

UNIVERSITY OF CALIFORNIA

Santa Barbara

Remote Sensing and Geostatistical Analysis of Anthropogenic Biomass Burning and  
Forest Degradation in Madagascar

A Thesis submitted in partial satisfaction of the  
requirements for the degree Master of Arts  
in Geography

by

Nicholas Joseph Matzke

Committee in charge:

Professor Dar Roberts, Chair

Professor Joel Michaelsen

Professor Phaedon Kyriakidis

December 2003

The thesis of Nicholas Joseph Matzke is approved.

---

Joel Michaelsen

---

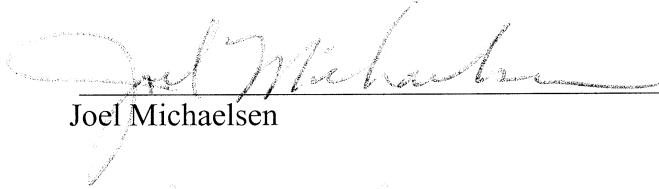
Phaedon Kyriakidis

---

Dar Roberts, Committee Chair

November 2003

The thesis of Nicholas Joseph Matzke is approved.



---

Joel Michaelsen



---

Phaedon Kyriakidis



---

Dar Roberts, Committee Chair

November 2003



GRADUATE DIVISION  
SANTA BARBARA, CALIFORNIA 93106-2070

Telephone: (805) 893-2277  
Facsimile: (805) 893-8259  
<http://www.graddiv.ucsb.edu>

December 10, 2003

To Whom It May Concern:

This is to certify that **Nicholas Joseph Matzke** has successfully completed all the requirements for the degree of Master of Arts in Geography at the University of California at Santa Barbara. The date of degree conferral—December 11, 2003—is reflected on the student's official transcript. In some cases, this letter of degree verification may precede finalization of the transcript. Transcripts may be obtained from the Office of the Registrar, University of California, Santa Barbara, CA 93106.

Sincerely,

A handwritten signature in cursive script that reads "Lynn Wilcoxon".

Lynn Wilcoxon  
Coordinator of Academic Services  
Graduate Division



December 10, 2003

GRADUATE DIVISION  
SANTA BARBARA, CALIFORNIA 93106-2070

Telephone: (805) 893-2277  
Facsimile: (805) 893-8259  
<http://www.graddiv.ucsb.edu>

NICHOLAS JOSEPH MATZKE  
29598 NE WESLINN  
CORVALLIS, OR 97333

Dear Mr. Matzke:

Congratulations on your completion of the requirements for the degree of Master of Arts in Geography. The degree will be awarded as of the end of Fall Quarter 2003 (December 11, 2003).

The degree will be posted to your official UCSB transcript in the near future. If you need letters to certify to employers or credentialing agencies that you have completed all degree requirements, request them from this office, Graduate Academic Services section. If you need official transcripts, order them from the Registrar (\$6.00 per copy) and mark your order "hold for posting of graduate degree, December 2003."

Diplomas take a long time to print. December 2003 diplomas will probably not be available until June 2004. If you have not already done so, please return your diploma request with the appropriate fee to the Cashier's Office. The request will be forwarded to the Registrar's Office; they will mail your diploma as soon as it arrives. Be sure the address you use is one for someone who will know where you are next June.

The Graduate Division staff joins me in congratulating you on your fine achievement, and in wishing you success in your future endeavors.

Sincerely,

A handwritten signature in black ink that reads "Charles N. Li".

Charles N. Li, Dean  
Graduate Division

CNL/lw

Remote Sensing and Geostatistical Analysis of Anthropogenic Biomass Burning  
and Forest Degradation in Madagascar

Copyright © 2003

by

Nicholas Joseph Matzke

## ACKNOWLEDGEMENTS

This work could not have been completed without the help of Jim Rowland and Jim Verdin of the U.S. Geological Survey, Jean-Michel Dufils of PACT Madagascar, Christian Kull, Ned Horning, the teaching and comments of committee members Joel Michaelsen and Phaedon Kyriakidis, and the continual encouragement of family and friends. My advisor Dar Roberts deserves special thanks for his teaching, advice, and his forbearance. This work was funded by a USGS Research Assistantship (1 year) and a NASA Earth System Science fellowship (2 years).

## ABSTRACT

### Remote Sensing and Geostatistical Analysis of Anthropogenic Biomass Burning and Forest Degradation in Madagascar

by

Nicholas Joseph Matzke

Anthropogenic biomass burning is a long-standing concern in Madagascar. Available estimates of annual burned area consist of unsupported but widely cited guesses that 25%-33% of Madagascar burns annually, and unreliable official statistics that put the figure at only about 2.2%. Therefore, an assessment was made of the utility of Defense Meteorological Satellite Program, Operational Linescan System (DMSP-OLS) fire product for prediction of burn area. Burn scars were mapped using Landsat Enhanced Thematic Mapper (ETM) imagery and a coarsened map was compared to a DMSP-derived fire index map. Traditional linear regression indicates low correlation on a per-pixel basis, but the datasets showed extremely skewed distributions and significant auto- and cross-correlation, mandating a geostatistical approach. A linear model of correlation (LMC) shows low correlation between DMSP and burn area at fine spatial scales but increasing correlation at coarser scales. The uncertainty of burn area predictions for polygons of differing

area was estimated using sequential Gaussian simulation of burn area maps, using DMSP as a secondary dataset and the variogram model of the true burn area map as a constraint. For a given time period, 100 simulations were generated and processed to produce mean and standard deviation estimates of burn area for polygons in the following GIS layers: the entire country, level 1 provinces, level 2 provinces, a JERS-derived landcover map, a protected forests map, and the outlines of the Landsat scenes used in the study. The true burn area for the calibration and validation ETM scenes was within one standard deviation of the simulation-estimated mean burn area for each scene. The annual burn area estimates for the entire island were  $7.0 \pm 1.1\%$  and  $6.0 \pm 1.1\%$  for August-December of 1999 and 2000, respectively. Uncertainty relative to the mean increases at finer spatial scales; Madagascar preserves appear to be small enough that annual changes in burn area will not be statistically significant given the large relative uncertainty in the local DMSP-derived burn area estimate. However, as a predictor of burn area at regional (province-level) or national scales, DMSP appears to have potential.

## TABLE OF CONTENTS

Chapter 1: Introduction .....	1
A geographic conflagration.....	1
Summary of study .....	8
Chapter 2: Background .....	11
Geography of fire in Madagascar.....	11
Importance of fire mapping.....	11
Physical geography .....	15
The cultural geography of fire in Madagascar .....	17
Remote Sensing of Fire .....	19
General considerations .....	19
Mappable fire-related variables.....	20
Chapter 3: Methods .....	23
Data sources .....	23
DMSP-OLS .....	23
Landsat 7 Enhanced Thematic Mapper (ETM).....	34
JERS L-band SAR mosaic .....	46
GIS data layers .....	49
Statistical analysis .....	49
Traditional statistical analysis .....	49
Geostatistical analysis .....	50
Chapter 4: Results .....	62

Landsat classification .....	62
JERS classification.....	70
Traditional statistical analysis .....	72
Exploratory data analysis .....	72
Histograms .....	73
Quantile-quantile plots .....	77
Linear regression analysis .....	78
Geostatistical analysis .....	81
Variogram maps and models.....	81
Stochastic simulation .....	87
Chapter 5: Discussion .....	105
Would the DMSP-OLS fire index be expected to correlate with actual fire counts rather than burn area?.....	105
Review of Hypotheses.....	110
Chapter 6: Conclusions .....	116
References .....	121
Appendices.....	133
Appendix A: Review of geostatistical terminology and theory .....	133
Introduction .....	133
Random variables and random functions.....	133
Stationarity .....	135
Semivariograms.....	137

Permissible semivariogram models.....	137
Linear model of regionalization (LMR).....	138
Linear model of coregionalization (LMC).....	140
Structural correlation coefficients .....	141
Simple kriging.....	142
Sequential Gaussian simulation .....	146
Appendix B: Description of programs written for this study.....	150
Appendix C: Fire-related terminology .....	156

## LIST OF FIGURES

Figure 1: Political map of Madagascar. ....	3
Figure 2: Estimated annual area burned in Madagascar, by province, 1968-1997. ....	4
Figure 3: Physical geography of Madagascar. ....	17
Figure 4: No simple theoretical relationship between DMSP-OLS active fire detections and either the total true number of fires or true burn area. ....	22
Figure 5: Three fire-detection maps from the 2000 Madagascar fire season. ....	30
Figure 6: Histogram of the DMSP-OLS fire product DN values for the 1999 fire season. ....	31
Figure 7: DMSP-OLS-derived maps, 8/1/00-10/28/00. ....	33
Figure 8: Flow chart depicting the processing of the Landsat-ETM images in order to produce burnscar maps for comparison to DMSP. ....	35
Figure 9: Footprints and dates of the Landsat ETM+ scenes used in this study. ....	36
Figure 10: Image endmembers used for spectral unmixing of the Landsat ETM reference image. ....	37
Figure 11: Path 158, row 73, October 28, 2000. ....	38
Figure 12: Enlargement of region delineated in Figure 11. ....	40
Figure 13: JERS L-band SAR color composite, 1997. ....	47
Figure 14: Outline of sequential Gaussian simulation setup. ....	58
Figure 15: The simulation procedure. ....	59
Figure 16: Final binary decision tree. ....	64
Figure 17: Land cover classification, path 158, row 73, October 28, 2000. ....	65

Figure 18: Land cover classification, enlargement of region delineated in Figure 11. .....	65
Figure 19: JERS-derived landcover map, and percentages of mapped classes. ....	72
Figure 20: Comparison of fire products at 1 km resolution over the extent of Landsat path 158, row 73.....	73
Figure 21: The distribution of the burn area fractions per 3-km pixel.....	74
Figure 22: The distribution of the DMSP-derived fire index values per 3-km pixel. .....	75
Figure 23: Distribution of DMSP-derived fired index for the entire island of Madagascar, 3 km resolution, August 1, 2000 – October 28, 2000.....	76
Figure 24: Normal probability plots of the DMSP-derived fire index values. ....	77
Figure 25: Unscaled and log-scaled quantile-quantile plots comparing the distribution of the DMSP-derived fire index values over just the extent of Landsat path 158 row 73 to the distribution of the DMSP-derived fire index values for the entire spatial extent of Madagascar DMSP data. ....	78
Figure 26: Histograms of ETM burn area, DMSP fire index, and regressions of DMSP predicting ETM burn area. ....	79
Figure 27: Histograms of ETM burn area, DMSP fire index, and regressions of DMSP predicting ETM burn area. ....	80
Figure 28: Plots of regression statistics (DMSP fire index predicting ETM burn area) as pixel resolution coarsens from 1 km – 65 km. ....	81

Figure 29: Variogram maps and variograms showing anisotropy in spatial correlation in 2-D. . . . .	84
Figure 30: Empirical ( <i>dots</i> ) and modeled ( <i>lines</i> ) variograms for the normal-score transformed datasets. . . . .	86
Figure 31: Example simulated burn area maps for Madagascar. . . . .	88
Figure 32: Example histograms showing the distribution of burn area values on a single simulated realization of a burn area map. . . . .	89
Figure 33: Variogram reproduction. . . . .	90
Figure 34: Distribution of whole-country percent burn area estimates from 100 simulated burn area maps for the 2000 fire season. . . . .	90
Figure 35: Percent contribution of each JERS-mapped landcover (Figure 19) to total burn area for 2000 fire season. . . . .	93
Figure 36: Negative correlation of JERS-mapped forest and DMSP fire index. . . . .	96
Figure 37: Monthly DMSP-estimated area burned, by landcover class. . . . .	97
Figure 38: Relative contribution of each level 1 province to total Madagascar burn area in 2000. . . . .	99
Figure 39: Distribution of percentage burn areas for 133 Madagascar districts in 2000. . . . .	100
Figure 40: Estimated monthly fraction burn area for Madagascar districts, 1999 and 2000. . . . .	101
Figure 41: Estimated burn area for protected areas polygons in the AIRE_PRO database. . . . .	102

Figure 42: Distribution of burn area estimates for the year 2000 fire season for 280 protected areas polygons. ....	103
Figure 43: Z-score map testing for significant differences in burn area in protected areas polygons between 1999 and 2000. ....	103
Figure 44: Comparison of actual burn counts, actual burn area fraction, and DMSP burn count maps. ....	108
Figure 45: Coefficient of Variation in burn area plotted against polygon area for all polygons used in the study. ....	114

## LIST OF TABLES

Table 1: Madagascar government statistics of annual percent area burned, by Madagascar Level 1 province. ....	6
Table 2: Linear regression equations and r-squared values for intercalibration of images to the reference image. ....	41
Table 3: Descriptions of mapped landcover classes. ....	43
Table 4: Number of nights and observations for the time periods over which burn area was simulated. ....	60
Table 5: Area of mapped landcover classes for the entire Landsat image. ....	66
Table 6: Confusion matrix for decision tree classification. ....	69
Table 7: Confusion matrix for six classes only. ....	69
Table 8: Confusion matrix for 144 GPS-located ground reference points. ....	69
Table 9: Confusion matrix for classification of JERS dataset. ....	71
Table 10: Comparison of primary rainforest area derived from JERS classification to other published estimates. ....	71
Table 11: Parameters for Linear Model of Correlation variogram structures, and calculated of structural correlation coefficients. ....	85
Table 12: Parameters for LMC structures, and calculated of structural correlation coefficients, for normal-score transformed datasets. ....	87
Table 13: Burn areas estimated from the DMSP-based simulations, compared to the reference burn areas as mapped by classification of Landsat ETM image pairs. .....	92

Table 14: Total area burned for Madagascar, estimated from 100 DMSP-based simulations, for the DMSP observation periods in 1999 and 2000.....	93
Table 15: Total area burned by JERS-classified landcover, estimated from 100 DMSP-based simulations, for the DMSP observation periods in 1999 and 2000. ....	95
Table 16: Estimated annual burn area for 1999 and 2000, by level 1 province. ....	99
Table 17: Top ten protected areas polygons for 1999 and 2000 as ranked by estimated burn area. ....	104

## **Chapter 1: Introduction**

### **A geographic conflagration**

It is difficult to determine which is more persistent in Madagascar: raging fires, or the raging controversy about fires. Fires have been common at least since the arrival of humans on the island ca 1500 years ago. Since precolonial times, it has been recognized that anthropogenic fires are a prime factor contributing to the environmental destruction that has occurred on Madagascar, but burning is also considered a useful tool and a right by many rural Malagasy farmers and herders (*tantsaha*). There is thus a long history of conflict in Madagascar between central authorities and *tantsaha* over burning, a conflict that has only increased with the recent influx of the money and recommendations of western conservation organizations. The on-the-ground debate has been paralleled by academic dispute, with the conventional wisdom advocating an almost completely anti-fire viewpoint, but with several geographers recently presenting a much more complex picture.

However, reliable numbers have yet to become part of this debate. How much area is burning? The most commonly cited figure comes from primatologist Alison Jolly. Jolly (1980), in a discussion lamenting the effects of fire in Madagascar, puts the figure at one-fourth; in Jolly (1990) the annual countrywide burn area is given a range of one-fourth to one-third. Similar figures are repeated in popular literature on Madagascar, e.g. Greenway (1997) puts the number at 25%-30% without citing a source, and Tyson (2000) cites Jolly putting the number at one-third, although

without citing a publication. Jolly's estimates lack support as she gives no quantitative or even qualitative data.

The most recent assessment of fire in Madagascar comes from geographer Christian Kull whose (2000) Ph.D. thesis *Isle of Fire* explores in detail the historical, cultural, political, and ecological facets of the topic of fire in Madagascar. Kull compiled all available government statistics on Madagascar provinces (Table 1) on the question of annual area burned. As reviewed by Kull in a 2002 article in *Environment*, these estimates do not have a high degree of certainty:

The precise area burned each year remains unknown...Government statistics report annual burned areas varying between 0.5 and 5 percent of the island's surface. In highland Antananarivo province, where grasses dominate, these statistics show annual burned areas ranging from 0.5 to 43% of provincial surface area. Unfortunately, these statistics are unreliable and probably underestimate the actual extent of fires. (Kull 2002)

The statistics in question are shown in Figure 2 and Table 1. The percentages given here have been calculated based on the area burned statistics given by Kull. Kull compiled them from government sources, mostly the annual reports of the Madagascar forest service (*Service des Eaux et Forêts* or *Direction Eaux et Forêts*, or DEF),

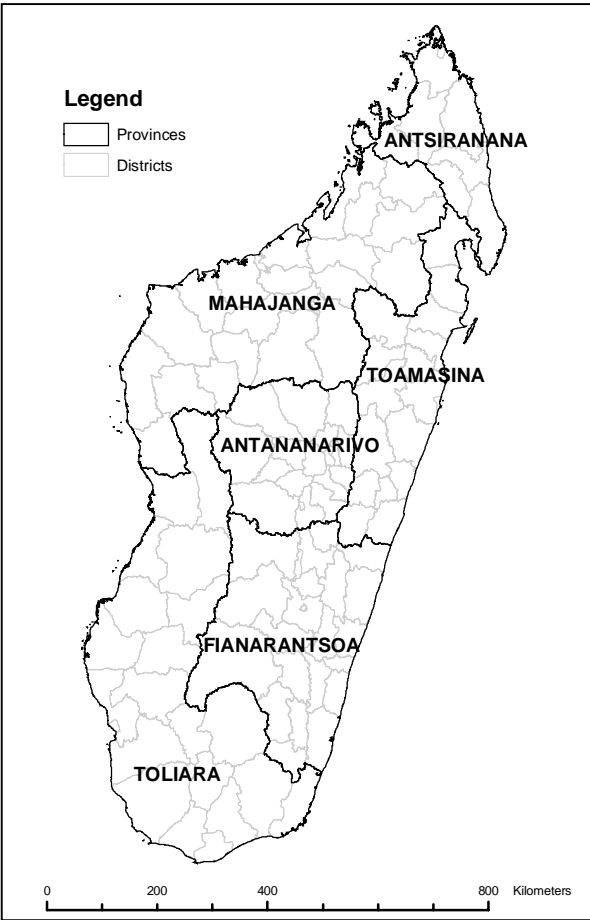


Figure 1: Political map of Madagascar

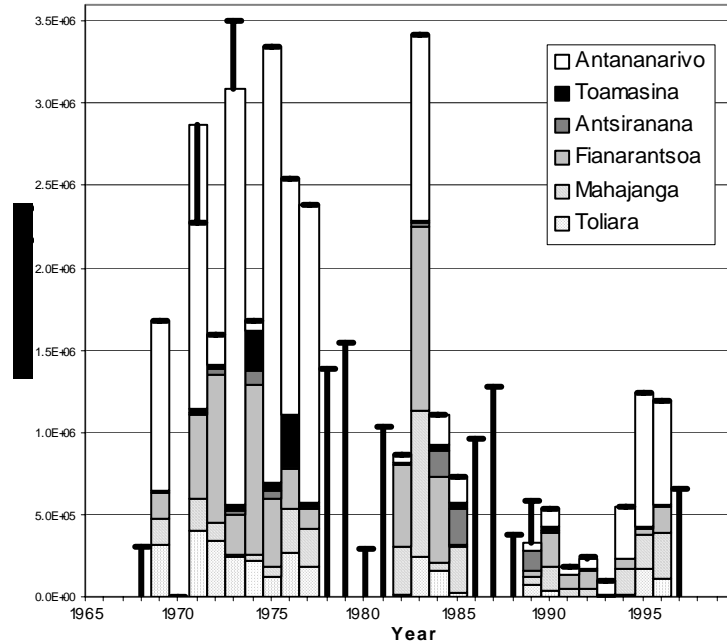


Figure 2: Estimated annual area burned in Madagascar, by province, 1968-1997. Data from Madagascar government statistics assembled by Kull (2000). The single error bar represents the annual area burned for years when no province-level statistics were available, or the difference between the published national estimate and the sum of the available province-level data. The difference is sometimes substantial. Total area of Madagascar = 58,700,000 ha.

although sometimes from citations of these reports when the originals were not available. The government reports themselves are theoretically based on field reports from forest service agents, although Kull writes, “[i]t is unclear exactly how the figures are arrived at in each forest agent’s territory”. Some figures result from the sum of burn areas documented in citations for illegal burning and authorized burns, while others are based on field surveys, which may be somewhat rigorous or may be “eyeball” estimates.

The reports of field agents are also heavily influenced by political events, for example a six-month general strike in 1991 resulted in a very low estimate for that

year as the agents were not working during the strike. Kull also lists statistics on national area burned; from Table 1 it can be seen that for some years, only national statistics were available, while in other years the total of the estimates from the provinces yields a different figure than the national estimate. Kull cites several examples where different sources give wildly different burn area estimates for a given year.

Table 1: Madagascar government statistics of annual percent area burned, by Madagascar Level 1 province (locations shown in Figure 1). After Kull (2000), Appendix A. As Kull emphasizes repeatedly these statistics are plagued by inconsistent and/or inappropriate measurement procedures, are “absolutely unreliable” and should be treated with extreme caution. The mean and standard deviation of the available statistics are calculated in order to gauge multi-annual variability in estimates. The highest percent burn area ever reported is also listed; as the primary problem with these statistics is considered to be burn area underestimation, the largest values may represent years when the data were most complete.

Year	Antsiranana (Diego Suarez)	Fianarantsoa	Mahajanga (Majunga)	Toamasina (Tamatave)	Antananarivo (Tananarive)	Toliara (Tulear)	TOTAL	Kull (2000) Island-wide Totals
1968								0.5
1969	0.1	1.5	1.0	0.1	17.7	2.0	2.8	2.8
1970								
1971	0.4	5.0	1.3	0.3	29.4	2.4	4.9	3.8
1972	1.0	8.9	0.7	0.3	3.2	2.1	2.7	2.7
1973	0.5	2.4	0.1	0.5	43.0	1.5	5.2	5.9
1974	2.2	10.2	0.3	3.3	1.1	1.3	2.8	2.8
1975	1.1	4.1	0.4	0.7	45.2	0.8	5.7	5.7
1976	0.4	2.3	1.8	4.5	24.4	1.7	4.3	4.3
1977	0.4	1.2	1.6	0.3	30.8	1.1	4.0	4.0
1978								2.3
1979								2.6
1980								0.5
1981								1.8
1982	<i>na</i>	4.9	1.9	0.3	0.8	0.0	1.5	1.5
1983	0.6	11.2	5.8	0.1	19.4	1.5	5.8	5.8
1984	3.6	5.3	0.3	0.4	3.2	1.0	1.9	1.9
1985	4.8	0.2	1.8	0.5	2.9	0.1	1.2	1.2
1986								1.6
1987								2.2
1988								0.6
1989	2.8	0.3	0.4	0.0	0.8	0.4	0.6	1.0
1990	0.1	2.0	1.0	0.4	1.8	0.2	0.9	0.9
1991	<i>na</i>	0.9	0.3	0.0	0.8	0.0	0.3	0.3
1992	0.3	1.1	0.3	0.1	1.0	0.0	0.4	0.4
1993	0.1	0.0	0.0	0.0	1.5	<i>na</i>	0.2	0.2
1994	0.1	0.5	1.1	0.1	5.4	0.1	0.9	0.9
1995	0.1	0.4	1.3	0.0	13.9	1.0	2.1	2.1
1996	0.2	1.6	1.8	0.1	10.9	0.7	2.0	2.0
1997								1.1
<b>Mean of available statistics</b>	<b>1.0 ± 1.4</b>	<b>3.2 ± 3.4</b>	<b>1.2 ± 1.3</b>	<b>0.6 ± 1.2</b>	<b>12.9 ± 14.6</b>	<b>0.9 ± 0.8</b>	<b>2.5 ± 1.9</b>	<b>2.2 ± 1.7</b>
<b>Highest recorded burn area</b>	<b>4.8</b>	<b>11.2</b>	<b>5.8</b>	<b>4.5</b>	<b>45.2</b>	<b>2.4</b>	<b>5.8</b>	<b>5.9</b>

At best these statistics can only yield an order-of-magnitude sense of annual burn area. But they do represent the baseline of present knowledge about burn area in Madagascar and thus should be kept in mind when the results of the present study are given. Finally, despite the fact that the statistics are acknowledged to be unreliable and incomplete by the DEF itself, according to Kull (2000) the “statistics are the only ones in existence, and they continue to be widely and uncritically used in documents and reports.” Kull did map burns in his primary study-site in 1998 (a 18.45 km<sup>2</sup> region located approximately on the border of Antananarivo and Fianarantsoa), reporting 38% burned (54% of pasture areas, 37% of *tapia* woodlands, and 16% of intensively-managed areas). For his part, Kull reserves judgment on a national burn area estimate, but for the highland grasslands in which he worked (western Antananarivo and Fianarantsoa provinces) he estimates the annual area burned as approximately 50%.

The lack of reliable data is not unique to Madagascar, but is endemic throughout Africa; for example, Lock (1998) calculates an estimate of annual biomass burned in Africa based on White’s 1983 vegetation map of Africa (Huxley *et al.* 1998). But Lock is forced to guess (“as a working assumption”) that 75% of grassland area burns annually across Africa. Even in the best of situations, obtaining accurate estimates of burn area, pattern, return interval, and other fire-related variables is not a simple matter. For example, in the chaparral of southern California, a region where resources, technology, and data are not in short supply, Minnich (2001) and Keeley (2001a; 2001b) are conducting a vociferous debate over fire history and

policy that is in significant part a dispute about geographic datasets and data analysis. Much of the difficulty is due to the nature of fire. Fires are extremely variable in time and space, and the nature of the variability changes radically depending on the scale of analysis. Active fires are a very short term, ephemeral phenomenon, and even the signals left after a fire decay relatively rapidly.

### **Summary of study**

Remote sensing provides the only likely possibility for improving the estimates of burn area in Madagascar and other developing countries. Remote sensing of fires and burn area now has an extensive literature (e.g., Levine 1996), much of it focused on AVHRR-derived datasets (e.g., Chuvieco 1999). However, the continuing emergence of new fire products has created a need for a common methodology for quantitatively assessing the utility of these products, and for comparing and synthesizing fire-related information from disparate sources. The goal is to take *data*, interpretable by remote-sensing specialists, and produce *maps* of well-understood variables (such as burn area), with well-defined estimates of uncertainty, that can be utilized by users from a wide range of backgrounds. This work is particularly crucial if the new fire sensors and products are going to be applied where they are most needed, in the developing world where conservation and development needs are most intense, and where resources are scarce.

An example of such a study is presented here. One of the longest-operating orbital sensors capable of fire detection is the U. S. Air Force Defense Meteorological

Satellite Program – Operational Linescan System (henceforth DMSP-OLS). DMSP-OLS is essentially a military weather sensor, but the ability of DMSP-OLS to detect fires (as well as city lights, gas flares, and lightning) has been recognized since the 1970s (Elvidge *et al.* 2001a). DMSP-OLS data in digital form became available in 1992, and Elvidge and colleagues, working at the National Geophysical Data Center (NGDC) of the US National Oceanic and Atmospheric Administration (NOAA), have produced a fire product for Madagascar in addition to many other areas (NGDC 1998). However, there is only a very limited literature assessing the accuracy of the DMSP-OLS fire product for quantitative estimates of burn area (Elvidge *et al.* 2001a, 2001b; Fuller and Fulk 2000).

The present study compares the DMSP-OLS fire product to a reference dataset of mapped burn scars derived from Landsat 7 image pairs acquired during the 1999 and 2000 fire seasons in Madagascar. Geostatistical methods (Deutsch and Journel 1998; Goovaerts 1997) are used to assess the spatial structure of the two fire datasets, and to quantify the cross-correlation of the two fire datasets at varying spatial scales. Building on this work, stochastic simulation is used to develop equiprobable alternative realizations of burn area maps for Madagascar, using the DMSP-OLS product as the input dataset. A Geographic Information System is then used to process the alternative realizations to generate mean burn area estimates, along with standard deviations, for the whole island and for provinces, preserves, and land cover.

Several hypotheses are tested in this study: (1) that the DMSP-derived fire product exhibits significant ability to predict a reference burn area map derived via classification of Landsat; (2) that this capability will vary at different spatial scales, depending on the spatial auto- and cross-correlation of the DMSP and reference datasets; (3) that a DMSP-derived fire product will be useful for estimating burn area and changes in burn area over different levels of interest, including nationally, the province and district level, national landcover types, and preserves.

## **Chapter 2: Background**

### **Geography of fire in Madagascar**

#### **Importance of fire mapping**

Accurate measures of fire in Madagascar are important for many reasons:

*Biodiversity:* Among western academics, Madagascar has been recognized as a biologically unique country since the first European explorers visited it. French explorer Phillippe de Commerson commented in 1771,

Nature seems to have retreated there into a private sanctuary, where she could work on different models from any she has used elsewhere. There you meet bizarre and marvelous forms at every step... (Kew Research Botanical Garden 2000)

Growing concern about biodiversity loss worldwide (e.g., Wilson 1985) has focused international attention on the island as a biodiversity hotspot (Conservation International 2000). Madagascar's unusual biology results from a number of converging factors: 1) Madagascar's large size (Green and Sussman 1990); 2) Latitudinal extent; 3) Physical diversity (Battistini and Richard-Vindard 1972); 4) The Gondwanaland-derived flora and fauna (Schatz 1996); 5) Extreme length of isolation from Africa (~120 million years) and India (~90 million years) (Scotese 2000); and 6) Long-term climatic stability in or near the tropics (Patterson *et al.* 1995). As a result, the island contains unparalleled levels of endemism in terms of species, genera (80-90%, Kew Research Botanical Garden 2000), and to a lesser

extent families. As the continental African rainforests are species-poor relative to South American and Asian rainforests, Madagascar, with only 2% of the area of Africa, contains approximately 25% of Africa's plant species, along with at least 66 endemic terrestrial mammal species and numerous other endemic vertebrates and invertebrates (Battistini and Richard-Vindard 1972; Ganzhorn *et al.* 2001). This is despite a rash of extinctions of insular megafauna and the almost complete clearing of the western deciduous forest by some combination of intentional and unintentional burning by the original immigrants in the 1,000 or so years before European contact (Battistini and Richard-Vindard 1972).

Due to the unique evolutionary and geological history of the island, Madagascar's native ecosystems are far more vulnerable to fire than fire-adapted continental African ecosystems. Although early scientific accounts of human-induced ecological change -- arguing that the original primary deciduous forest of the western part of the island was unbroken and undisturbed prior to human contact, and that this forest was converted wholesale in an almost catastrophic manner into the species-poor grasslands of today -- have been criticized as overly simplistic, recent research appears to support the general idea of significant conversion of the Western plateau's dry forest/woodland/shrubland mosaic to the currently ubiquitous secondary grassland (for discussion see Guillaumet 1984; Jolly *et al.* 1984; Lowry *et al.* 1997; Paulian 1984) which can only be used (although it is used intensively) as low-grade pasture by the Malagasy. As most of this change occurred before Western contact, this is one of the most dramatic examples of the possible impact of a

technologically modest culture (in this case, the immigrant iron-age culture) on land-cover change at landscape scales (Pyne 1997). Remaining endemic biodiversity is focused in the remnant rainforests of the rugged eastern mountains, of which only about 17% remains (Green and Sussman 1990; Kew Research Botanical Garden 2000; Sussman *et al.* 1994). Anthropogenic burning in the form of *tavy* (the local term for slash-and-burn activity) remains an oft-cited threat to the rainforests. There is, however, also significant concern about remnant patches of dry deciduous forests on the western plateau and coasts, and about the spiny forest in the south (Ganzhorn *et al.* 1999; 2001; Goodman and Rakotondravony 2000), both of which are vulnerable to cutting for fuel wood and (mostly in the northwest) grassland fires spreading into forest edges.

*Erosion:* A further impact of fire is the contribution of deforestation to massive soil erosion from the Madagascar highlands. Huge gullies known as *lavaka* scar the landscape. One author (Bourgeat 1972) estimated the rate of soil loss in the highlands at 3,000 tons per km<sup>2</sup> during the rainy season. Gade (1996) calls deforestation the "key cause" of erosion. Offshore sediment plumes have been observed from space (NASA 2001), with unknown impacts on offshore ecosystems. Continuing deforestation and re-burning of secondary vegetation is one of the factors priming the landscape for erosion in Madagascar (Wells 1995) and elsewhere in the tropics (Dwyer *et al.* 1998).

*Biogeochemistry:* Africa is the source of 20-30% of anthropogenic biomass burning worldwide (Justice *et al.* 1996), and global tropical deforestation is estimated to have been responsible for 20-30% (Watson *et al.* 1997) of annual global carbon emissions. Anthropogenic biomass burning as a whole is estimated to produce about 40% of anthropogenic CO<sub>2</sub> emissions annually (Fuller 2000; Levine 1996). Many other trace gasses are also produced (e.g., CO, CH<sub>4</sub>, N<sub>2</sub>O, HCN, CH<sub>3</sub>CN, CH<sub>3</sub>Cl, and assorted hydrocarbons) although there is much uncertainty in the estimates of trace gas fluxes due to fire (Crutzen and Andreae 1990; Dwyer *et al.* 1998). Although Madagascar is just a small part of this larger picture, the prevalence of both tropical deforestation and large-scale anthropogenic burning in Madagascar make it an excellent site for the testing of remote sensing and statistical techniques for measuring these processes. In addition, the many conservation efforts in the country provide an opportunity to study the effect of interventions.

*Impacts on humans:* The impacts of fire-related ecosystem degradation and deforestation on the people of Madagascar have been severe as well. The near-complete loss of the western deciduous woodlands in pre-colonial times, and the loss of most of the eastern rainforest during the 20<sup>th</sup> century, has had a cascade of impacts. While more modest erosion rates may have helped support valley-floor rice farming in the past, Gade (1996) argues that current erosion and deposition of coarser material is choking productivity in some areas, and that the denuded landscape contributes to alternating water excess and deficiency, resulting in flooding due to rapid runoff and subsequent premature drying of springs as the soil

fails to hold as much rain. The silting of major estuaries has required the relocation of a port in one case (NASA 2001). The loss of wood for cooking and heating has required many to rely on more expensive imports from plantations. Unknown opportunities are being lost along with the island's biodiversity; the Rosy Periwinkle (*Catharanthus roseus*) is the classic example of an invaluable endemic plant that is the source of a highly successful treatment for childhood leukemia (NWF 2000), but it and other endemic organisms are threatened by ecosystem change on Madagascar. Gade (1996) attributes one-half of the poverty of Madagascar to the impoverished resource base.

### **Physical geography**

Madagascar is the fourth largest island in the world with an area of approximately 587,000 km<sup>2</sup> and a length of approximately 1600 km. It is almost entirely located in the tropics, ranging from latitude 12° S to 26° S. The dominant topographic feature is a mountain chain running the length of the eastern side of the island (Figure 3). Latitude and topography divide Madagascar into three major geographic regions: the eastern slopes, the high plateau and plains to the west of the mountains, and the arid south. The east coast and mountains intercept the prevailing trade winds and are therefore much wetter due to orographic effects (1000 to >2000 mm/year precipitation). Evergreen tropical rainforest originally occupied the entire eastern slope and coast. The drier plateau to the west of the mountains has distinct wet and dry seasons, and originally supported a mosaic of tropical savanna and dry deciduous forest, and is currently mostly grassland. The south is semi-arid, with some regions

as dry as 350 mm/year, and supports a unique ecosystem of succulent and drought-resistant species, the spiny forest.

In Madagascar, current anthropogenic burning occurs in two main modes: the drier side of the island, west of the main mountain chain, experiences burning of the savanna and grassland from April-September, with maximum frequency near the end of the dry season (Randriambelo *et al.* 1998). On the moist eastern slope, burning is mostly restricted to a brief dry season in October and November, during the annual decline in trade winds and the concomitant Foehn effect (Donque 1972). Eastern-slope burning consists primarily of field re-burning but includes a significant component of rainforest clearing (*tavy*). It appears to be widely accepted that most fires in Madagascar are started by humans rather than lightning. This would conform with global picture of fires in the tropics (Dwyer *et al.* 1998; Stott 2000) and in Africa (Lock 1998). However, there is some evidence for fires predating human occupation (Kull 2000) and maps of global lightning distribution (GHCC 2001) indicate that lightning is not a rare phenomenon in Madagascar and should not be ignored as a potential source of ignitions, although this would be most applicable specifically in tropical dry forest and savanna (Middleton *et al.* 1997; Roberts 2000).

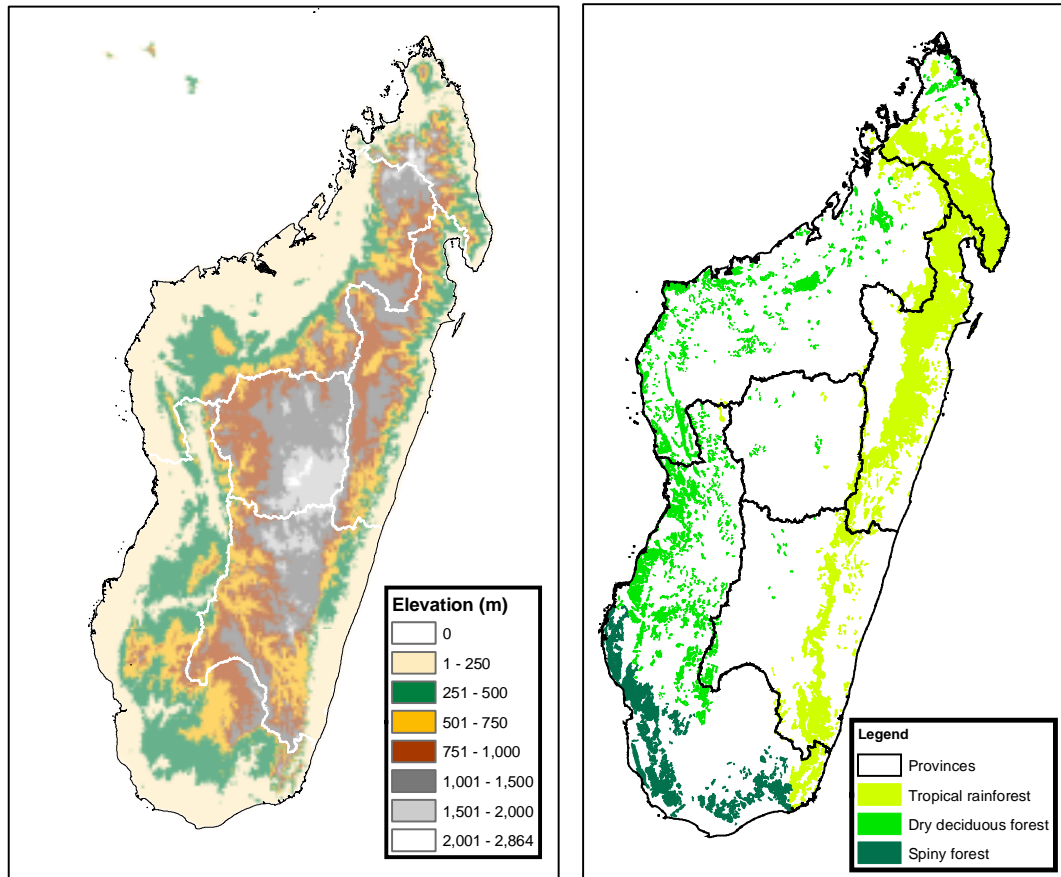


Figure 3: Physical geography of Madagascar. *Left*: Digital elevation model, 1 km resolution (data source: USGS GTOPO30 DEM). *Right*: Simplified version of Faramala's (1981) "Remaining Primary Vegetation" of Madagascar. Forest subclasses have been aggregated into the three major classes. GIS data made available by the Kew Royal Botanic Gardens at <http://www.rbkew.org.uk/gis/projects/madagascar/>. Data reproduced with the permission of the Trustees of the Royal Botanic Gardens, Kew.

