool for indexing the phenological changes in a Mediterranean climate during a drought period. It is limited by the coarseness of the sensor footprint to delineating gross vegetative differences. It's major advantage at the current time is that the precursor data, 37 GHz satellite images, are readily available at nominal charge.

MVI does appear capable of discriminating gross vegetation differences. MVI is a relatively new topic of study and, as such, much future research needs to be done to better understand how to best employ MVI as an additional tool in understanding the earth.

References

Becker, F. and Choudhury, B. (1988), Relative Sensitivity of Normalized Difference Vegetation Index (NDVI) and Microwave Polarization Difference Index (MPDI) for Vegetation and Desertification Monitoring, Remote Sensing Environ. 24:297-311.

Box, E. O., Holben, B. N., and Kalb, V., (1989), Accuracy of the AVHRR Vegetation Index as a Predictor of Biomass, Primary Productivity, and Net CO2 Flux, Vegetatio 80: 71-89.

Brady, N. C., (1984), The Nature and Properties of Soils, Macmillan Publishing Company, New York.

Choudhury, B. J., (1988a), Microwave Vegetation Index: A New Long Term Global Data Set For Biospheric Studies, Int. J. Remote Sensing, 9: 185-186.

Choudhury, B. J. (1988), Relating Nimbus-7 37 GHz Data to Global Land-Surface Evaporation, Primary Productivity and the Atmospheric CO_2 Concentration, Int. J. Remote Sensing, V. 9, 1: 169-176.

Choudhury, B. J. (1989), Monitoring Global Land Surface Using Nimbus-7 37 GHz Data: Theory and Examples, Int. J. Remote Sensing, V. 10, 10:1579-1605.

Choudhury, B., Wang, J., Hsu, A., Chien, Y., (1990), Simulated and Observed 37 GHz Emission Over Africa, Int. J. Remote Sensing, 11: 1837-1868.

Choudhury, B. J., Major, E., Smith, E., Becker, F., (1992), Atmospheric Effects on SMMR and SSM/I 37 GHz Polarization Difference Over the Sahel, Int. J. Remote Sensing, V. 13, 18: 3443-3463.

Choudhury, B. and Tucker, C. et al. (1987), Monitoring Vegetation Using Nimbus-7 Scanning Multichannel Microwave Radiometer's Data, Int. J. Remote Sensing, V. 8, 3: 533-538.

Davis, J. C., (1986), Statistics and Data Analysis in Geology, John Wiley and Sons, New York.

DMSP, (1989), DMSP Special Sensor Microwave/Imager Calibration/Validation: Volume I, Naval Research Laboratory, Washington, D. C.

Donley, M. W., Allan, S., Patricia, C., Patton, C., (1979), Atlas of California, Pacific Book Center, Culver City, CA.

Durrenberger, R. W., (1972), Patterns on the Land: Geographical, Historical and Political Maps of California, National Press Books, Palo Alto, CA.

Fung, A. and Ulaby, F. (1983), Chapter 4: Matter-Energy Interaction in the Microwave Region, Manual of Remote Sensing, American Society of Photogrammetry, Falls Church, VA., pp. 115-164.

Henderson-Sellers, A. (1987), Chapter 20: Effects of Change in Land Use on Climate in the Humid Tropics, The Geophysiology of Amazonia, New York, John Wiley & Sons.

Heymsfield, G. and Fulton, R. (1992), Modulation of SSM/I Microwave Soil Radiances by Rainfall, Remote Sens. Environ., 39: 187-202.

Hollinger, J. P., (1990), Introduction, IEEE Trans. Geosci. Remote Sensing, 28: 779.

Hornbeck, D., (1983), California Patterns: A Geographical and Historical Atlas, Mayfield Publishing Company, Mountain View, CA.

International Geosphere Biosphere Programme (IGBP), (1992), Improved Global Data for Land Applications, Townshend, J. (ed), IGBP Global Change Report #20, Intl. Council of Scientific Unions, Stockholm.

