

ABSTRACT

Title of Dissertation: INTEGRATED USE OF LANDSAT AND
CORONA DATA FOR LONG-TERM
MONITORING OF FOREST COVER
CHANGE AND IMPROVED
REPRESENTATION OF ITS PATCH SIZE
DISTRIBUTION

Danxia Song, Doctor of Philosophy, 2016

Dissertation directed by: Dr. John R. Townshend and Dr. Chengquan
Huang
Department of Geographical Sciences

Forest cover change has profound impact on global carbon cycle, hydrological processes, energy balance, and biodiversity. The primary goal of this dissertation is to improve forest cover change characterization by filling a number of knowledge gaps in forest change studies. These include use of Corona data to extend satellite based forest cover change mapping back to pre-Landsat years in the 1960s, quantification of forest cover change over four decades (1960s – 2005) for a major forested province in China using Corona and Landsat data, and development of more accurate patch size-

frequency modeling methods for improved representation of forest disturbances in ecosystem and other spatially explicit models.

With comprehensive data coverages in the 1960s, Corona data can be used to extend Landsat-based forest change analysis by up to a decade. The usefulness of such data, however, is hindered by poor geolocation accuracy and lack of multi-spectral bands. In this study, it was demonstrated that combined use of texture features and the advanced support vector machines allowed forest mapping with accuracies of up to 95% using Corona data. Further, a semi-automated method was developed for rapid registration of Corona images with residual errors as low as 100 m. These methods were used to assess the forest cover in the 1960s in Sichuan, a major forest province in China. Together with global forest cover change products derived using Landsat data, these results revealed that the forest cover in Sichuan Province was reduced from 45.19% in the 1960s to 38.98% by 1975 and further down to 28.91% by 1990. It then stayed relatively stable between 1990 and 2005, which contradicted trends reported by inventory data. The turning point between sharp decreases before 1990 and the stable period after 1990 likely reflected transitions in forest policies from focuses on timber production to forest conservation.

Representation of forest disturbances in spatially explicit ecosystem models typically relies on patch size-frequency models to allocate an appropriate amount of disturbances to each patch size level. Existing patch size-frequency models, however, do not provide accurate representation of the total disturbance area nor the patch sizes at each frequency level. In this study, a hierarchical method was developed for modeling patch size-frequency distribution. Evaluation of this method over China

revealed that it greatly improved the accuracy in representing the patch size at different frequency levels and reduced error in total disturbance area estimation over existing methods from around 40% to less than 10%.

The significance of this dissertation is the contribution to improve the characterization of forest cover change by extending the satellite-based forest cover change monitoring back to the 1960s and developing a more accurate patch size distribution model to represent the forest disturbance in ecosystem models. The work in the dissertation has a broader impact beyond developing methods and models, as they provide essential basis to understand the relationship between the long-term change of forest and the socioeconomic transitions. They also improve the capacities of ecosystem and other spatially explicit models to simulate the vegetation dynamics and the resultant biodiversity and carbon dynamics.

INTEGRATED USE OF LANDSAT AND CORONA DATA FOR LONG-TERM
MONITORING OF FOREST COVER CHANGE AND IMPROVED
REPRESENTATION OF ITS PATCH SIZE DISTRIBUTION

by

Danxia Song

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2016

Advisory Committee:
Dr. Shunlin Liang, Chair
Dr. Chengquan Huang, Co-chair
Dr. Tatiana V. Loboda
Dr. Mengxue Li
Dr. Michael S. Kearney

© Copyright by
Danxia Song
2016

Foreword

Materials presented in Chapter 3 have been published as a peer-reviewed journal article. Chapter 4 and 5 are now under internal review before being submitted for publication. Part of the materials presented in Chapter 5 have been presented at an international conference and been included in the conference proceeding. Materials presented in Chapter 6 have been submitted to a peer-reviewed journal and is now under review. The research presented in this dissertation was carried out in its entirety by Danxia Song. Following papers and conference proceedings were appeared in this dissertation:

Song, D.-X., C., Huang, J.O., Sexton, S., Channan, M., Feng, J. R., Townshend, Use of Landsat and Corona data for mapping forest cover change from the mid-1960s to 2000s: Case studies from the Eastern United States and Central Brazil. *ISPRS J. Photogram. Remote Sensing* (2015). 103, 81-92.

Song, D.X., C., Huang, P., Noojipady, S., Channan, J.R., Townshend, 2014, Comparison of remote sensing based forest area and change estimation with national forestry inventory between 2000 and 2005 in China, *Geoscience and Remote Sensing Symposium (IGARSS), 2014 IEEE International* , 4268-4271

Song, D.X., C., Huang, J.O., Sexton, T., He, A., Li, J.R., Townshend, (2016), Forest disturbance in China from 2000 to 2005: models of trend and variance in size-frequency distributions, *Landscape Ecology*, under-review

During my Ph.D. study, I have also participated in several research projects, which were very relevant to the topics of this dissertation. Following co-authored papers have been published or submitted for publication in peer-reviewed journals:

Feng, M., J.O., Sexton, C. Huang, A. Anand, S. Channan, X.P., Song, D.X., Song, D.H. Kim, P. Noojipady, J.R. Townshend, (2016), Earth Science Data Records of Global Forest Cover and Change: Assessment of Accuracy in 1990, 2000, and 2005 epochs, *Remote Sensing of Environment*, 184, 73-85

Sexton, J.O., P. Noojipady, X.P. Song, M. Feng, D.X., Song, D.H. Kim, A. Anand, C. Huang, S. Channan, S.L. Pimm, J.R. Townshend, (2015), Conservation policy and the measurement of forests, *Nature Climate Change*, 5

Channan, S., M., Feng, D.H., Kim, J.O., Sexton, X.P., Song, D.X., Song, P., Noojipady, K., Collins, A., Anand, J.R., Townshend, (2015), The GLS+: An Enhancement of the Global Land Survey Datasets, *Photogrammetric Engineering & Remote Sensing*, 81 (7): 521- 525.

N. M. Haddad, L. A. Brudvig, J. Clobert, K. F. Davies, A. Gonzalez, R. D. Holt, T. E. Lovejoy, J. O. Sexton, M. P. Austin, C. D. Collins, W. M. Cook, E. I. Damschen, R. M. Ewers, B. L. Foster, C. N. Jenkins, A. J. King, W. F. Laurance, D. J. Levey, C. R. Margules, B. A. Melbourne, A. O. Nicholls, J. L. Orrock, D.-X. Song, J. R. Townshend, (2015), Habitat fragmentation and its lasting impact on Earth's ecosystems. *Science Advances*, 1, e1500052.

Townshend, J.R., J.G. Masek, C. Huang, E.F. Vermote, F. Gao, S. Channan, J.O. Sexton, M. Feng., R. Narasimhan, D.-H. Kim, K. Song, D. Song, X. Song, P. Noojipady, B. Tan, M.C. Hansen, M. Li, R. Wolfe. 2012. Global characterization and monitoring of forest cover using Landsat data: opportunities and challenges. *International Journal of Digital Earth* 5:373-397.

Katharina Brinck, Rico Fischer, Jürgen Groeneveld , Sebastian Lehmann , Mateus Dantas De Paula , Sandro Pütz , Joseph Sexton , Dan-Xia Song , Andreas Huth, (2016), High Resolution Analysis of Tropical Forest Fragmentation and its Impact on the Global Carbon Cycle, *Nature Communications*, accepted

Acknowledgements

First and foremost, I would like to express my sincere gratitude to my advisor Dr. John Townshend for introducing me to the world's top department of geography and for guiding me throughout my study here. His insight and dedication to Earth's system science have always been the inspiration for my research and career in the future. Although Dr. Townshend was not able to chair my dissertation defense due to a family issue, he continued advising me until the dissertation was finished. I am also very grateful to my co-advisor, Dr. Chengquan Huang for his conscientious advices and encouragement to me to accomplish this dissertation.

I would also like to thank Dr. Shunlin Liang and Dr. Mengxue Li for helping me frame the research focus in China, Dr. Tatiana Loboda for encouraging me to seek for a better scientific question, and Dr. Michael Kearney for introducing ideas from a broader scale.

I am particularly thankful to my colleagues from the Global Land Cover Facility for providing a free but effective research environment. Many collaborations that I have made with colleagues outside the department when working in GLCF are valuable for my career. I also thank my colleagues from Dr. Huang's group for their great suggestions to my work and the inspiration I've got from the discussions. I thank Katie Collins for proofreading, Dr. Huiran Jin and Dr. Xiaopeng Song for giving valuable suggestions to my dissertation.