### The cultural geography of fire in Madagascar

Any discussion of anthropogenic fire in Madagascar must take into account the complexity of the historical and cultural dynamics present in the country. The preceding discussion, following the physical science and conservation literature, focuses on the negative impacts of fire on the environment in Madagascar, and identifies it as one of the major foes that conservation efforts must combat. This represents the "conventional wisdom," associated with conservation groups and

western media (as well as many conservation-minded Malagasy), and is well-represented in professional and popular literature (e.g., BBC 2002; Goldammer *et al.* 1996; Hoeltgen 1994; Jolly 1980; Pyne 1997; Terborgh 1992; Tyson 2000; Yirdaw 1996). Jolly (1980), for example, entitles a chapter “Suicide by Fire” after a phrase used by a Malagasy botanist who is also quoted as saying, “We are laying waste our own lands...Every year the land burns over and over again.” Similarly, Jolly *et al.* (1984) writes an impassioned plea for conservation in Madagascar, lest the world “let the libraries and museums of nature burn.”

This kind of evocative imagery has likely done much good in terms of attracting Western support for conservation and development efforts. Recent work by geographers, however, has criticized the standard academic/conservationist view of fire in Madagascar as over simple (Jarosz 1993; Kull 1996, 1998a, 1998b, 2002). The conclusion of Kull’s work is that “there has always been fire in Madagascar and there always should be” (Kull 2002), and that the only way to achieve sustainable conservation and a well-regulated fire regime would be to involve local people in the management process, something that neither the central government nor conservation groups considered until recently. Kull (2002) however notes that the situation is beginning to improve, with the devolution of power to local governments and the increasing emphasis of conservation groups on community-based conservation projects such as GELOSE (Kull 2000).

## **Remote Sensing of Fire**

### **General considerations**

Numerous sensors have some capability to detect active fires. Radiance emitted by an active fire depends on the temperature of the fire. The Planck equation describes how emission varies as a function of temperature and wavelength (Elachi 1987).

The temperature of forest fires has been estimated to range from a minimum temperature of 570 K (the minimum temperature for flaming combustion) through approximately 1800 K for heavy fuels (Robinson 1991). Wien's displacement law specifies the wavelength of maximum emission as a function of temperature; for the temperature range of wildfires this corresponds to a spectral range between about 2 and 6  $\mu\text{m}$ , the middle infrared (MIR), and as a result this is often considered the best wavelength range for fire detection. Several space-born sensors, though not designed specifically for fire detection, have channels in the MIR and have been exploited for fire detection. The most work has been done with AVHRR (beginning with Dozier 1981; see Dwyer *et al.* 1998 and Martin *et al.* 1999 for recent reviews).

Other sensors used for active fire detection in addition to DMSP-OLS and AVHRR include the geostationary weather satellite GOES (Alfaro *et al.* 1999; Menzel and Prins 1996; Prins and Menzel 1994; Weaver *et al.* 1995) and the recently launched MODIS (Owens 2001). Meteosat can detect smoke plumes, but not active fires (Fuller 2000). Sensors used for fire scar detection include AVHRR, SPOT (Eva and Lambin 1998b), Landsat, ATSR (ESA 2001, 2002; Eva and Lambin 1998a), and

MODIS. The field of fire remote sensing is reviewed by Levine *et al.* (1996) and Fuller (2000).

### **Mappable fire-related variables**

There are several fire-related variables which have the potential to be remotely sensed with a space born sensor sensitive to VNIR emissions. Care must be taken with terminology, as there are several potential variables which might be directly measured in some fashion by an active fire detector like DMSP-OLS, namely:

*Active fire intensity.* Fires can be hotter or cooler, releasing the chemical energy stored in vegetation at a faster or slower rate. Fires produce radiant energy, measured in radiance ( $\text{W m}^{-2} \text{sr}^{-1} \mu\text{m}^{-1}$ ). If burning vegetation is assumed to act as a black body, then it would be expected that an increase in fire temperature would result in a shorter wavelength of maximum radiance due to Wien's displacement law. For any given wavelength, emittance (emitted radiance) will also increase with temperature according to Planck's equation (Elachi 1987, pp. 114-116). Thus, for two fires where areal extent and other variables are equal, the hotter fire should be brighter, emitting more radiance at a given wavelength.

*Active fire size.* Emitted radiance mixes linearly within a pixel, meaning that radiance detected from a pixel would double if a fire of constant temperature were to double in areal extent. Large, low intensity fires will therefore be indistinguishable from smaller, high-intensity fires at a particular wavelength. If multiple wavelengths are available then the effects of temperature and size can be distinguished. The

percentage of a pixel occupied by active fire can be unmixed if the temperatures of the background and of the fire are independently known, but this is not usually the case.

*Number of fires.* Within a single pixel, a single fire cannot be distinguished from multiple fires; having more fires is equivalent to having a single fire of larger extent, as above. However, the number of separate pixels containing at least one active fire in a given scene could easily be estimated, and this would presumably have some relationship with the actual number of fires burning at the time of overpass.

In practice, distinguishing the influence of these different variables with the single-band, coarse resolution DMSP-OLS sensor is impossible. Furthermore, the main biophysical variable of interest is none of the above, but burn area, which can be measure by mapping burn scars with Landsat. Burn area does not necessarily correlate well with the above active fire variables. The expected relationship of the DMSP-OLS fire detections to actual mapped fire scars should therefore be carefully considered. At first glance, it would appear that a DMSP-derived cumulative fire count map would correlate best with the actual fire count on the ground. Every fire active during an overpass would have a certain chance of being detected, based on cloud cover, brightness, size, etc. However, careful consideration of the many complications involved in active fire detection (see Discussion) leads to the expectation that an active fire detection map would have a better correlation of

DMSP fire count with *burned ground area* rather than *number of individually mapped fire scars*. See Figure 4 for a simple illustration.

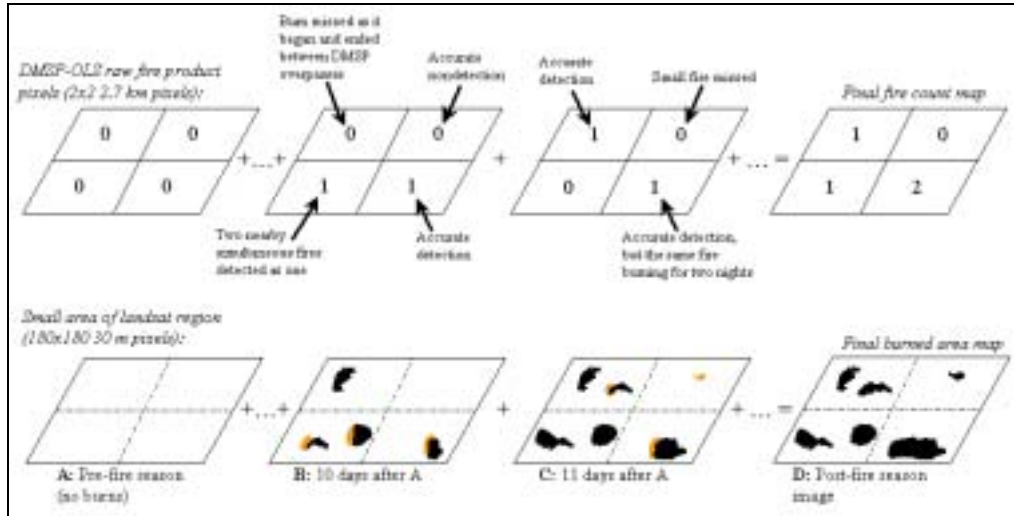


Figure 4: A simple illustration of the fact that there is no simple theoretical relationship between DMSP-OLS active fire detections and either the total true number of fires or true burn area. Some of the many possible situations that would cause fires to be missed, counted twice, etc., are shown. However, if large fires are easier to detect and more often burn for multiple nights, then some correlation between DMSP active fire detections and reference burn area can be expected, as ends up being the case in this example (compare final maps). Similar factors would apply to other coarse resolution active fire detection products. See discussion for a more detailed analysis.

## **Chapter 3: Methods**

### **Data sources**

#### **DMSP-OLS**

The Defense Meteorological Satellite Program, Operational Linescan System (DMSP-OLS) is primarily designed for use as a cloud imaging military weather satellite. The OLS sensor detects numerous nighttime sources of visible-near infrared (VNIR) radiation, including city lights, gas flares, lightning, moonlit clouds, and glare from the sun, as well as fires (Elvidge *et al.* 2001b). Cities in particular present large, stable targets that are easy to study quantitatively with the OLS, resulting in a large literature on the mapping of population, energy use, light pollution, economic activity and environmental impact (Doll *et al.* 2000; Elvidge *et al.* 1997, 2000; Imhoff *et al.* 1997a; 1997b; NGDC 1998; Sutton *et al.* 2001).

The OLS sensor is an oscillating scan radiometer and collects data day and night in two channels, VNIR and TIR. The VNIR channel is sensitive to the wavelength region from 580-910 nm (Fuller 2000). The TIR channel spans 10.3-12.9  $\mu\text{m}$  but saturates at 310 K, and is therefore useful for mapping clouds but has no real utility for fire mapping (Elvidge *et al.* 1998b, but see Kihn 1996). The Defense Meteorological Satellite Program keeps two satellites in sun-synchronous, nearly polar orbits, one in a dawn-dusk orbit, the other in a day-night orbit (Elvidge *et al.* 1998b). The current satellites are designated F-13 and F-12, respectively (Prasad *et al.* 2002). The orbital altitude is  $\sim 833$  km. The sensor swath is  $\sim 3000$  km and the

satellite makes 14 102-minute orbits per day, allowing daily global coverage (Prasad *et al.* 2002).

The VNIR channel can be used for fire detection. While this wavelength region is useless for fire detection during the daytime due to the reflection of solar radiation from the surface, at night only active radiation sources will be producing significant emission in these wavelengths, and thus a sensitive sensor has the potential to detect radiation sources far below pixel size. A primary advantage of DMSP-OLS over other coarse resolution sensors is its high sensitivity to low radiances, being able to measure quantities as low as  $10^{-9} \text{ W cm}^{-2} \text{ sr}^{-1} \mu\text{m}^{-1}$  (Elvidge *et al.* 2001a). However, the OLS radiometric resolution is limited to 6 bits, the digital numbers (DNs) ranging from 0 to 63. Thus the sensor has a significant potential to be saturated by bright emission sources. DMSP-OLS also has relatively coarse resolution: data can be collected in a “fine” mode with 0.56 km resolution, but most data are collected in “smooth” mode, averaging 25 fine pixels into a coarse 2.7 km pixel. An additional complication is that the gain on the DMSP-OLS VNIR channel is variable and is changed according to the phases of the moon. The gain is raised during new moon conditions, resulting in higher sensitivity to fires. Under a full moon the gain is lowered, resulting in lower sensitivity to fires and a greater chance of confusion of fire signal with moonlight reflected off of clouds, water bodies, or snow (Elvidge *et al.* 1998b). This precludes the use of an absolute DN scale to estimate fire brightness or to identify fires via an unchanging DN threshold.

*The NOAA-Elvidge fire product*

The DMSP product is hosted online at NOAA's National Geophysical Data Center website (<http://www.ngdc.noaa.gov/dmsp/dmsp.html>). The method by which this fire product was produced from the raw data was devised and published by Elvidge and colleagues (Elvidge *et al.* 1996, 1998a, 2000, 2001b). First, clouds are identified on the basis of thermal infrared (TIR) band thresholds (cloud tops usually being colder than the ground). Sun glare is removed by flagging pixels that are near to the local dawn/dusk portions of the planet or that are part of a large region of saturated 'lit' pixels. Remaining VNIR light sources are then identified. Elvidge *et al.* (2000) note that variability between observations (in gain, orbital geometry, sun and moon position, etc.) results in brightness variations that preclude the setting of a single DN threshold to identify VNIR signals. To overcome this, a contextual thresholding algorithm is employed: a local background is defined using a 50 x 50 block of pixels, and then lights in a central 20 x 20 pixel block rising 5 DNs above this background level are retained. Of the VNIR light source detections, stable light sources were screened out by tagging pixels in which light was detected in 10% or more of cloud-free observations. Lightning illuminating a cloud is recorded as a linear feature (as the linescan sensor gathers pixel observations in a line) and were identified manually. Remaining lit pixels are active fire detections. Two DMSP satellites are currently operating, so the number of observations available for a particular night varies from 0-2, usually 1. The result is an approximately nightly product for fire observations. A cloud product is generated as a companion to each fire product.

The DMSP fire product is produced as a ~1 km product (based on a 30 arc second reference grid), although the raw data are typically the 2.7 km-resolution "smoothed" acquired data. Due to the low radiometric resolution of OLS and the large number of factors that can influence VNIR at-sensor radiance from nighttime fires, DMSP fire mapping studies to date (e.g., Elvidge *et al.* 2001a) employ a thresholding technique to gauge fire activity. DN values equal to or above the threshold are considered detections. A fire count map is then developed by counting the number of active fire detections occurring at each particular pixel over the time series of interest.

#### *DMSP-OLS Fire Mapping*

Although the potential for remote sensing of fires with DMSP-OLS has been known for nearly 30 years (Croft 1973, 1978), serious systematic study of fires with DMSP began with Cahoon *et al.* (1992). Elvidge and colleagues (Elvidge and Baugh 1996; Elvidge *et al.* 1996, 2000, 2001b) developed DMSP-OLS as a practical tool for remote sensing of fires, although most of the assessments of fire-mapping capability have been relatively qualitative. Quantitative work attempting to correlate DMSP-OLS to ground burn area estimates and to other fire sensors has only recently begun.

Scholes *et al.* (1996a) argue that in general, maps of active fires are not good surrogates for burn area without calibration against high-resolution fire scar datasets. Many studies have examined active fire mapping with coarse-resolution sensors, but most of them produce maps of fire activity or fire detection frequency, and few attempt to produce quantitative maps of actual ground variables such as burn area.

Such studies are useful for mapping seasonal changes in the distribution of burning and for monitoring relative fire activity on a year-to-year basis (until the sensor goes offline), but have limited biogeochemical or ecological utility. Due to the sensitivity and low saturation limits of sensors, many active fire products appear to show burning occurring everywhere, resulting in large overestimates (Fuller 2000; Kihn 1996).

Quantitative assessments of coarse-resolution active fire products have been published, although the results are mixed. Pereira and Setzer (1996) were able to confirm 74% and 36% of AVHRR fire detections in Brazilian savanna with Landsat TM. Scholes *et al.* (1996a) reported successful calibration of AVHRR to Landsat MSS, by using a spatially varying calibration factor that declined exponentially with mean annual precipitation, such that the influence of the numerous, but small, fires in wet regions on estimated burn area was reduced relative to the scarcer but larger fires in drier areas. Scholes *et al.* then used this burn area map to estimate trace gas fluxes. Eva and Lambin (1998b), on the other hand, reported no significant relationship between AVHRR fire count and SPOT-mapped burn area in central Africa. These studies of coarse resolution active fire products should not be confused with coarse resolution burn scar products derived from e.g. AVHRR or ATSR-1, which have had more consistent success, at least for large fire scars (see e.g. Fuller 2000; Fuller and Fulk 2001; Pereira *et al.* 1999).

Quantitative work assessing DMSP-OLS correlation with burn area or other sensors is just beginning, with the studies published to date being rather limited. Fuller and Fulk (2000) compared AVHRR and DMSP-OLS-derived fire products in Indonesia and reported a 77% correspondence. Ehrlich *et al.* (1997) used both DMSP-OLS and AVHRR to gauge relative fire activity finding a weak correlation between fire activity and land cover change, but reported that DMSP and AVHRR derived fire maps had less than 10% of their fires in common. Kihn (1996), using one night of data, identified fire pixels using the OLS VNIR channel via a method similar to that of Elvidge, and also identified fire pixels with the TIR channel, counting pixels more than one standard deviation above the local background mean. If the entire pixel was assumed to be on fire, then both of these methods grossly overestimated burn area, especially the VNIR channel due to the smoothing effect described above. Kihn explored the TIR channel further, using the assumption that all fire pixels containing fires would be saturated, and obtained subpixel estimates of burn area that correlated reasonably well with actual burn area (84% of variance explained) even though just a single night of data was used. Elvidge *et al.* (1998a), on the other hand, did not attempt to map burn area, but did compare the active fire detection capability of GOES, AVHRR, and DMSP-OLS in southwest New Mexico in June 1996, and reported fire detection success rates of 22.2% for GOES, 37.5% for AVHRR, and 66.7% for DMSP-OLS. Elvidge *et al.* did report that the smallest fires that DMSP-OLS detected in the study were ~2 hectares. In contrast to Kihn, Elvidge *et al.* reported that while the TIR channel did show spikes in regions of active fires,

similar spikes were seen in areas that were not burning. Elvidge *et al.* (2001a) mapped burn severity in Roraima, Brazil using Landsat. They then selected a DMSP fire count threshold that would provide a conservative estimate of burn area. The count threshold was selected under the constraint of having a commission error with the Landsat-mapped unburnt forest of less than 10%, and minimized omission error of missed burn areas under that constraint. Using a minimum fire count threshold of five, they produced a DMSP-derived map that accurately mapped 79% of the heavily burn area. They attributed most of the commission error (misclassifying unburned forest as burned) to the very large footprint of the DMSP-OLS smooth data relative to Landsat, and argued that over large areas this effect would be minimal.

In conclusion, while there is abundant evidence that DMSP-OLS correlates in some fashion with actual active fires and with burn area, there is no standard method for extrapolating one from the other, and it is clear that simply assuming that a 2.7 km pixel that has detected a fire has been completely burned is a faulty assumption that will lead to massive overestimates of burn area. A standard method for assessing DMSP-OLS, as well as other coarse resolution fire products, would be a large improvement in the situation – particularly when very recent studies of trace gas emissions from anthropogenic fires are basing estimates on the assumption that a DMSP-OLS pixel with a fire detection has been completely burned (Prasad *et al.* 2002, p. 2845).

*Production of a fire index map*

The raw DMSP-OLS fire product and cloud product for the 1999 and 2000 fire seasons were downloaded from NOAA. The NOAA fire product

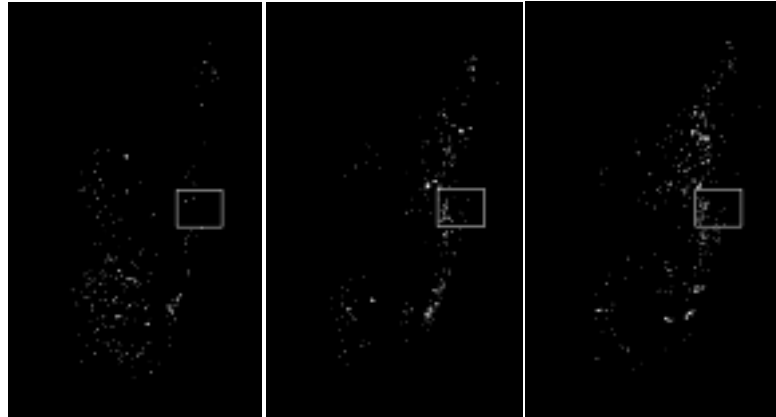


Figure 5: Three fire-detection maps from the 2000 Madagascar fire season. Dates: October 25, 26, 27. The white box represents the approximate extent of the Landsat reference scene.

for Madagascar spans the fire season, covering the five months (153 nights) from July 31-December 31. The initial format of each observation was a 1-km raster in the Geographic projection. 142 and 169 observations were available for 1999 and 2000, respectively (see Table 4). For each year, the raw fire and cloud observations were assembled into multiband time series. Each time series (essentially a data cube) was projected into Hotine Oblique Mercator (HOM), the standard projection for Madagascar. Fire count and cloud count maps were generated from the time series through thresholding each band and summing the number of detections over the time span of interest.

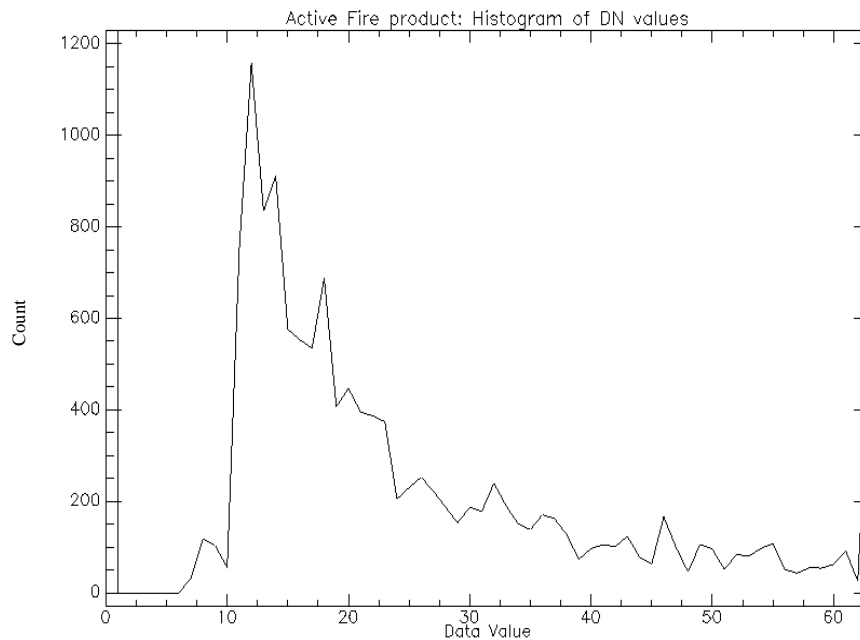


Figure 6: Histogram of the DMSP-OLS fire product DN values for the 1999 fire season. The cutoff at DN=5 is due to the contextual thresholding algorithm that generates the NOAA product. The frequency spike at DN=63 is due to saturation of the sensor.

The decision regarding the appropriate DN threshold to use for counting fire and cloud detections was based on the histograms of DN frequency for all bands in the time series. The DN frequency histogram for the year 1999 fire season time series is shown in Figure 6.

Although a number of DN thresholds were experimented with for generating fire count maps, the final threshold employed was  $DN \geq 5$  so that all fires detected by the NOAA processing algorithm were counted in the fire count map (Figure 7). For the cloud count map (Figure 7), the DN frequency threshold was set at  $DN \geq 160$  as this corresponded with vast majority of obvious clouds in the cloud product.

To account for the screening influence of clouds on DMSP-OLS active fire detections, a *fire index* map was generated from the fire and cloud count maps. The DMSP fire index has previously been used to gauge relative fire activity in Madagascar (Rowland, personal communication 2001; Kull 2002). A fire index takes into account the fact that, for a given grid location  $\mathbf{u}$ , some nights of observation will have been obscured by cloud cover. The simple fire index is therefore:

$$x_{si}(\mathbf{u}) = a(\mathbf{u})/[t - b(\mathbf{u})] \quad (1)$$

where:

$x_{si}(\mathbf{u})$  = simple fire index at grid location  $\mathbf{u}$ .

$a(\mathbf{u})$  = number of detected fires over the time period of interest

$b(\mathbf{u})$  = number of detected clouds over the time period of interest

$t$  = total number of observations over the time period of interest

This fire index will vary between 0 and 1. However, this index is time-independent: a value of  $x_{si}(\mathbf{u})=0.5$  could mean that one fire was detected over two cloud-free observations, or it could mean that 50 fires were detected over 100 cloud-free observations. This product can be assessed for correlation with a burned-area map produced for a given time period, but any transformation of the simple fire index (such as a linear regression) will not be applicable to a simple fire index generated for a different length of time (e.g. monthly vs. entire-fire-season fire maps, or different-length calibration periods). An improved fire index was the "cumulative" fire index, denoted  $x_{cu}(\mathbf{u})$ . It is calculated as follows:

$$x_{\text{cu}}(\mathbf{u}) = a(\mathbf{u}) / [T - c(\mathbf{u})] \quad (2)$$

where:

$x_{\text{cu}}(\mathbf{u})$  = cumulative fire index at grid location  $\mathbf{u}$ .

$a(\mathbf{u})$  = number of detected fires over the time period of interest

$c(\mathbf{u})$  = number of detected clouds for the entire fire season

$T$  = total number of observations for the entire fire season

Thus, for a given grid location  $\mathbf{u}$ , the number of observed fires  $a(\mathbf{u})$  will usually change depending on the time span over which observations are counted; however, the denominator  $T - c(\mathbf{u})$  is based on the entire time series of cloud products for a fire season, and therefore remains invariant. This index has the advantages of being additive and comparable between time periods.

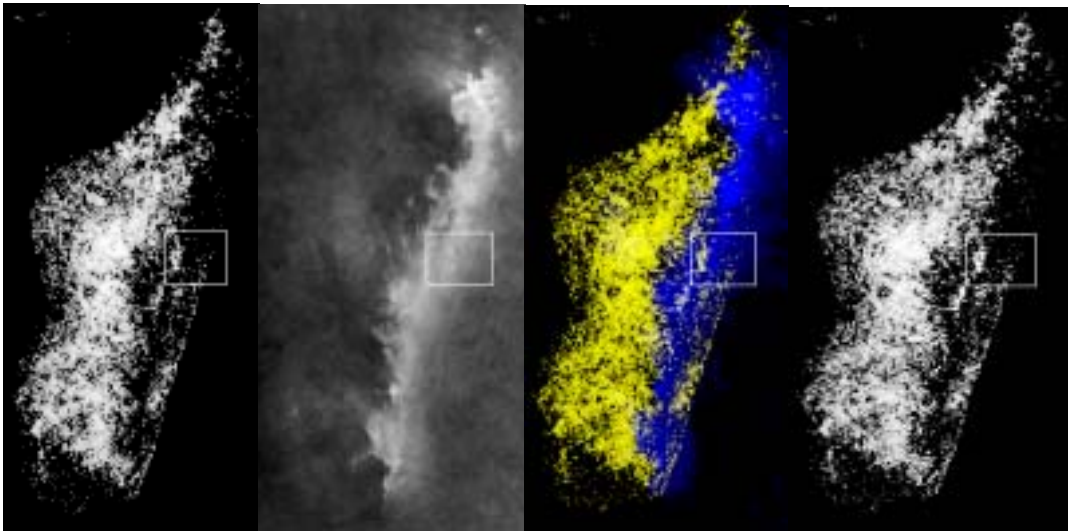


Figure 7: DMSP-OLS-derived maps, 8/1/00-10/28/00. From left to right: Cumulative fire count map; cumulative cloud count map; color composite of fire count (yellow) and cloud count (blue); and cumulative fire index map. The cumulative fire index serves to weight fire detections in cloudy regions more than in clear regions. The white box represents the approximate extent of the Landsat reference image.

The cumulative fire index was generated from the fire and cloud products. Cumulative fire index maps were produced for each of the five months in the 1999 and 2000 fire seasons, for the total fire seasons of 1999 and 2000, and for the three periods of the fire season spanned by the ETM image pairs. All five cumulative fire index (henceforth, just “fire index”) maps were generated at 1 km resolution. The region of the DMSP fire index map corresponding to the reference ETM dataset (path 158, row 73) was extracted for statistical comparison to ETM burn area. These subset maps were then coarsened (via averaging) to produce 3 km, 5 km, and coarser resolution fire index maps. The whole-island DMSP fire index maps used as inputs for geostatistical analysis and simulation were coarsened to 3 km, as the resolution approximated the 2.7 km resolution of the original raw DMSP imagery, and as the coarser datasets eased storage and processing requirements for geostatistical simulations.

### **Landsat 7 Enhanced Thematic Mapper (ETM)**

The Landsat program now has a thirty-year record of being used for landcover mapping and studies of tropical deforestation. Landsat time series are also commonly used for mapping fire scars in the tropics (e.g., Cochrane *et al.* 1999). Imagery from Landsat 7 Enhanced Thematic Mapper (ETM) became available in 1999. In order to assess the utility of DMSP-OLS for fire mapping, Landsat image pairs bracketing portions of the Madagascar fire season were acquired and classified in order to map fresh burn scars. The classification procedure is outlined in Figure 8.

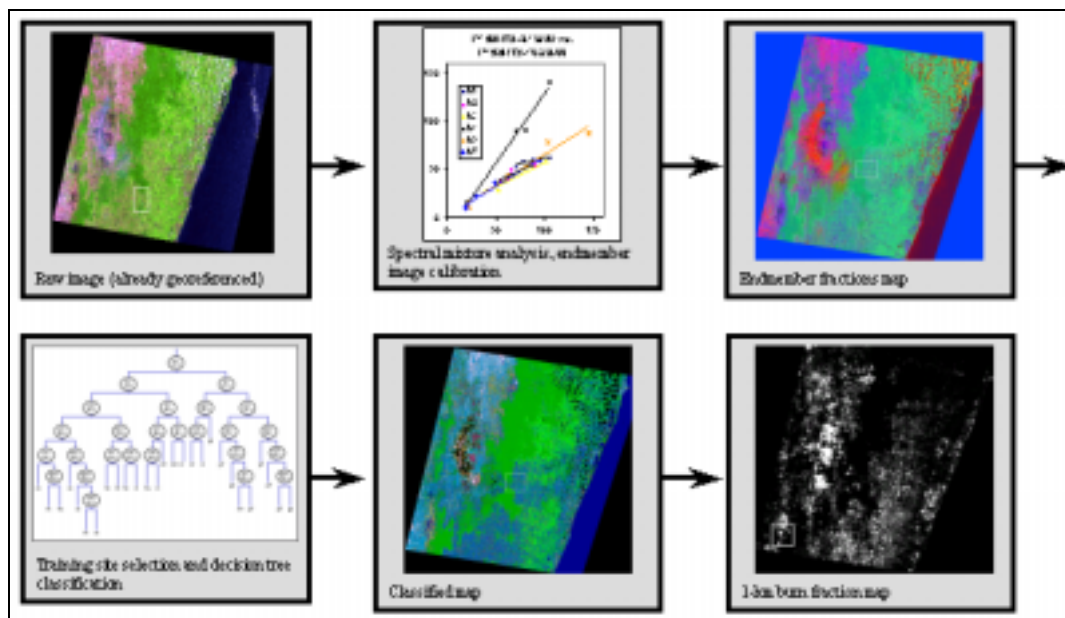


Figure 8: Flow chart depicting the processing of the Landsat-ETM images in order to produce burnscar maps for comparison to DMSP. The image shown is path 158 row 73, 10/28/00, and this was designated as the reference image for intercalibration and training purposes. The image endmembers for the other five Landsat images were regressed against the endmembers for the reference image, producing endmember fractions maps that were intercalibrated with the reference image, resulting in fractions that were directly comparable and that could all be classified with the same decision tree. New mapped burnscars were then extracted, and a coarsened image was produced where each 1 km pixel is assigned a fraction burn area.

#### *Landsat image pair selection*

In order to bracket the Madagascar fire season, pairs of early and late Landsat ETM images were acquired. Three image pairs were selected on the basis of the following criteria:

- Relative lack of cloud cover
- Availability of a pair of images from that calendar year
- Diversity of land cover types (spanning both sides of the main mountain range)

- Evidence of ongoing fire activity (fire scars and smoke plumes could be seen in even in many preview images)
- Areas of conservation interest (preserves in the tropical rainforests)

The six selected scenes are shown in Figure 9 and represent the only image pairs matching the criterion along the cloudy east coast of Madagascar in 1999 and 2000. Of the three image pairs, path 158, row 73 had the most fire activity as well as the largest range of burn scar sizes. Thus, this became the primary image pair used as the reference ('truth') image for comparison to DMSP-OLS/ETM. The other two image pairs were held in reserve for validation purposes. The image classification was also developed primarily with training sites taken from path 158, row 73.

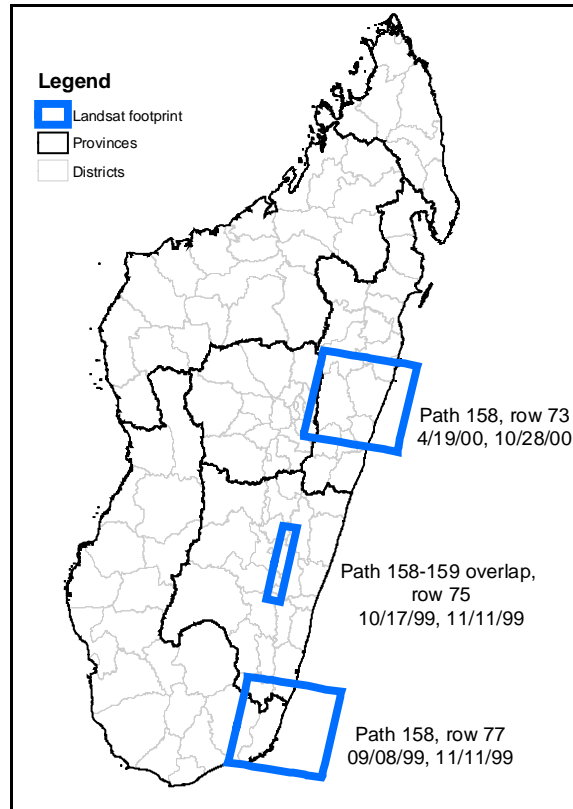


Figure 9: Footprints and dates of the Landsat ETM+ scenes used in this study. Path 158, row 73 was the reference dataset, the other two scenes were used for validation.

### Registration

ENVI 3.5 (RSI, 2001) was used to process all remote-sensing derived datasets in this study. All Landsat images were georeferenced according to the original metadata (Laborde projection) and were then projected into HOM. For the path 158-159 overlap image pair, both images were cropped to include just the portion of overlap, and the October 17 image was warped via control points to match the November 11 scene. All image pairs, when overlain, appeared to be coregistered to within 1 pixel.

### Spectral mixture analysis

Spectral mixture analysis

(SMA) is a

multidimensional analysis

technique wherein the

spectrum of every pixel in

an image is “unmixed”

into fractions of user-

specified spectral

endmembers (described in

Adams *et al.* 1995). In

certain applications, this procedure itself can produce a map of interest, as

endmember fractions may represent areal fractions of vegetation types, species, or

exposed rock or soil, which can be related to biophysical variables (e.g., Riaño *et al.*

2002). Spectral mixture analysis is not a required step for producing a remotely-

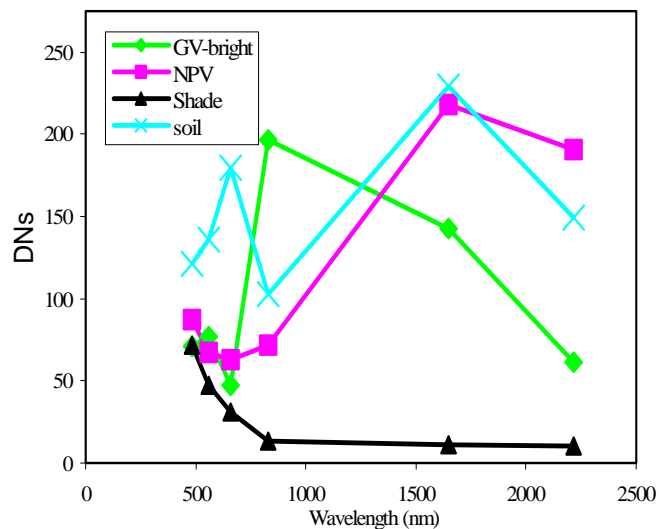


Figure 10: Image endmembers used for unmixing the Landsat ETM reference image.

sensed map, as an automated classifier can easily be applied to raw Landsat imagery. However, conducting spectral mixture analysis prior to classification serves several purposes, as in one step it provides a method to reduce dimensionality of the dataset, intercalibrates images from multiple dates and locations, and improves image interpretability.

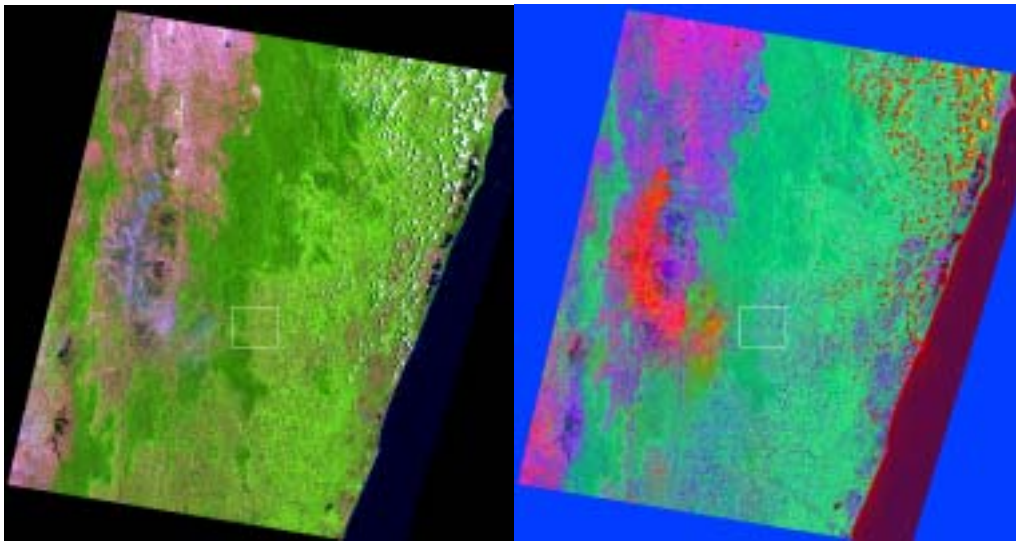


Figure 11: Path 158, row 73, October 28, 2000. *Left:* Landsat ETM false-color composite, RGB: 543. The dark areas in the center-left portion of the image are large burnscars. *Right:* Endmember fractions image: RGB: Soil, green vegetation, non-photosynthetic vegetation. The white box is shown enlarged in Figure 12. The presence of haze/smoke is revealed in the fractions images, represented by the red (high soil fraction) smudge just to the west of the location of the large burnscars.