Justice, C., Townshend, J., Choudhury, B., (1989), Comparison of AVHRR and SMMR Data for Monitoring Vegetation Phenology on a Continental Scale, Int. J. Remote Sensing, V. 10, 10: 1607-1632.

Kirby, K.N., (1993), Advanced Data Analysis with Systat, Van Nostrand Reinhold, New York.

Montgomery, D., and Peck, E., (1982), Introduction to Linear Regression Analysis, John Wiley & Sons, New York.

Munz, P. A. (1968), A California Flora, Los Angeles, U. C. Press.

Neale, C. et al. (1990), Land-Surface-Type Classification Using Microwave Brightness Temperatures from the Special Sensor Microwave/Imager, IEEE Trans. Geosci. Remote Sensing, V. 28, 5: 829-837.

RESEARCH SUPPORT OF THE WETNET PROGRAM

Final Technical Report for Research Grant # NAG8-929

Section 2

USING THE INVERSE RADIATIVE TRANSFER MODEL FOR TEMPORAL AND SPATIAL ANALYSIS OF SSM/I DATA

Authors:

MICHAEL LAWLESS
JOSEPH SCEPAN

1 October 1995

Remote Sensing Research Unit
Department of Geography
University of California
Santa Barbara, CA 93106-4060 U.S.A.

RSRU

TABLE OF CONTENTS

1. introduction

As a part of the research support of the Wetnet Program, we tested an Inverse Radiative Transfer Model (IRTM) for spatial and temporal analysis of SSM/I (Special Sensor Microwave/Imager) data. This work follows directly the development of an IRTM for use with SSM/I data by G.A. Vassiliades (1993). The model was testing using SSM/I data covering a site in the San Jacquin Valley, California for a 40-day period in 1988.

We expected to find that this IRTM would provide a useful tool in multi-temporal, multi-spatial detection primarily of soil moisture and surface temperature based on the SSM/I Satellite data. Our hope was that a model such as this could be automated to provide accurate and up to date information.

1.1. The SSM/I Sensor

The SSM/I sensor flies aboard the Defense Meteorological Satellite Program (DMSP) Block 5D-2 Spacecraft F8 and was launched June 19th, 1987. This sensor represents a joint effort between the U.S. Navy and Air Force to obtain synoptic maps of atmospheric, oceanographic, and selected land parameters on a global scale.

The Block 5D-2 spacecraft orbits in a sun-synchronous near-polar orbit at an altitude of 833 km, an inclination of 98.8 degrees, and an orbit period of 102.0 minutes. This orbit produces 14.1 full orbit revolutions per day, with a local ascending time of 06:12am. The SSM/I swath width is 1400 km and often results in repetitive coverage on consecutive days.

The SSM/I is a seven channel, four frequency, polarized, passive microwave radiometric sensor. The sensor measures atmospheric/ocean surface brightness temperatures at 19.35, 22.235, 37.0, and 85.5 Ghz. The SSM/I rotates about an axis parallel to the spacecraft vertical at 31.6 rpm and records the upwelling brightness over and angle of 102.4 degrees, thus when viewed from the aft, it scans left to right with an active scene measurement of +/- 51.2 degrees. This results in the swath width of 1400km. Each scan produces 128 discrete uniformly spaced radiometric readings in the two 85.5 GHz channels, and on alternate scans, 64 readings are taken in the other five lower frequency channels. This results in an instantaneous field of view (IFOV) diameter of approximately 12.5km for the 85.5 Ghz channels. The effective fields of view differ for each channel as well as along or cross track and is shown in the Table 1. (Hollinger 1987)

Table 1

Channel Frequency	Polarization	Along-track EFOV (km)	Across-track EFOV (km)
19.35 GHz	Vertical	69km	43km
19.35 GHz	Horizontal	69km	43km
22.235 GHz	Vertical	50km	40km
37.0 GHz	Vertical	37km	28km
37.0 GHz	Horizontal	37km	29km
85.5 GHz	Vertical	15km	13km
85.5 GHz	Horizontal	15km	13km

1.2. Radiative Transfer Modeling

When radiation comes in contact with a body that has dimensions on the order of the wavelength of the radiation, large amounts of random scattering will take place where the radiation is redistributed in all directions, although not necessarily uniformly. This scattered radiation takes two forms, coherent and non-coherent. The coherent component is that which is scattered in the normal direction of reflection. The other radiation component is called diffuse or non-coherent, and is scattered in every direction.