Sincere gratitude also goes to all of my friends in Maryland, who directly or indirectly help me complete my study and dissertation.

Lastly, I owe the deepest gratitude to my parents and my husband for their unconditional support and love to me. All of this would have been impossible without you.

Table of Contents

Foreword.....	ii
Acknowledgements.....	iv
Table of Contents.....	vi
List of Tables.....	ix
List of Figures.....	xi
Chapter 1: Introduction.....	1
1.1 Information needed for forest cover and change.....	1
1.2 Changes in China’s forest and its forest policies.....	1
1.3 Approaches and issues in forest cover change monitoring.....	3
1.4 Need for accurate representation of the patch-size frequency distribution of forest change.....	5
1.5 Research questions and objectives.....	6
1.6 Dissertation organization.....	7
Chapter 2: A review of remote sensing-based forest cover change monitoring.....	10
2.1 Introduction.....	10
2.2 Landsat-based forest cover change monitoring.....	11
2.2.1 Global forest cover change monitoring at Landsat resolution.....	11
2.2.2 The multi-epoch GLCF-GFCC product from 1975 to 2005.....	11
2.2.3 Issues in using Landsat data for forest cover change monitoring.....	14
2.2.4 Needs for forest cover monitoring before the 1970s.....	15
2.3 Corona data in forest cover estimation.....	16
2.3.1 Characteristics of Corona data.....	16
2.3.2 Applications of Corona data in land cover studies.....	17
2.3.3 Bottleneck problem in using Corona data for large area forest assessment: image registration.....	18
2.4 Summary.....	19
Chapter 3: Use of Landsat and Corona data for mapping forest cover change from the mid-1960s to 2000s: case studies from the eastern United States and central Brazil. 20	
3.1 Introduction.....	20
3.2 Study area and data.....	21
3.2.1 Study area.....	21
3.2.2 Landsat Data.....	22
3.2.3 Corona Data.....	22
3.3 Methods.....	23
3.3.1 Landsat Forest Cover Change Mapping.....	24
3.3.2 Corona forest cover mapping.....	24
3.3.3 Accuracy assessment of classification result.....	30
3.3.4 Forest cover change rate calculating.....	32
3.4 Results.....	32
3.4.1 Geometric Correction of Corona Data.....	32
3.4.2 Effectiveness of Textures for Corona Forest/non-forest Classification ..	33
3.4.3 Accuracy of Landsat Forest Cover Change Map.....	36
3.4.4 Forest Cover Change Rate.....	36
3.5 Discussion.....	39

3.6 Conclusions	41
Chapter 4: Towards automated registration of historical satellite images for land cover land use change applications	43
4.1 Introduction	43
4.1.1 Background	43
4.1.2 Objectives	45
4.2 Data	46
4.2.1 Corona data.....	46
4.2.2 Landsat and elevation data	46
4.3 Methodology	49
4.3.1 Data preparation (1): Automated mosaicking of Corona data	49
4.3.2 Data preparation (2): Preliminary registration of Corona data.....	50
4.3.3 Transformation model	53
4.3.4 Tie points extraction.....	57
4.3.5 Error estimation.....	61
4.4 Results	62
4.4.1 Manual point selection and collinear function accuracy	62
4.4.2 Automated tie point (feature) extraction and refinement	63
4.4.3 Comparison of model fitting accuracies.....	67
4.4.4 Accuracy assessment.....	68
4.5 Discussion and conclusions.....	72
Chapter 5: Remote sensing-based estimation of forest cover and change since the 1960s in Sichuan, China	78
5.1 Introduction	78
5.2 Study area and data.....	80
5.2.1 Study area	80
5.2.2 Data	81
5.3 Methods.....	83
5.3.1 Sampling strategy for Corona data.....	83
5.3.2 Block level forest cover estimate from Corona.....	87
5.3.3 Region level forest cover estimate for the 1960s	90
5.3.4 Estimation of forest cover and change for 1975 and after	91
5.4 Results	94
5.4.1 Accuracies of forest cover maps in 1975, 1990, 2000, and 2005.....	94
5.4.2 Estimates of forest cover from the 1960s to 2005.....	95
5.4.3 Change in forest cover from the 1960s to 2005	98
5.5 Discussion	99
5.5.1 Differences between satellite- and NFI-based forest cover estimates.....	99
5.5.2 Sources of difference in forest cover estimates.....	100
5.5.3 Relationships of long-term forest cover change to China's forest policies	102
5.6 Conclusions	107
Chapter 6: Improved modeling of the size-frequency distribution of forest disturbances based on Landsat forest cover change product	109
6.1 Introduction	109
6.2 Data and study region.....	114

6.2.1 Data	114
6.2.2 Study region	114
6.3 Methodologies	116
6.3.1 Modeling process	116
6.3.2 Model validation.....	119
6.3.3 Sensitivity analysis of model parameters and sampling intensities	120
6.3.4 Model utilization	121
6.3.5 Regional disturbance analysis	122
6.4 Results	122
6.4.1 Model assessment.....	122
6.4.2 Spatial pattern of forest disturbance in China	131
6.5 Discussion	134
6.6 Conclusions	137
Chapter 7: Conclusion	139
7.1 Major findings of the dissertation	139
7.2 Significance and implications of this dissertation.....	141
7.3 Implications for future research	143
Bibliography	146

List of Tables

TABLE 2.1 GLOBAL FOREST COVER AND CHANGE PRODUCTS AT LANDSAT OR SIMILAR RESOLUTION. ACCURACY MEASUREMENTS INCLUDE OVERALL ACCURACY (<i>OA</i>), PRODUCER'S ACCURACY (<i>PA</i>), USER'S ACCURACY (<i>UA</i>) AND STANDARD ERROR (<i>SE</i>).	12
TABLE 3.1 WRS-2 AND WRS-1* PATH/ROW NUMBERS, SENSORS AND ACQUISITION DATES OF LANDSAT IMAGES USED IN THIS CHAPTER.	22
TABLE 3.2 PROPERTIES OF CORONA IMAGES (GALIATSATOS 2009B; NPIC 1967) USED IN THIS CHAPTER.	23
TABLE 3.3 TEXTURE MEASURES USED IN THIS CHAPTER AND THEIR SOURCES OF LITERATURE.	28
TABLE 3.4 NUMBER OF TRAINING AND VALIDATING PIXELS USED FOR FOREST AND NON-FOREST CLASSIFICATION USING CORONA DATA IN EACH STUDY SITE	29
TABLE 3.5 TEXTURE FEATURES USED IN THE CLASSIFICATIONS, WINDOWS SIZES TO DERIVE TEXTURES, SCALES AT WHICH CLASSIFICATIONS AND CORRESPONDING ACCURACY ASSESSMENTS WERE CARRIED OUT AT THE VM STUDY AREA. *NOTE: AT MTP STUDY AREA, WINDOW SIZES ARE 5X5, 7X7, 9X9, 11X11, AND CLASSIFICATION RESOLUTION IS 13.7M, 19.04M, 24.66M AND 30.14M, OTHER PARAMETERS ARE THE SAME AS THE VM STUDY SITE.	30
TABLE 3.6 NUMBER OF VALIDATION PIXELS FOR ASSESSING LANDSAT-BASED FCC (1980S-2000S) PRODUCTS AT EACH STUDY AREA	31
TABLE 3.7 RMSE OF MANUAL IMAGE GEOMETRIC CORRECTION FOR CORONA DATA AT TWO STUDY AREAS.....	33
TABLE 4.1 BASIC INFORMATION OF THE THREE CORONA IMAGES USED IN THIS STUDY.....	47
TABLE 4.2 RMSE FOR COLLINEAR TRANSFORMATION TEST USING MANUALLY SELECTED TIE POINTS	63
TABLE 4.3 FITTING ACCURACY OF POLYNOMIAL AND COLLINEAR FUNCTION SEPARATELY BASED ON REFINED SIFT MATCHING POINTS.....	67
TABLE 4.4 ACCURACY ASSESSMENT USING MANUALLY SELECTED VALIDATION POINTS (VPTS) FOR TWO TRANSFORM MODELS.....	70