Endmembers for green vegetation (GV), soil, non-photosynthetic vegetation (NPV), and shade were chosen from the reference image (Path 158, row 73, 10/28/00). Potential image endmember spectra were identified through examination of 2-D plots of DN values (e.g., band 3 vs. band 4, 5 vs. 4, 5 vs. 3, etc.) and several combinations were rejected after trial-and-error unmixing runs. The final image endmember spectra are shown in Figure 10 and were chosen as follows:

*Shade* – The shade endmember was the lowest-reflectance deep water target available

*Green Vegetation (GV)* -- The brightest secondary-vegetation endmember available

*Soil* -- A bare soil target. The northeast portion of the image contained large regions of exposed soil as part of the rapidly eroding central highlands and rice agriculture in the valleys. An interesting spectral feature was the very strong absorption feature in band 4, probably due to iron (hematite) in the soil (see Figure 10; this would be in accord with descriptions of the red soil in the central plateau of Madagascar.

*Non-photosynthetic vegetation (NPV)* – This was the most difficult target to identify, but after significant effort an endmember distinct from soil and burned-area signals was found on the boundary of primary forest and secondary vegetation, probably representing slash from downed trees. Without field data this was unverifiable however, so the NPV endmember was treated skeptically and the identification of training sites in the fractions images primarily made use of the soil, GV, and shade fractions.

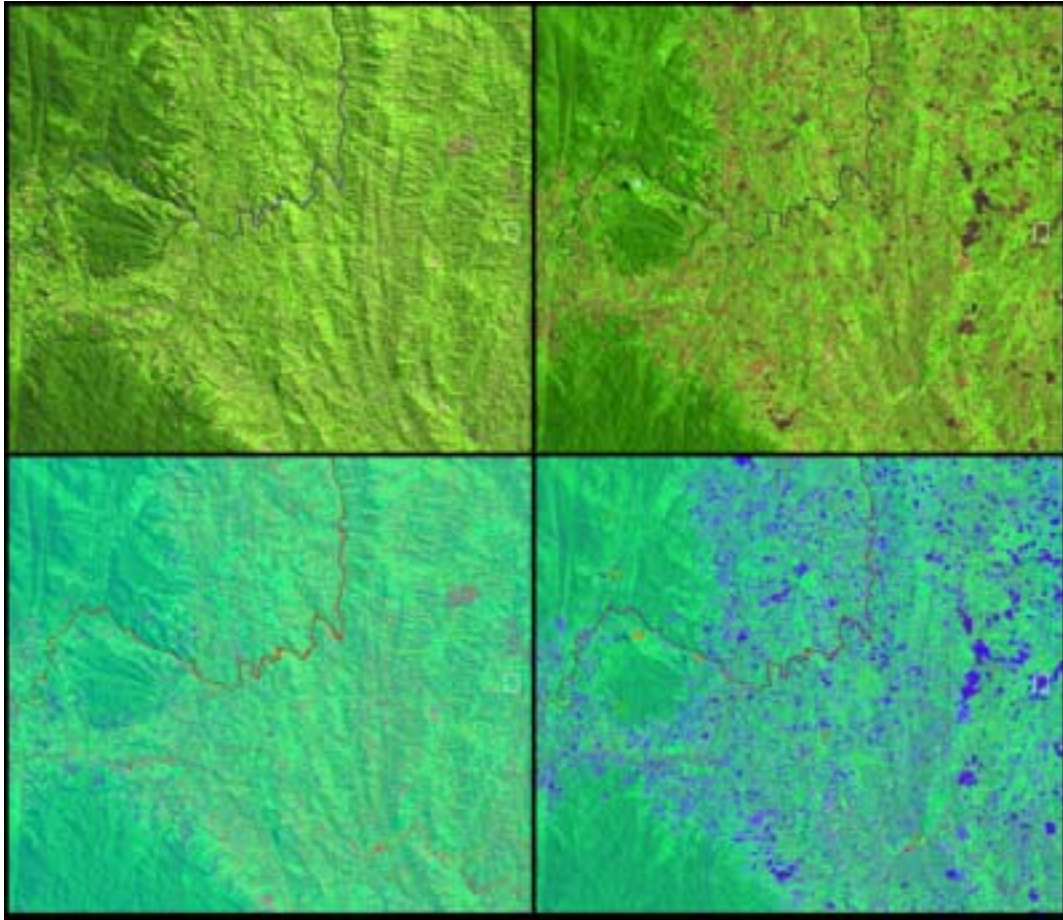


Figure 12: Enlargement of region delineated in Figure 11. Top row: Landsat images. Bottom row: Endmember fractions. Left column: Pre-fire season, April 19, 2000. Right column: End fire season, October 28, 2000.

### *Intercalibration*

Image intercalibration is a technique used to convert image values into values of a common reference image, thereby allowing direct intercomparison (Furby and Campbell 2001). Intercalibration of the five non-reference images to the reference image was accomplished by identifying large calibration regions (hundreds of pixels) of relatively invariant and uniform landcover types. In each image, significant regions of primary forest, grassland, water (lakes or oceans), and secondary vegetation were identified. For each image, the average spectrum of each

calibration region was extracted, resulting in four average DN values for any given ETM band. These four values were then regressed against the corresponding values for targets in the reference image, resulting in a linear equation that transforms image DNs into the equivalent DNs of the non-reference image. The R-squared values of these linear regressions are shown in Table 2.

Table 2: Linear regression equations and r-squared values for intercalibration of images to the reference image (path 158, row 73, 10/28/00). E.g., in order to convert DNs from band 7 of the first listed image (path 158, row 73, 4/19/00) into DNs from the reference image, the new DN values would equal  $0.62 \cdot \text{band } 7 + 1.6$ , and the r-squared of the regression would be 0.98. Equations derived by regression of pixels from calibration regions (described in text). “75e” and “75w” refer to the overlapping portions of P159/R75 and P158/R75 that were used as one early-late image pair.

Path/Row	Date	ETM band					
		1	2	3	4	5	7
158/73	04/19/00	$0.25 \cdot \text{band} + 36.6$ , rsq=0.84	$0.46 \cdot \text{band} + 14.4$ , rsq=0.78	$0.51 \cdot \text{band} + 6.8$ , rsq=0.94	$1.43 \cdot \text{band} - 14.7$ , rsq=0.95	$0.64 \cdot \text{band} + 1.4$ , rsq=0.95	$0.62 \cdot \text{band} + 1.6$ , rsq=0.98
159/75e	10/17/99	$0.72 \cdot \text{band} + 14.5$ , rsq=0.90	$0.95 \cdot \text{band} - 7.1$ , rsq=0.99	$1.07 \cdot \text{band} - 12.0$ , rsq=0.99	$1.43 \cdot \text{band} - 5.9$ , rsq=0.94	$1.17 \cdot \text{band} - 18.6$ , rsq=0.94	$1.23 \cdot \text{band} - 12.7$ , rsq=0.97
158/75w	11/11/99	$0.60 \cdot \text{band} + 4.2$ , rsq=0.91	$0.77 \cdot \text{band} - 8.7$ , rsq=0.98	$0.82 \cdot \text{band} - 10.6$ , rsq=0.99	$1.08 \cdot \text{band} - 0.8$ , rsq=0.95	$0.82 \cdot \text{band} - 6.6$ , rsq=0.95	$0.86 \cdot \text{band} - 5.0$ , rsq=0.97
158/77	09/08/99	$0.65 \cdot \text{band} + 10.9$ , rsq=0.96	$0.85 \cdot \text{band} - 5.4$ , rsq=0.99	$1.06 \cdot \text{band} - 16.2$ , rsq=0.97	$1.28 \cdot \text{band} - 2.8$ , rsq=0.99	$0.97 \cdot \text{band} - 10.8$ , rsq=0.96	$1.10 \cdot \text{band} - 10.1$ , rsq=0.99
158/77	11/11/99	$0.69 \cdot \text{band} + 0.7$ , rsq=0.90	$0.80 \cdot \text{band} - 6.4$ , rsq=0.99	$0.91 \cdot \text{band} - 10.4$ , rsq=1.00	$0.93 \cdot \text{band} + 9.6$ , rsq=0.98	$0.77 \cdot \text{band} - 1.0$ , rsq=0.95	$0.95 \cdot \text{band} - 6.1$ , rsq=0.98

The fits tended to be poorer for ETM bands 1 and 2, which are influenced more strongly by atmospheric haze. The R-squared values were highly satisfactory in all cases except the intercalibration between the 4/19/2000 image and its pair from later in the year, the reference image. The larger difference in season (April 19 is the middle of the wet season, the rest of the images are in the dry season) between this image and the others probably contributed to this difference, particularly due to the difference in soil reflectance properties between wet and dry seasons. Even so, only

bands 1 and 2 fell below R-squared values of 0.90, and this was attained using exactly the same calibration locations on the April and October images, so it was decided that this intercalibration was adequate as well. Essentially no burn scars were seen on the April image, so any minor errors in the resulting classification of the April scene would have little impact on the final analysis in any case.

The regression equations were used to convert the image endmember spectra from the reference image into equivalent endmember spectra for each image. These endmembers were then used to unmix each image. The output for each input ETM scene was a five band fractions image: Shade (sh), Green Vegetation (GV), Non-photosynthetic Vegetation (NPV), and Soil, plus an RMS error image) that was directly comparable across dates and locations. shows an example of the input ETM and output fractions images after spectral mixture analysis.

#### *Binary decision tree classification*

A supervised classification was performed on the endmember fractions images via a binary decision tree (Roberts *et al.* 1998, 2002). Training sites for ten classes were selected based on interpretation of the endmember fractions and ancillary landcover and map data (Battistini and Richard-Vindard 1972; CI 2000; Du Puy and Moat 1998; Green and Sussman 1990; Lowry *et al.* 1997; Sussman *et al.* 1996; Young and Axel 1999). The classes mapped are described in Table 3.

Table 3: Descriptions of mapped landcover classes.

Map class and abbreviation	Description of class
Bare soil (bs)	Exposed soil in the form of tilled, not-yet flooded rice fields; exposed soil due to severe erosion (Wells 1995, 1997); exposed bedrock
Fresh burn (bu)	Fresh burn scars with high shade fraction, reflecting the presence of char
Clouds (cl)	Clouds and cloud shadows, thick haze, mountain shadows.
Dark vegetation (dv)	Dry-deciduous forest or plantation forest (often pines or eucalyptus, Wells 1997, also personal observation). In the southeastern corner of Madagascar (path 158, row 77) there is a significant cover of spiny forest which was included in this class.
Dark water (dw)	Deep water: ocean, lakes, resevoirs.
Green savanna (gs)	Secondary vegetation with a lower green vegetation fraction than sv and a lower soil fraction than ss. Corresponds with "savoka," a local term for scrub that grows on areas unable to support forest or agriculture due to repeated clearing or farming (Young and Axel 1999).
Light burn (lb)	Burns identified via pre/post fire season image comparison, with lower shade fractions
Primary forest (pf)	Primary rainforest
Dry savanna (ss)	Grassland/savanna/brush of the central highlands of Madagascar. High non-photosynthetic vegetation signal.
Secondary vegetation (sv)	High-green vegetation-fraction secondary vegetation. Agriculture and regenerating rainforest.

The binary decision tree is implemented with the S-PLUS 2000 (Mathsoft, Inc., 1988-2000) `tree` function. This produces a multilevel conditional decision tree where the input data (in this case, the endmember fraction values for each training pixel) are used to predict the response variable (in this case, the map class identified by the map producer). The function employs binary recursive partitioning of the training site data, such that at each bifurcation in the tree, the resulting output data groups have maximum difference (Belmont *et al.* 1991, referenced in the S-PLUS documentation for `tree`). The algorithm repeats this procedure until a group of data reaches a minimum size (in this case, 300 training pixels) or minimum node deviance (the S-PLUS default is 0.01, by experimentation 0.03 was determined to be a more useful setting resulting in fewer small branches). At this point the data are assigned a class, representing a "leaf" on the tree.

An iterative methodology was used to classify the images. The endmember reference image (path 158, row 73, October 28, 2000, shown in Figure 11) was the primary image from which training sites were selected. The classification of the reference image was gradually improved by addition of training sites to areas where the map classification did not match image interpretation. The presence of smoke and haze in the image often required additional training in partially obscured regions, and the commonly-confused classes dry savanna and light burn also received extra training in order to avoid the hazard (evident in the first iterations of the classification) of significantly overmapping light burn. Confusion matrices and accuracy statistics for the classification were generated for each version of the classification.

When the classification of the reference image was satisfactory, the decision tree classifier was applied to all of the endmember images. While this classification worked reasonably well on all of the endmember images, a few additional training sites from the other images were added to improve mapping of hazy regions or in one case a vegetation type not included in the reference image (the spiny forest of southeastern Madagascar, found in the eastern regions of path 158, row 77). The resultant final classified images were passed through a 3x3 majority filter, accuracy assessment was performed, and this final map was used in subsequent analysis. In order to create burn area maps for comparison to the DMSF fire index, the pixels classified as burned (bu and lb) on each image were extracted. Each late fire season burned area map was then subtracted from the burned area map from early fires

season, ensuring that only burns that occurred between the image dates would be used for analysis. The resulting 30 m-resolution burn product was then coarsened to 1 km, where each 1 km pixel was assigned a fraction burn area based on the area burned within the 1 km block. If a 1 km pixel consisted of more than 50% cloud and nodata pixels, then the entire 1 km pixel was assigned to nodata. If the 1 km pixel was less than 50% obscured, then the fraction of visible 30 m pixels mapped as burn was assigned to the coarse pixel.

The accuracy of the classification was assessed relative to the training data with a confusion matrix and kappa statistic (Congalton 1987; Congalton and Mead 1983; Hudson and Ramm 1987). A second assessment of the classification was made by comparison of the classification to reference points collected in 2001 (data made available by James Rowland, USGS). These data points were collected in a USGS-funded forest-cover change study in the Mantadia-Zahamena corridor (the region of primary forest in the central-northern part of Landsat path 158 row 73) that focused on forest loss in fragmented forest edge areas, primarily due to *tavy*. 144 GPS ground points fell within the cloud-free regions of the Landsat scene. As the data were collected independently of the present study, with a different landcover scheme, time period, etc., the confusion matrix was created by lumping the map and reference classes into two categories, primary forest and non-primary forest. Reference points where forest had been burned in 2001 were assigned to the forest class for comparison to the 2000 Landsat data. Older clearings were assigned to the non-primary forest class.

### **JERS L-band SAR mosaic**

In order to assess the extent of burning within different land cover classes, a relatively recent land cover map of Madagascar was required. The only land cover map widely available during this study was that of Faramala (1981), which is significantly out of date (Conservation International has recently produced a vegetation map based on year 2000 Landsat data, but this was not available at the time of the analysis). Therefore, recent Japanese Earth Resources Satellite synthetic aperture radar imagery was used to produce a basic land cover map. The JERS L-band SAR mosaic was assembled as part of the Global Rainforest Mapping (GRFM) project (see NASDA/MITI 2000; De Grandi *et al.* 2000b). The GRFM project assembled continent-level mosaics of JERS data covering the Brazilian Amazon, southeast Asia and northern Australia, central and eastern Africa, and Madagascar. The Madagascar dataset was acquired by the GRFM project in January and early February 1997.

The production of the JERS mosaic is described in De Grandi *et al.* (2000a) and Rosenqvist *et al.* (2000). The resulting Madagascar 100 m product consists of three tiles, each 25-50 Mb. Included among the GRFM products is a coarsened 500 m resolution mosaic of the entire island, as well as a coefficient of variation (CV) image, also at 500 m, which gives a measure of the “texture” of the backscatter signal in terms of variability between nearby (100 m) pixels. The 500 m products were the only products used in this study. The 500 m backscatter and CV products

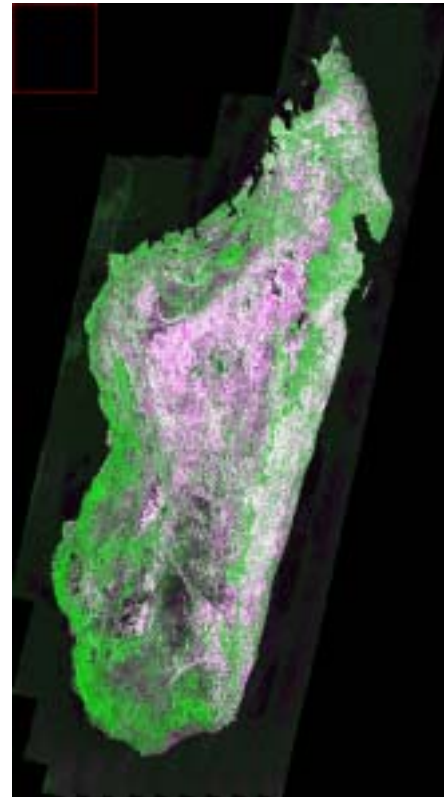


Figure 13: JERS L-band SAR color composite, 1997. *Red/blue*: coefficient of variation; *Green*: mean amplitude.

were georeferenced according to the product metadata into Albers Equal Area projection. It was then reprojected to HOM and cropped to 780 x 1550 km to match the common data frame used for all datasets in the project. Mauritius, included in the northwest corner of this grid, was masked out of the dataset. The 500 m mosaic was resampled (via averaging) to 1 km prior to classification. SAR backscatter depends on surface roughness and geometry. Smooth water will simply bounce the pulse up and directly away from the sensor, and so will have very low backscatter. Rougher surfaces, such as much of the rugged and eroded topography of Madagascar’s eastern mountains and central highlands, will have higher backscatter.

L-band backscatter also correlates with forest biomass although the signal will saturate at higher biomass values (Castel *et al.* 2002). However, it is apparent from Figure 13 that forests tend to have less variation between nearby pixels, resulting in a lower CV. Thus, these two bands contain enough information to produce a forest/nonforest/water map with a binary decision tree classifier. A similar technique, employing both backscatter and texture data, was employed by Simard *et al.* (2000).

#### *Classification*

Classification followed the same method as used for the Landsat classification.

Early attempts to distinguish subclasses of forest with the automated classifier had some success but resulted in substantial confusion between the forest classes that could only be remedied with manual correction. Therefore, forest, spiny forest, nonforest, and water were mapped with the automated classifier. Training sites with known landcover were selected on the basis ancillary knowledge derived from examination of the Landsat scenes and other landcover maps (e.g., Lowry *et al.* 1997). The centers of large forest patches were chosen for training purposes to minimize the possibility of landcover change between the acquisition times of the different datasets. After classification, the forest class was manually subdivided into two classes on the basis of geographic location (using Lowry *et al.* 1997 as a reference), western dry deciduous forest and eastern tropical rainforest. This map underwent accuracy assessment using the training sites. The mapped forest area was compared to previous estimates of Madagascar forest area in the literature order to

further assess the classification. The 1-km map was resampled (3x3 majority rule) to 3-km resolution for processing of simulations.

### **GIS data layers**

Access to the GIS layers used in the study was granted by the U.S. Geological Survey. These included Level 1 and Level 2 Madagascar Provinces (Figure 1), and a layer of protected areas (the AIRE\_PRO database). All of these data layers were georeferenced to the HOM projection. For the purposes of simulation processing, the GIS layers were converted to 3 km resolution grids and then converted to image files for input into ENVI. Comparing the coastlines of the GIS-derived rasters to the rasters derived from JERS, ETM, and DMSP indicated a good match (within one 3 km pixel).

### **Statistical analysis**

#### **Traditional statistical analysis**

The DMSP- and ETM-derived fire products first underwent exploratory data analysis in order to assess the distributions and linear correlation of the datasets. The representativeness of Landsat path 158, row 73 study site of the whole island was assessed by comparing the distribution of the DMSP fire product subset to the distribution of the DMSP fire product for the whole island via a quantile-quantile plot. Scatterplots and ordinary least squares (OLS) linear regression were employed to assess the pixel-to-pixel correlation of the DMSP- and ETM-derived datasets. If the true burn area is assigned to the variable  $z$ , then the uncertain estimate of  $z$  is modeled by a random variable (RV)  $Z$ .  $Z$  is random in that it can take a range of

values according to a probability density function (pdf). The OLS linear regression for predicting  $Z$  is described by:

$$Z = \alpha + \beta(X) + \varepsilon \quad (3)$$

where  $Z$  is the response variable (ETM-derived burn area),  $X$  is the predictor (DMSP-derived cumulative fire index),  $\alpha$  is the regression intercept,  $\beta$  is the slope, and  $\varepsilon$  is the residual (Spiegel *et al.* 2000).

As the scale of the analysis can be a crucial factor influencing results (Quattrochi and Goodchild 1997), the regression procedure was repeated after differing amounts of coarsening. The 1 km datasets were resampled (via averaging, and excluding nodata pixels) to progressively coarser resolutions, and regression analysis was repeated at each resolution. A more sophisticated spatial analysis was then performed using geostatistics.

### **Geostatistical analysis**

Traditional statistical techniques have several substantial weaknesses for assessing fire-related datasets. Visual comparisons of DMSP- and ETM-derived maps reveals some obvious spatial cross-correlation that is not measured in linear regression analysis where each pixel-pixel comparison is treated as an independent data point. It is also evident that the DMSP-derived product exhibited a considerable amount of spatial smoothing relative to the ETM-derived reference maps (i.e., DMSP exhibits more spatial autocorrelation). Histograms and log histograms of the datasets (see

Results) indicate that both the DMSP-derived and ETM-derived datasets have highly skewed, approximately lognormal distributions with a very large number of zero and near-zero pixels. These characteristics violate the assumptions of traditional linear regression analysis. In addition, the correlation obtained from a linear regression represents only an average correlation across all spatial scales; it is possible for correlation to be low or even negative at one scale, obscuring a high correlation that may exist at another scale (Clifford *et al.* 1989; Dobermann *et al.* 1995; Goovaerts 1997).

*Modeling the variograms and co-regionalization*

Geostatistical techniques and algorithms (Deutsch and Journel 1998; Goovaerts 1997) have been developed that deal with these difficulties on a regular basis. An overview is given here; a more thorough review of the theory and equations behind the geostatistical methods employed is given in Appendix A. The first step is to measure spatial autocorrelation via calculation of the empirical semivariogram plots and maps (essentially 2-D variograms). The semivariogram is calculated as the squared semidifference between data points separated by a particular lag distance:

$$\gamma_{zz}(\mathbf{h}) = \frac{1}{2N(\mathbf{h})} \sum_{i=1}^{N(\mathbf{h})} [z(\mathbf{u}_i) - z(\mathbf{u}_i + \mathbf{h})]^2 \quad (4)$$

where  $\gamma_{zz}(\mathbf{h})$  is the semivariance of the ETM-derived burn area  $z$  as a function of lag  $\mathbf{h}$ .  $N(\mathbf{h})$  is the number of pairs at lag distance  $\mathbf{h}$ , and  $z(\mathbf{u}_i)$  is the burn area value  $z$  at the  $i^{\text{th}}$  location  $\mathbf{u}_i$ . The semivariance at lag  $\mathbf{h}=\boldsymbol{\varepsilon}$ , where  $\boldsymbol{\varepsilon}$  is a very small but

nonzero lag (see Appendix A), is known as the nugget and represents microscale variability. The range is the lag distance at which all spatial autocorrelation is lost (semivariance approximately equals dataset variance). The semivariance value at the range is known as the sill (Goovaerts 1997). The cross variability between two different datasets of similar extent is measured by the cross-semivariogram:

$$\gamma_{zx}(\mathbf{h}) = \frac{1}{2N(\mathbf{h})} \sum_{i=1}^{N(\mathbf{h})} [z(\mathbf{u}_i) - z(\mathbf{u}_i + \mathbf{h})][x(\mathbf{u}_i) - x(\mathbf{u}_i + \mathbf{h})] \quad (5)$$

where  $z(\mathbf{u}_i)$  is the ETM burn area value at  $\mathbf{u}_i$  and  $x(\mathbf{u}_i)$  is the DMSP-derived cumulative fire index value at  $\mathbf{u}_i$ .

Empirical semivariograms for the DMSP- and ETM-derived maps, and the corresponding cross-variogram, were produced in order to measure auto- and cross-correlation between the DMSP- and ETM-derived products. Empirical variograms are discontinuous and therefore must be modeled by continuous functions where semivariance  $\gamma(\mathbf{h})$  is described by a semivariogram model  $g(\mathbf{h})$ . In order for kriging weights to be solvable, semivariogram models must be permissible. Several commonly used variogram structures are known to be permissible (Deutsch and Journel 1998; Goovaerts 1997). In order to model variograms with complex features and anisotropies, multiple permissible variogram structures  $g_l(\mathbf{h})$  can be combined additively, with each of the  $L$  model structures being assigned a proportion  $b^l$  of the sill. In this study, 4 structures were used for each variogram: a nugget effect, a

short-range spherical structure, a medium-range spherical structure, and a long-range

Gaussian structure:

$$\begin{aligned}
 \text{Nugget effect: } g_0(\mathbf{h}) &= \begin{cases} 0 & \text{if } \mathbf{h} = 0 \\ 1 & \text{otherwise} \end{cases} \\
 \text{Spherical models: } g_{1,2}(\mathbf{h}) = \text{Sph}\left(\frac{\mathbf{h}}{a_{1,2}}\right) &= \begin{cases} 1.5 \cdot \left(\frac{\mathbf{h}}{a_{1,2}}\right) - 0.5 \cdot \left(\frac{\mathbf{h}}{a_{1,2}}\right)^3 & \text{if } \mathbf{h} \leq a_{1,2} \\ 1 & \text{otherwise} \end{cases} \\
 \text{Gaussian model: } g_3(\mathbf{h}) &= 1 - \exp\left(\frac{-3\mathbf{h}^2}{a_3^2}\right)
 \end{aligned}$$

(6)

where  $\mathbf{h}$  equals the lag and  $a_{1,2,3}$  are the ranges. The structures were linearly combined to produce the auto- and cross-semivariogram models:

$$\begin{aligned}
 \gamma_{ZZ}(\mathbf{h}) &= b_{ZZ}^0 g_0(\mathbf{h}) + b_{ZZ}^1 g_1(\mathbf{h}) + b_{ZZ}^2 g_2(\mathbf{h}) + b_{ZZ}^3 g_3(\mathbf{h}) \\
 \gamma_{XX}(\mathbf{h}) &= b_{XX}^0 g_0(\mathbf{h}) + b_{XX}^1 g_1(\mathbf{h}) + b_{XX}^2 g_2(\mathbf{h}) + b_{XX}^3 g_3(\mathbf{h}) \\
 \gamma_{ZX}(\mathbf{h}) &= b_{ZX}^0 g_0(\mathbf{h}) + b_{ZX}^1 g_1(\mathbf{h}) + b_{ZX}^2 g_2(\mathbf{h}) + b_{ZX}^3 g_3(\mathbf{h})
 \end{aligned} \tag{7}$$

This joint modeling of all three variograms with the same spatial structures produces a Linear Model of Coregionalization (LMC). The LMC is said to be permissible if the variogram models are individually permissible and the coregionalization matrices  $\mathbf{B}^j$  are positive definite. This was checked using the following tests given by Goovaerts (1997):

$$\begin{aligned}
b_{ZZ}^l &\geq 0 \\
b_{XX}^l &\geq 0 \\
\left| \begin{array}{cc} b_{ZZ}^l & b_{ZX}^l \\ b_{XZ}^l & b_{XX}^l \end{array} \right| &= b_{ZZ}^l b_{XX}^l - [b_{ZX}^l]^2 \geq 0 \Rightarrow b_{ZX}^l \leq \sqrt{b_{ZZ}^l b_{XX}^l}
\end{aligned} \tag{8}$$

which were applied for all  $L$  structures. A major advantage of the LMC approach is that a by-product of the modeling procedure is the *structural correlation coefficient*,  $\rho_{ZX}^l$ , which describes the cross-correlation between the datasets for each particular spatial structure  $l$ , which can reveal scale-specific correlation that is obscured in the overall correlation figure (Clifford *et al.* 1989). The structural correlation coefficients were calculated from the sill proportions  $b$  (Dobermann *et al.* 1995):

$$\rho_{ZX}^l = \frac{b_{ZX}^l}{\sqrt{b_{ZZ}^l b_{XX}^l}} \tag{9}$$

The variogram modeling procedures described above were repeated on normal-score transformed versions of the ETM- and DMSP-derived datasets, as the transformed variables produced better-behaved variograms, and as sequential Gaussian simulation (described below) requires that the input datasets be Gaussian in any case.

#### *Sequential Gaussian Simulation*

The LMC procedure gives the analyst an estimate of the scale-dependent correlation between the ETM-derived reference dataset and the DMSP fire index. In order to estimate uncertainty in burn area predictions made by using the DMSP product as a secondary variable, across a large number of polygons and time intervals, sequential

Gaussian simulation (Deutsch and Journel 1998) was employed. This technique produces multiple realizations of equiprobable maps of burn area that are conditional upon the secondary data (the DMSP cumulative fire index) as well as the distribution and semivariogram of the ETM-derived reference dataset (for a similar use of simulations, but in a land-cover classification setting see Kyriakidis and Dungan 2001).

Simulation makes use of kriging, a method designed to produce the best-possible estimate of unknown value  $z(\mathbf{u})$  at location  $\mathbf{u}$ , by narrowing the *a priori* distribution  $Z(\mathbf{u})$  to a better-informed, narrower distribution  $Z_{SK}^*(\mathbf{u})$ . Strictly speaking then, a map resulting from kriging is only a map of best estimates at individual points, *not* a potentially accurate map of the phenomenon as a whole. In particular, if used to produce a map, kriging produces an artificially “smoothed” surface that does not reproduce the variogram structure of the input data. An additional problem is that the kriging variance for a particular point is simply a function of distance from locations with known data, and does not take into account the actual magnitude of the data points or their tendency to co-vary with each other. The uncertainty of estimates will thus be underestimated if only a kriging approach is used.

In this study, an additional point against kriging is that primary data (ETM-estimated burn-area values) are limited only to specific time periods spanned by Landsat image pairs and are not available for an entire fire season. For most time-spans the only data available is the secondary DMSP-derived fire index raster. Performing kriging

alone with just this secondary dataset would amount to nothing more than a regression transform of the DMSP-derived fire index into an unrealistically smooth burn area map. Stochastic simulation overcomes these difficulties by producing a number of *equiprobable realizations* of  $\{z(\mathbf{u}), \mathbf{u} \in A\}$ ,  $\{\mathbf{u} \in A\}$  being all of the locations in the raster  $A$ . Each realization draws values from a pre-specified distribution, and approximately replicates the variogram (and therefore the spatial structure) of the reference data.

The disadvantage of the simulation technique is that multiple simulated realizations of "truth" must be generated and processed to provide a mean area burned estimate for a given region. The corresponding advantage is that the distribution of burn area estimates for any region of choice can be displayed as a histogram and any skew in the distribution of simulated outcomes can be ascertained. This distribution of possible true burn area values, given the DMSP dataset, has utility for end-users:

The distribution of outcomes is useful for inputs into biogeochemical and land-use change models, especially those where the tails of a distribution of potential burn area outcomes are crucial (for instance, a catastrophic burn of a forest preserve).

Having the distribution of realization outcomes provides a ready method to statistically test for significant differences in burn area for a given region, for example between fire seasons.

Deutsch and Journel (1998) lay out the basic procedure for sequential Gaussian simulation. In particular, simulation of a Gaussian field requires that the variable being simulated can be modeled by adopting the multivariate Gaussian RF model. A number of steps and checks have to be performed in order to meet this requirement. The outline of the simulation setup is shown in Figure 14 and an outline of the simulation procedure is given in Figure 15 (summarized from Deutsch and Journel 1998; Goovaerts 1997). Simulation of multiple equiprobable alternative maps is accomplished by repeating steps 4-5. In this study, 100 simulated island-wide burn area maps were generated for each time-span of interest (listed in Table 4). The involvement of normal-score transforms in the procedure is depicted in more detail in Figure 14.

In the present study, the construction of the LMC was only used to calculate structural correlation coefficients in order to obtain an estimate of how correlation varied at differing scales. The LMC structures can be directly incorporated into simulations in a cokriging context, but this option was not implemented at the time of the study. Instead, stochastic simulation using simple kriging with a locally varying mean was employed. Although this option incorporated less of the available information in the simulation model, it proved relatively simple to implement and automate. Additionally, given the low correlations between DMSP- and ETM-derived datasets, it is not clear that a more elaborate simulation procedure would have substantially improved results.

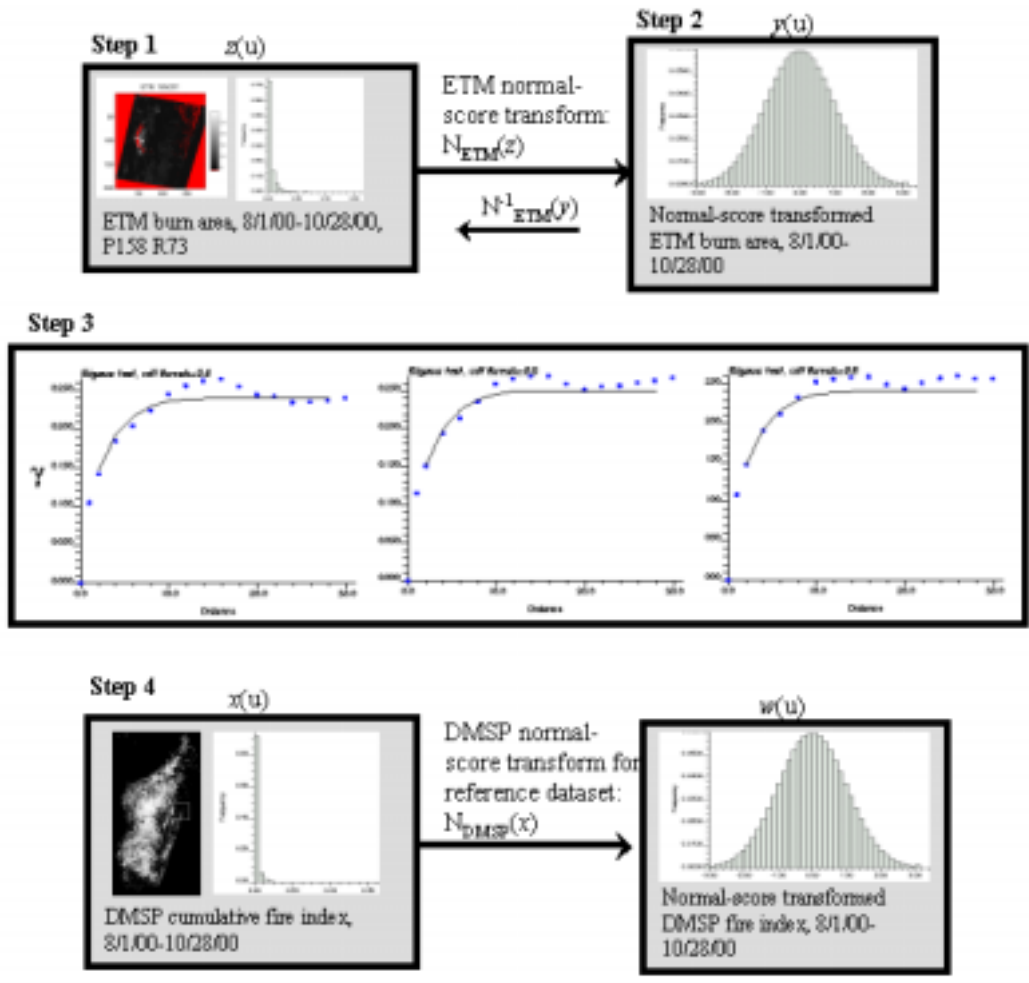


Figure 14: Outline of sequential Gaussian simulation setup. The steps: **(1)** Construct a distribution of burn-area values based on reference dataset (the ETM-derived burn area map). **(2)** Transform the reference distribution into the required standard normal distribution via the rank-preserving normal-score transform. The resulting pdf will have 0 mean and unit variance. **(3)** Check for bivariate normality by comparing the empirical indicator variograms of the normal-score transformed dataset to the theoretically expected indicator variograms for a bivariate normal dataset. If there is a reasonable match then the decision is made to adopt the multivariate normal model for the dataset and simulation can proceed. **(4)** Normal-score transform the DMSP-derived cumulative fire index dataset corresponding to the reference ETM dataset. This same transform will be used to transform each DMSP secondary dataset, so that simulation results are intercomparable. The labels  $x(u)$ , etc. refer to the RVs discussed in the text.

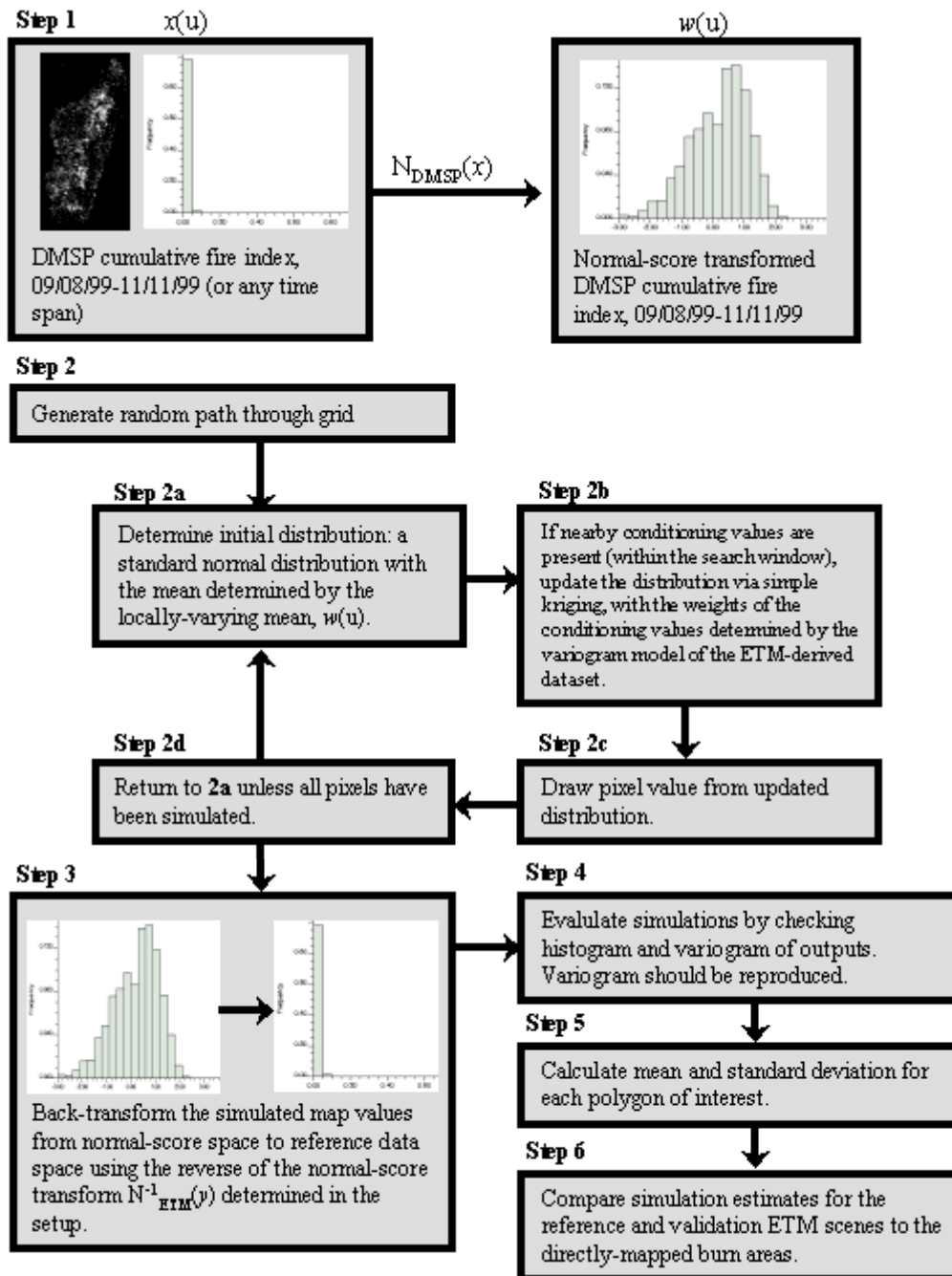


Figure 15: The simulation procedure. (1) Normal-score transform the DMSP fire index dataset for the time period of interest, using the transformation  $N_{DMSP}(x)$  derived in step 4 of the setup. (2) For each simulation run, steps 2a-2d are repeated. (3) Transforms simulation in normal-score space back into regular data space (burn area fraction). Steps 2-3 are repeated 100 times to produce 100 alternative realizations of a burn-area map. Steps 4-6 involve evaluation and processing of the simulation realizations.