In non-coherent models, the emission is only determined by dielectric properties of the surface of the medium, or each layer of the medium. It is assumed that the other component is reflected back into the soil, and only the non-coherent component is emitted into the atmosphere. The simplest of these models is the Zero-Order Non-Coherent Model. This model ignores all emissions from within the soil medium and only considers the radiation coming from the soil surface, it assumes that all reflected emissions are rescattered by the soil medium. A more complex model is the First-Order Non-Coherent Model by which the soil is stratified into homogeneous layers, and the emissions from the layers are integrated.

The coherent model considers emissions from the entire soil profile. It assumes that the coherent component of the radiation is reradiated from the surface into the atmosphere. It is composed of contributions that are incident upon the surface from below. All the contributions are integrated to give a coherent component.

The SSM/I sensor does not penetrate more than a few millimeters, and thus the complicated coherent model that considers multiple reflections deeper in the soil, is, according to the author of the model, not worth the trouble. Instead, the non-coherent model seems more appropriate where only the surface parameters are considered. (Vassiliades, 1993)

1.3. Inverse Radiative Transfer Model for Analysis of SSM/I Data

Initially, Vassiliades (ibid.) developed a regular Radiative Transfer Model in order to simulate the microwave radiation emitted from the earth's surface, as detected by the SSM/I sensor. This RTM simulates reflectance for any case of mixed spatial footprints in terms of soil type, vegetation type, at any moisture level, and with any fraction of water bodies present. This model was then inverted in order to retrieve surface moisture, temperature, and percent land cover values from SSM/I data recorded over a variety of footprints, using the Parabolic Golden-Section Search algorithm (Adby, 1974).

The purpose of our effort was to determine the usefulness of this model for multi-temporal analysis with spatially explicit data.

2. The Inverse Radiative Transfer Model

Vassiliades' model was written as a partner to the RTM and as a stand-alone executable program. The model was coded in Microsoft's QuickBasic. The model source code was not available to us, nor would it have been exportable to faster platforms, and it was therefore necessary that the program be run on an IBM-PC / compatible computer.

The Inverse-RTM uses three separate sets of input parameters. The first set is a *static set.* The static set defaults to a best-case variable; it is modifiable by the user prior to model iterations, although once these parameters are set, they must be used for the entire run of the data. The second variable set is the set of *output limits*; the user is prompted to give a range over which the output variables can fluctuate. The third variable set is the actual *input data*, fed into the model in the form of a ASCII table.

The initial (static set) parameters are constants and include information on soil particles, bound water, and vegetation. These may be the most important variables in terms of reducing the error factor of the model. They are:

- Soil Particles -
  Permittivity (E_s)
  Loss factor (jE_s)
  Volume fraction (V_s)

- Bound Water -
  Permittivity (E_{bw})
  Loss factor (jE_{bw})
  Volume fraction (V_{bw})

- Vegetation -

Permittivity of bulk (E_b)
Loss factor of bulk (jE_b)
Leaf water content (V_{wl})
Volume of leaves (V_l)
Canopy height (h)

The second set of parameters are the bounding and initial values of the output variables. They ask for a minimum and maximum value within which the output variables must fall. In addition, the model asks for a starting value from which the model will work. These values are:

- Soil Moisture (V_{fw})
- Temperature (T)
- Relative Humidity (RHM)
- Roughness Factor (RoughF)
- Slabs Factor (QF)
- Area Fraction of Bare Soil (A_{smx})
- Area Fraction of Vegetation Canopy (A_{can})
- Area Fraction of Water Bodies (A_{wb})
- Apparent Angle (ApA)
- Lapse Rate (Rt_{Lps})

Once these values are determined, the source values are input. They are entered either manually individually at the terminal, or, should the user wish to run more than one set of data through the model, in batch mode, by file. Since the goal of this project was to use the Inverse-RTM over a large area over a number of days, the file entry method was used.