TABLE 5.1 THE SUMMARY OF STRATIFIED SAMPLING DESIGN, INCLUDING THE BOUNDARY, SIZE OF TOTAL POPULATION, SIZE OF SAMPLED BLOCKS AND THE RATIO OF SAMPLING FOR EACH STRATUM.	87
TABLE 5.2 THE SIZE OF SAMPLE BLOCKS AND THE SAMPLING RATIO (%) EACH STRATUM BY USING ALL BLOCKS COVERED BY REGISTERED CORONA IMAGE. n_h AND N_h ARE DENOTED THE SAMPLE SIZE AND SIZE OF TOTAL POPULATIONS FOR STRATUM H	89
TABLE 5.3 ERROR MATRIX OF SAMPLE COUNTS, n_{ij} AND ESTIMATED AREA PROPORTION, \hat{p}_{ij} . MAP CLASSES ARE THE ROWS WHILE REFERENCE CLASSES ARE THE COLUMNS.....	93
TABLE 5.4 ACCURACY OF FOREST CLASS FOR GLCF FOREST COVER PRODUCT AT FOUR EPOCHS IN SICHUAN. THE BIOME WHERE EACH LANDSAT SCENE LOCATED IS ALSO SHOWN. ACCURACY MEASUREMENTS INCLUDE OVERALL ACCURACY (OA), PRODUCER'S ACCURACY (PA), AND USER'S ACCURACY (UA) OF THE FOREST CLASS.	94
TABLE 5.5 FOREST COVER (%) FROM 1960S TO 2005 ESTIMATED FROM REMOTE SENSING DATA AND STATISTICS OF FOREST INVENTORY. THE STANDARD ERROR ASSOCIATED WITH FOREST COVER IS REPORTED IN PARENTHESES.	95
TABLE 5.6 LIST OF KEY FOREST AND RELEVANT POLICIES IN CHINA FROM LATE 1950S TO 2000S.....	102
TABLE 6.1 MODEL PARAMETERS OF SIZE-FREQUENCY DISTRIBUTION FOR SELECTED PROVINCES IN CHINA. THEY ARE LISTED FOR REFERENCE OF THE VALUE RANGE OF MODEL PARAMETERS AT REGIONAL SCALE.....	134

List of Figures

FIGURE 1.1 GLOBAL COVERAGE OF CORONA IMAGES ACQUIRED BY THE FORWARD-LOOKING CAMERAS DURING THE KH-4A AND KH-4B MISSIONS. ALL COVERAGE MAPS HAVE BEEN RASTERIZED FOR DISPLAY PURPOSES FROM VECTOR FORMAT TO 1° GEOGRAPHIC GRID. (A) KH-4A MISSION BETWEEN 1962 AND 1965; (B) KH-4A MISSION BETWEEN 1966 AND 1969; (C) KH-4B MISSION BETWEEN 1967 AND 1969; AND (D) KH-4B MISSION BETWEEN 1970 AND 1972 (SOURCE: [HTTPS://LTA.CR.USGS.GOV/DECLASS_1](https://lta.cr.usgs.gov/DECLASS_1)). METADATA WAS DOWNLOADED FROM U.S. GEOLOGICAL SURVEY (USGS) EARTHEXPLORER WEBSITE. 4

FIGURE 3.1 CORONA IMAGE, AGGREGATED TO 27.5M RESOLUTION, IN THE LEFT WINDOW HAS MULTIPLE LAND COVER TYPES, INCLUDING FOREST, CROPLAND, WATER BODY AND SCATTERED BUILD-UPS. ZOOM-IN WINDOWS 1, 2 AND 3 SHOW NON-FOREST LAND COVER, WINDOW 4 AND 5 SHOW DENSE FOREST COVER AND SCATTER FOREST COVER IS PRESENTED IN WINDOW 6. 25

FIGURE 3.2 CLASSIFICATION ACCURACIES AND KAPPA COEFFICIENTS OF CORONA FOREST/NON-FOREST CLASSIFICATIONS USING SINGLE AND MULTIPLE TEXTURES IN THE VM AND MTP STUDY AREAS. ABBREVIATIONS ARE DEFINED IN TABLE 3.5. FOR THE MTP STUDY AREA, THE VALUE OF CLASSIFICATION USING MN FELL BELOW THE RANGE OF Y-AXIS, KAPPA COEFFICIENTS WERE 0.52, 0.57 AND 0.46 AT WINDOW SIZES OF 7×7 TO 11×11, AND CLASSIFICATION ACCURACIES WERE 88.0%, 89.6% AND 85.1% CORRESPONDINGLY. 35

FIGURE 3.3 FOREST COVER CHANGE RATES DURING THREE EPOCHS FOR TWO STUDY AREAS, FOREST LOSS (OR GAIN) PERCENTAGE = FOREST LOSS (OR GAIN) AREA/FOREST AREA OF BEGINNING YEAR * 100. 37

FIGURE 3.4 GEOREFERENCED CORONA, GLS 2000 IMAGES WITH BAND 4, 3, 2 IN COLOR R, G, B, AND SEQUENTIAL FOREST COVER CHANGE MAPS OF EACH TWO EPOCHS IN THE VM STUDY AREA. (1) IS MONTGOMERY COUNTY IN MARYLAND, (2) AND (3) IS LOUDOUN AND FAIRFAX COUNTY IN VIRGINIA SEPARATELY. 38

FIGURE 3.5 GEOREFERENCED CORONA, GLS 2000 IMAGES WITH BAND 4, 3, 2 IN COLOR R, G, B, AND SEQUENTIAL FOREST COVER CHANGE MAPS OF EACH TWO EPOCHS IN MTP STUDY AREA. 39

FIGURE 4.1 THE ACTUAL CORONA COVERAGE (B/W IMAGE) OVERLAID BY THE ESTIMATED FOOTPRINT PROVIDED BY USGS (RED FRAME, DOWNLOADED FROM [HTTP://EARTHEXPLORER.USGS.GOV/](http://earthexplorer.usgs.gov/)). 48

FIGURE 4.2 IMAGE MOSAICKING FOR SCENE DS1105-2248DF067. FOUR SUBSET IMAGES ARE SHOWN IN THE UPPER PANEL AND THE FINAL MOSAIC IMAGE IS SHOWN IN THE LOWER PANEL. APPROXIMATE OVERLAPPING AREAS ARE SHADED IN BLUE IN THE SUBSETS AND MOSAIC IMAGE. 50

FIGURE 4.3 PRELIMINARY REGISTERED CORONA (LEFT) AND REFERENCE LANDSAT (RIGHT) IMAGES IN SICHUAN. EACH PAIR OF WINDOWS (WINDOWS IN THE SAME ROW) REPRESENTS THE SAME RANGE ACCORDING TO THE GEOINFORMATION OF CORONA AND LANDSAT IMAGES. 52

FIGURE 4.4 METHOD OF THE SEMI-AUTOMATED REGISTRATION OF CORONA IMAGE WITH LANDSAT IMAGE AS REFERENCE. SIFT: THE SCALE INVARIANT FEATURE TRANSFORM. USGS: THE UNITED STATES GEOLOGICAL SURVEY. 61

FIGURE 4.5 REFINED MATCHING POINTS SELECTED BY SIFT SHOWN ON CORONA (LEFT) AND REFERENCED LANDSAT (RIGHT) IMAGE FOR SICHUAN. 65

FIGURE 4.6 REFINED MATCHING POINTS SELECTED BY SIFT SHOWN ON CORONA (LEFT) AND REFERENCED LANDSAT (RIGHT) IMAGE FOR ST. LOUIS. 66

FIGURE 4.7 THE DISTRIBUTION OF MANUALLY SELECTED VALIDATION POINTS IN SICHUAN (UPPER) AND ST. LOUIS (LOWER) OVERLAID ON A REGISTERED CORONA IMAGE. 69

FIGURE 4.8 CORRECTED CORONA IMAGES (RED FRAMED) AFTER A) PRELIMINARY REGISTRATION, AND REGISTRATION USING B) 2ND ORDER POLYNOMIAL FUNCTION AND C) COLLINEAR FUNCTION BASED ON THE REFINED SIFT MATCHING POINTS, OVERLAID ON LANDSAT REFERENCE IMAGE IN SICHUAN. EACH WINDOW IS COMPOSED BY 200-BY-200 PIXELS AT 30M RESOLUTION. 71

FIGURE 4.9 CORRECTED CORONA IMAGES (RED FRAMED) AFTER A) PRELIMINARY REGISTRATION, AND REGISTRATION USING B) 2ND ORDER POLYNOMIAL FUNCTION AND C) COLLINEAR FUNCTION BASED ON THE REFINED SIFT MATCHING POINTS, OVERLAID ON LANDSAT REFERENCE IMAGE IN ST. LOUIS. EACH WINDOW IS COMPOSED BY 100-BY-100 PIXELS AT 30M RESOLUTION. 72