Table 4: Number of nights and observations for the time periods over which burn area was simulated. For each run (a specific time) period 100 simulations were produced and processed.

Burn area estimates for	Start date	End date	Number of days	Number of DMSP observations used in generation of DMSP fire index
Calibration to P158 R73 reference image pair	8/1/00	10/28/00	89	106
Validation with P158-9 R75 image pair	10/17/99	11/11/99	26	26
Validation with P158 R77 image pair	9/8/99	11/11/99	65	64
Aug-99	7/31/99	8/31/99	32	31
Sep-99	9/1/99	9/30/99	30	29
Oct-99	10/1/99	10/31/99	31	31
Nov-99	11/1/99	11/30/99	30	30
Dec-99	12/1/99	12/31/99	31	21
1999 total burn area	7/31/99	12/31/99	154	142
Aug-00	8/1/00	8/31/00	31	42
Sep-00	9/1/00	9/30/00	30	38
Oct-00	10/1/00	10/31/00	31	30
Nov-00	11/1/00	11/30/00	30	29
Dec-00	12/1/00	12/31/00	31	30
2000 total burn area	8/1/00	12/31/00	153	169

For each of the 15 simulation runs, the listed DMSP fire index map specified the locally-varying mean. It is important to note that the same forward normal-score transform (the one used for transformation of the calibration dataset, 8/1/00-10/28/00) was used to transform each input DMSP fire index map. Similarly, the reverse normal score transform (converting the simulation from normal-score space back to burn area space) always used the transform derived for conversion of the reference ETM dataset into normal score space. This ensures comparability between the outputs (see for illustration). The empirical variograms of selected simulation outputs were compared to the variogram of the ETM reference dataset in order to check for reproduction of the spatial structure. As different locally-varying mean datasets were used for each simulation, reproduction of the reference dataset

histogram will not occur, however the histograms were still inspected to check reproduction of the expected logarithmic shape.

Each data cube of 100 simulations was then processed for the data layers of interest (calibration and validation regions, country outline, level 1 provinces, level 2 provinces, protected areas, and JERS-derived landcover). For each polygon, the mean and standard deviation of burn area was calculated. For the calibration and validation regions (where burn area had been mapped with ETM), the mean simulated burn area was compared to the mapped burn area. As the potential for monitoring changes in burn area was of particular interest, statistical tests for significant difference in burn area were performed on the burn area estimates for protected areas.

## **Chapter 4: Results**

### **Landsat classification**

The final binary decision tree is displayed in Figure 16. This tree is more complex than the trees generated in the first iterations of the classification as the “fine-tuning” of the tree adds more branches to the lower levels. A major advantage of decision trees is the easy interpretability of the classifier. In Figure 16, the basic division is between forest and nonforest on the basis of high or low GV endmember. On both resultant branches, the next bifurcation occurs based on high or low shade endmembers; in the primarily forest group, high shade includes all of the primary forest (pf) leaves as well as two green savanna (gs) and a dark vegetation (dv) leaves. The low shade branch includes the only secondary vegetation leaf as well as a green savanna leaf and a (bright) cloud leaf.

On the left, low-vegetation side of the decision tree, the shade decision rule results in two groups. High shade includes fresh burn (bu), dark vegetation (dv), water (dw), and two cloud shadow (cl) leaves. Low shade includes all of the dry savanna (ss), bare soil (bs), and light burn (lb) leaves, as well as a (bright) cloud leaf and a final fresh burn (bu) leaf.

The lower branches of the tree are all involved in distinguishing spectrally similar classes. For example, reading from left-to-right across the bottom of Figure 16 shows that light burn (lb) is distinguished from fresh burn (bu) by a lower NPV fraction; light burn (lb) is distinguished from dry savanna (ss) on the basis of higher

shade and NPV fractions and a lower soil fraction; light burn (lb) is distinguished from bare soil (bs) on the basis of higher shade fraction; dark vegetation (dv) is distinguished from fresh burn (bu) and cloud shadow (cl) by a higher GV; and primary forest (pf) is distinguished from dark vegetation (dv) by a lower shade fraction. Clouds (cl) are often distinguished from other classes by high soil fractions. The classified images were processed with a 3x3 majority filter to remove isolated pixels of distinct landcover types resulting from speckle in the input images. The resulting classification of the area in Figure 12 (a region with many small burn scars) is shown in Figure 18. The percentages of each landcover mapped in the reference scene are shown in Table 5. The total percentage of burn area mapped in the scene was 2.54%. For the other two image pairs, P158-9, R75 and P158, R77, the respective mapped burn area percentages were 0.44% and 0.90%, respectively.

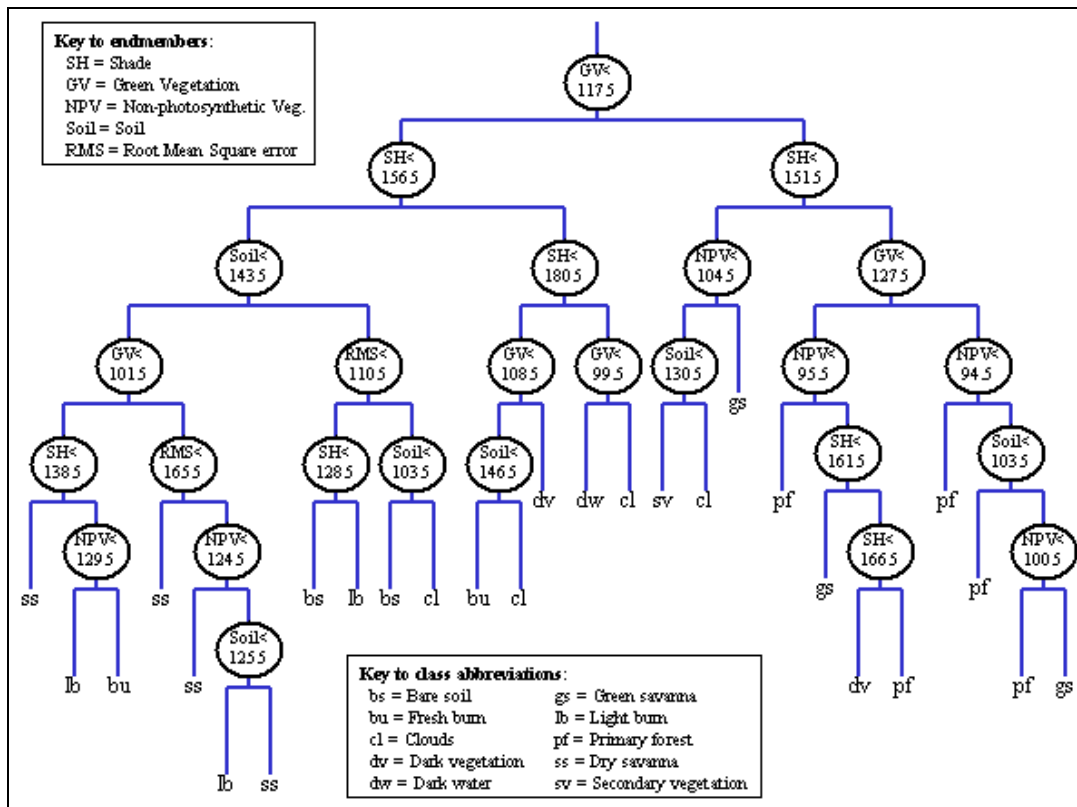


Figure 16: Final binary decision tree. The decision rules are shown in ovals, the output class abbreviations (described in text) are the "leaves" of the tree. The decision rules are based on the endmember fractions (see key). For every decision, the left branch is followed if the pixel is less than the specified value, and the right branch is followed if the pixel value is higher. The values given are DN values representing the pixel fractions of each image, where 100 = 0% fraction and 200 = 100% fraction (a convention allowing for the encoding of fraction values <0% or >100% which are occasionally produced by the SMA algorithm).

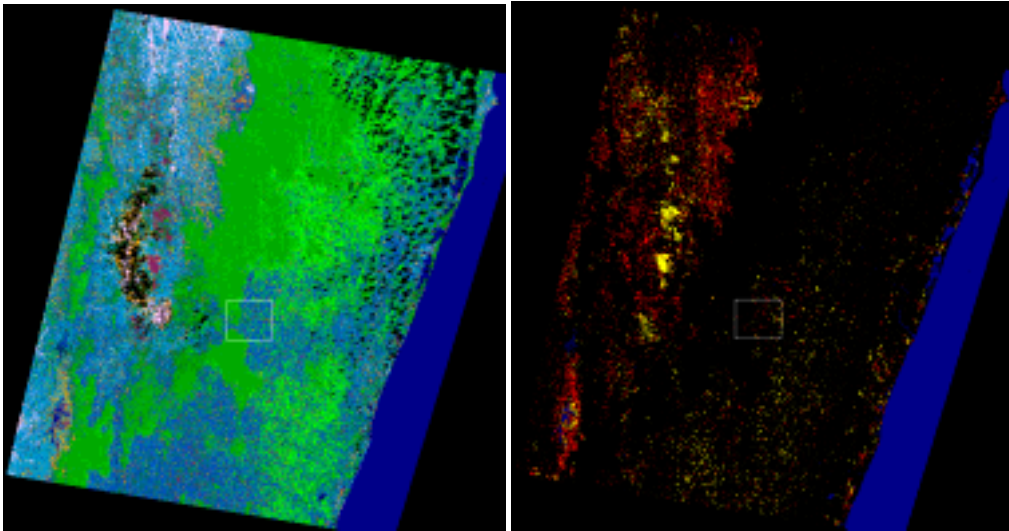


Figure 17: Land cover classification, path 158, row 73, October 28, 2000 (compare to Figure 11). *Left*: Ten-class landcover map. *Right*: Extracted burn scars. Map key is in Figure 19. The white box is shown magnified in Figure 19.

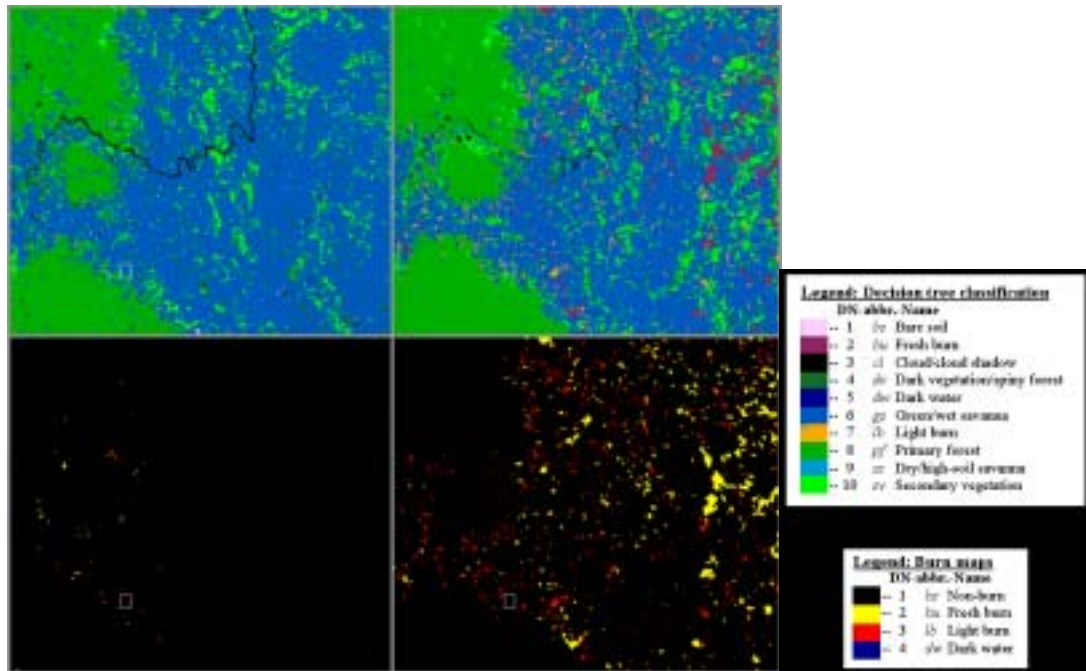


Figure 18: Land cover classification, enlargement of region delineated in Figure 11 (same subset region as Figure 12). Top row: landcover map resulting from final decision tree classification. Bottom row: Extracted burn scars (yellow = fresh burn, red = light burn, black = unburned). Left column: April 19, 2000. Right column: October 28, 2000.

Table 5: Area of mapped landcover classes for the entire Landsat image path 158, row 73, 10/28/00. “CI” includes cloud, cloud shadow, and the nodata area outside of the bounds of the Landsat image, all of which were excluded as nodata in the last column.

Class	Number of 30 m <sup>2</sup> pixels	Area (km <sup>2</sup> )	Percent of total	Percent of total excluding nodata
bs	502868	452.6	1.18%	1.70%
bu	550174	495.2	1.29%	1.85%
cl	12879316	11591.4	30.27%	N/A
dv	1276774	1149.1	3.00%	4.30%
dw	3468941	3122.0	8.15%	11.69%
gs	5212190	4691.0	12.25%	17.57%
lb	203937	183.5	0.48%	0.69%
pf	9528104	8575.3	22.40%	32.12%
ss	6512871	5861.6	15.31%	21.95%
sv	2409825	2168.8	5.66%	8.12%
Total	42545000	38290.5	100.00%	100.00%

*Accuracy assessment*

The confusion matrix for the final classification is given in Table 6. Overall the confusion matrix indicates clear separability for most of the classes. However, significant confusion arises for pairs of spectrally similar classes: Primary forest (pf) and dark vegetation (dv); dry savanna (ss) and light burn (lb); and green savanna (gs) and secondary vegetation (sv). Green savanna, a heterogenous class from the start, also experienced some confusion with primary forest and dry savanna. The cloud class experienced some confusion with bare soil, dark vegetation, and fresh burn (bu), but much of this can be attributed to smoke and haze in the western portion of the path 158, row 73 image, which also happen to have large areas of dark vegetation (pine and eucalyptus plantations). The vast majority of fresh burn area was also in or adjacent to this hazy area (see Figure 11). The confusion between cloud and bare soil results from the high soil fractions that spectral mixture analysis assigns to clouds.

Considering the goals of the present study, the low accuracy statistics for light burn (lb) must be considered. While fresh burn scars are spectrally quite distinct (e.g.,

Figure 12), light burns can often be identified by comparison with a pre-fire season image, but remain spectrally indistinct from dry savanna. The resulting user's accuracy (Table 6) is only 45%, although the producer's accuracy is higher. However, even on the image with the most fire activity (path 158, row 73, 10/28/00), light burn was by far the smallest mapped class (<0.5% of the image) and constituted only about one-fourth of the total mapped burn area (see Table 7). Dry savanna (ss), on the other hand, was mapped as some 15% of the image, so a small proportion of savanna pixels mis-mapped as light burn will heavily influence the accuracy statistics for light burn. The overall accuracy for this classification (Table 6) is 80.5%; generally, an overall accuracy of 85% or higher is preferred in remote sensing studies. However, much of the interclass confusion discussed above was between related classes that can be merged into a broader class. When this is done (Table 7), the kappa climbs to 0.811 and the overall accuracy to 86%.

The confusion matrix for the 144 ground reference control points are shown in Table 8. The overall accuracy was 85.4%. Most sources of confusion between primary forest and non-primary forest are understandable as resulting from fine distinctions that are possible to make on the ground. Of the 15 pixels mapped as primary forest but identified as non-primary forest in the reference dataset, 5 were ground-identified as secondary forest (a class not separately mapped in the classification, due to the difficulty of distinguishing between various forms of agriculture and secondary vegetation), and 9 were identified as *tavy*, often very small clearings not always identifiable on Landsat. Of the 6 pixels mapped as non-primary forest in the

ground reference but identified on the ground as primary forest, 5 were mapped as green savanna, representing small patches of remaining trees. Only one pixel mapped as burned in the 2000 Landsat scene (in this case, light burn) had a corresponding reference point; the 2001 ground data for this location indicated clearing the previous year.

Table 6: Confusion matrix for decision tree classification. Columns are training classifications, rows are map classifications. Class descriptions are given in Table 3. Kappa coefficient = 0.773.

		Training data											
		bs	bu	cl	dv	dw	gs	lb	pf	ss	sv	Total	User's Acc.
Mapped class	bs	960	0	51	3	0	0	12	0	110	0	1136	84.5
	bu	15	1862	119	2	0	0	69	0	160	0	2227	83.6
	cl	143	134	1088	114	0	0	23	18	0	0	1520	71.6
	dv	0	1	14	1026	0	0	0	326	114	0	1481	69.3
	dw	0	0	0	0	820	0	0	0	0	0	820	100
	gs	0	0	0	56	0	1298	0	223	253	34	1864	70.0
	lb	4	156	0	4	0	9	908	0	933	0	2014	45.1
	pf	0	0	0	128	0	46	0	4303	1	5	4483	96.0
	ss	35	20	1	49	0	3	229	100	2760	0	3197	86.3
	sv	0	0	0	0	0	20	0	69	0	643	732	87.8
	Total	1157	2173	1273	1382	820	1376	1241	5039	4331	682	19474	
Producer's Accuracy		83.0	85.7	85.5	74.2	100	94.3	73.2	85.4	63.7	94.3		80.5

Table 7: Confusion matrix for six classes only. Grouped classes are nonforest (nf = ss+bs+gs), burn (bu + lb) and forest (fo, pf+dv). Kappa coefficient = 0.811.

		Training data							
		Bu	cl	dw	fo	nf	sv	Total	User's Acc.
Mapped class	bu	2995	119	0	6	1121	0	4241	70.6
	cl	157	1088	0	132	143	0	1520	71.6
	dw	0	0	820	0	0	0	820	100
	fo	1	14	0	5783	161	5	5964	97
	nf	261	52	0	431	5419	34	6197	87.4
	sv	0	0	0	69	20	643	732	87.8
	Total	3414	1273	820	6421	6864	682	19474	
Producer's Accuracy		87.7	85.5	100	90.1	78.9	94.3		86.0

Table 8: Confusion matrix for 144 GPS-located ground reference points. Kappa coefficient = 0.646.

		Reference data			
		Non-primary forest	Primary forest	Total	User's Accuracy
Mapped class	Non-primary forest	92	6	98	93.9
	Primary forest	15	31	46	67.4
	Total	107	37	144	
Producer's Accuracy		86.0	83.8		85.4

## **JERS classification**

The confusion matrix for the JERS classification is shown in Table 9. In general good separability was achieved; however, post-classification, the forest class was manually divided into western deciduous forest and eastern rainforest on the basis of ancillary maps as these were not readily distinguishable with the SAR data. The classification can be further assessed by comparison of the primary forest estimate with other published estimates (Faramala 1981; Green and Sussman 1990; Humbert and Cours Darne 1965; Mayaux *et al.* 2000; Nelson and Horning 1993). Table 10 shows that the JERS-derived estimate is slightly higher than other recent estimates, however given the inconsistency of published estimates, and the varying datasets and definitions employed by the different studies, rigorous comparison is difficult. The high JERS-derived estimate probably results from including secondary forest with primary forest; however it is similar to the most recent (1998) estimates, so for the purposes of gauging fire activity at a coarse scale it is adequate. The final JERS landcover map, and relative proportions of the landcover classes, are shown in Figure 19.

Table 9: Confusion matrix for classification of JERS dataset. dv = spiny forest, dw = water, nf = nonforest, pf = forest. Kappa coefficient = 0.853.

		Training data					
		dv	dw	nf	pf	Total	User's Accuracy
Mapped class	dv	4411	0	39	513	4963	88.9
	dw	0	4009	167	0	4176	96.0
	nf	130	212	9539	470	10351	92.2
	pf	917	0	463	6604	7984	82.7
	Total	5458	4221	10208	7587	27474	
Producer's Accuracy		80.8	95.0	93.4	87.0		89.4

Table 10: Comparison of primary rainforest area derived from JERS classification to other published estimates.

Dataset Year	Percent primary rainforest	Data source	Reference
Original cover	19.08%	none	Green & Sussman (1990)
1950	12.95%	Maps based on 1950 aerial photos (Humbert & Cours Darne, 1965)	Green & Sussman (1990)
1974-76	11.46%	Landsat MMS interp	Faramala (1981)
1985	6.47%	Landsat interp.	Green & Sussman (1990)
1991	5.82%	AVHRR-LAC, 1 km	Nelson & Horning (1993)
1997	10.18%	JERS 500 m	Current study
1998	9.34%	SPOT-4 Vegetation, 1.15 km	Mayaux, Gond, Bartholome (2000)

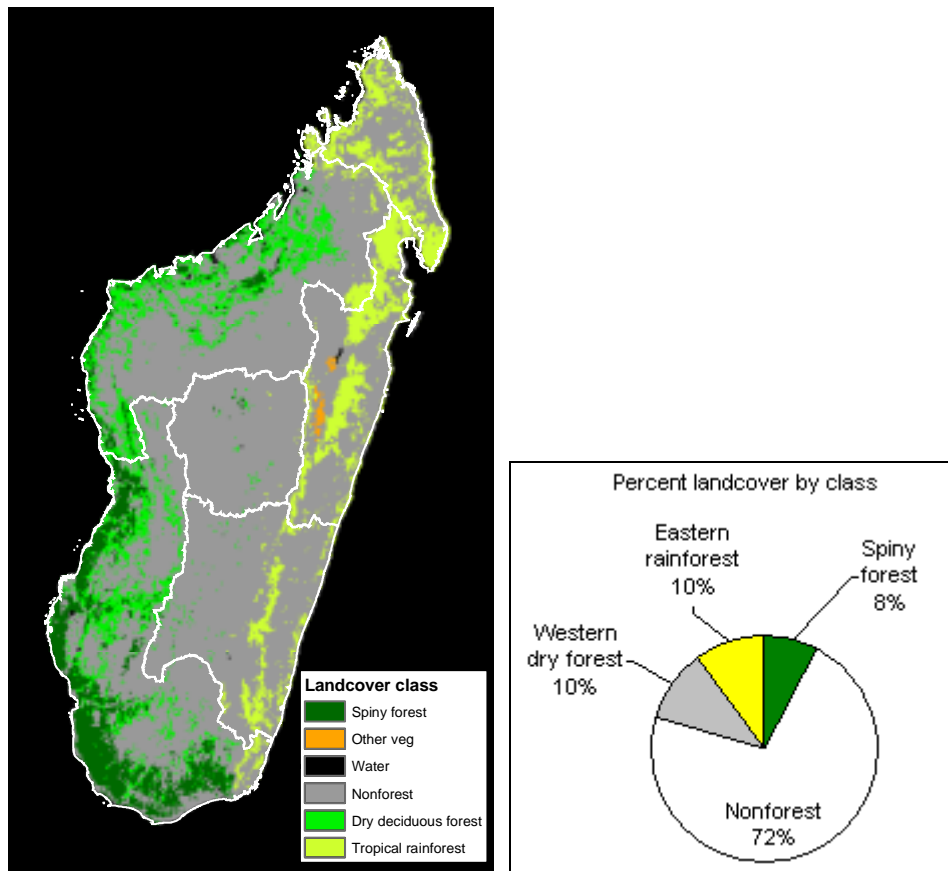


Figure 19: JERS-derived landcover map, and percentages of mapped classes. “Other vegetation” was a small region of anthropogenic vegetation (eucalyptus plantation) that was manually removed from the forest class (0.2 % of total island area).

## Traditional statistical analysis

### Exploratory data analysis

The ETM-derived burn area map and the DMSP-derived fire index map are compared at 1 km resolution in Figure 20a. Nodata pixels (those outside the data extent or with over 50% cloud cover) are mapped in red. The correspondence is very inexact, but it is evident that at a coarse scale there is some relationship. The

surface plots (Figure 20b) show the difference in “roughness” between the two datasets.

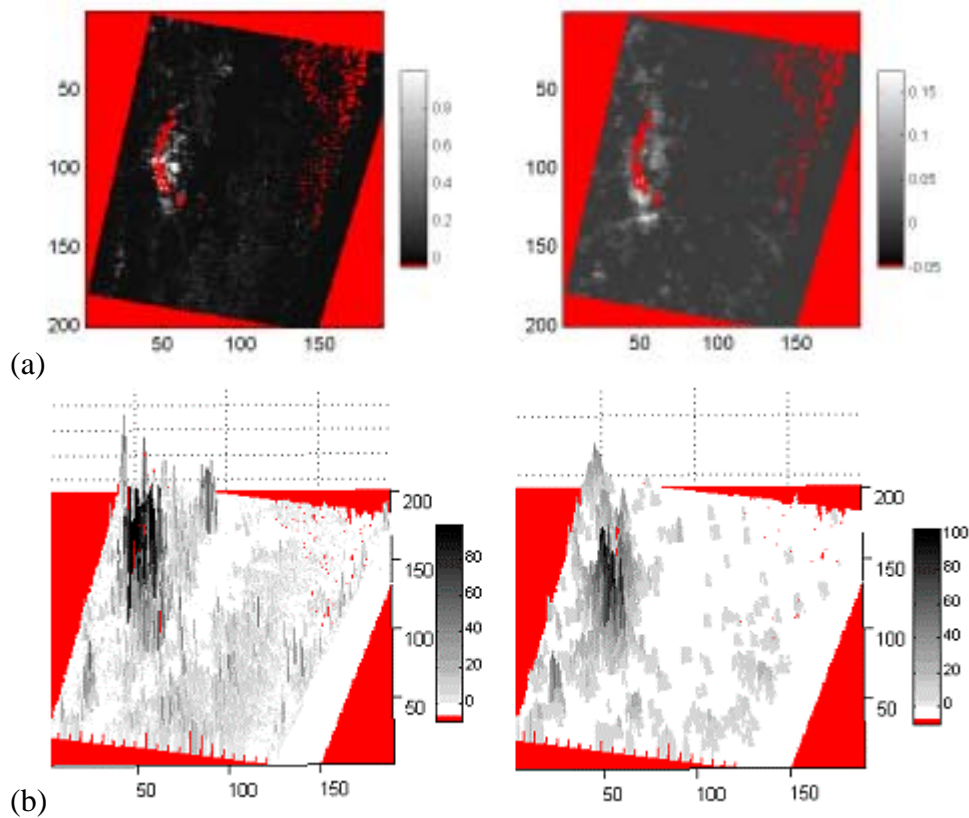


Figure 20: Comparison of fire products at 1 km resolution over the extent of Landsat path 158, row 73. (a) Standard view. *Left column*: ETM-derived burn area fraction map; *bottom*: surface representation of the same data (percentage burn area). *Right column*: DMSP-derived fire index map. (b) Surface representation of the same data, values rescaled to range 0-100. Red represents majority nodata areas of the classified Landsat scene.

## Histograms

Histograms of the initial DMSP- and ETM-derived maps are shown in Figure 21 -- Figure 23. The distributions of both datasets were extremely right-skewed, with a large number of zero and near-zero pixels, and a few pixels with high extreme values. Logarithm transforms produced approximately symmetrical distributions,

indicating that the datasets are approximately lognormal. This is confirmed by examination of normal probability plots (Figure 24).

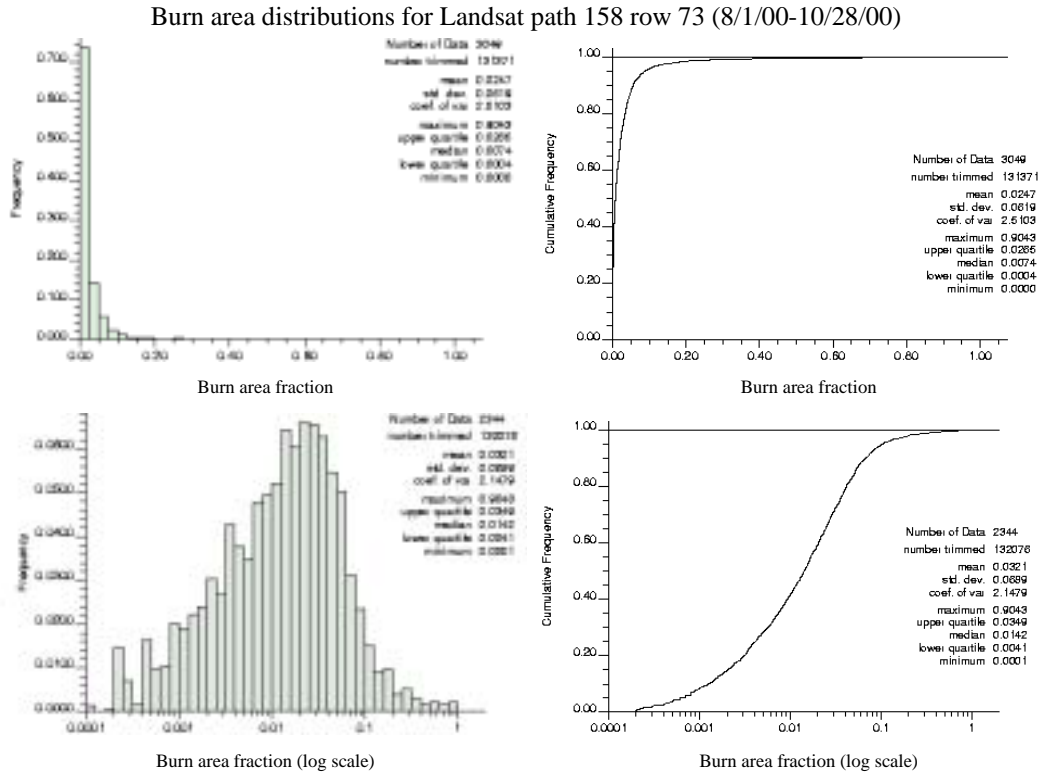


Figure 21: The distribution of the burn area fractions per 3-km pixel, derived by coarsening of the classified 30-m burn scar maps from Landsat ETM, path 158 row 73, 4/19/00 and 10/28/00. Left column: histograms. Right column: Cumulative frequency plots. *Top row*: no scaling. *Bottom row*: Log-scaling. The data follows a roughly log-normal distribution.

DMSP fire index distributions for Landsat path 158 row 73 (8/1/00-10/28/00)

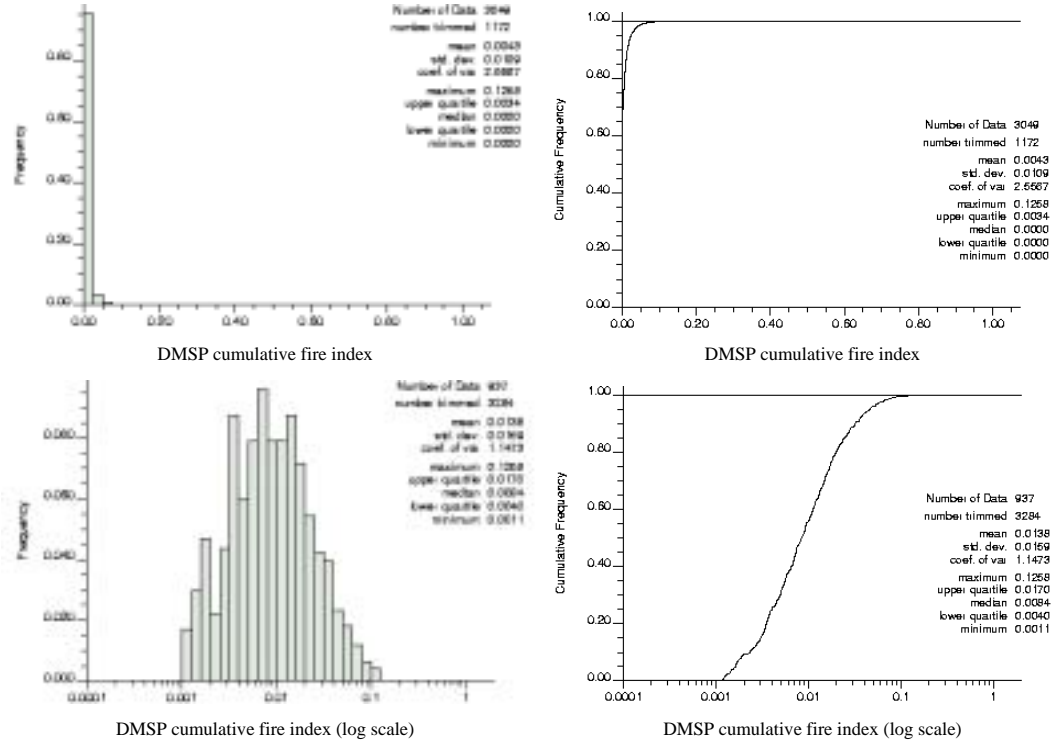


Figure 22: The distribution of the DMSP-derived fire index values per 3-km pixel, derived by coarsening of the 1-km DMSP-derived fire index map for August 1, 2000 – October 28, 2000 covering the same extent as Landsat path 158 row 73. Left column: histograms. Right column: Cumulative frequency plots. Top row: no scaling. Bottom row: Log-scaling. The data follows a roughly log-normal distribution.

DMSP fire index distributions for whole island (8/1/00-10/28/00)

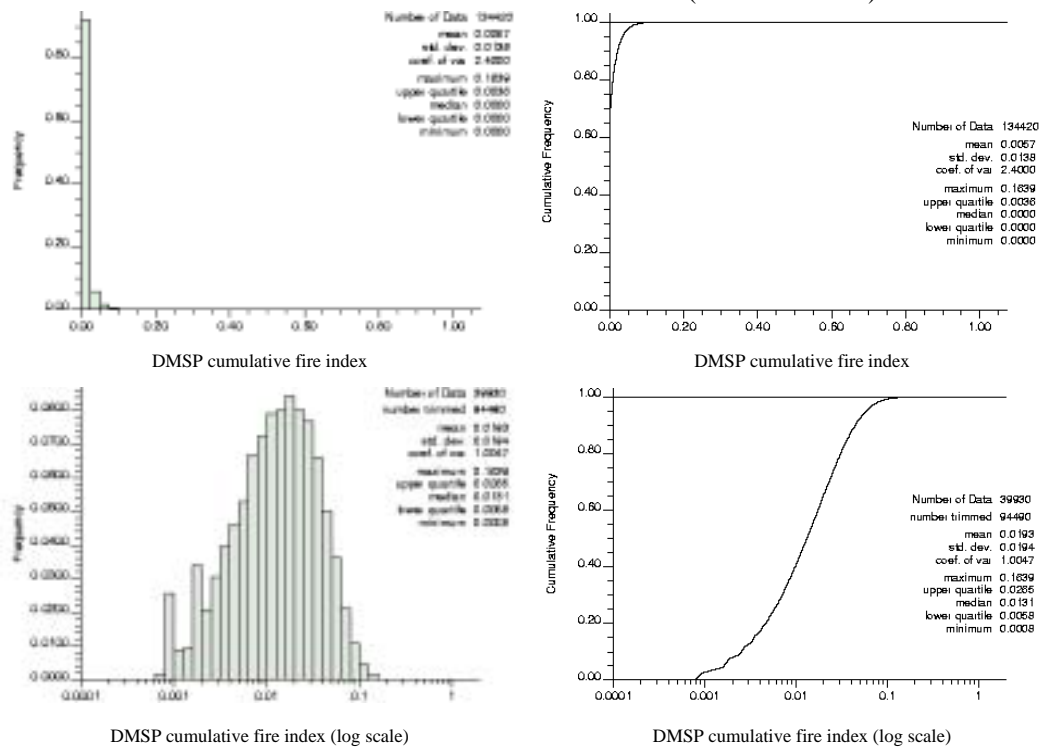


Figure 23: Distribution of DMSP-derived fire index for the entire island of Madagascar, 3 km resolution, August 1, 2000 – October 28, 2000. Left column: histograms. Right column: Cumulative frequency plots. Top row: no scaling. Bottom row: Log-scaling. The distribution is roughly log-normal and similar to the distribution (see Figure 22) of the DMSP fire index for the Landsat path158, row 73 footprint.

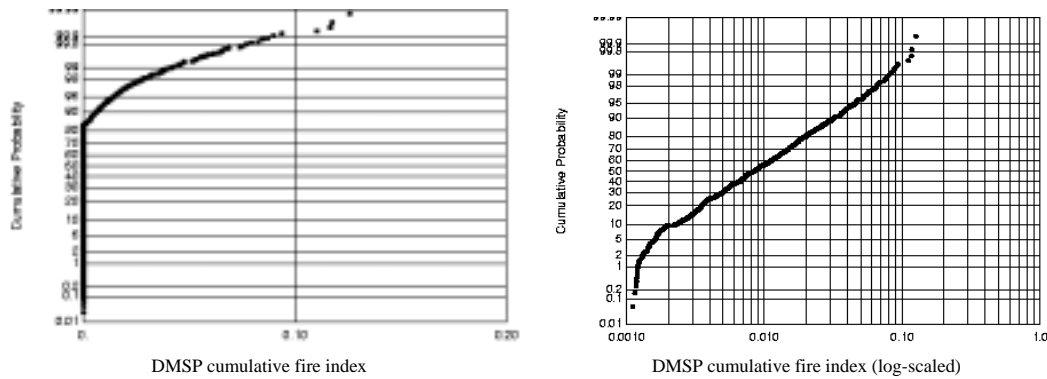


Figure 24: Normal probability plots of the DMSP-derived fire index values. A straight diagonal would indicate a normal distribution. Left: no scaling. The large number of zeros in the DMSP dataset is evident (dataset exhibits heavily right-skew). Right: log-scaling (zero-values excluded). The straight region on the log-scaled plot represents approximately log-normal behavior. The curves at each extremity indicate that the fire index distribution has truncated tails compared to a theoretical log-normal distribution.