The input file must strictly follow a standard format. The input file contains each set of parameters, one set per line. These are reported in the following sequence as follows:

1. Latitude specified in decimal degrees

2. Longitude specified in decimal degrees

3. Calendar day

4. Tb19v: the SSM/I brightness temperature for the 19.35 GHz vertically polarized channel;

5. Tb19h: the SSM/I brightness temperature for the 19.35 GHz horizontally polarized channel;

6. Tb22v: the SSM/I brightness temperature for the 22.235 GHz vertically polarized channel;

7. Tb37v: the SSM/I brightness temperature for the 37.0 GHz vertically polarized channel;

8. Tb37h: the SSM/I brightness temperature for the 37.0 GHz horizontally polarized channel;

9. Tb85v: the SSM/I brightness temperature for the 85.5 GHz vertically polarized channel;

10. Tb85h: the SSM/I brightness temperature for the 85.5 GHz horizontally polarized channel;

11. Neale (1990) classification code;

12. MLRA region code: The location code based on the Soil Conservation Services Major Land Resource Area of the United States (MLRA) map;

13. api2: Antecedent Precipitation Index calculated without any correction for maximum depth of water available for evaporation to one decimal place;

14. Precipitation (mm) for that day;

15. Physical Temperature (degrees Kelvin);

16. KR1: API values calculated with the maximum depth set to 5, mm.

17. KR2: API values calculated with the maximum depth set to 7.5mm.

18. KR3: API values calculated with the maximum depth set to 10mm.

19. KR4: API values calculated with the maximum depth set to 15 mm .

20. KR5: are the API values calculated with the maximum depth set to 20 mm.

21. Flag: set to 1 if there is any missing data, and the reported values are estimates;

22. Number of days since last rain.

2.1. Test Site and Dates

A single site was selected for testing the ITRM. Several criteria were used for choosing a test site and test dates. We desired a site located in the state of California. The 69 x 43 kilometer footprint of the SSM/I sensor necessitated a test site that covered tens of thousands of square kilometers and contained a large area of relatively homogeneous land cover, soil types, and terrain.

Three candidate test areas were identified: the Sacramento Valley, the San Joaquin Valley, and the San Bernardino Desert region.

The Sacramento Valley (located north of 37 degrees north latitude) was not ideal, since the rainfall values for the area were not always even, with the northern portion of the area receiving much more rain than the south. The San Joaquin Valley and the San Bernardino Desert were then evaluated as potential study sites.

Ground weather stations were insufficient in number and not uniformly distributed over the San Bernardino Desert area. A precipitation map was generated for the San Bernardino Desert using the kriging method based on weather station locations within the site. This analysis demonstrated that the weather stations within the area that were located in the San Bernardino mountain region distorted the precipitation values higher than was likely for desert regions.

The San Joaquin Valley, on the other hand, had many positive aspects. The area is extensive, topographically flat, and climatically homogeneous; without major precipitation gradations. For these reasons. the San Joaquin Valley was chosen as our test site. The test site area is a polygon bounded by the following Latitude / Longitude points:

- 121W 37N
- 119.5W 37N
- 119W 36N
- 119W 35N
- 119.5W 35N
- 120W 36N

In bounding the temporal domain for this study, it was decided to examine a period of dry days prior to and following a relatively large rain storm. For our purposes, it was necessary to identify a brief rain storm (of a few days duration, at most) which effected the entire test site as evenly as possible and occurred during a tine period for which both SSM/I data and weather station data are available.