FIGURE 4.10 EXAMPLE OF LAND COVER CHANGE BETWEEN 1960S AND CIRCA 2000 IN MGDS AND VM STUDY AREAS CAPTURED BY CORONA (LEFT) AND LANDSAT (RIGHT) DATA. 76

FIGURE 4.11 EXAMPLE OF LAND COVER CHANGE BETWEEN 1960S AND CIRCA 2000 IN SC AND SL STUDY AREAS CAPTURED BY CORONA (LEFT) AND LANDSAT (RIGHT) DATA. 77

FIGURE 5.1 BIOME MAP (UPPER-LEFT), DOWNLOADED FROM THE WORLD WILDLIFE FUND ([HTTP://WWW.WORLDWILDLIFE.ORG/PUBLICATIONS/TERRESTRIAL-ECOREGIONS-OF-THE-WORLD](http://www.worldwildlife.org/publications/terrestrial-ecoregions-of-the-world)), AND GLCF-GFCC MAPS FOR SICHUAN, INCLUDING FOREST COVER MAP IN THE 1975 (UPPER-RIGHT), AND CHANGE MAP FROM 1990 TO 2000 (LOWER-LEFT), AND FROM 2000 TO 2005 (LOWER-RIGHT), DOWNLOADED FROM GLOBAL LAND COVER FACILITY ([HTTP://LANDCOVER.ORG/DATA/LANDSATFCC/](http://landcover.org/data/landsatfccc/)). 82

FIGURE 5.2 FOOTPRINT OF CORONA DATA FOR 1960S IN SICHUAN THAT ARE FREELY AVAILABLE FOR DOWNLOAD, INQUIRED FROM EARTHEXPLORER WEBSITE ([HTTP://EARTHEXPLORER.USGS.GOV/](http://earthexplorer.usgs.gov/)). 84

FIGURE 5.3 STRATA MAP (LEFT) AND THE LOCATION OF SELECTED SAMPLE BLOCKS (RIGHT). 87

FIGURE 5.4 PATH/ROW SELECTED FOR THE ACCURACY ASSESSMENT OF GLCF-GFCC PRODUCT IN SICHUAN, DISPLAYED ON TOP OF THE BIOME MAP, DOWNLOADED FROM THE WORLD WILDLIFE FUND ([HTTP://WWW.WORLDWILDLIFE.ORG/PUBLICATIONS/TERRESTRIAL-ECOREGIONS-OF-THE-WORLD](http://www.worldwildlife.org/publications/terrestrial-ecoregions-of-the-world)). 92

FIGURE 5.5 EXAMPLES DEMONSTRATE FOREST COVER DYNAMICS OVER FIVE EPOCHS IN SICHUAN, INCLUDING A) FOREST COVER LOSS DUE TO COAL MINING; B) FOREST LOGGING AND AFFORESTATION; C) CONVERSION FROM FOREST TO AGRICULTURAL LAND. ESRI WORLD IMAGERY MAPS AT VERY HIGH RESOLUTION ARE SHOWN AS REFERENCES (THE LAST ROW). ... 97

FIGURE 5.6 COMPARISON OF FOREST COVER RATES FOR 2000 AND 2005 AND CHANGE RATES DURING THE FIVE YEARS DERIVED FROM GLCF-GFCC MAPS AND 6TH AND 7TH NFI STATISTICS RESPECTIVELY (SONG ET AL. 2014). 100

FIGURE 5.7 DEFINITIONS OF MAJOR FOREST LAND TYPES AND KEY VARIABLES IN FOREST INVENTORY AND STATISTICS. “FOREST COVER” IS USED FOR COMPARISON WITH REMOTE SENSING-BASED ESTIMATE OF FOREST COVER. 101

FIGURE 6.1 (A) SIZE-FREQUENCY DISTRIBUTION OF FOREST DISTURBANCES IN CHINA BASED ON GLOBAL LAND COVER FACILITY 2000-2005 FOREST COVER CHANGE DATA. B, C, AND D SHOW THE HORIZONTAL SLICES THROUGH PART (A) AT FREQUENCY LEVEL OF 1 ($=10^0$), 2 ($\approx 10^{0.3}$), AND 3 ($\approx 10^{0.48}$) RESPECTIVELY. NUMBER OF BINS ARE 50, 30 AND 10 FOR (B), (C) AND (D), RESPECTIVELY, BUT SOME BINS INCLUDE ZERO EVENT. THE HISTOGRAMS ARE FITTED USING GAMMA DISTRIBUTION (BLACK CURVES IN B, C AND D). NOTE THAT ALL DOTS IN (A) ARE SEPARATED WHEN ZOOMED ALL THE WAY INTO A RESOLUTION WHERE VERY SLIGHT DIFFERENCES IN PATCH SIZE COULD BE RESOLVED. 113

FIGURE 6.2 FOREST DISTURBANCE RATE AND PATCH SIZE MAP FOR CHINA AT 5-KM RESOLUTION. ADMINISTRATIVE BOUNDARY MAP IS PRODUCED BY NATIONAL GEOMATICS CENTER OF CHINA, AND TERRESTRIAL ECOREGION MAP IS DISTRIBUTED BY WORLD WILDLIFE FOUNDATION. 115

FIGURE 6.3 FLOWCHART FOR BUILDING A LOOK UP TABLE FOR THE PATCH SIZE $X_{N,I}$ AND FREQUENCY Y_I USING THE HIERARCHICAL MODEL. THE RANGE OF PATCH SIZE $[X_{MIN}, X_{MAX}]$ IS NEEDED AS INPUT (FOR THE RED BOX) TO INITIALIZE THE PROCEDURE AND THE OUTPUT IS A LUT. VARIABLES WITH SUPERSCRIPTS ARE AFTER LOGARITHMIC TRANSFORMATION AND THE ONES WITHOUT SUPERSCRIPTS ARE OF THEIR ORIGINAL VALUE. 121

FIGURE 6.4 COMPARISON OF FOREST-DISTURBANCE PATTERNS IN CHINA FROM 2000-2005 SIMULATED USING THE HIERARCHICAL METHOD (LEFT), SIMPLE ORDINARY LEAST SQUARES (OLS), (CENTER), AND MAXIMUM-LIKELIHOOD ESTIMATION (MLE) (RIGHT). SIMULATED

FREQUENCIES LESS THAN ONE WERE FORCED TO ONE, AND THOSE LARGER THAN ONE WERE ROUNDED TO THE CLOSEST INTEGER VALUE.	123
FIGURE 6.5 THE DIFFERENCE BETWEEN THE ESTIMATED TOTAL DISTURBANCE AREA USING DIFFERENT MODELING METHODS SEPARATELY AND THE TRUE TOTAL AREA FOR SELECTED PROVINCES (OR REGIONS) IN CHINA.....	124
FIGURE 6.6 MODEL ACCURACY ASSESSMENT MEASURED BY TOTAL AREA ESTIMATE AT DIFFERENT FREQUENCY INTERVALS, FI1 (10^0 - 10^1), FI2 (10^1 - 10^2), FI3 (10^2 - 10^3), FI4 (10^3 - 10^4), FI5 (10^4 - 10^5), AND FI6 (10^5 - 10^6) FOR SELECTED PROVINCES (OR REGIONS).....	126
FIGURE 6.7 MODEL ACCURACY ASSESSMENT MEASURED BY FREQUENCY ESTIMATE AT DIFFERENT PATCH-SIZE INTERVALS: SI1 (10^0 - 10^1 HA), SI2 (10^1 - 10^2 HA), SI3 (10^2 - 10^3 HA), SI4 (10^3 - 10^4 HA), AND SI5 (10^4 - 10^5 HA), FOR SELECTED PROVINCES (OR REGIONS).	128
FIGURE 6.8 THE PERCENTAGE OF TOTAL DISTURBANCE AREA ACCOUNTED FOR BY DIFFERENT SIZE OF PATCHES, SIMULATED BY INCREASING TWO PARAMETERS OF THE SIZE-FREQUENCY DISTRIBUTION, A_1 FROM 1.9 TO 2.3 WITH AN INTERVAL OF 0.1 AND B_1 FROM 3.2 TO 4.2 WITH A STEP OF 0.2 RESPECTIVELY.	129
FIGURE 6.9 SIMULATION OF THE PERCENTAGE OF TOTAL DISTURBANCE AREA ACCOUNTED FOR BY PATCHES WHEN VARYING GAMMA DISTRIBUTION PARAMETERS IN SIZE-FREQUENCY MODELING. THE HORIZONTAL AXES IS DISTURBANCE AND THE VERTICAL AXES IS THE PERCENTAGE OF DISTURBANCE AREA ACCOUNTED FOR BY DISTURBANCES WITHIN EACH SIZE INTERVAL.	130
FIGURE 6.10 DIFFERENCES IN TOTAL ESTIMATED AREA USING PROPOSED, HIERARCHICAL METHOD WHEN INCREASING SAMPLING PROPORTION FROM 10 TO 90% AT 10% INTERVALS.	131
FIGURE 6.11 FOREST DISTURBANCE AREAS (A) AND THE A ESTIMATES (B) FOR TERRESTRIAL FOREST ECOREGIONS IN CHINA. BIOME 1: TROPICAL AND SUBTROPICAL MOIST BROADLEAF FORESTS; BIOME 2: TEMPERATE BROADLEAF AND MIXED FORESTS; BIOME 3: TEMPERATE CONIFEROUS FORESTS.	133