### Quantile-quantile plots

An even sampling of reference data was not available for the entire island, so it was important to assess the representativeness of the reference region (path 158, row 73). A comparison of Figure 22 and Figure 23 shows the gross similarities between the distributions of the full extent of the DMSP-derived fire index and the subset covering just the study region co-extensive with Landsat reference scene. While the histograms appear similar, a more sensitive graphical comparison can be made by plotting the quantiles of the full and subsetted datasets against each other. This is shown by a quantile-quantile plot (Figure 25). Although the distribution of the full dataset is slightly more left-skewed than the distribution of the subset, it is evident that the difference is not extreme and the reference region is reasonably representative. This is important as all further assessments and extrapolations depend upon this assumption.

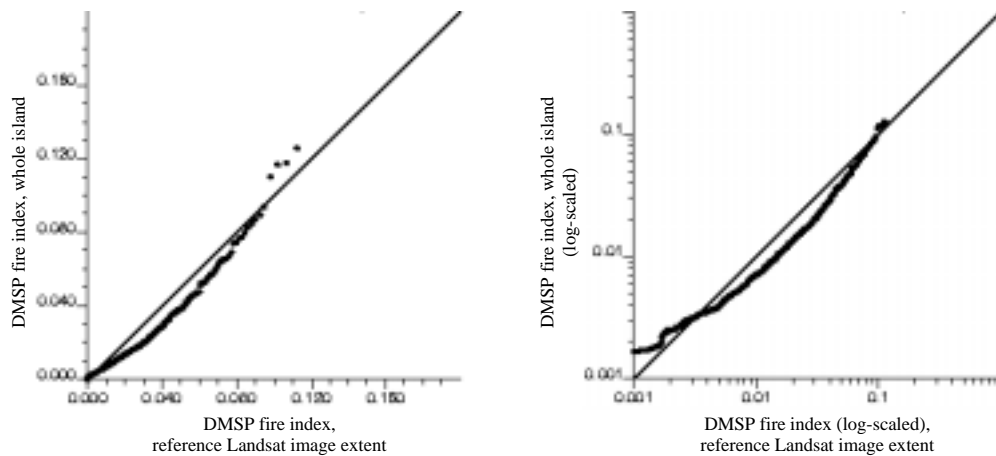


Figure 25: Unscaled and log-scaled quantile-quantile plots comparing the distribution of the DMSP-derived fire index values over just the extent of Landsat path 158 row 73 (x-axis) to the distribution of the DMSP-derived fire index values for the entire spatial extent of Madagascar DMSP data (y-axis). The diagonal line represents identically-shaped distributions. The log-scaled quantile-quantile plots exclude zero pixels, so the lower tail appears more skewed in the log-scaled plot.

### Linear regression analysis

Regressions of DMSP fire index vs. ETM-derived burn area fraction datasets at various resolutions are shown in Figure 26 and Figure 27. Slope, intercept, correlation, and significance values are included. A very gradual trend towards increasing correlation as resolution coarsens is observed, following a common pattern for spatial datasets (O'Sullivan and Unwin 2003). However, this result is highly dependent on small changes in resolution – if regression slope, r-squared, and p-value are plotted as a function of pixel resolution (Figure 28), it can be seen that drastic changes in regression parameters can occur with a “small” changes in pixel block size due to reconfiguration of the coarse pixels (e.g., moving from a 6x6 to a 5x5 coarse grid).

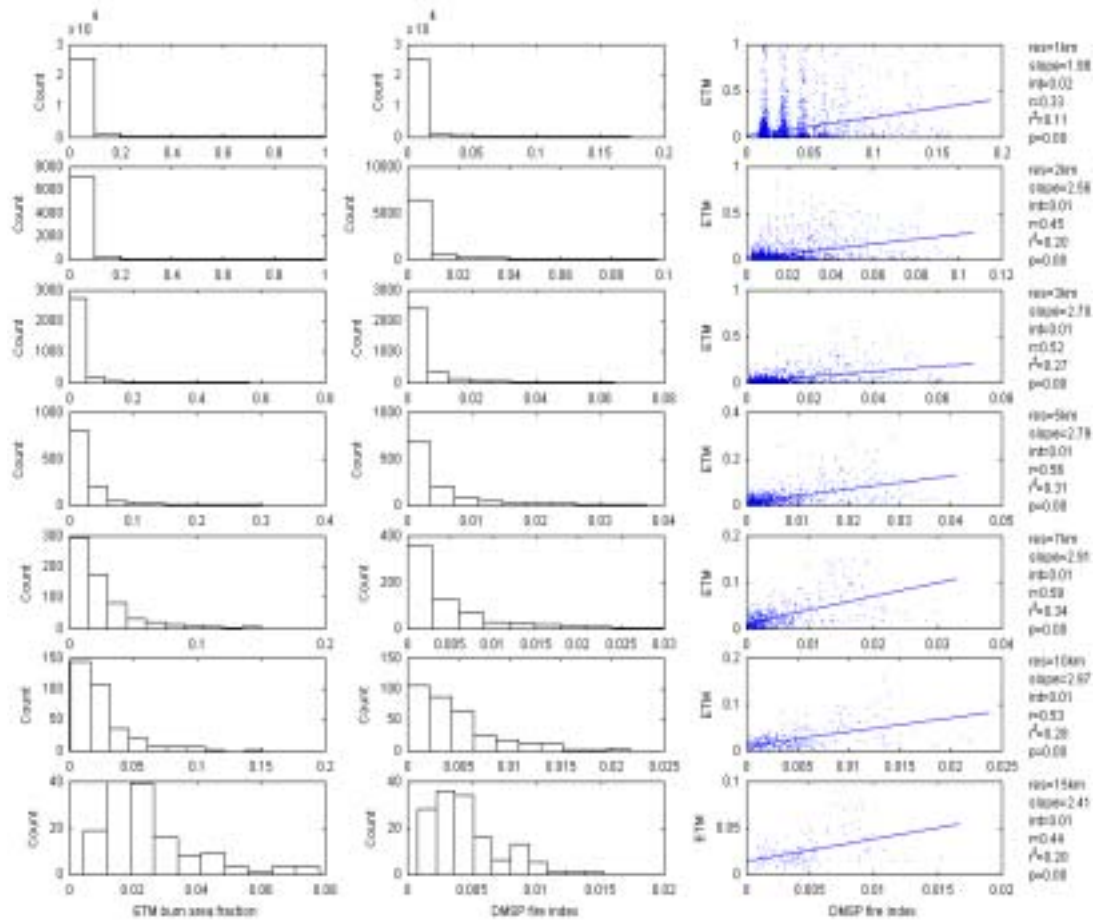


Figure 26: Histograms of ETM burn area, DMSP fire index, and regressions of DMSP predicting ETM burn area. Resolution ranges from 1-15 km.

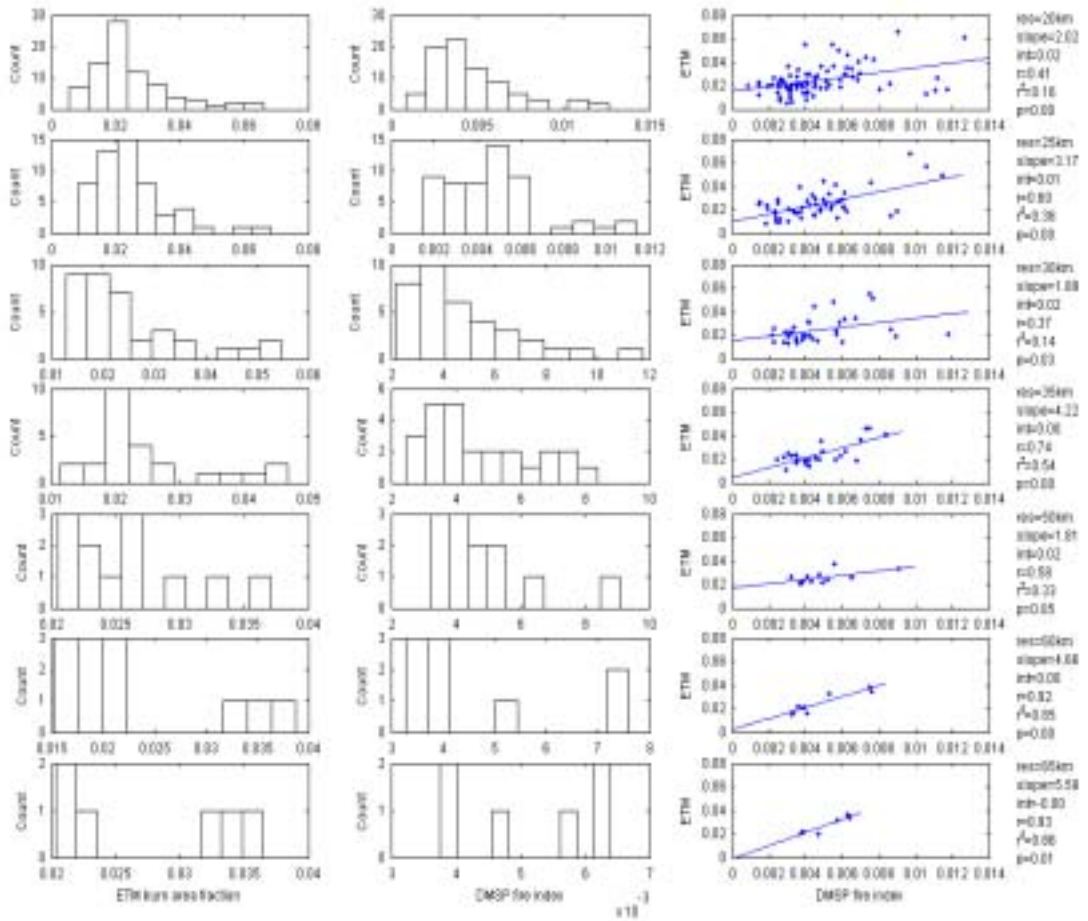


Figure 27: Histograms of ETM burn area, DMSF fire index, and regressions of DMSF predicting ETM burn area. Resolution ranges from 20-65 km.

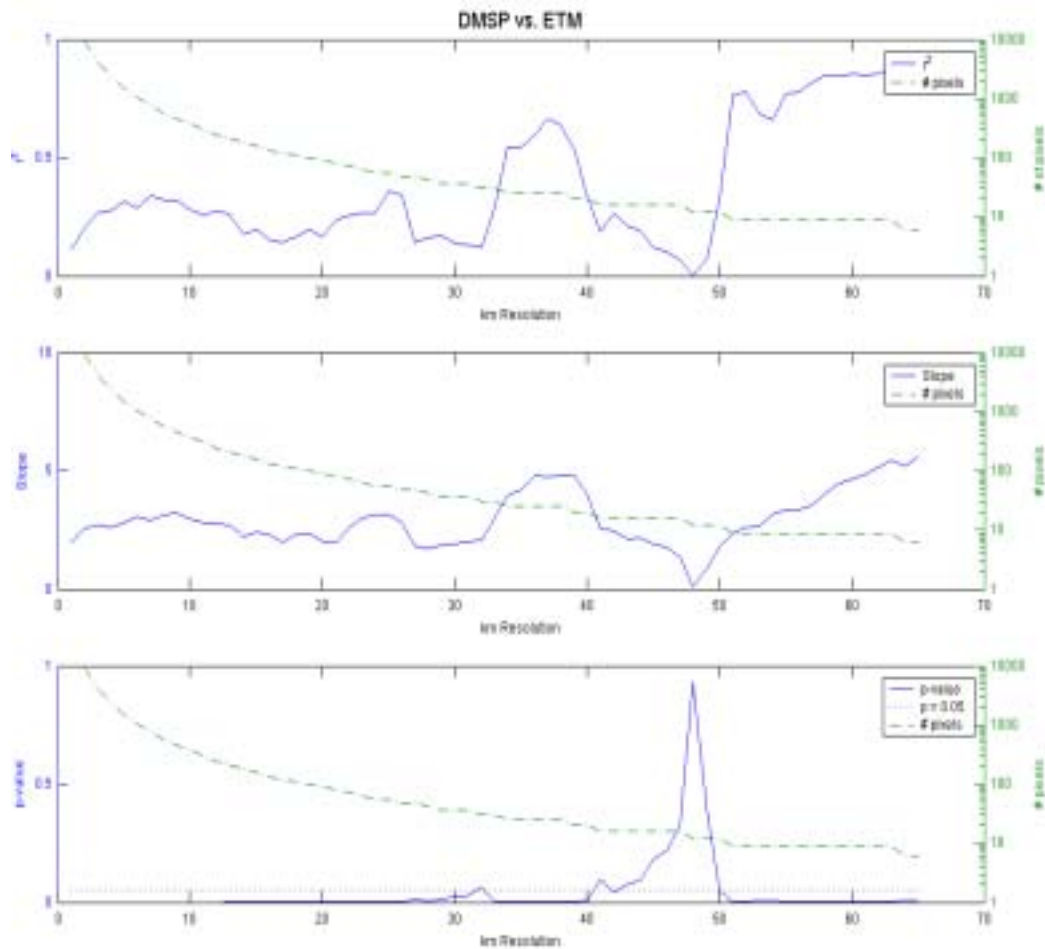


Figure 28: Plots of regression statistics (DMSP fire index predicting ETM burn area) as pixel resolution coarsens from 1 km – 65 km. **Top:** r-squared. **Middle:** slope. **Bottom:** p-value. The number of pixels is plotted with a dotted line on each plot using a log scale. Note that sudden changes in statistics co-occur with changes in pixel configuration (e.g., a switch from 4x4 pixels to 3x3 pixels). The spike in p-value (approaching 1.0, indicating no statistically significant relationship) occurs when no relationship is found in the regressions and the slope is flat. That such spikes occur as resolution coarsens is an example of the modifiable unit area problem.

## Geostatistical analysis

### Variogram maps and models

Variogram maps give a good impression of the spatial auto- and cross-correlation of the DMSP- and ETM- derived datasets in two dimensions. Figure 29 shows that both datasets exhibit significant anisotropy in a SSW-NNE direction. This is the

“direction of maximum continuity” and is about  $10^\circ$ . This reflects the orientation of the main mountain chain in Madagascar, as well as the parallel coastline. Fire occurrence is restricted by the ocean and by the remaining primary forest in the mountains, resulting in more variability perpendicular to these features.

Because of the anisotropy, variograms were calculated and modeled for the direction of maximum continuity (the direction of maximum spatial autocorrelation on variogram maps), and at increasing  $45^\circ$  increments ( $10^\circ$ ,  $55^\circ$ ,  $100^\circ$ , and  $145^\circ$ ). The empirical and modeled variograms for the direction of maximum continuity ( $10^\circ$ ) and minimum continuity ( $100^\circ$ ) are shown in Figure 29. The sill of all models is constrained to equal the dataset semivariance, therefore in some directions the model variogram will not track the empirical variogram at long lags (where the empirical variogram temporarily exceeds the dataset semivariance). The parameters of the four variogram structures used in the Linear Model of Coregionalization, and the resulting structural correlation coefficients, are shown in Table 11. The full equations for the models can be derived by plugging the range and sill contributions from Table 11 into equations (6) and (7).

Although the LMC structures were not incorporated into the simulation procedure, the structural correlation coefficients are useful results on their own accord.

Although overall correlation between datasets is consistently low, the structural correlation coefficients show that this is due to low or negative correlation at fine

scales obscuring the more impressive correlation at coarse scales. This corroborates the conclusion reached in the linear regression analysis, above.

As the simulations are conducted in normal-score space, the variograms were calculated and modeled for the normal-score transformed data. These are shown in Figure 30. These variograms were modeled with the same structures (nugget, 2 sphericals, Gaussian) that were used for the untransformed data. While the compromises required by the LMC procedure resulted in some inexactness for the models, a good overall fit was achieved. The transformed datasets show higher relative nugget contributions than the original datasets, primarily due to the normal-score transform assigning different normal scores to values that were identical in the original dataset, increasing relative variability. As the normal-score transform standardizes the dataset variance to  $\sim 1$ , the problem of the empirical variogram exceeding the sill at long ranges did not arise. As with the non-transformed datasets, the cross-variogram showed a much slower rise than the autovariograms, indicating increased cross-correlation at the very coarse scales. The Linear Model of Coregionalization parameters are shown in Table 12.

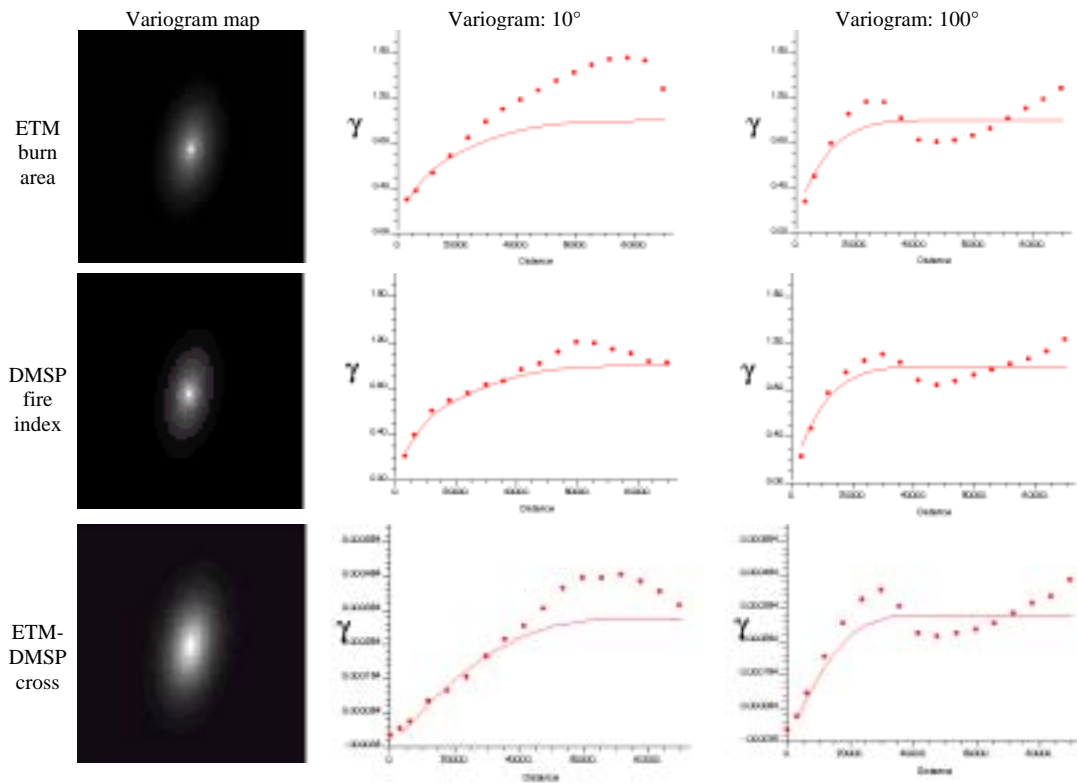


Figure 29: Variogram maps and variograms showing anisotropy in spatial correlation in 2-D. *Variograms*: Dots represent empirical variograms; lines represented the fitted Linear Model of Coregionalization. For the auto-semivariograms, the semivariance has been normalized so that the dataset variance = sill = 1.

Table 11: Parameters for Linear Model of Correlation variogram structures, and calculated of structural correlation coefficients. Overall raw data correlation coefficient = **0.52**.

		<b>Nugget</b>	<b>Short-range spherical structure</b>	Max range	Min range
<b>ETM burn area</b>	<i>% sill contrib.</i>	15%	12%	5	5
	absolute contrib.	0.000575	0.000460		
<b>DMSP fire index</b>	<i>% sill contrib.</i>	9%	20%	5	5
	absolute contrib.	0.000016	0.000036		
ETM-DMSP cross	<i>% sill contrib.</i>	-10%	0%	5	5
	absolute contrib.	-0.000038	0.000000		
Structural correlation coefficient		-0.40	0.00		

		<b>Medium-range spherical structure</b>	Max range	Min range	<b>Long-range Gaussian structure</b>	Max range	Min range
<b>ETM burn area</b>	<i>% sill contrib.</i>	43%	12	6	30%	20	10
	absolute contrib.	0.001648			0.001150		
<b>DMSP fire index</b>	<i>% sill contrib.</i>	39%	12	6	32%	20	10
	absolute contrib.	0.000070			0.000057		
ETM-DMSP cross	<i>% sill contrib.</i>	45%	12	6	60%	20	10
	absolute contrib.	0.000171			0.000228		
Structural correlation coefficient		0.51			0.89		

\* ranges are in terms of 3-km pixels

\*\* direction of maximum continuity = 10 degrees, direction of minimum continuity = 100 degrees

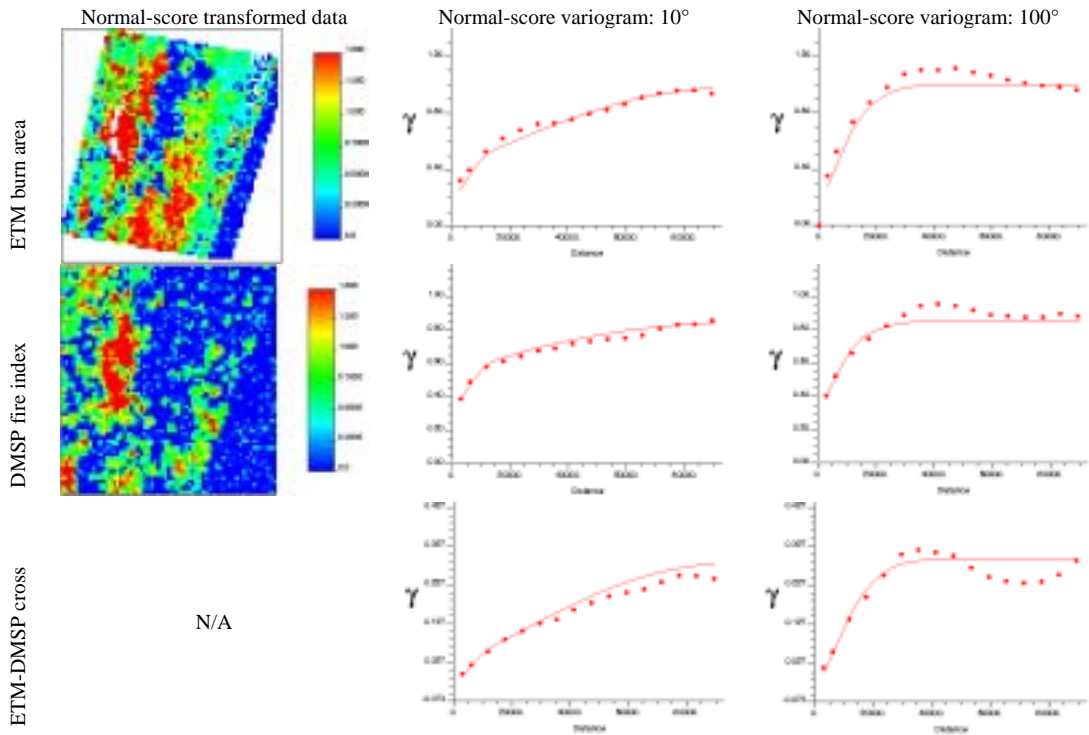


Figure 30: Empirical (*dots*) and modeled (*lines*) variograms for the normal-score transformed datasets. The maps of the normal-score transformed datasets are shown on the left. Normal-score values range from about  $-3$  to  $+3$ , the colormap has been scaled to ease interpretation.

Table 12: Parameters for LMC structures, and calculated of structural correlation coefficients, for normal-score transformed datasets. Overall normal-scored correlation coefficient = **0.42**.

		<b>Nugget</b>	<b>Short-range spherical structure</b>	<b>Max range</b>	<b>Min range</b>
<b>ETM burn area</b>	<i>% sill contrib.</i>	35%	30%	5	5
	absolute contrib.	0.3223	0.2762		
<b>DMSP fire index</b>	<i>% sill contrib.</i>	15%	30%	5	5
	absolute contrib.	0.1499	0.2999		
<b>ETM-DMSP cross</b>	<i>% sill contrib.</i>	-10%	20%	5	5
	absolute contrib.	-0.0403	0.0806		
<b>Structural correlation coefficient</b>		<b>-0.18</b>	<b>0.28</b>		

		<b>Medium-range spherical structure</b>	<b>Max range</b>	<b>Min range</b>	<b>Long-range Gaussian structure</b>	<b>Max range</b>	<b>Min range</b>
<b>ETM burn area</b>	<i>% sill contrib.</i>	10%	12	6	25%	30	10
	absolute contrib.	0.0921			0.2302		
<b>DMSP fire index</b>	<i>% sill contrib.</i>	10%	12	6	45%	30	10
	absolute contrib.	0.1000			0.4498		
<b>ETM-DMSP cross</b>	<i>% sill contrib.</i>	10%	12	6	60%	30	10
	absolute contrib.	0.0403			0.2418		
<b>Structural correlation coefficient</b>		<b>0.42</b>			<b>0.75</b>		

\* ranges are in terms of 3-km pixels

\*\* direction of maximum continuity = 10 degrees, direction of minimum continuity = 100 degrees

### Stochastic simulation

Examples of the simulated realizations of burn area maps are shown in Figure 31.

The simulations reproduce the coarse-scale pattern of the DMSP fire index, but have the spatial variability of the reference ETM-derived burn area map. The substantial variation between each simulation realization reflects the uncertainty in DMSP-based predictions. Histograms of selected simulation runs were generated in order to

check that the simulation algorithm reproduced the lognormal distribution of the reference ETM burn area map (an exact reproduction of the reference histogram is not expected due to the influence of the secondary DMSP dataset on the simulation); an example is shown in Figure 32. The simulation algorithm's reproduction of the spatial variability of the reference dataset is confirmed by comparison of a simulation variogram to the LMC-specified model variogram for the reference dataset (Figure 33).

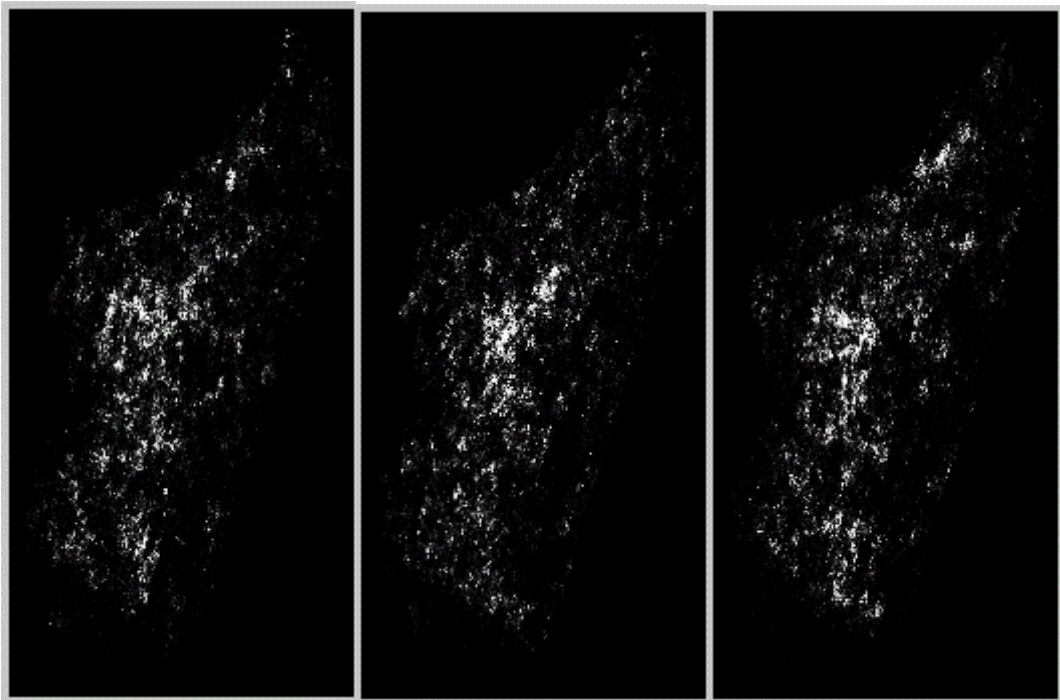


Figure 31: Example simulated burn area maps for Madagascar, 8/1/00-12/31/00.

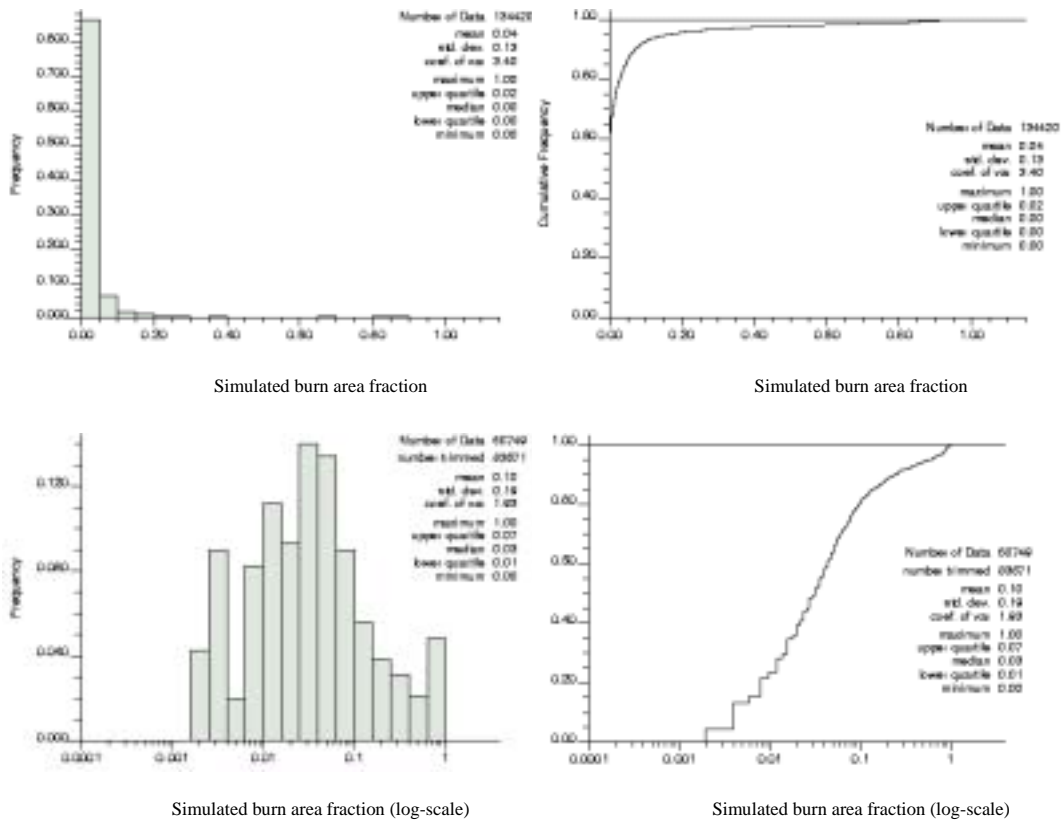


Figure 32: Example histograms showing the distribution of burn area values on a single simulated realization of a burn area map. *Top*: regular histogram and cumulative histogram. *Bottom*: Log-transformed histogram and cumulative histogram. This confirms that the distribution of simulated values on a burn area map realization are approximately lognormal, as expected from the reference data..

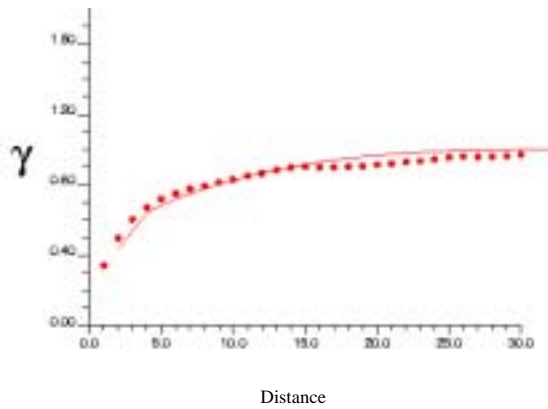


Figure 33: Variogram reproduction. The empirical variogram of a single simulated burn area map (*dots*) compared to the LMC-specified model (*line*). The variograms is calculated in the 0° direction as calculation of the variogram at 10° is very computationally expensive for the large simulated grid

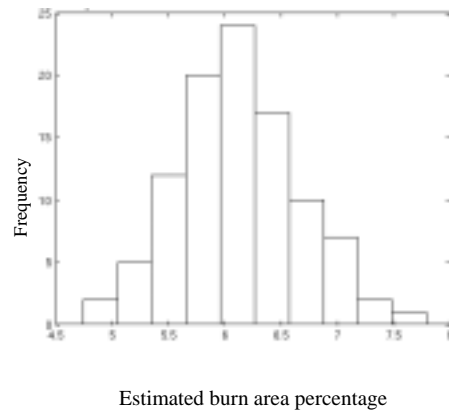


Figure 34: Distribution of whole-country percent burn area estimates from 100 simulated burn area maps for the 2000 fire season.

For each polygon of interest, a distribution of burn area estimates was produced; an example of such a distribution is shown in Figure 34. The distribution of burn area estimates gives a measure of the uncertainty of the prediction. These distributions of outcomes tended to be approximately normal; this is essentially an expression of the Central Limit Theorem, where the distribution of a sample statistic (such as the total burn area of a polygon) will be normal after repeated sampling (O'Sullivan and Unwin 2003). The normal distributions of simulated outcomes allow the burn area estimates to be adequately summarized by reporting their mean and standard deviation.

*Comparison of simulation results to validation scenes.*

The burn area estimates for the non-obscured regions of the Landsat scenes are shown in Table 13. Because of the uncertainty in DMSP-based estimates of burn area even for areas the size of a Landsat scene, the only comparisons that could reasonably be made were for the overall estimates for each scene. The DMSP-based simulated burn area falls well within 1 standard deviation of the ETM-mapped burn area for each of the two validation scenes, as well as the reference scene. This constitutes a limited validation of the simulation procedure. It is important to note that this result was achieved despite the fact that the mean simulated burn area for a Landsat footprint will be influenced not only by the DMSP fire index values within the scene, but also the surrounding DMSP data. However, the high uncertainty in the simulation estimates -- the coefficients of variation are 1.6, 0.43, and 0.59 (respectively, moving down the table) -- is cause for some concern. A coefficient of variation of 0.51 or greater indicates that the mean is not statistically significantly different than 0 at the 95% confidence level (if the standard deviation is  $> 0.51$  of the mean, then 1.96 times the standard deviation will equal or exceed the mean).

Table 13: Burn areas estimated from the DMSP-based simulations, compared to the reference burn areas as mapped by classification of Landsat ETM image pairs.

DMSP data date range	# of dmsp obs.	Time span	Landsat ETM+ image pair <i>path row</i>	ETM-mapped burn area	Total area of mapped ETM scene	DMSP-mapped burn area	DMSP Std. Dev.	ETM % area burned	DMSP % area burned	DMSP % standard deviation
10/17/99-11/11/99	25	25 days	158-159 75	20.3 km <sup>2</sup>	4617 km <sup>2</sup>	13.9	22.1 km <sup>2</sup>	<b>0.44%</b>	<b>0.30%</b>	± 0.48%
9/8/1999-11/11/99	63	64 days	158 77	154.9 km <sup>2</sup>	17208 km <sup>2</sup>	149.8	62.9 km <sup>2</sup>	<b>0.90%</b>	<b>0.87%</b>	± 0.37%
8/1/00-10/28/00	106	89 days	158 73	677.8 km <sup>2</sup>	27441 km <sup>2</sup>	573.4	337.2 km <sup>2</sup>	<b>2.47%</b>	<b>2.09%</b>	± 1.23%

*Estimated burn areas for Madagascar*

The estimated burn areas for the entire island of Madagascar in 1999 and 2000 are given in Table 14. The relative uncertainty in the estimates for this relatively large area is quite low – the coefficient of variation for the estimates equals 0.08 in 1999 and 0.06 in 2000. These estimates fall squarely between the average of government statistics (2.2% burned annually) and Jolly’s figure of 25% or higher. The estimate of 6-7% is similar to the highest burn area ever recorded in government statistics, 5.9% in 1973, which may reflect more complete data collection.

Table 14: Total area burned for Madagascar, estimated from 100 DMSP-based simulations, for the DMSP observation periods in 1999 and 2000.

DMSP data date range	# of obs.	Time span	Total area of Madagascar	DMSP-mapped burn area	DMSP Standard Deviation	DMSP % area burned	DMSP % standard deviation
07/31/1999-12/31/1999	142	154 days	591948.0 km <sup>2</sup>	41546.5	3365.8 km <sup>2</sup>	7.02%	0.57%
8/1/2000-12/31/2000	169	153 days	591948.0 km <sup>2</sup>	35781.4	3183.1 km <sup>2</sup>	6.04%	0.54%

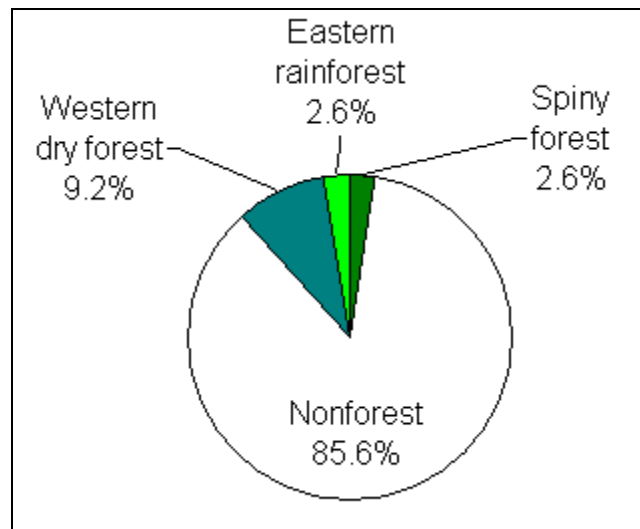


Figure 35: Percent contribution of each JERS-mapped landcover (Figure 19) to total burn area for 2000 fire season.

#### *Burn area by landcover type*

The simulated percent burn area for each broad landcover category mapped by JERS is shown in Table 15. As expected, burn area in the nonforest class, which represents 72% of the area of Madagascar, is the dominant influence on burn area totals for the island. The disproportionate influence of nonforest burning on the total

burn estimate for Madagascar is shown in Figure 35. The relatively tiny contribution of rainforest fires to the total is indicative of the difficulty of mapping the fires that are of greatest importance for conservation. Furthermore, given the above described results concerning the scale-dependent nature of DMSP predictions, the question arises of whether burn area estimate for rainforest reflects actual burn area in this class, or the influence of fires just outside of the rainforest. Rainforest in Madagascar is a very linear feature, with a high perimeter-to-area ratio. The coarse resolution of DMSP in many cases would fail to distinguish between fires inside or outside of the forest in many cases. However, despite this, overlaying the DMSP fire index map on the landcover map reveals clear negative correlation between forest and fires at a coarse scale (Figure 36).