Weather station data for the state of California were available for the period prior to 1991 on a single CD-ROM. SSM/I data were available to us covering 1987 and 1988, and 1992 through the present. (EarthInfo 1992)

Initially, a subset of all the weather station sites was extracted from the weather data CD-ROM. This subset included all weather stations that had complete continuous daily coverage of precipitation, and maximum and minimum temperatures that covered the potential study areas: between 119W and 121W, and 35N and 37N (San Jacquin Valley site); and between 114W and 118W, and 34N and 36N (San Bernardino Desert site).

This dataset was then further subset into 1987 and 1988 units. From here, the data were transformed to an ASCII grid format and imported into a spreadsheet software package (Microsoft Excel). Upon inspection, it was determined that the optimal dates for our analysis were prior to and following January 15th, 1988. A rain storm occurred in California on that date with relatively uniform coverage over the potential study areas. The precipitation for this storm lasted from two to three days depending on the location within the site. There was a relatively dry period prior to this date and an extended dry period of at least four weeks followed the storm.

2.2. Data Compilation

The ARC/INFO software package (Environmental Systems Research Institute, Redlands, California) was used for compiling the data layers needed for this project. While not especially well known for its raster-based GIS ability, its functionality for this type of project was considered preferable to alternative available products.

ARC/INFO will import ASCII text files, as well as a number of other raster and vector image file formats and will overlay raster, vector, and point coverages. It is well suited as a programming platform using its ARC Macro Language (AML). ARC/INFO includes the projection functions required for this study, and a raster layer may be interpolated from a point coverage (important for our point weather station sources).

Data compilation was performed in steps as described below.

1) Precipitation and temperature data was exported from the CD-ROM it into a spreadsheet.

2) From the CD-ROM software, the sites were subset by location and by data availability for the period of 1987 and 1988.

3) The appropriate date was determined by querying for precipitation values closely associated with by precipitation free days.

4) Data points identified storm activity days along with the following 40 days were exported into a set of ASCII files that were then imported into ARC/INFO.

5) The data were recompiled into a point coverage; with each weather station point coded by latitude and longitude.

6) The values of precipitation, minimum temperature, and maximum temperature were imported into an INFO file, and attached to the weather station location points coverage.

7) An interpolation (using the kriging technique) was performed to transform these point data to raster data layers. This procedure generated a best-fit raster of 0.1 degree cells. The 0.1 degree raster cell was chosen to accommodate locational uncertainties within the final 0.2 degree cell resolution sampling framework.

8) The values over the 3 day rain storm period for both the San Jacquin Valley and the San Bernardino Desert sites were added and a raster layer was compiled showing total rainfall at each 0.1 degree raster cell for the storm system. This yielded a total precipitation image.

9) Raster layers of minimum and maximum temperatures were created for each day starting on the 15th day of January for 40 days for both sites; creating 160 additional data layers.
10) The Antecedent Precipitation Index (API) was next generated using the equation:

$$API_i = (API_{(i-1)} + P_i)K_i$$

where:

$_i$ = calendar day,

P = precipitation in mm;

K = coefficient determined by;

$$K_i = \exp(E_i / W_m);$$

where;

E_i= potential evaporation on day I, and

W_m = the depth of soil water available for evaporation.

The Thornthwaite method (Vassiliades 1993) was used to estimate E_i. This method required the use of a Heat Index determined by:

I = (mean monthly temp / 5)^1.514

Once I values are known for each month, they are summed for the year to obtain an annual heat index.

11) Data on average minimum and maximum temperatures within California for each month for the past 30 years is archived in RSRU. This dataset exported into ARC/GRID and reprojected from its original lambert-azimuthal projection into the geographic projection.

12) The Heat Index was then determined for each month from the mean of the minimum and maximum averages in Degrees C, and then summed to create the Annual Heat Index, (I).

13) The unadjusted PE was determined based on the Annual Heat Index, and the Mean Monthly Temperature using the Thornthwaite model. The adjusted Potential Evaporation is the PE corrected for the variation in length of daylight depending on Latitude. The actual evaporation is simply the Potential Evaporation adjusted for any precipitation.

14) From the AE, the API was determined for the various depths of moisture availability, 5, 7.5, 10, 15, and 20 millimeters.