FIGURE 6.12 FOREST DISTURBANCE AREAS (LEFT) AND THE ESTIMATED A (RIGHT) FOR SELECTED
PROVINCES IN CHINA..... 134

Chapter 1: Introduction

1.1 Information needed for forest cover and change

Forest, covering about 30% of global land surface, is important in balancing carbon cycle, water exchange, and terrestrial energy budget (Defries et al. 1995). Global forest plays a role as terrestrial carbon sink sequestering 2.4 ± 0.4 petagrams of carbon per year (Pg C year^{-1}) from 1990 to 2007 (Pan et al. 2011). Land-use activities have caused a great loss of ~ 7 to ~ 11 million hectares of forest in the past three centuries, which affected the biogeophysical cycles, as well as habitat quality and human wellbeing (Foley et al. 2005). The change in forest cover, especially, the deforestation, is a major driver of anthropogenic climate change (Bonan 2008) and the major driver of biodiversity loss (Pimm et al. 2014). Until 2000, 70% of the remaining forest on Earth is within 1 km of the forest's edge, subject to the degrading effect of forest fragmentation (Haddad et al. 2015) and the degradation in carbon stock near forest edge (Chaplin-Kramer et al. 2015). Therefore, an accurate monitoring of forest cover change over a long-term period is desired to assist the evaluation of the impact of past forest change to ecosystem and climate change, to seek for solutions of many environmental problems and to assist the prediction of future changes.

1.2 Changes in China's forest and its forest policies

China, the most populous country on Earth, is one of the countries with more than 1 million km^2 of forest (Hansen et al. 2013) and is the largest emitter of fossil-fuel CO_2 into the atmosphere which increased to $1.5 \text{ Pg C year}^{-1}$ by 2006 (Piao et al. 2009). However, China's terrestrial ecosystem played a role of carbon sink in the

range of 0.19-0.26 Pg C yr⁻¹, which is comparable to that in the terrestrial biosphere of Europe (Janssens et al. 2003). It also absorbed 28-37 per cent of the fossil-fuel carbon emissions during the two decades before 2000 (Piao et al. 2009). Among the two carbon sink categories in terrestrial ecosystem, vegetation and soil, the average carbon balance in China's forest is 75.2±34.7 teragrams (Tg) C yr⁻¹ during 1982 and 2003, which is equivalent to the total carbon in soil (Fang et al. 2007; Piao et al. 2009). As a result, the history and future of forest cover changes in China have regional to global consequences considering the country's sheer magnitude and rapid development.

China has undergone frequent social and economic transitions since the founding of the current government in 1949, which at times led it to fast economic development and natural resource consumption. At the same time, the frequently changed forestry policies, with goals transitioned from providing timber products and fuel woods to afforestation and conserving natural resources, have likely caused drastic forest cover change in China over decades (Albers et al. 1998; Liu et al. 2008; Zeng et al. 2015). However, literature presents sharp controversy on China's long-term forest dynamics. Some studies acknowledged the success of China's afforestation programs by citing the official statistics that the national forest cover increased from 12.7% in 1976 to 20.36% in 2008 (Fang et al. 2001; FAO and JRC 2012; Nabuurs et al. 2007). Whereas, others remained critical by showing the irreversible destruction to the natural forests (Smil 1984) and the continuous deforestation in some key forest regions (Gao and Liu 2012). The rest kept conservative by claiming that the reported forest increase due to afforestation has not

yet been traceable by satellite observations (Hansen et al. 2013). Little is known about the actual forest cover change and hence the effectiveness of the changed policies (Vina et al. 2016). Therefore, an accurate and objective estimation on the long-term forest cover and change in China is needed to understand the past changes and to evaluate the effectiveness of forest policies and conservation programs.

1.3 Approaches and issues in forest cover change monitoring

Landsat-based forest cover change mapping

Remote sensing data have been used to monitor forest cover change from regional to global scales at various spatial and temporal resolutions, complementing issues in inventory-based estimates such as spatial inaccessibility or temporal inefficiency (Friedl et al. 2010; Hansen et al. 2000; Hansen et al. 2013; Kim et al. 2015; Townshend et al. 2012). Featured with a long-term acquisition and fine spatial resolution, remote sensing data acquired by Landsat satellite facilitate the global fine-resolution land cover mapping. Landsat-based estimates of forest cover in different time periods have been published at global scale from the 1970s (Chen et al. 2015; Feng et al. 2016; Gong et al. 2013; Hansen et al. 2013; Kim et al. 2014). However, there were many forest changes happened before the Landsat era such as in China or South America, thus additional datasets are needed to monitor these changes.

Declassified satellite data for forest cover mapping

Declassified satellite images acquired by Corona program during the U.S. Key Hole (KH) missions (denoted as “Corona” image or data hereafter) from 1960 to 1972 can potentially be utilized to extend historical land cover mapping back to the

1960s. Corona data have a global coverage during the 1960s with repeated coverage for most of the Earth surface (Figure 1.1). Acquired during the Key Hole mission, Corona data are featured with particularly frequent observations for Eastern Europe and Asia, which allows the possibility of regional to global land cover land use change (LCLUC) applications in combination with other satellite data obtained at a later time. Although Corona data have been declassified for over two decades, the application of this legacy data on a broader spatial scale was still hindered by the poor geolocation accuracy and lack of multi-spectral bands (Song et al. 2015). Therefore, an automated approach suitable for applying Corona data in large area estimation of forest cover is desired.

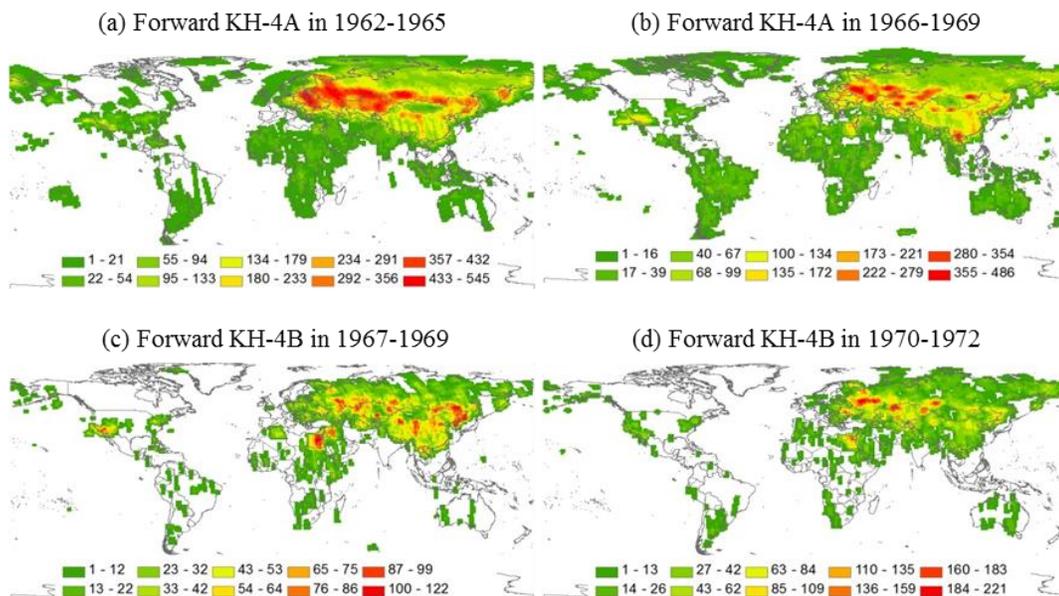


Figure 1.1 Global coverage of Corona images acquired by the forward-looking cameras during the KH-4A and KH-4B missions. All coverage maps have been rasterized for display purposes from vector format to 1° geographic grid. (a) KH-4A mission between 1962 and 1965; (b) KH-4A mission between 1966 and 1969; (c) KH-4B mission between 1967 and 1969; and (d) KH-4B mission between 1970 and 1972 (Source: https://ita.cr.usgs.gov/declass_1). Metadata was downloaded from U.S. Geological Survey (USGS) EarthExplorer website.