Table 15: Total area burned by JERS-classified landcover, estimated from 100 DMSP-based simulations, for the DMSP observation periods in 1999 and 2000.

JERS landcover classification	Total area of regions	1999 DMSP-mapped burn area	1999 DMSP Standard Deviation	1999 DMSP % area burned	1999 DMSP % standard deviation
Spiny forest	47178 km <sup>2</sup>	1211.1	338.6 km <sup>2</sup>	2.57%	0.72%
Nonforest	422478 km <sup>2</sup>	35192.4	2864.6 km <sup>2</sup>	8.33%	0.68%
Western dry forest	61182 km <sup>2</sup>	3937.6	613.4 km <sup>2</sup>	6.44%	1.00%
Eastern rainforest	59805 km <sup>2</sup>	1130.3	296.1 km <sup>2</sup>	1.89%	0.50%
<b>TOTAL (land)</b>	<b>590643 km<sup>2</sup></b>	<b>41471</b>	<b>2172.2 km<sup>2</sup></b>	<b>7.02%</b>	<b>0.37%</b>

JERS landcover classification	Total area of regions	2000 DMSP-mapped burn area	2000 DMSP Standard Deviation	2000 DMSP % area burned	2000 DMSP % standard deviation
Spiny forest	47178 km <sup>2</sup>	933.8	286.4 km <sup>2</sup>	1.98%	0.61%
Nonforest	422478 km <sup>2</sup>	30541.5	2700.2 km <sup>2</sup>	7.23%	0.64%
Western dry forest	61182 km <sup>2</sup>	3276.6	548.3 km <sup>2</sup>	5.36%	0.90%
Eastern rainforest	59805 km <sup>2</sup>	913.6	254.3 km <sup>2</sup>	1.53%	0.43%
<b>TOTAL (land)</b>	<b>590643 km<sup>2</sup></b>	<b>35666</b>	<b>2245.3 km<sup>2</sup></b>	<b>6.04%</b>	<b>0.38%</b>

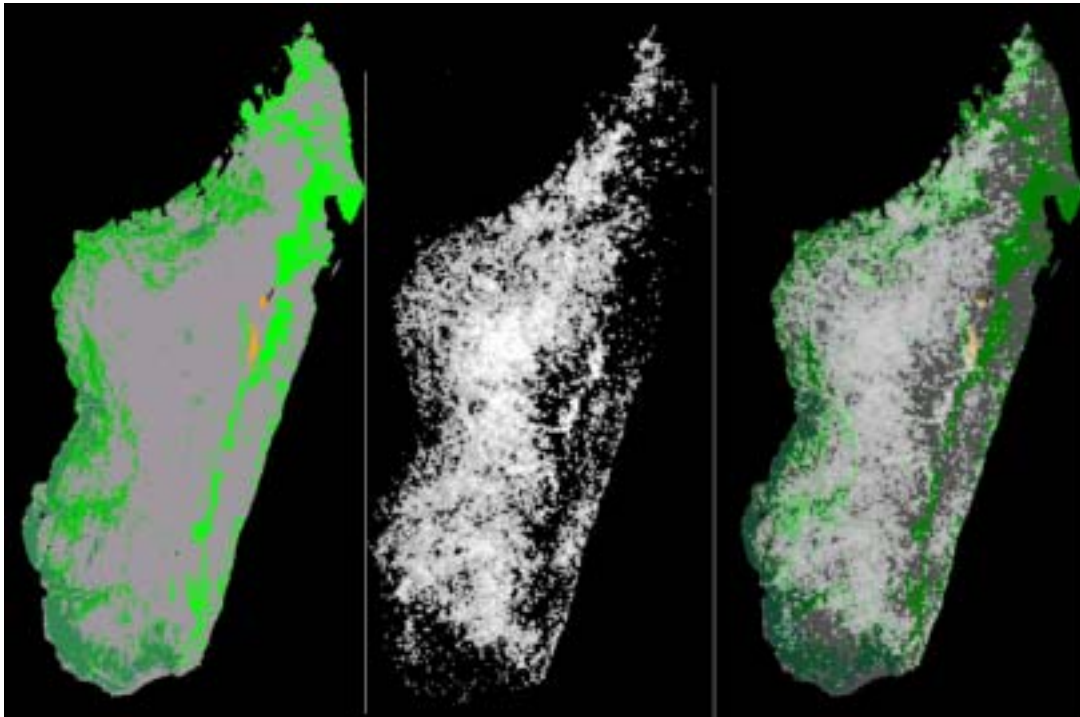


Figure 36: Negative correlation of JERS-mapped forest (*left*) and DMSP fire index (*center*). The two maps are shown overlaid at right.

Estimation of annual burn area also yields, via inversion, an estimate of fire return interval. If 7-8% of the nonforest landcover burns in a year, then the fire return interval would be about 13 years. This is a reasonable return interval for grassland, however it can be expected that the return interval is much shorter in the central highland provinces where DMSP-recorded fire activity is most intense. Naïve interpretation of a 1.5-1.9% annual burn rate for eastern rainforest (Table 15) would imply that the entire remaining rainforest would be burned over in 50-65 years. This is a very generous estimate according to much of the conservation literature, however such estimates are commonly incorrect (e.g. Terborgh (1992) predicted that Madagascar's forests would be gone within 10 years). In any case, burn area

estimates based on DMSP nighttime fire detections do not take into account forest loss due to number of other processes, e.g. logging, nor the influence of the many conservation projects implemented in the 1990's.

The monthly landcover burn area estimates derived from simulations using the monthly DMSP fire indices as inputs are shown in Figure 37. The monthly statistics reproduce the expected pattern, where the eastern rainforest experiences a late burning peak in October/November, whereas for other landcovers burning peaks in September-October.

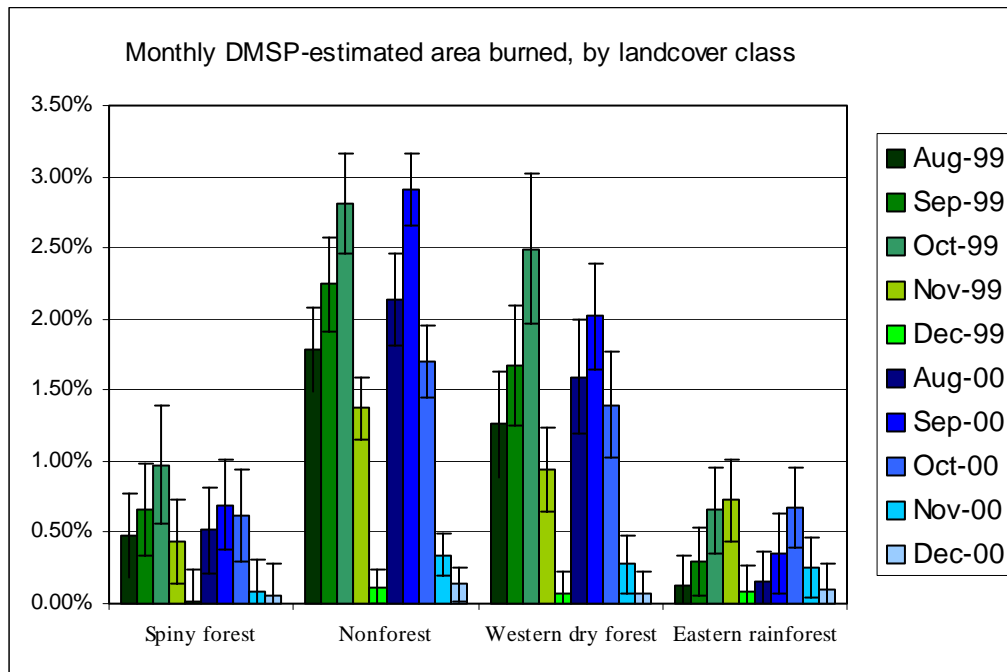


Figure 37: Monthly DMSP-estimated area burned, by landcover class. Bars indicate 1 standard deviation in the estimate.

*Burn area by province*

The burn area estimates for each of the six major (level 1) provinces are shown in Table 16. In both 1999 and 2000, the central highland province of Antananarivo had the highest burn area percentage; however, in terms of contribution to total annual burn area, the larger, western Toliara and Mahajanga provinces contribute about the same amount of burn area as Antananarivo (see Figure 38). The percent burn area estimates for 133 districts (level 2 provinces, see Figure 1) are shown in Figure 39. The monthly burn area by district is shown in map form (Figure 40). The west-to-east progression throughout the fire season can be seen at the district level, as well as the concentration of burning in the central highlands.

Table 16: Estimated annual burn area for 1999 and 2000, by level 1 province. See also Figure 38.

Level 1 province	Total area of regions	1999 DMSP-mapped burn area	1999 DMSP Standard Deviation	1999 DMSP % area burned	1999 DMSP % standard deviation
Mahajanga	153207 km <sup>2</sup>	11518.6	1639.2 km <sup>2</sup>	7.52%	1.07%
Antananarivo	58662 km <sup>2</sup>	11573.7	1764.8 km <sup>2</sup>	19.73%	3.01%
Toliara	163980 km <sup>2</sup>	7646.9	1335.9 km <sup>2</sup>	4.66%	0.81%
Fianarantsoa	100746 km <sup>2</sup>	6960.2	1220.2 km <sup>2</sup>	6.91%	1.21%
Toamasina	71370 km <sup>2</sup>	2520.2	748.1 km <sup>2</sup>	3.53%	1.05%
Antsiranana	43452 km <sup>2</sup>	1389.8	549.4 km <sup>2</sup>	3.20%	1.26%

Level 1 province	Total area of regions	2000 DMSP-mapped burn area	2000 DMSP Standard Deviation	2000 DMSP % area burned	2000 DMSP % standard deviation
Mahajanga	153207 km <sup>2</sup>	10090.9	1560.0 km <sup>2</sup>	6.59%	1.02%
Antananarivo	58662 km <sup>2</sup>	8519.6	1497.2 km <sup>2</sup>	14.52%	2.55%
Toliara	163980 km <sup>2</sup>	8392.6	1389.0 km <sup>2</sup>	5.12%	0.85%
Fianarantsoa	100746 km <sup>2</sup>	6223.4	1192.1 km <sup>2</sup>	6.18%	1.18%
Toamasina	71370 km <sup>2</sup>	1909.8	651.9 km <sup>2</sup>	2.68%	0.91%
Antsiranana	43452 km <sup>2</sup>	901.6	422.4 km <sup>2</sup>	2.07%	0.97%

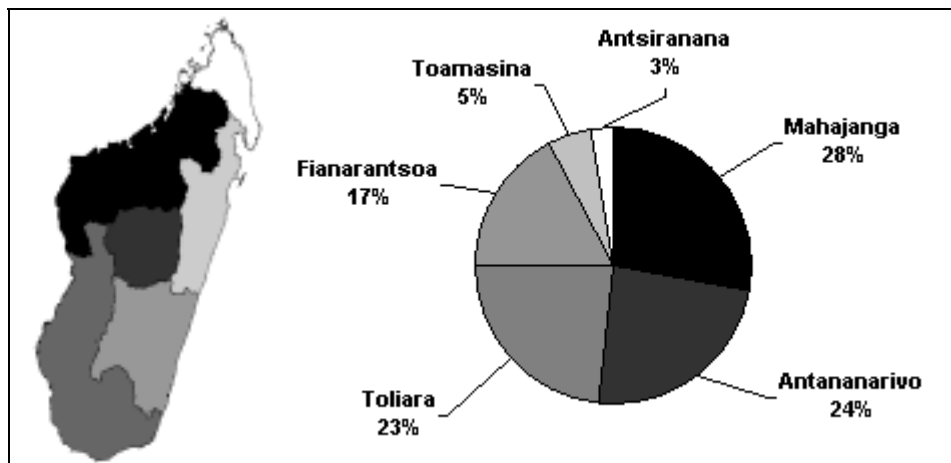
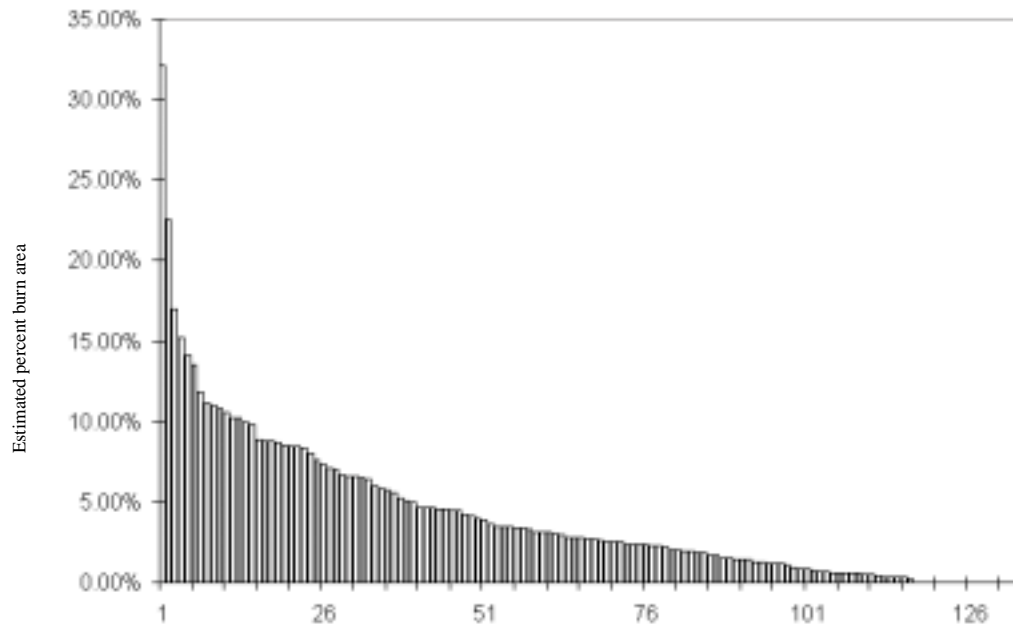


Figure 38: Relative contribution of each level 1 province to total Madagascar burn area in 2000. The shading of the each province on the map corresponds to the shading on the pie chart.



Ranking of district in terms of percent burn area

Figure 39: Distribution of percentage burn areas for 133 Madagascar districts (level 2 provinces, see Figure 1) in 2000.

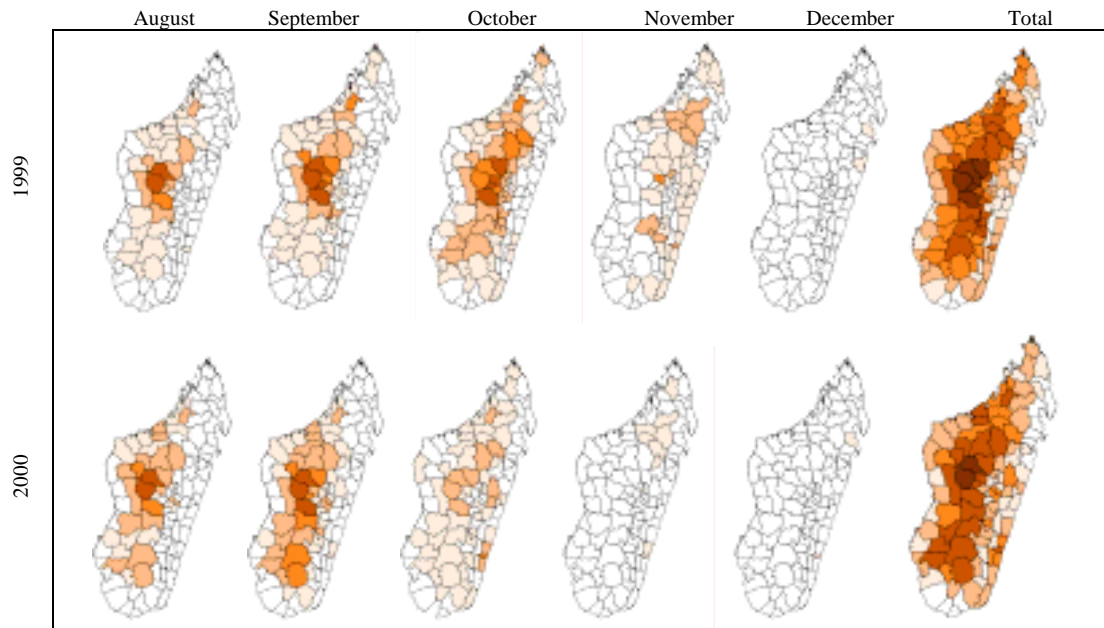
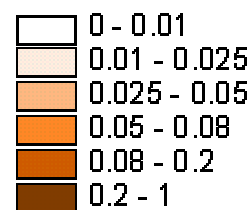


Figure 40: Estimated monthly fraction burn area for Madagascar districts, 1999 and 2000. *Right:* The cumulative fraction area burned for each year.



#### *Burn area by protected areas polygon*

The burn area estimates for 280 protected areas polygons (forests classified for some form of management, with varying degrees of protection) are mapped in Figure 41, and the distribution of burn area estimates in 2000 for all of the polygons is shown in Figure 42. This product has potential for identifying polygons of particular concern regarding burning. For example, the top ten protected areas polygons in terms of DMSP-estimated burn area are shown in Table 17. However, the coefficient of variation for polygons this small is often 1 or higher, and always higher than 0.51;

therefore, based on the simulations, the uncertainty in burn area estimates for these polygons precludes the drawing of any conclusions about which forest polygons actually experience the most burning. Similarly, based on the mean and standard deviation of burn area estimates, a z-score can be calculated for every protected areas polygon in the dataset, in order to test for significant changes in burn area for each polygon between 1999 and 2000 (Figure 43); however, due to the high uncertainty, none of the changes can be considered statistically significant. Preserve-size polygons are probably too small to be effectively monitored with the DMSP-based technique used in this study.

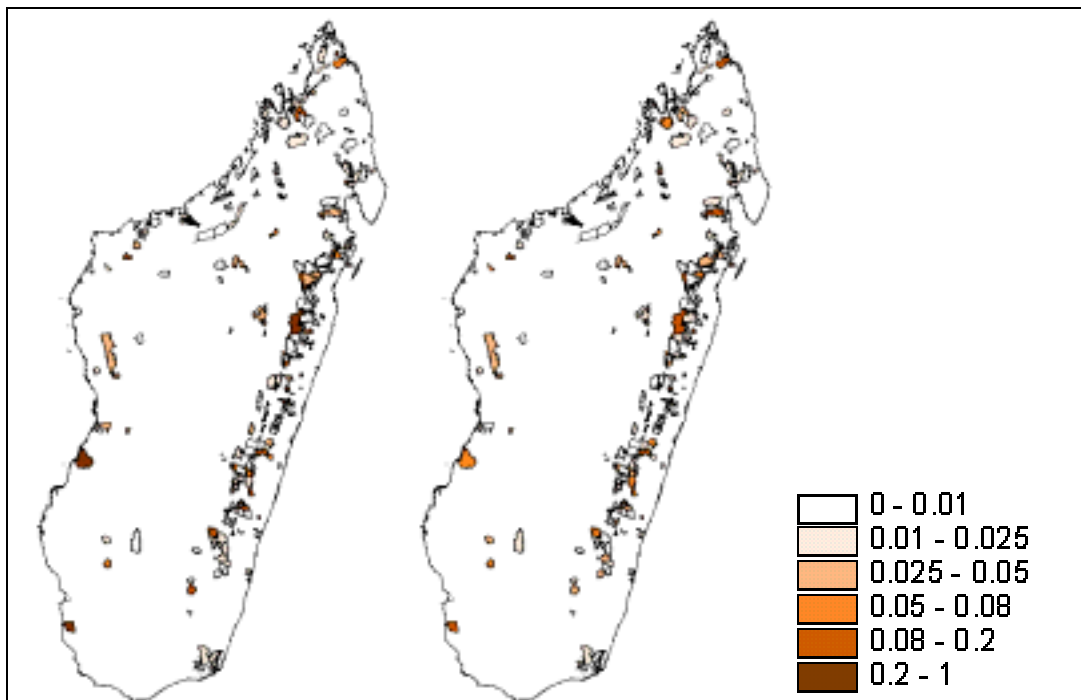


Figure 41: Estimated burn area for protected areas polygons in the AIRE\_PRO database (data layer supplied by James Rowland, USGS). *Left: 1999. Right: 2000.*

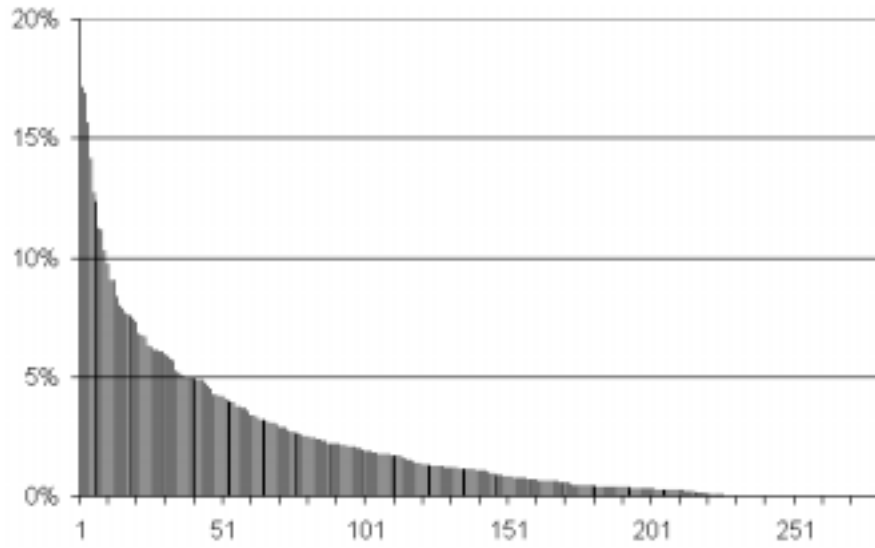


Figure 42: Distribution of burn area estimates for the year 2000 fire season for 280 protected areas polygons.

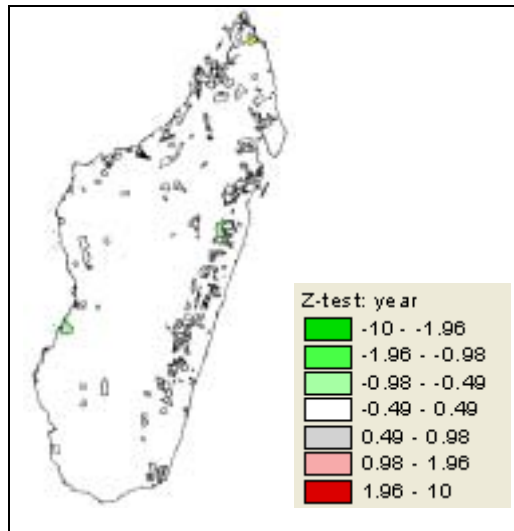


Figure 43: Z-score map testing for significant differences in burn area in protected areas polygons between 1999 and 2000. Z-scores  $> 1.96$  or  $< -1.96$  would be significant changes, but none of the differences are significant between these two years.

Table 17: Top ten protected areas polygons for 1999 and 2000 as ranked by estimated burn area.

Name	Classification	Area (sq. km)	1999 Burn area	Std. Dev.	1999 % burn area	% std. dev.	C.V.
Befato	classified forest	54.0	17.9	13.7	33%	25%	0.77
Bora	classified forest	36.0	10.4	10.3	29%	29%	0.99
Antsinjomorony	classified forest	180.0	51.4	33.6	29%	19%	0.65
Andongonambo	classified forest	72.0	19.4	17.1	27%	24%	0.88
Amboromailala	classified forest	396.0	101.7	62.1	26%	16%	0.61
Ankazomivady	classified forest	9.0	2.1	2.7	23%	30%	1.30
Masiapoza	classified forest	90.0	17.7	17.9	20%	20%	1.01
Andanivatsy	classified forest	27.0	4.4	5.7	16%	21%	1.29
Tsinjoarivo	reserve forest	153.0	24.0	25.1	16%	16%	1.04
Forêt d'Ambre	special reserve	36.0	5.6	7.2	16%	20%	1.30

Name	Classification	Area (sq. km)	2000 Burn area	Std. Dev.	2000 % burn area	% std. dev.	C.V.
Amboromailala	classified forest	396.0	73.0	50.5	18%	13%	0.69
Andongonambo	classified forest	72.0	12.3	13.3	17%	18%	1.08
Vohitrambo (1)	classified forest	18.0	3.0	4.0	17%	22%	1.32
Anjiamangirana	classified forest	144.0	22.5	22.9	16%	16%	1.02
Andavakoera	classified forest	117.0	16.6	18.5	14%	16%	1.12
Ambila et Sahamasina	classified forest	18.0	2.3	3.1	13%	17%	1.36
Bora	reserve forest	171.0	21.1	23.3	12%	14%	1.10
Bora	special reserve	54.0	6.1	9.3	11%	17%	1.53
Vohitrambo (2)	reserve forest	90.0	10.0	14.3	11%	16%	1.43
Betafo	classified forest	54.0	5.5	7.9	10%	15%	1.42

## Chapter 5: Discussion

### **Would the DMSP-OLS fire index be expected to correlate with actual fire counts rather than burn area?**

An analysis of what might be expected from a nighttime, coarse-resolution sensor such as DMSP-OLS will be useful in assessing the results of this study. The raw DMSP-OLS fire product is fundamentally a nightly map of detected active fires, which is then converted into a cumulative detected fire count map, and finally to a fire index. The naïve expectation might therefore be that a fire index would correlate with *fire counts* better than *burn area*. However, a consideration of the various expected biases of DMSP fire detection capability indicates that burn area is actually more likely to be mapped.

*Factors favoring overestimation of actual fire counts:* There are a number of reasons why an active fire detection sensor might overestimate actual fire count. These are well known and acknowledged in the literature, and support the conventional wisdom that DMSP drastically overmaps fire.

As discussed by Kihn (1996), the PMT used for the nighttime VNIR imagery actually measures photons from a ~1.75 km IFOV, essentially a 3x3 block of 'fine' (~0.54 km) pixels, and assigns this measurement to each 0.54-km pixel. As these data are generally smoothed into ~2.7 km 'smooth' pixel, a single fire on the edge of a 2.7 km pixel will be recorded on adjacent 2.7 km pixels. This is a significant factor 'blowing up' many detected fires.

Leaving aside the IFOV issue, a single large or bright fire could be detected in two or more pixels, due to several factors:

- Atmospheric scattering of VNIR light will spread out a signal.
- Haze, light clouds, and smoke (particularly important as where there is fire, there's smoke) will further scatter the signal.
- The same fire, burning over multiple nights, could be repeatedly detected on multiple DMSP overpasses. If the fire front moved several kilometers per day, certainly a possibility in grassland situations, the fire could be detected on multiple pixels. This situation could not be distinguished from multiple, separate fires occurring in nearby pixels on consecutive nights.

An additional factor is that the resampling of pixels from 2.7 km to 1 km will, if fire occurrence per 1 km pixel is tabulated, count one detected fire in each of several 1 km pixels.

*Factors favoring underestimation of actual fire counts:* Although this is less often considered in the literature and in discussions, many characteristics of the DMSP-OLS data would favor *underestimation* of fire occurrence. Active fires could be *missed* by DMSP for any number of reasons, including:

- Fires can be missed because they are burning for only a short period of time, and are therefore less likely to be detected during the DMSP satellite overpass.
- Fires could be missed because of cloud cover or (less likely) heavy smoke, haze, or fog cover. Fires under tree canopies may be similarly obscured.
- Fires will, for reasons of decreased temperature, increased atmospheric humidity, and human diurnal activity, die down at night, reducing the potential radiance signal.
- Multiple fires within a single pixel will only be recorded as a single active fire.
- Fires could simply be too small or cool to reach the detection threshold of the DMSP-OLS. Thus it would be expected that the DMSP-OLS would be biased against the detection of small fires.

An interesting point that emerges from considering the biases of the DMSP-OLS platform is that an argument can be made that a DMSP-OLS fire index would correlate better with mapped burn area, rather than mapped fire scar count.

- Fires of large spatial extent will give off a larger raw radiance signal.
- Fires burning across a large area over multiple nights may be detected on multiple adjacent DMSP pixels.
- A large fire, particularly one with an active front extending over kilometers, has an improved chance of detection due to the greater

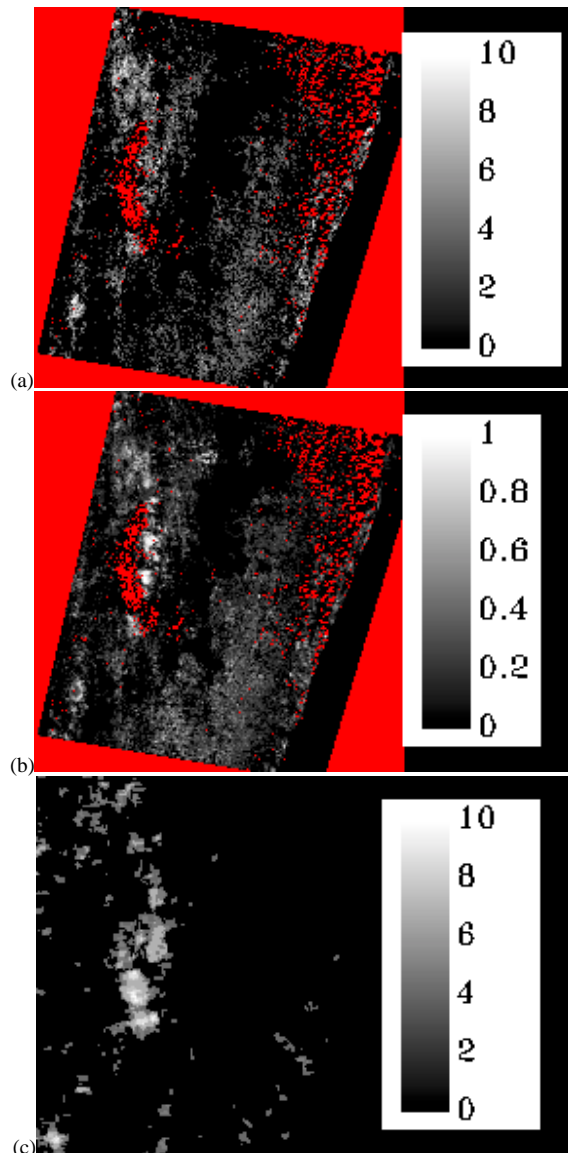


Figure 44: Comparison of actual burn counts, actual burn area fraction, and DMSP burn count maps. (a) Reference burnscar count map, created by counting the number of ETM-mapped burn scar centroids in each 1km pixel. (b) ETM-mapped burn area fraction map, 1km resolution. (c) DMSP fire count map. The count values for (a) and (c) have been given the same stretch so that the counts are comparable. Nodata regions are red.

spatial extent of the signal -- a large fire would be less likely to be obscured by clouds, as there would be an increased chance of having clear sky over at least some portion of the fire.

- A fire of large spatial extent is less likely to be an actively managed fire. As small-area fires are likely more-or-less controlled burns, for example to clear weeds or brush off of an area to facilitate the planting of crops. *Tavy* fires, in which a small amount of forest is burned, could fall into this observational category as well. Controlled fires may often be lit during the day but have been put out by nightfall, their purpose having been accomplished. These fires would thus be completely removed from the DMSP observation window. Uncontrolled fires, on the other hand, will almost by definition burn until they are extinguished by natural factors, and therefore would be more likely to be active at night and over multiple nights, increasing the odds of detection by DMSP.
- The vigor and flame height of an active fire is influenced by numerous factors, such as temperature, humidity, vegetation composition and water content, dew and precipitation, slope, and wind direction and velocity. Several of these factors combine to suppress fire activity and flame height at night: temperature decreases, humidity and dew will increase, and wind may die down. Thus, if other things were equal between night and day (which they of course are not), a fire detection sensor would be more likely to detect

a particular fire during the day than at night. Large fires, however, would probably be less sensitive to this, as they have more energy available to dry out fuels at the fire front (on the perhaps debatable assumption that burns of large spatial extent are produced by vigorous fires).

Of course, as previously noted DMSP would be much less likely to detect smaller fires, and so the above considerations favoring the detection of fires of large spatial extent (which would correlate with burnscars of large spatial extent) will not apply very well to small fires and small burnscars. However, if it is the case that most of the area burned in a given season occurs as a result of fires of large spatial extent (as is the case in systems as diverse as Canadian boreal forest and southern California chaparral; the subject is reviewed by Pereira *et al.* 1999), then the expectation would be that missing small fires would contribute relatively little error to the estimate of burn area. And indeed, a comparison of the DMSP-derived active fire count map to a burn scar count map and burn area map (Figure 44) shows that the regions with a few large burn scars will have a high burn area, but a relatively low actual burn scar count; however, the regions with large burn area are approximately the same regions as those with high DMSP active fire counts.

### **Review of Hypotheses**

The three hypotheses tested in this study were: (1) that the DMSP-derived fire index exhibits statistically significant ability to predict burn area; (2) that this predictive ability will increase at coarser scales; and (3) that this product will therefore be

useful for estimating burn area and changes in burn area at the national, province, district, broad landcover, and preserve levels. Regarding the first hypothesis, simple linear regressions between the 1 km DMSP fire index map and ETM burn area map showed a significant relationship, with highly significant p-values ( $p < 0.00$ ). However, statistical significance in this case primarily reflects the large number of observations in the regression; the predictive power of the regression at 1 km is very limited, explaining only 11% of the variance in burn area. Regressions at increasingly coarser resolutions usually showed improvement in r-squared, and were usually statistically significant even though the number of observations dropped precipitously at coarser resolutions.

However, plots showing how statistical significance and r-squared change as a function of resolution showed occasional deviations in the trend due to changes in the configuration of the aggregate pixels; this is essentially an expression of the Modifiable Unit Area Problem (O'Sullivan and Unwin 2003; Quattrochi and Goodchild 1997). Regression slope also showed some variability under coarsening resolution. There is a seemingly intractable problem of picking the "right" resolution for prediction. An additional difficulty with regressions are the violations of the statistical assumptions of linear regression by spatial datasets. Even though the datasets showed less skewed distributions at coarser resolutions, all regressions commit the sin of only assessing overall correlation, without distinguishing correlation at specific spatial scales. This was assessed by fitting a linear model of coregionalization; this analysis revealed that negative correlation at the nugget scale

(-0.40) could be distinguished from a very high correlation present for the largest spatial structure (0.89). LMC analysis thus usefully avoids the Modifiable Unit Area Problem discussed above. However, several criticisms can be made of the LMC technique. First, the modeling process is somewhat subjective: empirical variograms are modeled based on analyst judgment and matching by eye. Some algorithms for automated variogram modeling have been proposed (Jian *et al.* 1995; Pardo-Iqúizquiza 1999), but they generally allow for only fairly simple variogram models, and furthermore it is debatable whether substituting the judgment of the algorithm developer for the judgment of the analyst constitutes an improvement. Second, geostatistical studies making use of the structural correlation coefficients (Dobermann *et al.* 1995; Goovaerts 1997) were performed in a soil science context; the authors recommend that the choices for structures correspond to physical soil processes identified based on background knowledge. In a remote sensing study, the best that can be done is to identify the nugget as sub-pixel variability, and then label the other structures as modeling local, regional, and landscape-level variability. Third, calculation of the structural correlation coefficient gives a measure of co-variability between datasets, but gives none of the other useful outputs of regressions (statistical significance, slope and intercept, etc). Thus calculation of the structural correlation coefficient in LMC analysis serves to complement rather than replace the linear regression analysis discussed above.

Processing of geostatistical simulations gave the most readily interpretable results regarding the predictive ability of the DMSP fire index. Given a DMSP fire index

map for any given time period, 100 alternate equiprobable realizations of the "true" burn area map can be produced. If simulations are averaged into one map, then the result is identical to a kriging map -- and in this case, where the input data are an exhaustive secondary dataset specifying the local mean, the resulting map would just be a rank transform of the input DMSP fire index map into burn area space, essentially a nonlinear regression. The point of simulations is to accurately estimate uncertainty *over an area*; for a best estimate of the mean at a single pixel, kriging would be the appropriate technique. This allows for statistical testing for significant differences between the mean and zero, and between the mean estimates for different years. Having a distribution of burn area estimates could be especially useful if a user is interested in the *probability* that the burn area in a particular polygon exceeded a certain threshold (perhaps a maximum annual burn area set by regulations, carbon emissions credits, or by a conservation agency's target number in a management plan). The probability of threshold exceedance is easily calculated as the ratio of simulation outcomes that exceed the threshold divided by the total number of simulations.

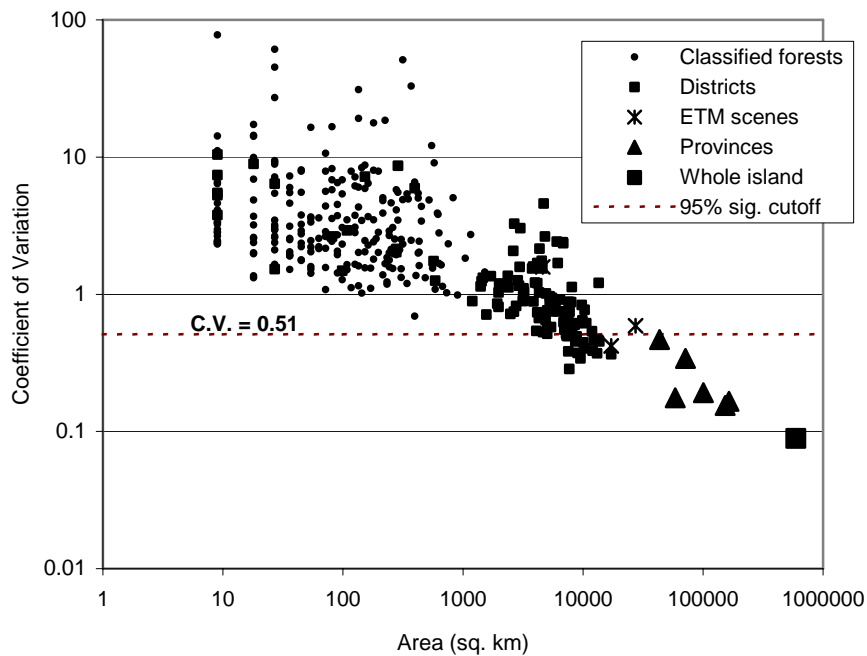


Figure 45: Coefficient of Variation in burn area plotted against polygon area for all polygons used in the study. C.V. = 0.51 represents the cutoff where a burn area estimate is statistically significantly different than zero. This can be achieved for large areas (approximately the size of a Landsat scene or larger), but not for smaller areas. The axes are log-scaled.