15) The Neale classifier code was not determinable from the data we had based on the Neale paper because this paper does not explicitly state how to generate this code. Since this variable is not actually used anywhere in the equation, but is specified simply as a reference point for the researcher. A dummy variable was used in its place.

16) The MLRA code (Major Land Resource Area) was determined from the MLRA handbook and map (1981). This code, as well, is not actually used by the model, but since it was easily accessible, the proper code was used.

The SSM/I datasets were sent by the NASA Wetnet group to RSRU on magneto-optical disks; each disk labeled with the coverage date for easy access. These data are displayed and viewed using McIDAS, the Wetnet SSM/I data accessing software.

The data are formatted in images known as (AREA). The McIDAS software also allows exporting of the image data. Images acquired during the January 15th through February 24th 40 day peried (which included California) were examined for coverage of the test site areas. Some dates or passes were not examined. The SSM/I sensor lacks complete planetary coverage for every date and, infrequently, image data covering California was not acquired. Additionally, images with only partial coverage of the site were not used.

The images were inspected using the McIDAS software and the datasets which included our study area were copied to the RSRU UNIX workstation network. An RSRU software program, MCERDAS (McGwire, 1994), was used to convert the MSS/I image AREA files from McIDAS format into ERDAS format (as .LAN files). The imagery was next georeferenced. Each individual MSS/I image spectral band was extracted, stored as a separate image file, and exported into ARC/GRID coverages. Each grid coverage of the SSM/I data is referenced by the, pass direction, and spectral band. As an example, file 88021ar_l1 refers to 1988, day 21, ascending pass, band 19v. (the first band).

Following georeferencing of all images, spectral values were systematically calculated at each 0.2 degree latitude / longitude point throughout the dataset. An ARC/Info AML was written for this purpose. These spectral values were automatically recorded in a script, or log using the UNIX script command which will capture everything that appears on the screen. A program was written in the C programming language to extract these data and reformat for input in the Inverse Radiative Transfer Model.