Inventory-based forest monitoring

Forest inventory serves as a valuable source of forest status for a country to make forest policies and management plans, but many countries lack the resources to conduct a regular survey. For countries where multiple inventories were conducted, the statistics derived from inventories are the only data accessible to researchers who do not have collaborations with the government (Zeng et al. 2015). However, the statistics have also been criticized for inconsistencies in forest definitions between epochs and inconsistent inventory approaches (Tomppo et al. 2010; Zeng et al. 2015). In addition, being reported at the regional scale every few years, the forest statistics also lack spatial and temporal details of forest cover, thus the locations and occurrence times of forest cover change are not sufficiently reflected (Tomppo et al. 2010). These insufficiencies underscore the necessity of complementing national forest inventory (NFI) statistics with forest cover change estimated from long-term satellite data record.

1.4 Need for accurate representation of the patch-size frequency distribution of forest change

Understanding of past forest cover change is also critical for the projection of future changes. In order to predict the future forest change and its ecological and carbon consequences, ecosystem or other spatially explicit models is usually used, in which the forest cover change is represented by its area and the distribution of patches at various sizes composing the total change (Asner et al. 2013; Espirito-Santo et al. 2014; Fisher et al. 2008; Hurtt et al. 2010). The patch size of forest change has been shown to impact carbon dynamics through affecting the forest recovery rate and tree species composition after forest disturbance (Putz 1983). The patch-size distribution

of forest change is also very important for fragmentation analysis to evaluate the edge effect (Haddad et al. 2015). Thus, an accurate representation of the patch size distribution is desired for the projection of the future change and its ecological and carbon impacts.

A power-law relationship is often used to model the patch-size frequency distribution that a small scaling exponent indicates a flat distribution dominated by large disturbance events while a large scaling exponent describes a steep distribution dominated by smaller events (Di Vittorio et al. 2014; Espirito-Santo et al. 2014; Fisher et al. 2008; Hurtt et al. 2010). The Ordinary Least Square (OLS) and Maximum Likelihood Estimator (MLE) are the two most used methods to estimate the parameters of a power-law relationship. However, they either underestimated the power-law exponent (Lloyd et al. 2009) or were skewed towards the lower-end of patch size values, lacking a representation of the “fat-tail” at low frequency levels (Clauset et al. 2009). Thus, the modeling of the patch size distribution of forest change needs to be improved. Since the remote sensing-based forest cover change map provides spatially explicit information of the past change, it could be used in developing a new model to describe the patch size distribution.

1.5 Research questions and objectives

The overarching question of this dissertation is **what is the long-term forest cover change in China and how to represent the change for ecosystem modeling?** To address this question, approaches for using historical satellite images need to be developed and the representation of forest cover change need to be improved. The following research objectives are sequentially addressed to achieve the goal.

Objective 1: Explore the feasibility of using Corona data for mapping forest cover and the use of Corona and Landsat data for forest cover change monitoring.

Objective 2: Address the bottleneck problem of the automatic image registration in using Corona data for large area forest assessment.

Objective 3: Derive forest cover estimate in the 1960s and assess the forest cover change over four decades for a major forest province in China.

Objective 4: Improve the representation of the patch size-frequency distribution of forest disturbances.

1.6 Dissertation organization

The dissertation consists of seven chapters, four of which correspond to the specific research objectives.

Chapter 1 introduces the topic, briefly reviews the methods for forest cover change monitoring, discusses issues in China's forest statistics and issues in the representation of forest disturbance in ecosystem models, presents the research questions, and sets the research objectives.

Chapter 2 reviews the satellite-based forest cover change monitoring, and presents the issues and challenges in using Landsat and Corona data for mapping the change.

Chapter 3 demonstrates the feasibility of applying Corona data in historical forest cover mapping and presents a method to separate forest and non-forest from

Corona data with high accuracy. Study areas located in the eastern USA and central Brazil were used to develop the algorithms in Chapter 3 considering the availability of local reference data.

Chapter 4 focuses on the registration of Corona data by developing a semi-automated registration method, leveraging the Corona data to large scale application. Four study areas located in the USA, Brazil and China were used featuring terrain complexities and land cover transition types.

Chapter 5 uses the Corona and Landsat data to estimate the forest cover change from the 1960s to 2005 in China. The methods developed in Chapter 3 and 4 for Corona data are adopted and the relevance of the forest cover change to forest policies is then discussed. Sichuan province was selected as the study area considering the fact that Sichuan hosts both state-run and collective-run forests thus it is affected by both government policies and the timber market. It was also a pioneer region of multiple afforestation programs. The processing time and the cost of Corona data are the other major reasons for conducting the analysis at the province scale.

In Chapter 6, the modeling of the patch size-frequency distribution for forest disturbances is improved by using Landsat-based forest cover change map. The spatial variation of the distribution in ecosystems and provinces in China is then analyzed to further examine the impact of frequent change in forest policies on forest cover change.

Chapter 7 synthesizes the results of the previous chapters and presents the conclusions and implications of the results. The dissertation concludes with a

discussion of the limitations of current work and the direction to move the research forward.

Chapter 2: A review of remote sensing-based forest cover change monitoring

2.1 *Introduction*

The launch of the first Landsat satellite (originally named the “Earth Resource Technology Satellite”, ERTS-1) in 1972 opened an era of monitoring the Earth’s terrestrial surface by space-borne, remotely sensed imagery. To date, Landsat archive provides the only globally available and internally consistent data record for the past four decades. The medium spatial resolution, global coverage, and potential for time-series analysis of the Landsat archive record have enabled assessments of forest cover change from local to global scales (Brandt et al. 2012; Huang et al. 2009b; Sexton et al. 2013b). However, forest cover change happened before the Landsat era remains unknown.

Designed for a reconnaissance purpose, the declassified satellite images acquired by Corona program during the U.S. Key Hole (KH) missions from 1960 to 1972 can extend the global satellite observations effectively by at least a decade (Song et al. 2015).

In this chapter, Landsat-based forest cover change monitoring is reviewed and potential issues are discussed. The characteristics of Corona data are then introduced and the potential issues hindering the application of Corona data to large area are also discussed. The chapter concludes with a recommendation of a combined use of Landsat and Corona data to extend remote sensing-based land cover change monitoring back to the 1960s.

2.2 *Landsat-based forest cover change monitoring*

2.2.1 Global forest cover change monitoring at Landsat resolution

Benefiting from the public opening of the Landsat archive, Landsat datasets have facilitated forest change analysis at the global scale (Townshend et al. 2012). Four global forest/land cover products have been developed using Landsat imagery or data at a similar resolution (Table 2.1) (Chen et al. 2015; Feng et al. 2016; Gong et al. 2013; Hansen et al. 2013; Kim et al. 2014). Although forest was defined differently by different datasets, almost all products have reached an overall accuracy of above 80% for the forest class. Three out of four products characterized forest cover after 2000, with one product produced by Global Land Cover Facility mapped forest cover for epochs of 1975 and 1990.

2.2.2 The multi-epoch GLCF-GFCC product from 1975 to 2005

The global forest cover change data produced by the Global Land Cover Facility (GLCF-GFCC) mapped forest/non-forest cover at 60 m resolution for 1975, and the change of forest between 1990, 2000 and 2005 also at 30 m resolution. The GLCF-GFCC map is produced by using enhanced Global Land Survey (GLS) data collection for epoch 1990, 2000 and 2005 (Channan et al. 2015; Feng et al. 2016; Kim et al. 2014; Sexton et al. 2013a). The entire MultiSpectral Scanner (MSS) collection during the 1970s are used to make the best pixel composite for removing gaps caused by clouds or shadows; thus, the GFCC 1975 spans a range of years but represents the nominal year of 1975 (Feng 2015).

Table 2.1 Global forest cover and change products at Landsat or similar resolution. Accuracy measurements include overall accuracy (*OA*), producer's accuracy (*PA*), user's accuracy (*UA*) and standard error (*SE*).