The areal cutoff for having a mean estimate significantly different than zero was approximately the size of half a Landsat scene (see Figure 45). This is a sobering result for attempts to utilize DMSP products for assessing changes in annual burn area at finer scales, e.g. for preserves in Madagascar, all of which are smaller than a Landsat scene. While the high sensitivity of the DMSP-OLS sensor implies that under ideal conditions very small fire events may sometimes be detectable (e.g. Kihn 1996), in practice it appears that the multitude of additional factors that can interfere with the light from a fire getting detected at the space borne platform (clouds, intermittent overpass, nighttime dieback in flames, fires being extinguished at dusk, tree canopy obscuring fire signal, etc.) as well as confounding factors (multiple

possible fires at subpixel resolution, multiple detections of the same fire due to DMSP footprint overlap, etc.) limit predictive utility to quite coarse scales, at which point the law of large numbers and the various factors favoring a relationship between burn area and DMSP fire index take over. These results indicate that DMSP can only be reliably used to monitor changes in annual burn area at regional and landscape scales (as would be expected based on the regression and LMC analyses). However, it is possible that over long time periods a trend could be detected at somewhat finer scales.

## **Chapter 6: Conclusions**

The factors that limit the burn area estimation capability of DMSP should not be viewed in isolation. Similar factors will apply to any of the multitude of coarse-resolution sensors capable of detecting active fires or burn scars (AVHRR, MODIS fire, ATSR-1 GLOBSCAR, GOES, etc.). One of the most useful products of this study is the creation of a common method by which the utility of all of these coarse-resolution products can be assessed. A crucial question that has not yet been addressed is whether any one of these fire products is superior to the others, or whether some combination may prove to be superior to any single product. To the extent that all coarse resolution sensors have similar limitations, it seems most likely that all of the sensors will be better at detecting large, long-burning fires that leave large fire scars, and that each sensor will exhibit a rapid decay in ability to map smaller, shorter-burning fires. However, the sensors have varying sensitivities, and different wavelengths and algorithms are used to identify burning (Fuller 2000; Levine *et al.* 1996). It might therefore be expected that one fire product would be superior to the others, but that the addition of more fire products as additional predictors would add little improvement in burn area mapping capability. This was the conclusion reached in a preliminary comparison of DMSP, MODIS, and ATSR fire products (Matzke 2003). In this study, DMSP-OLS was the superior predictor, followed by MODIS and ATSR, neither of which contributed substantial additional prediction capability when added to the DMSP fire product.

It is important to note what has been achieved even with the somewhat limited capabilities of DMSP. This study presents the first scientific estimate of annual burn area for Madagascar (6-7%), and disagrees with both the governmental statistics (which were severe underestimates) and with the overestimates deriving from Jolly (1980) which have been much-repeated in popular and conservation literature. Similarly, annual burn area estimates have been calculated for the major landcover types of Madagascar, and for the provinces. Highland Antananarivo province, which contains the capital and which seems to disproportionately influence scientific and political opinion concerning the extent and severity of fire in Madagascar, experiences an approximately 15-20% annual burn rate from August-December. This is less than half of Kull's informed guess of approximately 50%, however Kull's estimate was made specifically for grassland land cover, whereas areal units such as provinces will have numerous areas that would burn less or not at all -- e.g. urban areas and lowland rice fields -- lowering the overall estimate. Furthermore, all DMSP-based burn area estimates presented in this study are based on the USGS-purchased DMSP dataset, which extends from August-December. While this clearly covers the narrow burning season of eastern Madagascar (which peaks in October), the burning season in the west is more extended, and while it peaks in September-October (Randriambelo *et al.* 1998), Kull recorded some substantial burns (~10% of the total) as early as June in his study site (Kull 2002). Kull (personal communication, 2003) suspects that pre-August burning increases further as one

moves west. For these reasons, as an estimate of burn area in pure grasslands, an annual burn rate somewhat higher than 25% may be closer to the truth.

The major source of concern regarding the present burn area estimates derives from the limitations of the reference and validation datasets. The Landsat regions selected for these purposes were originally chosen on the basis of being areas of interest to USGS-funded conservation projects in the mountainous primary forest regions of eastern Madagascar. While fire in the form of slash-and-burn clearings is of primary conservation concern, in the course of the project it became clear that DMSP primarily measured large grassland fires. The reference Landsat image did contain both rainforest and grassland regions, and did contain a few large areas (1 km and 3 km pixels) that were almost completely burned ( $> 90\%$ ); however, the overall burn area of the image was only 2.54%, substantially smaller than the total burn area estimates for many provinces, or indeed for the whole island. The validation scenes had even lower percent burn areas. The present results therefore represent extrapolation beyond the data, at least at the coarse scale of a Landsat scene.

A satisfying test of these results would require at least one Landsat image pair covering the highland fire season. With the present results a DMSP-based burn area estimate could be made for any Landsat footprint for a given time span within the DMSP time series. Given the LMC and regression results, which indicate that high correlation is only obtained at very coarse scales, an even better assessment could be made with 4+ image pairs from the drier regions of central-west Madagascar. Cloud

cover is much reduced in western Madagascar, so acquisition of appropriate images would be comparatively simple. However, the peak of burning in the west is in September, while the available DMSP fire product begins August 1, so the range of appropriate image dates may be narrower for western regions as it cannot be safely assumed that negligible fires occur before August 1.

The resolution-dependence of the DMSP capabilities means that the kind of burns that DMSP can reasonably predict -- primarily large grassland fires -- are basically not the fires of ecological and conservation interest. Madagascar's grasslands are essentially made up of low diversity, exotic, pan-tropical grasses. However, grassland fires should not therefore be ignored. Grassland burning is a topic of interest for a multitude of local reasons (reviewed by Kull) and national/international reasons (such as carbon and trace gas emissions, as well as particulate release -- smoke pollution is a complaint in the capital city Antananarivo, for instance). Furthermore, although the problem of anthropogenic savanna fires impacting forest edges (and thereby changing the ecological balance between savanna and forest) has not been identified as a significant issue in Madagascar, in many parts of Africa it is a considerable issue, and DMSP may have significant ecological applications in such regions. Finally, it cannot be emphasized enough that the fate of Madagascar's forest-embedded biodiversity is not entirely a matter of simply protecting forests. For example, grassland fires are sometimes implicated as causing unintentional loss of plantation forests and woodlots; such events increase the demand for fuel and building wood, which can increase the pressure on native forests.

Like Kull's work, this study should help to increase realism in political and scientific discussions of fire in Madagascar. Simple maps of DMSP fire counts (e.g. Figure 7) give the impression that virtually the entirety of Madagascar burns annually, and may have helped to reinforce the anti-fire conventional wisdom. Accurate burn area estimates are necessary before well-informed judgments can be made about what, if anything, to do about burning in various regions of Madagascar. This work, and similar future work with the many new fire products becoming available, will hopefully contribute to a resolution of the fire controversy that serves the interests of both the biodiversity of Madagascar and the Malagasy who are responsible for it.

## References

- Adams, J. B., Sabol, D. E., Kapos, V., Almeida, R., Roberts, D. A., Smith, M. O. and Gillespie, A. R. (1995). "Classification of Multispectral Images Based on Fractions of Endmembers - Application to Land-Cover Change in the Brazilian Amazon." *Remote Sensing of Environment* **52**(2): 137-154.
- Alfaro, R., Fernandez, W. and Connell, B. (1999). "Detection of forest fires of April 1997 in Guanacaste, Costa Rica, using GOES-8 images." *International Journal of Remote Sensing* **20**: 1189-1195.
- Battistini, R. and Richard-Vindard, G., Eds. (1972). *Biogeography and Ecology in Madagascar*. Monographiae Biologicae. The Hague, Dr. W. Junk B.V., Publishers.
- BBC (2002). "Madagascar biodiversity threatened." Accessed online: 8/15/2002. Source: <http://www.tandroyconservation.org.uk/bbcnewsarticle.htm>.
- Belmont, C. A., Chambers, J. M. and Hastie, T. J., Eds. (1991). *Statistical Models in S*, CRC Press.
- Bourgeat, F. (1972). "Les paturages naturels de Madagascar." *Memoires de l'Institut Scientifique de Madagascar* **5 (ser. B)**: 65-86.
- Cahoon, D. R., Stocks, B. J., Levine, J. S., Cofer, W. R. and O'Neill, K. P. (1992). "Seasonal distribution of African savanna fires." *Nature* **359**: 812-815.
- Castel, T., Guerra, F., Caraglio, Y. and Houllier, F. (2002). "Retrieval biomass of a large Venezuelan pine plantation using JERS-1 SAR data. Analysis of forest structure impact on radar signature." *Remote Sensing of Environment* **79**: 30-41.
- Chuvieco, E., Ed. (1999). *Remote Sensing of Large Wildfires in the European Mediterranean Basin*. Berlin, Springer.
- CI (2000). "Where we work: Madagascar." Accessed online: 12/12/2000. Source: <http://www.conservation.org/web/fieldact/regions/afrireg/Madagasc.htm>.
- Clifford, P., Richardson, S. and Hemon, D. (1989). "Assessing the Significance of the Correlation between Two Spatial Processes." *Biometrics* **45**(1): 123-134.
- Cochrane, M. A., Alencar, A., Schulze, M. D., Souza, C. M., Nepstad, D. C., Lefebvre, P. and Davidson, E. A. (1999). "Positive Feedbacks in the Fire Dynamic of Closed Canopy Tropical Forests." *Science* **284**(5421): 1832-1835.

- Congalton, R. G. (1987). "Comments on the Remote Sensing Brief entitled "Correct Formulation of the Kappa Coefficient of Agreement"." *Photogrammetric Engineering and Remote Sensing* **53**(4): 422.
- Congalton, R. G. and Mead, R. A. (1983). "A Quantitative Method to Test for Consistency and Correctness in Photointerpretation." *Photogrammetric Engineering and Remote Sensing* **49**(1): 69-74.
- Conservation International (2000). "Where we work: Madagascar." Accessed online: 12/12/2000. Source:  
<http://www.conservation.org/web/fieldact/regions/afrireg/Madagasc.htm>.
- Croft, T. A. (1973). "Burning waste gas in oil fields." *Nature* **245**: 375-376.
- Croft, T. A. (1978). "Nighttime images of the Earth from space." *Scientific American* **239**(1): 86-98.
- Crutzen, P. J. and Andreae, M. O. (1990). "Biomass burning in the tropics: impact of atmospheric chemistry and biogeochemical cycles." *Science* **250**: 1669-1678.
- De Bano, L. F., Neary, D. G. and Ffolliott, P. F. (1998). *Fire's Effects on Ecosystems*. New York, John Wiley & Sons.
- De Grandi, G., Mayaux, P., Rauste, Y., Rosenqvist, A., Simard, M. and Saatchi, S. S. (2000a). "The Global Rain Forest Mapping Project JERS-1 radar mosaic of tropical Africa: Development and product characterization aspects." *Ieee Transactions on Geoscience and Remote Sensing* **V38**(N5 PT1): 2218-2233.
- De Grandi, G. F., Mayaux, P., Malingreau, J. P., Rosenqvist, A., Saatchi, S. and Simard, M. (2000b). "New perspectives on global ecosystems from wide-area radar mosaics: flooded forest mapping in the tropics." *International Journal of Remote Sensing* **V21**(N6-7): 1235-1249.
- Deutsch, C. V. and Journel, A. G. (1998). *GSLIB geostatistical software library and user's guide*. New York, Oxford University Press.
- Dobermann, A., Goovaerts, P. and George, T. (1995). "Sources of soil variation in an acid Ultisol of the Philippines." *Geoderma* **68**: 173-191.
- Doll, C. N. H., Muller, J.-P. and Elvidge, C. D. (2000). "Night-time imagery as a tool for global mapping of socioeconomic parameters and greenhouse gas emissions." *Ambio* **29**(3): 157-162.

- Donque, G. (1972). The Climatology of Madagascar. *Biogeography and Ecology in Madagascar*. R. Battistini, and Richard-Vindard, G. The Hague, Dr. W. Junk B.V., Publishers. **21**: 87-144.
- Dozier, J. (1981). "A method for satellite identification of surface temperature fields of subpixel resolution." *Remote Sensing of Environment* **11**: 221-229.
- Du Puy, D. J. and Moat, J. (1998). Vegetation mapping and classification in Madagascar (using GIS): implications and recommendations for the conservation of biodiversity. *Chorology, Taxonomy and Ecology of the Floras of Africa and Madagascar*. C. R. Huxley, J. M. Lock and D. F. Cutler. Kew, Royal Botanic Gardens: 97-117.
- Dwyer, E., Gregoire, J. M. and Malingreau, J. P. (1998). "Global analysis of vegetation fires using satellite images: spatial and temporal dynamics." *Ambio* **27**: 175-181.
- Ehrlich, D., Lambin, E. F. and Malingreau, J. P. (1997). "Biomass burning and broad-scale land-cover changes in western Africa." *Remote Sensing of Environment* **61**: 201-209.
- Elachi, C. (1987). Introduction to the Physics and Techniques of Remote Sensing. New York, John Wiley & Sons.
- Elvidge, C., Baugh, K., Hobson, V. R., Kihn, E. A. and Kroehl, H. W. (1998a). Detection of fires and power outages using DMSP-OLS data. *Remote Sensing Change Detection: Environmental Monitoring Methods and Applications*. R. S. Lunetta and C. Elvidge, Ann Arbor Press: 123-135.
- Elvidge, C., Pack, D. W., Prins, E. M., Kihn, E. A., Kendall, J. D. and Baugh, K. (1998b). Wildfire detection with meteorological satellite data: results from New Mexico during June of 1996 using GOES, AVHRR, and DMSP-OLS. *Remote Sensing Change Detection: Environmental Monitoring Methods and Applications*. R. S. Lunetta and C. Elvidge, Ann Arbor Press: 103-122.
- Elvidge, C. D. and Baugh, K. E. (1996). Survey of fires in Southeast Asia and India during 1987. *Biomass Burning and Global Change*. J. S. Levine. Cambridge, MIT Press. **1**: 73-85.
- Elvidge, C. D., Baugh, K. E., Hobson, V. R., Kihn, E. A. and Kroehl, H. W. (2000). "Detection of Fires and Power Outages At Night Using DMSP-OLS Data." Accessed online: February 10, 2002. Source.
- Elvidge, C. D., Baugh, K. E., Kihn, E. A., Kroehl, H. W. and Davis, E. R. (1997). "Mapping of city lights with nighttime data from the DMSP Operational

- Linescan System.” *Photogrammetric Engineering and Remote Sensing* **63**(6): 727-734.
- Elvidge, C. D., Hobson, V. R., Baugh, K. E., Dietz, J. B., Shimabukuro, Y. E., Krug, T., Novo, E. M. L. M. and Echavarria, F. R. (2001a). “DMSP-OLS estimation of tropical forest area impacted by surface fires in Roraima, Brazil: 1995 Versus 1998.” *International Journal of Remote Sensing* **22**(14): 2661-2673.
- Elvidge, C. D., Kroehl, H. W., Kihn, E. A., Baugh, K. E., Davis, E. R. and Hao, W. M. (1996). Algorithm for the Retrieval of Fire Pixels from DMSP Operational Linescan System Data: Remote Sensing, Modeling and Inventory Development, and Biomass Burning in Africa. *Biomass Burning and Global Change*. J. S. Levine. Cambridge, MIT Press. **2**: 73-84.
- Elvidge, C., Nelson, I., Hobson, V. R., Baugh, K. and Safran, J. (2001). Detection of Fires At Night Using DMSP-OLS Data. *Global and Regional Vegetation Fire Monitoring from Space: Planning a Coordinated International Effort*. Ahern, F. J., Goldammer, J. G. and Justice, C. O. The Hague, The Netherlands, SPB Academic Publishing: 125-144
- ESA (2001). “Ionia Fire.” Accessed online: 5/29/2001. Source: <http://shark1.esrin.esa.it/ionia/FIRE/>.
- ESA (2002). “GLOBSCAR: ATSR World Burned Surface Atlas.” Accessed online: February 10, 2002. Source: <http://shark1.esrin.esa.it/ionia/FIRE/BS/ATSR/>.
- Eva, H. and Lambin, E. F. (1998a). “Burnt area mapping in central Africa using ATSR data.” *International Journal of Remote Sensing* **19**: 3473-3497.
- Eva, H. and Lambin, E. F. (1998b). “Remote Sensing of Biomass Burning in Tropical Regions: Sampling Issues and Multisensor Approach.” *Remote Sensing of Environment* **64**: 292-315.
- Faramala, M. H. (1981). Etude de la végétation de Madagascar à l’aide de données spatiales. Toulouse, France, Université Paul Sabatier.
- Fuller, D. O. (2000). “Satellite remote sensing of biomass burning with optical and thermal sensors.” *Progress in Physical Geography* **24**(4): 543-561.
- Fuller, D. O. and Fulk, M. (2000). “Comparison of NOAA-AVHRR and DMSP-OLS for operational fire monitoring in Kalimantan, Indonesia.” *International Journal of Remote Sensing* **21**(1): 181-187.

- Fuller, D. O. and Fulk, M. (2001). "Burned are in Kalimantan, Indonesia mapped with NOAA-AVHRR and Landsat TM imagery." *International Journal of Remote Sensing* **22**(4): 691-697.
- Furby, S. L. and Campbell, N. A. (2001). "Calibrating images from different dates to 'like-value' digital counts." *Remote Sensing of Environment* **77**(2): 186-196.
- Gade, D. W. (1996). "Deforestation and Its Effects in Highland Madagascar." *Mountain Research and Development* **16**(2): 101-116.
- Ganzhorn, J. U., Fietz, J., Rakotovao, E., Schwab, D. and Zinner, D. (1999). "Lemurs and the regeneration of dry deciduous forest in Madagascar." *Conservation Biology* **13**(4): 794-804.
- Ganzhorn, J. U., Porter, P. L. I., Schatz, G. E. and Sommer, S. (2001). "The biodiversity of Madagascar: one of the world's hottest hotspots on its way out." *Oryx* **35**(4): 346-348.
- GHCC (2001). "GHCC Scientists Study Lightning From Space." Accessed online: 12/01/2001. Source: [http://thunder.nsstc.nasa.gov/bookshelf/docs/ghcc\\_scientists\\_driscoll.html](http://thunder.nsstc.nasa.gov/bookshelf/docs/ghcc_scientists_driscoll.html).
- Goldammer, J. G., Pfund, J. L., Helfert, M. R., Lulla, K. P. and Crew, S.-M. (1996). Use of the Earth Observation System in the Space Shuttle Program for Research and Documentation of Global Vegetation Fires: A Case Study from Madagascar. *Biomass Burning and Global Change: Remote Sensing, Modeling and Inventory Development, and Biomass Burning in Africa*. J. S. Levine. Cambridge, MIT Press. **1**: 236-240.
- Goodman, S. M. and Rakotondravony, D. (2000). "The effects of forest fragmentation and isolation on insectivorous small mammals (Lipotyphla) on the Central High Plateau of Madagascar." *Journal of Zoology* **250**(PT2): 193-200.
- Goovaerts, P. (1997). *Geostatistics for natural resources evaluation*. New York, Oxford University Press.
- Green, G. M. and Sussman, R. W. (1990). "Deforestation History of the Eastern Rain Forests of Madagascar from Satellite Images." *Science* **248**(4952): 212-215.
- Greenway, P. (1997). *Madagascar & Comoros*. Hawthorn, Australia, Lonely Planet Publications.

- Guillaumet, J.-L. (1984). The Vegetation: An Extraordinary Diversity. *Madagascar*. A. Jolly, P. Oberle and R. Albignac. Oxford, Pergamon Press: 1-26.
- Hoeltgen, D. (1994). Where the hills catch fire. *Ceres*. **26**: 42.
- Hudson, W. D. and Ramm, C. W. (1987). "Correct Formulation of the Kappa Coefficient of Agreement." *Photogrammetric Engineering and Remote Sensing* **53**(4): 421-422.
- Humbert, H. and Cours Darne, G. (1965). Carte internationale du tapis végétal et des conditions écologiques. 3 coupures au 1/1,000,000 de Madagascar. Toulouse, French Institute of Pondichéry.
- Huxley, C. R., Lock, J. M. and Cutler, D. F., Eds. (1998). *Chorology, Taxonomy and Ecology of the Floras of Africa and Madagascar*. Kew, Royal Botanic Gardens.
- Imhoff, M. L., Lawrence, W. T., Elvidge, C. D., Paul, T., Levine, E., Privalsky, M. V. and Brown, V. (1997a). "Using nighttime DMSP/OLS images of city lights to estimate the impact of urban land use on soil resources in the United States." *Remote Sensing of Environment* **59**(1): 105-117.
- Imhoff, M. L., Lawrence, W. T., Stutzer, D. C. and Elvidge, C. D. (1997b). "A technique for using composite DMSP/OLS "city lights" satellite data to map urban area." *Remote Sensing of Environment* **61**(3): 361-370.
- Jarosz, L. (1993). "Defining and Explaining Tropical Deforestation - Shifting Cultivation and Population Growth in Colonial Madagascar (1896-1940)." *Economic Geography* **69**(4): 366-379.
- Jian, X., Olea, R. A. and Yu, Y.-S. (1995). "Semivariogram modeling by weighted least squares." *Computers and Geosciences* **22**(4): 387-397.
- Jolly, A. (1980). *A World Like Our Own*. New Haven, Yale University Press.
- Jolly, A. (1990). On the edge of survival. *Madagascar: A World Out of Time*. F. Lanting. New York, Aperture: 110-121.
- Jolly, A., Oberle, P. and Albignac, R., Eds. (1984). *Madagascar*. Key Environments. Oxford, Pergamon Press.
- Justice, C. O., Kendall, J. D., Dowty, P. R. and Scholes, R. J. (1996). "Satellite remote sensing of fires during the SAFARI campaign using NOAA advanced very high resolution radiometer data." *Journal of Geophysical Research-Atmospheres* **101**(D19): 23851-23863.

- Keeley, J. E. and Fotheringham, C. J. (2001a). "Historic fire regime in Southern California shrublands." *Conservation Biology* **V15**(N6): 1536-1548.
- Keeley, J. E. and Fotheringham, C. J. (2001b). "History and management of crown-fire ecosystems: a summary and response." *Conservation Biology* **V15**(N6): 1561-1567.
- Kew Research Botanical Garden (2000). "Madagascar Plant Diversity." Accessed online: 12/12/2000. Source: <http://www.rbgekew.org.uk/herbarium/madagascar/>.
- Kihn, E. A. (1996). Forest Fire Detection from DMSP Operational Linescan System (OLS) Imagery. *Biomass Burning and Global Change: Remote Sensing, Modeling and Inventory Development, and Biomass Burning in Africa*. E. Levine. Cambridge, MIT Press. **1**: 86-91.
- Kull, C. A. (1996). Paysanneries Malgaches dans la Crise. *Geographical Review*. **86**: 638.
- Kull, C. A. (1998a). Chair de la Terre, Oeil De L'Eau. *Human Ecology: An Interdisciplinary Journal*. **26**: 168.
- Kull, C. A. (1998b). Leimavo revisited: agrarian land-use change in the highlands of Madagascar. *Professional Geographer*. **50**: 163.
- Kull, C. A. (2000). Isle of Fire: The Political Ecology of Grassland and Woodland Burning in Highland Madagascar. *Department of Environmental Science, Policy, and Management*. Berkeley, CA, University of California at Berkeley.
- Kull, C. A. (2002). "Madagascar's burning issue: the persistent conflict over fire." *Environment* **44**(3): 8-19.
- Kyriakidis, P. C. and Dungan, J. L. (2001). "A geostatistical approach for mapping thematic classification accuracy and evaluating the impact of inaccurate spatial data on ecological model predictions." *Environmental and Ecological Statistics* **V8**(N4): 311-330.
- Levine, J. S. (1996). Introduction. *Biomass Burning and Global Change: Remote Sensing, Modeling and Inventory Development, and Biomass Burning in Africa*. J. S. Levine. Cambridge, MIT Press. **1-2**: xxxv-xliii.
- Levine, J. S., Cahoon, D. R., Costulis, J. A., Couch, R. H., Davis, R. E., Garn, P. A., Jalin, A., McAdoo, J. A., Robinson, D. M., Roettker, W. A., Sasamoto, W. A., Sherril, R. T. and Smith, K. D. (1996). FireSat and the Global Monitoring

of Biomass Burning. *Biomass Burning and Global Change: Remote Sensing, Modeling and Inventory Development, and Biomass Burning in Africa*. J. S. Levine. Cambridge, MIT Press. **1**: 107-129.

Lock, J. M. (1998). Aspects of fire in tropical African vegetation. *Chorology, Taxonomy and Ecology of the Floras of Africa and Madagascar*. C. R. Huxley, J. M. Lock and D. F. Cutler. Kew, Royal Botanic Gardens: 181-189.

Lowry, P. P., II, Schatz, G. E. and Phillipson, P. B. (1997). "The Classification of Natural and Anthropogenic Vegetation In Madagascar." Accessed online: 12/12/2000. Source: <http://www.mobot.org/MOBOT/Madagasc/vegmad1.html>.

Martin, M. P., Ceccato, P., Flasse, S. and Downey, I. (1999). Fire detection and fire growth monitoring using satellite data. *Remote Sensing of Large Wildfires in the European Mediterranean Basin*. E. Chuvieco. Berlin, Springer-Verlag: 101-122.

Matzke, N. (2003). Comparative Analysis of DMSP, MODIS, and GLOBSCAR burn area estimation capabilities for Madagascar. *Paper and presentation*. New Orleans, LA, American Association of Geographers Annual Meeting, 2003.

Mayaux, P., Gond, V. and Bartholome, E. (2000). "A near-real time forest-cover map of Madagascar derived from SPOT-4 VEGETATION data." *International Journal of Remote Sensing* **V21(N16)**: 3139-3144.

Menzel, W. P. and Prins, E. M. (1996). Monitoring biomass burning with the new generation of geostationary satellites. *Biomass burning and global change*. J. S. Levine. Cambridge, MA, MIT Press: 56-64.

Middleton, B. A., Sanchez-Rojas, E., Suedmeyer, B. and Michels, A. (1997). "Fire in a tropical dry forest of Central America: a natural part of the disturbance regime." *Biotropica* **29**: 515-517.

Minnich, R. A. (2001). "An integrated model of two fire regimes." *Conservation Biology* **V15(N6)**: 1549-1553.

NASA (2001). "Shuttle image of Madagascar, 11/27/1989." Accessed online: 3/8/2001. Source: <http://science.ksc.nasa.gov/mirrors/images/images/pao/STS33/10063680.htm>

NASDA/MITI (2000). "The Global Rain Forest Mapping Project." Accessed online: 10/1/00. Source:

<http://southport.jpl.nasa.gov/GRFM/cdrom/africa/AFRI/AFR1b/docs/html/index.htm>.

- Nelson, R. and Horning, N. (1993). "Avhrr-Lac Estimates of Forest Area in Madagascar, 1990." *International Journal of Remote Sensing* **14**(8): 1463-1475.
- NGDC (1998). "Defense Meteorological Satellite Program (DMSP)." Accessed online: 3/01/01. Source: <http://www.ngdc.noaa.gov/dmsp/dmsp.html>.
- NWF (2000). "Rosy Periwinkle." Accessed online: 3/1/2001. Source: <http://www.nwf.org/wildalive/periwinkle/>.
- O'Sullivan, D. and Unwin, D. J. (2003). *Geographic Information Analysis*. Hoboken, New Jersey, John Wiley & Sons.
- Owens, J. (2001). "MODIS Fire and Thermal Anomalies Products." Accessed online: 5/28/2001. Source: <http://modis-fire.gsfc.nasa.gov/>.
- Pardo-Iqúizquiza, E. (1999). "VARFIT: a fortran-77 program for fitting variogram models by weighted least squares." *Computers and Geosciences* **25**: 251-261.
- Patterson, B. D., Goodman, S. M., Sedlock, J. L. and Field Museum of Natural History (1995). *Environmental change in Madagascar*. Chicago, Field Museum.
- Paulian, R. (1984). Madagascar: A Micro-Continent between Africa and Asia. *Madagascar*. A. Jolly, P. Oberle and R. Albignac. Oxford, Pergamon Press: 1-26.
- Pereira, A. C. and Setzer, A. W., Jr. (1996). "Comparison of fire detection in savannas using AVHRR's channel 3 and TM images." *International Journal of Remote Sensing* **17**(10): 1927-1937.
- Pereira, J. M. C., Sousa, A. M. O., Sa, A. C. L., Martin, M. P. and Chuvieco, E. (1999). Regional-scale burnt area mapping in southern Europe using NOAA-AVHRR 1 km data. *Remote Sensing of Large Wildfires in the European Mediterranean Basin*. E. Chuvieco. Berlin, Springer: 139-140.
- Prasad, V. K., Kant, Y., Gupta, P. K., Elvidge, C. and Badarinath, K. V. S. (2002). "Biomass burning and related trace gas emissions from tropical dry deciduous forests of India: A study using DMSP-OLS data and ground-based measurements." *International Journal of Remote Sensing* **23**(14): 2837-2851.

- Prins, E. M. and Menzel, W. P. (1994). "Trends in South American biomass burning detected with the GOES visible infrared spin scan radiometer atmospheric sounder from 1983 to 1991." *Journal of Geophysical Research* **99**(16): 719-735.
- Pyne, S. J. (1997). *Vestal fire : an environmental history, told through fire, of Europe and Europe's encounter with the world*. Seattle, University of Washington Press.
- Quattrochi, D. A. and Goodchild, M. F. (1997). *Scale in Remote Sensing and GIS*. Boca Raton, CRC Lewis Publishers.
- Randriambelo, T., Baldy, S. and Bessafi, M. (1998). "An improved detection and characterization of active fires and smoke plumes in south-eastern Africa and Madagascar." *International Journal of Remote Sensing* **19**(14): 2623-2638.
- Riaño, D., Chuvieco, E., Ustin, S., Zomer, R., Dennison, P., Roberts, D. and Salas, J. (2002). "Assessment of vegetation regeneration after fire through multitemporal analysis of AVIRIS images in the Santa Monica Mountains." *Remote Sensing of Environment* **79**: 60-71.
- Roberts, D. A., Batista, G., Pereira, J., Waller, E. and Nelson, B. (1998). Change Identification using Multitemporal Spectral Mixture Analysis: Applications in Eastern Amazonia. *Remote Sensing Change Detection: Environmental Monitoring Methods and Applications*. R. S. Lunetta and C. Elvidge, Ann Arbor Press: 137-161.
- Roberts, D. A., Numata, I., Holmes, K., Batista, G., Krug, T., Monteiro, A., Powell, B. and Chadwick, O. A. (2002). "Large area mapping of land-cover change in Rondônia using multitemporal spectral mixture analysis and decision tree classifiers." *Journal of Geophysical Research* **107**.
- Roberts, S. J. (2000). "Tropical fire ecology." *Progress in Physical Geography* **24**(2): 281-288.
- Robinson, J. M. (1991). "Fire from Space - Global Fire Evaluation Using Infrared Remote Sensing." *International Journal of Remote Sensing* **V12**(N1): 3-24.
- Rosenqvist, A., Shimada, M., Chapman, B., Freeman, A., De Grandi, G., Saatchi, S. and Rauste, Y. (2000). "The Global Rain Forest Mapping project - a review." *International Journal of Remote Sensing* **V21**(N6-7): 1375-1387.
- Schatz, G. E. (1996). "Malagasy/Indo-Australo-Malesian Phytogeographic Connections." Accessed online: 12/12/2000. Source: <http://www.mobot.org/MOBOT/Madagasc/biomad1.html>.

- Scholes, R. J., Kendall, J. D. and Justice, C. O. (1996a). "The quantity of biomass burned in southern Africa." *Journal of Geophysical Research* **101**(23): 667-676.
- Scholes, R. J., Ward, D. E. and Justice, C. O. (1996b). "Emissions of trace gases and aerosol particles due to vegetation burning in Southern Hemisphere Africa." *Journal of Geophysical Research* **101**(23): 677-682.
- Scotese, C. R. (2000). "Paleomap Project." Accessed online: 12/12/2000. Source: <http://www.scotese.com/>.
- Simard, M., Saatchi, S. S. and De Grandi, G. (2000). "The use of decision tree and multiscale texture for classification of JERS-1 SAR data over tropical forest." *Ieee Transactions on Geoscience and Remote Sensing* **V38**(N5 PT1): 2310-2321.
- Spiegel, M. R., Schiller, J. and Srinivasan, R. A. (2000). *Theory and Problems of Probability and Statistics*. New York, McGraw-Hill.
- Stott, P. (2000). "Combustion in tropical biomass fires: a critical review." *Progress in Physical Geography* **24**(3): 355-377.
- Sussman, R. W., Green, G. M. and Sussman, L. K. (1994). "Satellite Imagery, Human Ecology, Anthropology, and Deforestation in Madagascar." *Human Ecology* **22**(3): 333-354.
- Sussman, R. W., Green, G. M. and Sussman, L. K. (1996). The Use of Satellite Imagery and Anthropology to Assess the Causes of Deforestation in Madagascar. *Tropical Deforestation: The Human Dimension*. L. E. Sponsel, T. N. Headland and R. E. Bailey. New York, Columbia University Press: 296-315.
- Sutton, P., Roberts, D., Elvidge, C. and Baugh, K. (2001). "Census from heaven: An estimate of the global human population using night-time satellite imagery." *International Journal of Remote Sensing* **22**(16): 3061-3076.
- Terborgh, J. (1992). *Diversity and the Tropical Rain Forest*. New York, Scientific American Library.
- Tyson, P. (2000). *The Eighth Continent : Life, Death, and Discovery in the Lost World of Madagascar*. New York, Avon Books.
- Watson, R. (1997). "IPCC Special Report: Land Use, Land-Use Change, and Forestry: Summary for Policymakers." Accessed online: 3/01/01. Source: <http://www.ipcc.ch/pub/srlulucf-e.pdf>.

- Weaver, J. F., Purdom, J. F. W. and Schneider, T. L. (1995). "Observing forest fires with the GOES-8 3.9  $\mu\text{m}$  imaging channel." *Weather and Forecasting* **10**: 803-808.
- Wells, N. A. (1995). Extreme Gullying in Madagascar and its Cultural and Natural Causes. *Environmental Change in Madagascar*. B. D. Patterson, S. M. Goodman and J. L. Sedlock. Chicago, The Field Museum: 47-48.
- Wells, N. A. (1997). Extreme Gullying in Madagascar and its Cultural and Natural Causes. *Natural Change and Human Impact in Madagascar*. B. D. Patterson and S. M. Goodman. Smithsonian Institution Press, Washington and London: 44-74.
- Wilson, E. O. (1985). "The Biological Diversity Crisis: A Challenge to Science." *Issues in Science and Technology* **2**(1): 20-29.
- Yirdaw, E. (1996). Deforestation in Tropical Africa. *Sustainable forestry challenges for developing countries*. M. Palo and G. Mery. Dordrecht ; Boston, Kluwer Academic. **10**: 291-310.
- Young, R. and Axel, A. (1999). "Applying GIS and satellite remote sensing to endangered species conservation: a study of the distribution and habitat status of the Indri indri in Madagascar." Accessed online: 11/15/2001. Source: <http://www.yale.edu/ceo/Projects/Students/indri/EndangeredIndri.html>.

## **Appendices**

### **Appendix A: Review of geostatistical terminology and theory**

#### **Introduction**

This discussion follows closely the introduction to the topic given by Deutsch and Journel (1998), condensed to focus on the case where variable of interest (burn area) is continuous and is being predicted at specific pixel centroid.

In predictive statistics, the goal is to characterize an unknown value of interest, here denoted as  $z$ . Any datum or unknown value is modeled as an outcome of a random variable (RV)  $Z$  specific to a location. A RV is “random” in the sense that it can produce a variety of outcomes described by a histogram or probability density function (pdf). The key advantage of geostatistics over nonspatial statistics is that spatial autocorrelation is taken into account: measurements that are near to each other in space will tend to be more similar than measurements widely separated in space. Uncertainty in predictions can be therefore be reduced because a prediction for a particular pixel will be further constrained by information from nearby locations.

#### **Random variables and random functions**

In the present study,  $z$  is a continuous variable representing *fraction (or percent) area burned*. The unknown value  $z(\mathbf{u})$  is modeled as an outcome of the RV  $Z(\mathbf{u})$  for a particular pixel within a raster. The characteristics of  $Z(\mathbf{u})$  will vary in space depending on pixel location ( $x, y$  coordinate). The vector of coordinates describing a

pixel location is denoted  $\mathbf{u}$ . The integral of the pdf is the cumulative distribution function (cdf), and the function describing the cdf of  $Z$  at location  $\mathbf{u}$  is:

$$F(\mathbf{u};z) = \text{Prob}\{Z(\mathbf{u}) \leq z\} \quad (10)$$

Equation (10) describes the *a priori* probability that the RV  $Z(\mathbf{u})$  attains outcomes less than or equal to a given cutoff  $z$ , as  $Z$  increases. When additional information is added, such as  $n$  nearby measurements at locations  $\{\mathbf{u}_\alpha, \alpha = 1, \dots, n\}$ , the cdf becomes conditional on this new information, and the resulting conditional cumulative distribution function (ccdf) is described using the “conditional to  $n$ ” ( $|n$ ) notation:

$$F(\mathbf{u}; z | n) = \text{Prob}\{Z(\mathbf{u}) \leq z | n\} \quad (11)$$

This ccdf is therefore dependent on the location  $\mathbf{u}$ , the additional information at measurements  $n$ , and the configuration and distance of the locations  $\{\mathbf{u}_\alpha, \alpha = 1, \dots, n\}$  of  $n$  measurements.

As in most geostatistical applications, in this study the goal is predicting  $z(\mathbf{u})$  at not just one pixel, but over the entire raster of the study area ( $\mathbf{u} \in \text{study area}$ ). A set of random variables (RVs) is a random function (RF). The RF containing all of the RVs in the study area ( $K$  pixels at locations  $\{\mathbf{u}_k, k = 1, \dots, K\}$ ) is modeled by the multivariate ( $K$ -variate) cdf of the RVs in the study area:

$$F(\mathbf{u}_1, \dots, \mathbf{u}_k; z_1, \dots, z_k) = \text{Prob}\{Z(\mathbf{u}_1) \leq z_1, \dots, Z(\mathbf{u}_K) \leq z_K\} \quad (12)$$

While the univariate cdf of the RV  $Z(\mathbf{u})$  models the uncertainty in predicting the unknown truth  $z(\mathbf{u})$ , the multivariate cdf of the RF  $\{Z(\mathbf{u}_k), \mathbf{u}_k, k = 1, \dots, K\}$  models the *joint* uncertainty of predicting  $\{z(\mathbf{u}_k), k = 1, \dots, K\}$ , taking into account the influence of all of the RVs in the study region on each other.