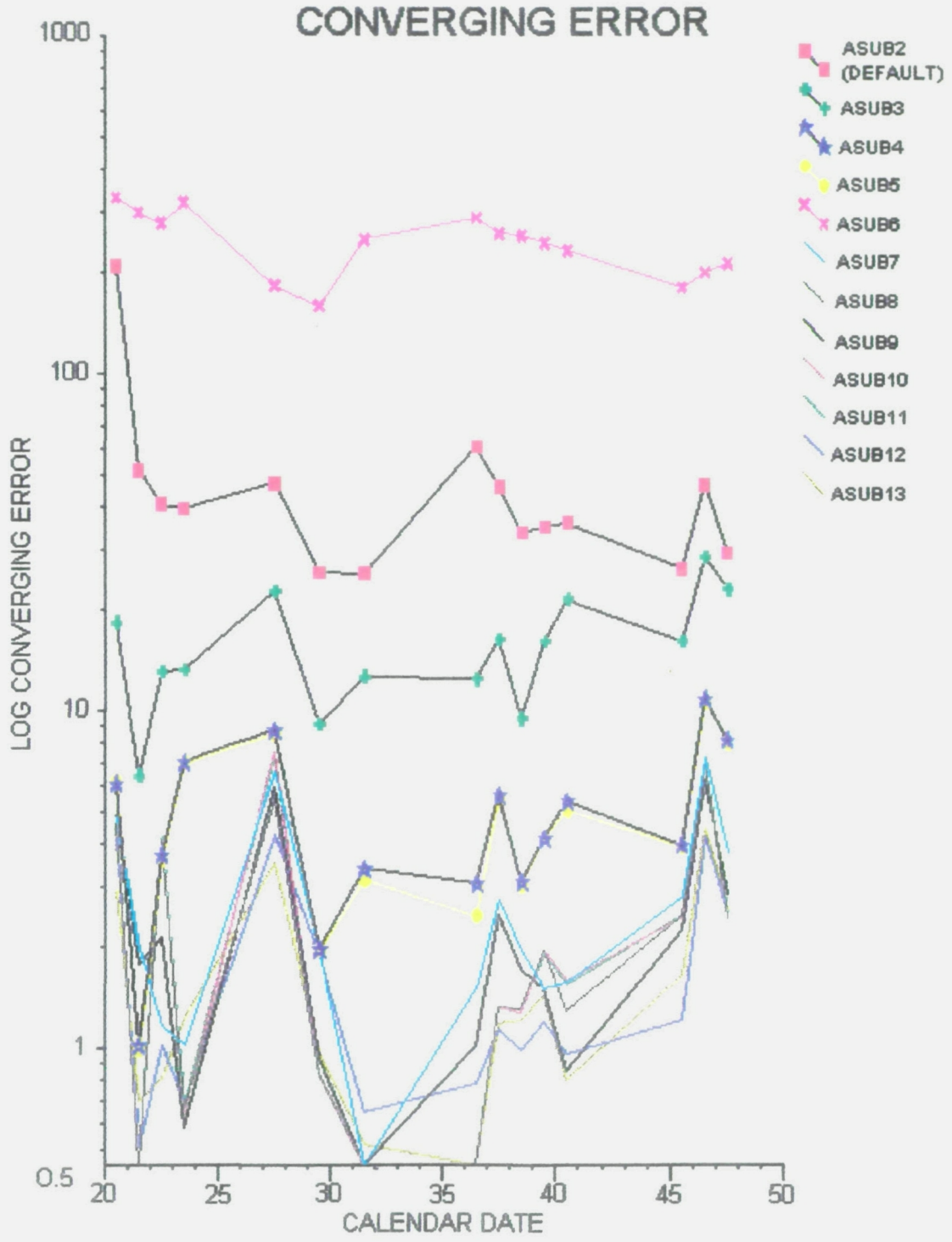
2.3. Model Optimization

Without complete documentation, the IRTM model required some experimenting before it could be used with the final dataset. Due to the time restrictions, it was determined that a subset of the data would be used to optimize the model. It was determined that this subset should be a single randomly selected location, using all the dates available at that location. This turned out to be the point at 35.6 degrees North Latitude, 119.0 degrees West Longitude and contains 15 days worth of data. After a number of runs and attempts to optimize the parameters, it was determined that the afternoon passes would have to be separated from the morning passes as the differences between the two were significant enough that they would have to be run separately.

Using the 6pm subset of our test data, we were able to bring about as close to optimal results as we felt possible within a reasonable amount of time. This process consists of 12 basic steps, or combinations of input parameters. The data was labeled asub1 through asub13, with asub1 being the input parameter, and asub2 through asub13 being the 12 iterations. Each combination of parameters was a modification of the previous one, starting with the default, until there was minimal gain in optimization. There is a converging error factor determined by the model that is displayed with the output, and it was decided that this was the

appropriate determinate for optimization. As the converging error approaches zero, the model is presumed to be more accurate. While it could not be determined how close to zero one had to be in order to have reliable data, it could be shown that data with high converging error showed no correlation with the ancillary data while the data with a low converging error showed a stronger correlation.

Figure 1 shows the converging error (log scale) plotted against the individual days. The parameters used for asub13 had the lowest average converging error, and the highest ground truth correlation. These values were used for the final run of the afternoon data.



2.4. Model Run, Afternoon Dataset

The model was run using the entire 6pm dataset for the test site. Due to some unknown reason, the model would not accept the entire dataset and hold onto the preset parameters, instead resetting to the defaults. Broken up into smaller sectors based on latitude and longitude, this problem was fixed. The mean converging error was 7.61, which is a bit higher than expected based on the subset, but still within the same orders of magnitude. A statistical analysis was done on the output of the IRTM model, and the results were not promising.