Product (Reference)	Temporal coverage	Spatial resolution	Map classes	Highest accuracy	Forest definition
Global Forest Change (Hansen et al. 2013)	Annual map between 2000 and 2014	30 m	Tree cover in 2000, forest cover loss and gain	UA = 87% PA = 88% for forest loss	All trees height > 5m
GLCF-GFCC* (Feng et al. 2016; Kim et al. 2014)	Circa-1975, 1990, 2000-2005	60m for 1975, 30 m after 1990	Forest and non-forest for 1975 and 1990; tree cover in 2000; forest loss and gain after 2000	OA = 91% (SE = 1%) after 1990	Tree cover > 30%
GlobeLand30 (Chen et al. 2015)	2000, 2010	30 m	Water bodies, wetland, artificial surfaces, cultivated land, permanent snow/ice, forest, shrubland, grassland, bareland, tundra	UA = 84% PA = 92% for forest	Not provided
FNF* maps (Shimada et al. 2014)	Annual map between 2007 and 2010	25 m	Forest and non-forest	OA = 85% - 95%	Crown cover > 10%, area > 0.5 ha
FROM-GLC* (Gong et al. 2013)	Circa 2010	30 m	Cropland, forest, grassland, shrubland, wetland, water bodies, tundra, impervious surface, barren land, snow and ice	UA = 80% PA = 76% for forest	Tree cover > 15% and height > 3m
*Note: GLCF-GFCC: Global Land Cover Facility, Global Forest Cover Change; FNF: Forest and Non-Forest; FROM-GLC: Finer Resolution Observation and Monitoring of Global Land Cover.					

The GFCC collection is featured with consistent forest definition, consistent training and classification algorithms across different epochs (Kim et al. 2014; Sexton

et al. 2013a). Thirty meter resolution tree cover layers circa 2000 and 2005 were generated by integrating Landsat surface reflectance and the Moderate Resolution Imaging Spectrometer Vegetation Continuous Field (MODIS VCF) tree cover data with a regression tree algorithm (Sexton et al. 2013a). Subsequently, forest cover in the 2000 and 2005 epochs were derived by translating percent tree data to categorical forest/non-forest cover with the International Geosphere-Biosphere Program (IGBP) forest definition (Belward 1996). Forest cover loss and gain between 2000 and 2005 were estimated using a probability-based bi-temporal change detection algorithm (Sexton et al. 2015). For the 1990 epoch, stable pixels that were either persistent forest or persistent non-forest between 2000 and 2005 were identified and used as training to classify the forest cover with 1990 Landsat surface reflectance as input (Kim et al. 2014). A similar algorithm was developed for the 1975 epoch by adding training samples collected from 1990 (Feng 2015). Forest cover data in the 1990, 2000 and 2005 as well as the forest cover change data between 2000 and 2005 are publicly available from the Global Land Cover Facility (GLCF) (www.landcover.org), while the production of forest cover circa 1975 epoch is still in progress.

A comprehensive accuracy assessment for the forest cover and change layers was carried out based on a global, design-based sample of 27,988 points which were marked as forest/non-forest by visual interpretation based on Landsat and high-resolution satellite images, vegetation index profiles, and field photos. The overall accuracies (*OAs*) was consistent across epochs, equaling 91% and the overall accuracy of forest cover change was over 88% (Feng et al. 2016). Both commission

error and omission error were low for static forest class in each epoch, but errors were larger for forest change classes (Feng et al. 2016).

2.2.3 Issues in using Landsat data for forest cover change monitoring

As techniques improve during the past four decades, the Landsat collection comprises data acquired by multiple sensors including the MultiSpectral Scanner (MSS), Thematic Mapper (TM), Enhanced Thematic Mapper Plus (ETM+), and the Operational Land Imager (OLI). The multi-sensor characteristic also brings issues in monitoring forest cover change when using data comprised from different sensors.

First, the varying resolution from 60 m MSS data to 30 m TM/ETM+/OLI makes the forest cover and change be mapped at different resolutions. It also indicates that changes at sub-pixel level with size smaller than 0.36 ha or 0.09 ha may be omitted when using MSS data or data from the other sensors separately (Lechner et al. 2009; Townshend and Justice 1988).

Second, the bandwidths as well as the calibration of spectral bands are not consistent among different sensors, making it difficult to assess changes by directly comparing the differences in images (Holden and Woodcock 2016). Images acquired by different sensors are usually converted to a common variable, such as tree cover or vegetation indices, to estimate the change (Kim et al. 2014; Sexton et al. 2013a).

Third, explicit atmospheric correction of MSS data is hindered by the lack of concurrent aerosol parameters. Instead, a radiometric matching approach or the top-of-atmospheric (TOA) reflectance data are usually adopted when using a MSS data together with a TM/ETM+ data to detect forest change (Feng 2015; Townshend et al.

2012). Another difficulty in using MSS data is the problems in cloud detection and separating cloud from snow due to its lack of shortwave infrared and thermal bands (Dozier 1984; Jorgensen 2000).

Fourth, the geolocation accuracy varies among different sensors, particularly between MSS and the other sensors. A difference of up to two MSS pixels (~120 m) in geolocation between MSS and TM both onboard of Landsat 5 was observed (He et al. 2016). The misregistration among images poses risks of detecting false change or obscuring small changes (Roy 2000; Townshend et al. 1992). Given the inconsistent geolocation accuracy, an additional image-matching process is usually adopted (Huang et al. 2010a). However, the Global Land Survey dataset consisted of precisely registered Landsat image for five epochs, 1975, 1990, 2000, 2005 and 2010, provides variable option for mapping forest cover change over a long-term period (Gutman et al. 2013; Kim et al. 2014).

Last, the spatial coverage especially for MSS and TM in the 1980s and 1990s is limited due to data acquisition strategy and sharing policy in the early years of the Landsat program (Gutman et al. 2013; Townshend et al. 2012), hindering a spatially complete mapping for those time periods.

2.2.4 Needs for forest cover monitoring before the 1970s

Some of the technical issues in using Landsat data for forest cover change monitoring discussed above may be solved if appropriate approaches are adopted or developed. However, another issue unsolvable by Landsat is that for many parts of the world, large-scale forest changes were already occurring prior to the Landsat era.

For instance, eastern Paraguay was once covered by intact forests, but only $73.4 \pm 4.9\%$ of the original Atlantic forests were left by the 1970s (Huang et al. 2009b; Nagel 1991; Nickson 1981). In southern Brazil, agricultural developments in the 1960s and 1970s resulted in the consolidation of small farms and a shift from labor-intensive crops to extensive ranching and soy production (Richards 2011), which resulted in a large forest area being cleared. In these circumstances, knowledge of land cover prior to the Landsat era is important for a complete understanding of the impact of socioeconomic activities on natural resources.

2.3 Corona data in forest cover estimation

2.3.1 Characteristics of Corona data

Images acquired by the U.S. Key Hole (KH) missions, which consisted of the Corona, Argon, and Lanyard satellites that operated from 1960 to 1972 (McDonald 1995), have the potential to extend historical land cover mapping from the Landsat era into the 1960s. The declassified Corona imagery has worldwide spatial coverage, especially in Eastern Europe and Asia (Figure 1.1) (Song et al. 2015).

KH-4A and KH-4B missions accomplished between the 1960s and early 1970s acquired most of the high-quality images among all KH missions, and satellites launched during these two missions shared similar camera, platform, and image parameters. For instance, both the KH-4A and KH-4B cameras are dual panoramic camera system, with one pointing 15° forward from the nadir and the other pointing 15° afterward. This configuration resulted in a 30° angle between the two cameras and stereo images were acquired as the satellite moved along the flight direction

(Sohn et al. 2004). Both the forward- and backward-looking camera are characterized with focal length of 61 cm and wide view angle of 71.16° along the scan direction of panoramic image (Galiatsatos 2009a). The images were originally recorded on film by panchromatic cameras and then digitized by USGS to 8-bit radiometric precision (Galiatsatos 2009a). The panoramic camera worked in the way of scanning rather than taking a whole image at a time. It took the camera 0.5 second to scan an entire Corona image (Sohn et al. 2004). The flying altitudes were around 180 km for KH-4A satellites and 150 km for KH-4B, respectively, so the ground resolution of the image varies across the scan direction, ranging from 2.7 to 7.6 m for KH-4A image and from 1.8 to 7.6 m for KH-4B image. Along the scan direction, the pixels with the best resolution are located at the image centers. The nominal ground coverage per frame is about $17 \times 232 \text{ km}^2$ and $13.8 \times 188 \text{ km}^2$ for the two satellite separately (Galiatsatos 2009a).