However, except for the special condition discussed below, for any large  $K$  the problem of modeling the influence of all RVs on all other RVs is intractable; as result, only the bivariate cdf modeling the joint uncertainty of two RVs  $Z(\mathbf{u}), Z(\mathbf{u}')$  at locations  $\mathbf{u}$  and  $\mathbf{u}'$  is usually used:

$$F(\mathbf{u}, \mathbf{u}'; z, z') = \text{Prob}\{Z(\mathbf{u}) \leq z, Z(\mathbf{u}') \leq z'\} \quad (13)$$

The covariance between  $Z$  at two locations  $\mathbf{u}$  and  $\mathbf{u}'$ , a moment of the bivariate pdf, is given as:

$$C(\mathbf{u}, \mathbf{u}') = E\{Z(\mathbf{u})Z(\mathbf{u}')\} - E\{Z(\mathbf{u})\}E\{Z(\mathbf{u}')\} \quad (14)$$

where  $E$  denotes “expected value of”, i.e., the mean (for a more complete discussion, see Goovaerts 1997, pp. 13-21, 68-71). When  $\mathbf{u} = \mathbf{u}'$ ,  $C(\mathbf{u}; \mathbf{u}') = C(0)$  is the variance of the RF.

### **Stationarity**

In practice, multiple observations of the RV  $Z(\mathbf{u})$  at a given location  $\mathbf{u}$  are not available. Nearby measurements are often available, however, and can be substituted for measurement replications at  $\mathbf{u}$  if the decision of stationarity is made.

Strict stationarity entails that over a spatial field  $A$  the multivariate cdf (12) of the RF  $\{Z(\mathbf{u}), \mathbf{u} \in A\}$  is invariant under translation of all  $K$  locations in the locations vector  $\{\mathbf{u}_1, \dots, \mathbf{u}_K\}$ :

$$F(\mathbf{u}_1, \dots, \mathbf{u}_k; z_1, \dots, z_K) = F(\mathbf{u}_1 + \mathbf{h}, \dots, \mathbf{u}_k + \mathbf{h}; z_1, \dots, z_K), \quad (15)$$

$\forall$  translation vector  $\mathbf{h}$

Stationarity thus allows the inference of statistics from a group of samples within field  $A$ . Under second-order stationarity any cdf of  $z$  at any location  $\mathbf{u}$  within field  $A$  is the same:

$$F(z) = F(\mathbf{u}; z), \forall \mathbf{u} \in A \quad (16)$$

and the stationary covariance for any two RVs  $Z(\mathbf{u})$  and  $Z(\mathbf{u} + \mathbf{h})$  at any locations  $\mathbf{u}$ , separated by approximate lag vector  $\mathbf{h}$ , is a slight modification of equation (14)

$$C(\mathbf{h}) = E\{Z(\mathbf{u} + \mathbf{h})Z(\mathbf{u})\} - [E\{Z(\mathbf{u})\}]^2, \quad (17)$$

$\forall \mathbf{u}, \mathbf{u} + \mathbf{h} \in A$

If the RF  $\{Z(\mathbf{u}), \mathbf{u} \in A\}$  has a multivariate Gaussian distribution in addition to being stationary, then knowledge of the covariance function (14) also defines any multivariate cdf, allowing inference of the joint uncertainty in predicting  $\{z(\mathbf{u}_k), k=1 \dots K\}$ . This property makes the multivariate Gaussian model statistically tractable; it serves as the basis for the stochastic simulation procedure discussed later.

### **Semivariograms**

The covariance at lag  $\mathbf{h}=0$  equals the variance, and as  $\mathbf{h}$  increases, typically the covariance decreases. The semivariance is the squared semidifference between pairs of data separated by distance  $\mathbf{h}$ , and will increase as  $\mathbf{h}$  increases. The covariance and semivariance are directly related:

$$\gamma(\mathbf{h}) = C(\mathbf{0}) - C(\mathbf{h}), \forall \mathbf{h} \quad (18)$$

and the plot of the semivariance  $\gamma(\mathbf{h})$  against lag  $\mathbf{h}$  is referred to as the *semivariogram*, or simply *variogram*. The empirical variogram, calculated from a dataset via an algorithm, is the traditional method of estimating the spatial autocorrelation of measured sample  $z$  values.

The semivariance value at  $\mathbf{h} = \varepsilon$ ,  $\gamma(\varepsilon)$ , where  $\varepsilon$  denotes a very small distance vector, is called the *nugget* and represents micro-scale variability in the data (variability over a distance shorter than the sample interval). As the lag  $\mathbf{h}$  increases, the correlation between  $Z(\mathbf{u})$  and  $Z(\mathbf{u} + \mathbf{h})$  will decrease to 0; when this point is reached, the semivariance  $\gamma(\mathbf{h})$  approximates the sample variance, and  $\gamma(\mathbf{h})$  will remain approximately constant as  $\mathbf{h} \rightarrow \infty$ . This value is known as the *sill* and the lag  $\mathbf{h}$  where it is reached is known as the *range*.

### **Permissible semivariogram models**

Empirical variograms are not continuous ( $\mathbf{h}$  is discrete), because available data values will never exist at all possible lags; in addition, algorithms must pool data

across a range of lags (e.g.,  $\mathbf{h}_1=0-5$ ,  $\mathbf{h}_2=5-10$ , etc.) because of small numbers of pairs at small lags, and computational efficiency at large lags.

In order for a variogram to be usable for prediction of a point at any given lag from known points  $\mathbf{u}_k \in A$ , the variogram must be modeled as a continuous function.

Therefore, a model must be fit to the empirical variogram. The variogram model must meet the condition of being positive definite (more correctly, nonnegative definite), such that for any linear combination of  $K$  random variables

$\{Z(\mathbf{u}_k), \mathbf{u}_k \in A\}$ , the variance of the resulting RV  $Y$  is  $\geq 0$ :

$$\text{Var}\{Y\} = \text{Var}\left\{\sum_{k=1}^K \lambda_k Z(\mathbf{u}_k)\right\} \geq 0 \quad (19)$$

Otherwise, between all combinations of RVs at two locations  $\mathbf{u}_k$  and  $\mathbf{u}_l$ , separated by lag  $\mathbf{h} = (\mathbf{u}_k - \mathbf{u}_l)$ , the covariance model  $C(\mathbf{h})$  must be positive definite so that  $\text{Var}\{Y\}$  is nonnegative:

$$\text{Var}\{Y\} = \sum_{k=1}^K \sum_{l=1}^K \lambda_k \lambda_l C(\mathbf{u}_k - \mathbf{u}_l) \geq 0 \quad (20)$$

for any choice of the  $K$  locations and the weights  $\lambda_k$  (Goovaerts 1997, p. 87). A variogram model that meets this condition is said to be permissible.

### **Linear model of regionalization (LMR)**

As testing the permissibility of a model after its construction is onerous, variogram models are typically built by combining standard variogram functions that are known

to be permissible (e.g. those given in Goovaerts 1997, p. 88)). Not all combinations of permissible functions result in permissible models, however, so a standard technique is typically used, the *linear model of regionalization* (LMR), wherein the semivariogram is modeled by a linear combination of several permissible functions  $g_l(\mathbf{h})$  with unit sill,

$$\gamma(\mathbf{h}) = \sum_{l=0}^L b^l g_l(\mathbf{h}), \quad b^l \geq 0 \quad (21)$$

where each component variogram function models a *proportion*  $b^l$  of the maximum semivariance (equal to the sill of the empirical variogram).

In addition to each component variogram model accounting for a proportion  $b^l$  of the total sill, each variogram will have its own range  $a^l$ . Anisotropic variogram models will have two ranges ( $a_\theta^l$  and  $a_\phi^l$ ) and an azimuth specified; the azimuth denotes the direction of maximum continuity,  $a_\theta^l$  denotes the range in that direction, and  $a_\phi^l$  denotes the range in the perpendicular direction. Different component variogram models can have different ranges and different anisotropies.

Variogram modeling is a rather subjective procedure, relying not only on the empirical semivariogram but also on ancillary information about sources of anisotropy and spatial structure in the dataset, on the modeller's goals (e.g., predicting short-range or long-range structures), and also on parsimony (getting a good fit to the empirical variogram with as few models as possible).

### Linear model of coregionalization (LMC)

Modeling a *coregionalization* is simply the *joint* modeling of the auto-variograms of two datasets *and* the cross-variogram between the two datasets (more than two datasets can be jointly modeled, but in this study this was not necessary), where the auto- and cross-variograms are all modeled with the same combination of component variograms (or “structures”). Thus each variogram  $\gamma(\mathbf{h})$  will be composed of  $L$  component variogram structures  $g_l(\mathbf{h})$ , and each structure will be given a proportion  $b^l$  of the empirical variogram sill semivariance. By convention,  $l=0$  represents the nugget structure (range=0). If three variogram structures in addition to the nugget are used to model the auto-variograms of the two datasets ( $\gamma_{11}(\mathbf{h})$  and  $\gamma_{22}(\mathbf{h})$ ) and cross-variogram between the two datasets ( $\gamma_{12}(\mathbf{h})$ ), the variogram models are represented as:

$$\begin{aligned}\gamma_{11}(\mathbf{h}) &= b_{11}^0 g_0(\mathbf{h}) + b_{11}^1 g_1(\mathbf{h}) + b_{11}^2 g_2(\mathbf{h}) + b_{11}^3 g_3(\mathbf{h}) \\ \gamma_{22}(\mathbf{h}) &= b_{22}^0 g_0(\mathbf{h}) + b_{22}^1 g_1(\mathbf{h}) + b_{22}^2 g_2(\mathbf{h}) + b_{22}^3 g_3(\mathbf{h}) \\ \gamma_{12}(\mathbf{h}) &= b_{12}^0 g_0(\mathbf{h}) + b_{12}^1 g_1(\mathbf{h}) + b_{12}^2 g_2(\mathbf{h}) + b_{12}^3 g_3(\mathbf{h})\end{aligned}\quad (22)$$

This can be represented in matrix notation as:

$$\begin{bmatrix} \gamma_{11}(\mathbf{h}) & \gamma_{12}(\mathbf{h}) \\ \gamma_{21}(\mathbf{h}) & \gamma_{22}(\mathbf{h}) \end{bmatrix} = \sum_{l=0}^L \begin{bmatrix} b_{11}^l & b_{12}^l \\ b_{21}^l & b_{22}^l \end{bmatrix} g_l(\mathbf{h}) = \sum_{l=0}^L \mathbf{B}_l g_l(\mathbf{h}) \quad (23)$$

The LMC is permissible if all coregionalization matrices  $\mathbf{B}_l$  are nonnegative definite. This can be checked with the following simple tests (Goovaerts 1997, pp.

113-114): the diagonal coefficients must be nonnegative, and the determinant of  $\mathbf{B}_l$  is nonnegative:

$$\begin{aligned} b'_{11} &\geq 0 \\ b'_{22} &\geq 0 \\ \begin{vmatrix} b'_{11} & b'_{12} \\ b'_{21} & b'_{22} \end{vmatrix} &= b'_{11}b'_{22} - [b'_{21}]^2 \geq 0 \Rightarrow b'_{21} \leq \sqrt{b'_{11}b'_{22}} \end{aligned} \quad (24)$$

The LMC method is widely used to model the auto- and cross-variograms between datasets, because permissibility is easily checked and yet sufficient flexibility is maintained to allow reasonable approximation of empirical variograms. The models can then be employed in cokriging or cokriging-based simulation exercises (although this was not done in the present study, simple kriging with a locally-varying mean was used). The modeling becomes increasingly cumbersome if the number of datasets increases to 3 or larger (the number of checks increases exponentially) and so a semi-automated approach is often employed.

### **Structural correlation coefficients**

An additional usage of the LMC is the simple calculation of the *structural correlation coefficient*,  $\rho'_{12}$ , where  $l$  denotes the spatial scale (or process) modeled by a given component variogram structure. The structural correlation coefficient is calculated as follows (Dobermann *et al.* 1995, pp. 177-178):

$$\rho'_{12} = \frac{b'_{12}}{\sqrt{b'_{11}b'_{22}}} \quad (25)$$

Dobermann *et al.* (1995, pp. 177-178) use the structural correlation coefficient in a geostatistical study of soil variation, with dramatic results as compared to a comparison of raw correlation coefficients.  $\rho_{12}^l$  provides, then, an estimate of the correlation between datasets at the differing spatial scales of each of the component data structures used in the model. This provides a direct method for quantifying how correlation between datasets changes between fine- and coarse-scale comparisons. This is an improvement over calculation of the raw correlation  $\rho$  commonly employed in remote sensing studies (in pixel-pixel comparisons between rasters), which averages out the covariability across all scales. In addition, calculation of the structural correlation coefficient may detect correlation between rasters that is obscured for other reasons such as registration error and differences in raw data resolution. It appears that LMC analysis has not been appreciably applied in remote sensing studies or other raster comparisons.

### **Simple kriging**

Simple kriging (SK) is the basis for the sequential Gaussian simulation technique described below. The term 'kriging' covers a group of least-squares regression algorithms designed to provide the best-possible estimate of a value at a particular location, given nearby known data. An important point to emphasize at the outset is that kriging is designed to give the best estimate of a value at single location, *not* to provide the best-possible map of a field of locations.

Formally, the goal of kriging is to estimate an unknown value  $z(\mathbf{u})$  at location  $\mathbf{u}$ , given  $n(\mathbf{u})$  nearby known  $z$  values  $\{z(\mathbf{u}_\alpha), \alpha = 1, \dots, n(\mathbf{u}_\alpha)\}$ . As discussed above, the uncertainty in the estimation of a random variable is given by a distribution  $Z(\mathbf{u})$ . The expected value of the RV  $Z(\mathbf{u})$  is the mean of that distribution:

$$E\{Z(\mathbf{u})\} = m(\mathbf{u}) \quad (26)$$

Kriging incorporates information from the  $n(\mathbf{u})$  known nearby data points,  $\{z(\mathbf{u}_\alpha), \alpha = 1, \dots, n(\mathbf{u}_\alpha)\}$ , in order to produce an improved estimator of  $z(\mathbf{u})$ ,  $Z^*(\mathbf{u})$ .  $Z^*(\mathbf{u})$  is the basic linear regression estimator at location  $\mathbf{u}$  (Deutsch and Journel 1998, p. 64; Goovaerts 1997, p. 126):

$$Z_{SK}^*(\mathbf{u}) - m(\mathbf{u}) = \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_\alpha(\mathbf{u}) [Z(\mathbf{u}_\alpha) - m(\mathbf{u}_\alpha)] \quad (27)$$

$\lambda_\alpha(\mathbf{u})$  is the weight assigned to known value  $z(\mathbf{u}_\alpha)$  at location  $\mathbf{u}_\alpha$ . The known and the unknown values  $z(\mathbf{u})$  are treated as realizations of their respective RVs.  $n(\mathbf{u})$  describes the number of data points to be used in the estimation of the unknown  $z$  at location  $\mathbf{u}$ . It is a function of the location  $\mathbf{u}$ , and can change from point to point as different known data fall within the local neighborhood.

Kriging attempts to minimize the error variance  $\sigma_E^2(\mathbf{u})$

$$\sigma_E^2(\mathbf{u}) = \text{Var}\{Z^*(\mathbf{u}) - Z(\mathbf{u})\} \quad (28)$$

under the constraint that the estimator is unbiased. The unbiasedness constraint simply means that the expected value of the RV of the improved kriging estimator,  $Z^*(\mathbf{u})$  is equal to the expected value of the unimproved estimator RV  $Z(\mathbf{u})$ , in other words:

$$\begin{aligned} E\{Z^*(\mathbf{u}) - Z(\mathbf{u})\} &= 0 \\ m^* - m &= 0 \end{aligned} \quad (29)$$

In simple kriging, it is assumed that the mean of the RF  $Z(\mathbf{u})$  is known and invariant across the entire area of study, meaning that the RF  $Z(\mathbf{u})$  is second-order stationary. This is often a poor assumption (especially across large areas), but adjustments such as subtracting out the coarse-scale trend or specifying the mean based on secondary data can justify the decision of stationarity.

Given a known mean  $m$  and  $n(\mathbf{u})$  nearby data, the SK estimator is described by re-writing equation (27):

$$Z_{SK}^*(\mathbf{u}) = \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_{\alpha}(\mathbf{u})Z(\mathbf{u}_{\alpha}) + \left[1 - \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_{\alpha}(\mathbf{u})\right]m \quad (30)$$

The optimal weights  $\lambda_{\alpha}(\mathbf{u}), \alpha = 1, \dots, n(\mathbf{u})$  that minimize the error variance  $\sigma_E^2(\mathbf{u})$  are determined by solving the simple kriging system. This is usually represented in terms of covariance functions (Goovaerts 1997, p. 128):

$$\sum_{\beta=1}^{n(\mathbf{u})} \lambda_{\beta}^{SK}(\mathbf{u}) C(\mathbf{u}_{\alpha} - \mathbf{u}_{\beta}) = C(\mathbf{u}_{\alpha} - \mathbf{u}) \quad (31)$$

where  $\alpha = 1, \dots, n(\mathbf{u})$

or in matrix notation:

$$\mathbf{K}_{SK} \boldsymbol{\lambda}_{SK}(\mathbf{u}) = \mathbf{k}_{SK} \quad (32)$$

where  $\mathbf{K}_{SK}$  is the  $n(\mathbf{u}) \times n(\mathbf{u})$  matrix of data-point to data-point covariances,  $\boldsymbol{\lambda}_{SK}(\mathbf{u})$  is the  $n(\mathbf{u}) \times 1$  vector of data-point weights, and  $\mathbf{k}_{SK}$  is the  $n(\mathbf{u}) \times 1$  vector of data-point to unknown-point covariances.

$$\begin{bmatrix} C(\mathbf{u}_1 - \mathbf{u}_1) & \cdots & C(\mathbf{u}_1 - \mathbf{u}_{n(\mathbf{u})}) \\ \vdots & \ddots & \vdots \\ C(\mathbf{u}_{n(\mathbf{u})} - \mathbf{u}_1) & \cdots & C(\mathbf{u}_{n(\mathbf{u})} - \mathbf{u}_{n(\mathbf{u})}) \end{bmatrix} \begin{bmatrix} \lambda_1^{SK}(\mathbf{u}) \\ \vdots \\ \lambda_{n(\mathbf{u})}^{SK}(\mathbf{u}) \end{bmatrix} = \begin{bmatrix} C(\mathbf{u}_1 - \mathbf{u}) \\ \vdots \\ C(\mathbf{u}_{n(\mathbf{u})} - \mathbf{u}) \end{bmatrix} \quad (33)$$

The optimal weights are obtained simply by inverting the data-to-data covariance matrix and multiplying it by the data-unknown covariance matrix:

$$\boldsymbol{\lambda}_{SK}(\mathbf{u}) = \mathbf{K}_{SK}^{-1} \mathbf{k}_{SK} \quad (34)$$

There will be a unique solution to the optimal weights if the covariance matrix  $\mathbf{K}_{SK}$  is positive definite, which will be the case if the variogram models employed were permissible by construction as is the case when the LMC method is used.

### **Sequential Gaussian simulation**

In the following sections, “z-data” refers to the reference data of known values, in this case, burn area as mapped by ETM. “y-data” refers to z-data transformed with the normal score transform.

#### *Constructing a z-data distribution*

Realizations of  $z$  are drawn from a cumulative distribution function (cdf)  $F_Z(z)$ . Care must therefore be taken to make this cdf as representative as possible of the true distribution of  $z$  (which is of course unknown as a regular matter). If the available  $z$ -data are clustered, a declustering algorithm may be employed; rough or truncated distributions may be smoothed and/or extrapolated to ameliorate artefacts due to limited sampling.

#### *Normal score transform*

The first requirement of a multivariate Gaussian RF model is that the univariate (single-point) distribution of the variable be normally distributed. In environmental datasets, the distribution of  $z$ -data is rarely normal. Right-skewed distributions, with many small or zero values and a few high values, are very common. One potential option to normalize the data is to perform a log transform, but this is not a universal solution and in particular is confounded by zero values in the dataset. The rank-preserving normal score transform is broadly used and is recommended by Deutsch and Journel (1998).

The normal score transform proceeds by first sorting the  $n$   $z$ -data values from smallest to largest and assigning each a rank  $k$  between 1 and  $n$ . In a spatial dataset,

ties (e.g., with zeros) are broken by ranking tied values based on the local averages (i.e. a zero value in a region of higher average values will be ranked higher). Each  $z$ -datum  $z(\mathbf{u}_\alpha)$  is assigned a cumulative frequency  $p_k^*$  ( $0 < p_k^* < 1$ ) in a standard normal cdf, using the following formula:

$$p_k^* = k / n - 0.5 / n \quad (35)$$

Each  $p_k^*$  is converted to its respective normal score  $y(\mathbf{u})$ ; for example, the middle-ranked  $z$  datum is assigned a  $p_k^* = 0.5$ , and a normal score value  $y(\mathbf{u}) = 0$ . Formula (35) cannot produce  $p_k^*$  values of 0 or 1, so the normal score transform never yields  $y(\mathbf{u}) = -\infty$  or  $y(\mathbf{u}) = \infty$ . Thus, by construction, any recalcitrant distribution of  $z$ -data can be converted into the statistically tractable standard normal distribution of  $y$ -data. The normal distribution satisfies the requirement of the multivariate Gaussian model that the simulated variable be univariate normal. Simulations are performed in normal score  $y$ -data space and the results are back-transformed to the original  $z$ -data space.

#### *Checking for bivariate normality*

The second requirement of a multivariate Gaussian RF model is that the variable be bivariate normal, that is that for any two values  $Y(\mathbf{u})$  and  $Y(\mathbf{u}+\mathbf{h})$  separated by any lag  $\mathbf{h}$ , the joint (two-point) cdf be normal. While there is not a readily usable formal statistical test for this property, a standard check exists (Deutsch and Journel 1998, p. 142).

The check is performed by comparing empirical variograms of the  $y$  data to the variograms theoretically expected for a multivariate normal RF. Theoretical and empirical indicator variograms are constructed for a series of cdf cumulative frequency cutoffs, for example  $p = 0.1, 0.2, \dots, 0.9$ . In a multivariate Gaussian RF, the middle of the distribution should exhibit the most spatial continuity, while the high and low extremes of the distribution should be more random and less clustered in space. In practical terms this means that the theoretical multivariate Gaussian RF indicator variograms for cutoffs  $p=0.1$  and  $p=0.9$  will have higher nuggets and shorter ranges than the theoretical indicator variograms for central cutoff  $p=0.5$ .

If the empirical indicator variograms approximately match the theoretically expected pattern, then the decision to adopt the bivariate normal model for the  $y$ -data is considered justified. Theoretically, in order to be assured that the multivariate Gaussian RF model is appropriate, the joint cdfs of all trivariate (3-point) and higher-variate ( $K$ -point) distributions should also be checked for normality. In practice, data limitations exclude this possibility, so the decision of multivariate normality is made if the checks for univariate and bivariate normality are passed.

*The simulation algorithm*

Once the multivariate Gaussian random function model has been adopted, a computer-implemented algorithm generates realizations of the  $y$  variable over the area of interest, field  $A$ .  $A$  is generally a rectangular raster. The simulated  $y$ -values are filled in pixel-by-pixel using the following procedure:

1. Using a user-specified seed value, the order of pixel simulation is determined such that the cells are filled in a pseudo-random manner.
2. If the pixel  $\mathbf{u}$  has no nearby conditioning data, then the  $y$ -value is drawn from the cdf of the RV  $Y(\mathbf{u})$ , a normal distribution.
3. If nearby conditioning data are present (in the form of original known  $y$ -values, nearby already-simulated pixels, or in the form of a local mean specified by secondary data), then the pixel  $y$ -value is drawn from the ccdf  $Y_{SK}^*(\mathbf{u})$ . This is the step that ensures reproduction of the variogram model in this Gaussian space.
4. Return to step 1 and repeat until all pixels are filled.

As many realizations should be generated as are necessary to adequately characterize the joint uncertainty in the output.

*Post-simulation processing*

Once the simulation run has been completed, the realization in  $y$  space must be converted to the data  $z$  space by using the back-transform of the originally employed normal score transform. Variograms and histograms for each  $z$  space realization can be generated to check reproduction of the model statistics. The data cube of simulation realizations can be processed in a GIS to obtain distributions of outcomes for particular areas of interest, e.g. specific polygons.

## **Appendix B: Description of programs written for this study**

This appendix contains a list of programs written and used in this study. This study required the processing of a very large number of gridded datasets. This could only be effectively carried out with extensive use of scripts for automating data conversion, map creation, simulation, and statistical analysis. ENVI was used for basic image processing, visualization, and output, tasks which were automated with IDL (ENVI/IDL produced by Research Systems, Inc.). Geostatistical analysis and simulation were performed using GSLIB, a public-domain Geostatistical Software Library written in FORTRAN and compiled for use in UNIX (Deutsch & Journel, 1998). GSLIB executables require parameter files that are extremely tedious to edit manually, so UNIX shell scripts were employed to write the parameter files, run GSLIB programs in loops, perform simple calculations, and display the output postscript graphics with XV. Finally, statistical analysis and sophisticated graphical display, as well as processing of the simulation data cube, was carried out using MATLAB (Mathsoft, Inc.). A few additional scripts used for similar processing of datasets used in other studies (MODIS data) are also included. The scripts will be made available online at the author's website (<http://www.geog.ucsb.edu/~matzke>), or by request; however no guarantees are made with regard to support or modification for a user's particular needs.

File name	Language	Approximate date written or revised	Purpose	Description
bm2000_fireindex.pro	ENVI/IDL	Jan-02	Calculate 2000 fire index	ENVI Band Math script that, for each pixel, divides fire count by number of cloud-free days; if number of cloud-free days = 0 (a rare event), then returns 0 for the pixel. First store in ENVI save_add directory and then restart ENVI; Band Math syntax to run: bm2000_fireindex(b1, b2, /check)
confmat1.pro	ENVI/IDL	May-02	Multiple confusion matrices	Computes confusion matrices for a series of input data files
generate_bmps2.pro	ENVI/IDL	Jun-01	Output image as bitmaps	Generate bytescaled bitmap output files for each band of an input image. Used for creating inputs for band-by-band animations, e.g. in Animation Shop
lotsa_bandmath_float_namelist_v2.pro	ENVI/IDL	Nov-02	Perform same bandmath operation on multiple images	This particular version performs binary thresholding on a series of 8-day MODIS fire product files
generate_from_modis_bmps4.pro	ENVI/IDL	Oct-02	Output bitmap series of MODIS fire detections	Display only fires from MODIS fire mask image series; rotate images to produce "upright" bitmaps
dmspcldoud_process3.pro	ENVI/IDL	May-02	Generate cloud count image from a list of files in a directory	No manual file selection; directory specified in-script
imagetoGSLIBpts_v3.pro	ENVI/IDL	Oct-01	Convert an ENVI image into a GSLIB points file.	GSLIB files come as points and grid files; this outputs the points version (can be used for both). ENVI image Hotine Oblique Mercator coordinates (in meters) are output for each pixel
dmspfirecount_filelist_v3.pro	ENVI/IDL	May-02	Generate fire count image from a list of files in a directory	No manual file selection; directory specified in-script; loading and closing images one-by-one prevents crashing and tedious construction of huge multiband images
dmspedit2_process2.pro	ENVI/IDL	Feb-01	Generate fire count image from multiband DMSP fire time series	Input ENVI meta file
process_modis1.pro	ENVI/IDL	May-02	Open a large series of MODIS HDF files into ENVI	Once the files are open they can be easily saved as an ENVI meta file for easy access.
imagetoGSLIBgrid.pro	ENVI/IDL	Feb-02	Convert an ENVI image into a GSLIB grid file.	Outputs to correctly-formatted GSLIB grid format, no coordinates output. Image is appropriately rotated etc. in-script

File name	Language	Approximate date written or revised	Purpose	Description
dmspcldoud_process2.pro	ENVI/IDL	Feb-01	DMSP cloud count map from multiband DMSP cloud image	For each night of observation, count as cloud any pixel with threshold > 160
dmspedit2_process.pro	ENVI/IDL	Feb-01	Calculate DMSP fire count images for a series of DMSP DN values	Fire count maps for thresholds DMSP DN = 5, 10, 20, 35, 50, 60, 63
sruleconv2_3.pro	ENVI/IDL	Apr-01	Convert decision tree rules and generate confusion matrix and accuracy statistics	Convert S-PLUS-generated decision tree rules (tree function) to simpler format for Dar Robert's Landsat classification program classifyd. Written with Kerry Halligan
imageroGSLIBpts_for MATLAB.pro	ENVI/IDL	Oct-01	Convert ENVI image into a MATLAB-readable GSLIB-style file	Image is appropriately rotated for use in ENVI
fft_middle_v2.pro	ENVI/IDL	Feb-02	Fast Fourier Transform on DMSP	Experimenting with FFT filters on DMSP to reproduce ETM-style variability
nan_replace_v1.pro	ENVI/IDL	Dec-01	Replace negative values with NaN	Useful as it inserts the ENVI NaN code: !VALUES.F_NAN
extractROI_v4.pro	ENVI/IDL	Dec-01	Extracts pixel values corresponding to ENVI ROIs	This fixed version only extracts the values for the desired bands, which the user selects. Written with Kerry Halligan
generate_burnimage_linregs6	ENVI/IDL	Oct-01	Perform linear regressions between large numbers of burn area, fire count, and DMSP fire index images at different resolutions	A truly massive number of comparisons, however ENVI's statistical capability is very limited, and it turned out that this was a very limited approach
generate_burnimages10	ENVI/IDL	Oct-01	Resample 30 m classified Landsat scenes to 1, 3, and 5 km fraction burn area images	The program is complex as the two burn area classes are extracted, nodata and cloud pixels are counted and subtracted from the total possible number of pixels; coarse pixels with >50% nodata were classified as nodata. Multiple versions of burn area, fire count, and average burn area maps produced for analysis with generate_burnimage_linregs6
imageroascii.pro	ENVI/IDL	Oct-01	Convert ENVI image into ASCII format	Prototype for output to various other formats (GSLIB, MATLAB, etc.)

File name	Language	Approximate date written or revised	Purpose	Description
totalimage.pro	ENVI/IDL	May-01	Total the values of pixels in an image	Loads image and uses TOTAL on it
generateMRTscript.m	MATLAB	Mar-03	Writes batch file for processing MODIS fire data	Writes a Windows Command Line (DOS) batch file that will run the MODIS Reprojection Tool mosaic and resample functions, thereby mosaicking, reprojecting, cropping, and outputting the MODIS data as flat binary files with the appropriate ENVI headers. For more details see: <a href="http://www.geog.ucsb.edu/%7Ematzke/mad/modis/MODIS_processing_notes.html">http://www.geog.ucsb.edu/%7Ematzke/mad/modis/MODIS_processing_notes.html</a>
writescript.m	MATLAB	Mar-03	output lines of batch file	Subroutine of generateMRTscript
filenaming_v1.m	MATLAB	Mar-03	Returns a filename with a constant number of digits	Subroutine of generateMRTscript
ma_figs_v2.m	MATLAB	Mar-03	Produces 2D and 2.5D figures of fire images	With re-scaled colorbar and red covering the nodata regions
surfimage.m	MATLAB	Mar-03	Input image, output surface	Subroutine of ma_figs
run_an1_justdmsp.m	MATLAB	Dec-02	Compare ETM and DMSP via regression	Outputs several plots of regression statistics at varying scales
runplotmatrix.m	MATLAB	Dec-02	Histograms and regression plots between two datasets	subroutine of run_an1
coarsen2.m	MATLAB	Jan-03	Resample 1 km images to coarser resolutions	Resamples via averaging, excluding nodata pixels
runregress.m	MATLAB	Nov-02	Runs the REGRESS function	Output results to a pre-defined structure. Allows the storage of the results of many regression runs in an easily-accessible format.
disp2image.m	MATLAB	Nov-02	Display ASCII matrix as image	Rotates and stretches the image as appropriate
sims_area_002f.m	MATLAB	Jun-02	Process simulation data cube	Calculate population statistics for the 100 simulations for various polygon layers of interest (preserves, districts, etc.); display histograms and summary tables
totaltheareas.m	MATLAB	Jun-02	Conducts totalling and optional displays	Subroutine of sims_area

File name	Language	Approximate date written or revised	Purpose	Description
loadsim.m	MATLAB	Mar-02	Loads the GSLIB input files	Subroutine of sims_area
readkey.m	MATLAB	Feb-02	Reads metadata for input polygon layers	Subroutine of sims_area
002f_pxphst1.csh	UNIX k-shell script	Apr-02	Displays pixel plot and histogram of input GSLIB file	*.CSH k-shell script used for running GSLIB executables PIXELPLT, HISTPLT
002c_sim_vg.csh	UNIX k-shell script	Feb-03	Calculates variogram of a simulation file	*.CSH k-shell script used for running GSLIB executables PIXELPLT, GAM, VARGPLT
compare_dmsh.csh	UNIX k-shell script	Feb-02	Compares DMSP spatial subset to full DMSP dataset	Compares datasets via histograms (HISTPLT), log-histograms, and quantile-quantile plots (QQPLT)
cv4_2cl_d1e1a-3km.csh	UNIX k-shell script	Feb-02	Calculate cross-variogram of DMSP and ETM	Data in regular (non-normal-scored) space; variograms (GAMV, VMODEL, VARGPLT) in 4 directions; option to rescale by total cross-variance
cvg4_nsc_ed01a-3km.csh	UNIX k-shell script	Feb-02	Calculate cross-variogram of DMSP and ETM	Data in regular normal-scored space; variograms in 4 directions.
eda_sgshim_002f_full-3km.csh	UNIX k-shell script	Feb-02	Exploratory data analysis of full datasets	Normal-score transforms of datasets also calculated according to reference distributions
eda_sgshim_002f_sub-3km.csh	UNIX k-shell script	Feb-02	Exploratory data analysis of subset datasets	Normal-score transforms of datasets also calculated according to reference distributions
run_2cl_vgs.csh	UNIX k-shell script	Feb-02	Variograms and variogram models	Data in regular (non-normal-scored) space; variograms (GAMV, VMODEL, VARGPLT) in 4 directions; option to rescale by total semivariance
run_nsc_vgs.csh	UNIX k-shell script	Feb-02	Variograms and variogram models	Data in regular normal-scored space; variograms in 4 directions.
runsgshims-3km.csh	UNIX k-shell script	Mar-02	Run simulations and transformations	Runs NSCORE, SGSHIM, BACKTR
sct_2cl_d1e1a-3km.csh	UNIX k-shell script	Jan-02	Scatterplots of datasets	Data in regular space, log space, normal-score space
setupsgshims-3km.csh	UNIX k-shell script	Mar-02	Create normal-score transforms from reference datasets; double-check with histograms	Uses NSCORE, HISTPLT

File name	Language	Approximate date written or revised	Purpose	Description
vmap_2cl_dm10a-3km.csh	UNIX k-shell script	Jan-02	Variogram maps and variogram model maps	Uses VMAPMODEL (VARMAP is too slow on these gridded datasets)
modis_sumburns.pro	ENVI/IDL	Sep-02	Detected active fire count map	Counts MODIS fire classes 7-9 (Low, nominal, and high-confidence fires, respectively)
modis_sumnodata.pro	ENVI/IDL	Apr-03	Count map of number of nodata events per pixel in a time series	Counts MODIS fire classes 0, 1, 2 (Unclassified/unprocessed) and 4 (clouds)
modis_sumunknown.pro	ENVI/IDL	Oct-03	Count map of number of "unknown" classifications per pixel in a time series	Counts MODIS fire class 6 (unknown)
modis_sumclouds.pro	ENVI/IDL	Sep-02	Count map of cloud detections in a time series	Counts MODIS fire class 4 (cloud)

## Appendix C: Fire-related terminology

As terminology regarding fire and various fire variables can be ambiguous, the terms as used in this thesis are defined below. Except where other references are cited, De Bano *et al.* (1998) is being followed. A review from a specifically tropical perspective is Stott (2000).

*Fire* is the rapid release of the chemical energy stored in living and dead vegetation. As a chemical reaction, the process can be summarized as the oxidation of carbohydrate (sugar) molecules (cellulose and hemicellulose, the major constituents of plant cell walls, are made of long chains of simple sugar molecules). The reaction requires ignition heat, oxygen, and fuel, and releases carbon dioxide, water, and heat:



The detailed physical process of fire is divided into the stages of *pre-ignition*, *flaming*, *smoldering*, *glowing*, and finally *extinction*. In addition to CO<sub>2</sub>, H<sub>2</sub>O, and heat, fires produce smoke, char, aerosols such as black carbon, and diverse trace gases such as NO<sub>x</sub>, CO, and CH<sub>4</sub>.

*Flame* is the rapid oxidation of combustible gases. The gases are produced by the heating of fuel which results in an endothermic reaction that breaks down the large organic molecules found in vegetation (such as cellulose and lignin) into small, volatile compounds, a process known as *pyrolysis*. These volatile compounds ignite in the presence of oxygen at a minimum temperatures of about 570 K (297 ° C),

resulting in an exothermic reaction producing heat -- in the form of fast moving gas molecules which will radiate electromagnetic energy which can be detected from appropriate remote sensing platforms. The maximum temperature of flaming combustion, with heavy fuels, is about 1800 K (Robinson 1991). *Fire intensity* refers to rate at which a fire produces thermal energy. *Fire severity* is the degree to which a site has been altered by a fire, and is roughly a product of fire intensity and fire residence time, the latter being dependent on the rate of spread of a fire. Rigorously measuring fire severity is difficult given the variability of the environment in terms of fuels, fire conditions, etc. Something that is *burning* is currently on fire; a *burn* is the area over which a fire has spread, including active fire and burn scar.

A *burn scar*, in this thesis, refers to the combination of post-fire char and altered vegetation that results from a fire. A burn scar has different reflective properties than unburned vegetation, and thus can be mapped with sensors such as Landsat. *Fire frequency* refers to the number of actual individual fires over a specified time and area. This should not be confused with a *fire count*. A *fire count* is an attempted measurement of fire frequency based on the remotely-sensed detection of active fires over a given area and time interval. As the pixel size of the sensor under consideration is much larger than most active fires, one 'lit' pixel could contain multiple distinct fires, so a count of pixels showing active fires is not necessarily a good estimate of the fire frequency. *Burned area* is the area mapped as burn scar in a given area over a given time interval. This can be accurately mapped with high-

resolution optical sensors such as Landsat, but coarse-resolution sensors that detect active fires have the potential to overestimate burned area because an active fire need only take up a small portion of a pixel for that pixel to register a fire.