Although the converging error had been kept fairly low throughout the run, the correlation between the calculated surface temperature from the model, and the approximate surface temperature from the weather data shows absolutely no correlation. While you would not expect these values to be exact, since the satellite measures surface temperature and the weather information measures air temperature, you would expect these values to be correlated to some extent. In addition to a full dataset correlation matrix, the correlation was determined for all temperature values whose converging error was below 1, 2, 4, 6, and 10. By removing the iterations that had higher error, it was thought that this would leave the most accurate data, and thus show the expected correlation. This was not the case. In addition to direct correlation between the expected and calculated values, it was attempted to determine a correlation between: Converging error and the difference between the two temperature value, converging error and spatial aspects of the data, soil moisture and API, and soil moisture and converging error.

3. Conclusion

As a part of his work developing the IRTM, Vassiliades makes a number of observations and suggestions. We must briefly address two of these before we summarize our own conclusions about our work here.

1. Vassiliades suggests that such aspects such as Roughness Factor, Slab Factor, and apparent angle could be parameterized and estimated using more real world factors such as slope, aspect, degree of undulation, and soil texture and structure. This would most definitely have helped us. It was very difficult to come up with proper estimates of the current variables, and substituting this kind of data would have been much more appropriate, easily attainable, and would presumably have been much more accurate.

2. Vassiliades also states that software improvements would be made to include an on-line help where users can get descriptions of parameters, instructions on how to proceed, and some form of data quality control. This capability would have been enormously helpful in our work. Additionally, some form of software manual (either on-line or hard copy) is required as there is not adequate information about the software itself nor the collection of

model parameters. An online reference for each variable, its significance, and possible ways of attaining such information would be a much added bonus.

Our work here does not demonstrate that this model can be used for spatially and temporally distributed data, even on a relatively similar areas. There are a number of problems and possible solutions.

Our primary problem had to do with the speed at which the model ran. Taking most of a day to run a dozen data points through, it was difficult to find the correct set of parameters by experimentation, even when approximations or reliable values were known. Timely modification of parameters and model optimization was impossible.

Model optimization may have been possible using another programming language, and a compiler optimized for specific machines. Vassiliades' model was coded in Micrsoft Basic. Coding this model in C, C++, or another language would allow the user to take advantage of cross platform compilation, optimized code compilers that would consider CPU type, floating point processors, and possibly the inclusion of assembly language that would speed processing.

Our second problem was the lack of software documentation. While we have provided here some useful information regarding the working of the model, we can provide little explanation of the software or real world application. This emphasis on the internal model (at the expense of the interface) was the largest factors in our somewhat unsatisfactory results.

Although every attempt was made to use correct and reliable model parameters input based on the available documentation, we feel that we cannot definitively rule out the possibility that one or more of these parameters was misinterpreted causing the lack of data correlation.

References

Adby, P.R. and Dempster, M.A. Introduction to optimization methods, 1974 Chapman and Hall, New York.

EarthInfo Inc., National Climatic Data Center Summary of the Day dataset, 1992.

Gerard, B.G. 1990. "Development of Surface Moisture Algorithms using Special Sensor Microwave/Imager (SSM/I) Signatures." thesis submitted in partial fulfillment of the requirements for the degree of Master's of Science in Agriculture and Irrigation Engineering, Utah State University, Logan, Itah.

Hollinger, J., Lo, R.. Poe, G. Savage, R., Peirce, J. 1987 Special Sensor Microwave/Imager user's guide. Naval Research Laboratory, Washington, D.C.

Neale, C.M.U., McFarland, M.J., Chang, K. 1990 Land-surfacetype classification using microwave brightness temperatures from the Special Sensor Microwave/Imager (SSM/I). IEEE Trans. Geosci. Remote Sens. 28(5):829-838.

Soil Conservation Services, 1981. Land Resource Regions and Major Land Areas of the United States (MLRA). United States Department of Agriculture, SCS, Washington, D.C.

Vassiliades, Georghios A. 1993. Development of a Radiative Transfer Model for the Special Sensor Micorwave Imager (SSM/I) and Its Application for Retrieving Surface Moisture and Temperature. Doctoral Dissertation, Utah State University, Logan UT.