2.3.2 Applications of Corona data in land cover studies

Corona images have been used from studying the boreal forest decline (Rigina 2003), vegetation dynamics (Kadmon and Harari-Kremer 1999), land resource change (Tappan et al. 2000), forest fire carbon emissions (Isaev et al. 2002), ice sheet change (Bindschadler and Vornberger 1998), to analyzing the archaeological features (Beck et al. 2007; Casana and Cothren 2008; Challis et al. 2002). However, in these localized studies, Corona images were analyzed mainly through visual interpretation and manually tuned histogram segmentation or were used as stereo image pairs for extracting digital elevation models (DEMs). Although the spatial coverage of available Corona data may allow large-area applications for many parts of the world,

such applications would be possible only using more automated digital image analysis methods.

2.3.3 Bottleneck problem in using Corona data for large area forest assessment: image registration

The biggest challenge hindering the use of Corona data for large area assessment remains in the accurate automated georegistration. Since most of the satellite parameters are nominal values and vary more or less during the satellite operation. The initial geolocation accuracy of Corona images provided by USGS is usually very poor (USGS 2009). There are three major factors affecting the procedure of an automated image matching based on finding adequate number of tie points between Corona and the reference data. First, the accurate satellite orbit information is absent and the gesture control is not stable for the historical satellite, so there is not an accurate estimation of the geolocation of the Corona scene. The scan time (~ 0.5 s) of the image's panoramic projection brings an additional source of errors in determining the exterior orientation parameters due to an unknown flight height and velocity. Second, there are four subsets composing a Corona scene distributed by USGS which features a moving image center. The panoramic camera also makes the resolution of off-nadir pixel different from the pixel at image edge. Third, cumulative land cover changes since 1960s bring a lot of uncertainties in matching Corona image with other satellite images which are obtained at least a decade later. The poor image quality due to the low signal-to-noise (SNR) further makes image interpretation and feature selection more difficult.

2.4 *Summary*

Enabled by multiple Landsat-based forest cover and change products, the global forest cover change happened since the 1970s can be effectively estimated. But there are still issues in Landsat-based mapping that will affect the change detection, including the omission of small patches, inconsistent bandwidths and calibration, atmospheric correction for MSS data, geolocation accuracy and limited spatial coverage. Specific approaches should be adopted to solve the issue according to the mapping purpose. Corona data provides a great change to extend the forest cover change mapping back to the 1960s. But before achieving this goal, challenges need to be addressed including developing an automated image classification and registration method for Corona data. In summary, it is worth to explore the feasibility of building a forest cover change record from the 1960s by combining the Corona and Landsat data.

Chapter 3: Use of Landsat and Corona data for mapping forest cover change from the mid-1960s to 2000s: case studies from the eastern United States and central Brazil¹

3.1 *Introduction*

Building on the conclusion of Chapter 2, this chapter aims to explore the feasibility of extending forest change monitoring back to the 1960s using Corona data and Landsat data. Focusing on increasing automation relative to earlier approaches, major methodological components of this approach include extraction of texture features from Corona images, classification of Corona and Landsat images, and change detection based on the classification results. The effectiveness of this approach is demonstrated by mapping forest cover change between four epochs—1960s, 1980s, 1990s, and 2000s—in two study areas that have experienced major anthropogenic forest changes: an urbanized landscape in the eastern United States and a forested area in central Brazil characterized by recent agricultural expansion. The following sections provide a brief description of the study areas and datasets, followed by a detailed description of the various methodological components and results derived using these methods. The chapter closes with a discussion of potential improvements and applications of the approach developed in this study.

¹ This chapter has been published in 2015: Song, D.-X., C., Huang, J.O., Sexton, S., Channan, M., Feng, J. R., Townshend, Use of Landsat and Corona data for mapping forest cover change from the mid-1960s to 2000s: Case studies from the Eastern United States and Central Brazil. *ISPRS J. Photogram. Remote Sensing* (2015). 103, 81-92.

3.2 *Study area and data*

3.2.1 Study area

Two study areas with widely different vegetation properties and human land uses were selected to determine the feasibility of using paired Corona and Landsat images for detecting forest changes. The sites were selected to represent the range of forest types and change trajectories that impact forest cover classification of high resolution remote sensing images and change detection. The Virginia-Maryland (VM) study area in the eastern United States, centered at 39°02'34.63'' N, 77°23'35.45'' W and spanning 32.6-by-23.6 km, lies within the temperate mixed forest biome and experienced forest change due to urbanization and managed forest planting. Forests in this region comprise a mix of deciduous (e.g. oak and hickory) and evergreen (e.g. loblolly pine) species. The leaf-on season for deciduous forest normally starts in early April and ends in October (White et al. 2002). Forest phenology was considered when choosing appropriate remote sensing image acquired in leaf-on season. Areal coverage of forest and non-forest classes is approximately equal in the VM study area.

Located in central Brazil, centered at 9°36'52.86'' S, 50°12'34.69'' W and covering 47.5-by-26.1 km area, the Mato Grosso-Tocantins-Pará (MTP) study area lies within the southeastern frontier of the Amazon rain forest and has experienced widespread forest loss due to agricultural expansion. The region's tree species diversity is exceptionally high and the predominant evergreen forest phenology is not a dominant factor when selecting satellite image dates. Instead, cloud cover is a stronger limitation on data quality. Before these forest losses took place, the MTP

study area was mainly covered by forest with a very small proportion of other land cover classes.

3.2.2 Landsat Data

Landsat images were collected from three epochs: mid-1980s, circa-1990, and circa-2000 (Table 3.1). For each site, the mid-1980s image was downloaded from the USGS. The latter two images were part of the Global Land Survey (GLS), a collection of images optimally selected for land cover change detection and orthorectified to within one pixel geolocation accuracy (Gutman et al. 2008; Tucker et al. 2004). A Landsat Multi-Spectral Scanner (MSS) image, representing a circa-1970 epoch, was also included for the VM study area but not for the MTP study area due to lack of high-quality MSS image.

Table 3.1 WRS-2 and WRS-1* Path/row numbers, sensors and acquisition dates of Landsat images used in this chapter.

VM study area				MTP study area			
Path	Row	Acq. date	Sensor	Path	Row	Acq. date	Sensor
16*	33*	10/11/1972	MSS	-	-	-	-
15	33	09/15/1985	TM	223	67	07/17/1986	TM
15	33	05/16/1987	TM	223	67	07/25/1992	TM
15	33	10/05/2001	ETM+	223	67	09/01/2000	ETM+

3.2.3 Corona Data

Corona images used in this chapter were acquired in KH 4A and 4B missions in 1966 and 1967 and the major characteristics are listed in Table 3.2. Due to variations in pixel size and geometric distortions resulting from the wide view angles of the cameras, geometric correction of entire Corona images was very challenging. In order to achieve satisfactory geometric correction results, I divided each Corona

image into eight subsets along the cross-track direction, and one subset from each of the two study areas was selected for analysis.

Table 3.2 Properties of Corona images (Galiatsatos 2009b; NPIC 1967) used in this chapter.

	VM study area	MTP study area
Acquisition Date	09/25/1967	06/22/1966
Spatial Resolution	6 ft. (~ 1.83m)	9 ft. (~ 2.74m)
Camera	Forward looking	Afterward looking
Field of View	5° (along track)	5° (along track)
Scan Angle	70° (+/- 35° from track)	70° (+/- 35° from track)

3.3 *Methods*

Images gathered by sensors aboard the Corona satellites have different spectral coverage (e.g. panchromatic vs. multispectral) and ground resolutions (e.g. ~2 m vs. 30 m) than images acquired by Landsat Thematic Mapper (TM) or Enhanced Thematic Mapper Plus (ETM+) sensors. Change detection methods based on directly measuring differences of spectral responses between dates (Coppin et al. 2004; Macleod and Congalton 1998; Singh 1989) cannot be applied to data of such different radiometric and spatial characteristics. Spatial and radiometric differences thus necessitated independent image preparation and a post-classification approach to change detection.

Corona images were used to derive forest classifications for the 1960s, and the Landsat images were analyzed using an automated change-mapping algorithm to map forest-cover change in later periods. The maps derived from Corona and Landsat data were then combined to quantify forest-cover changes across the consecutive epochs from the 1960s to 2000s. The Landsat methods have been described in previous publications (Feng et al. 2012a; Huang et al. 2008b; Masek et al. 2006) and so are

only outlined here. The remainder of this section focuses on methods for processing Corona images.

3.3.1 Landsat Forest Cover Change Mapping

Landsat TM and ETM+ images were atmospherically corrected to estimates of surface reflectance using the Landsat Ecosystem Disturbance Adaptive Processing System (LEDAPS) implementation of the 6S atmospheric correction algorithm (Feng et al. 2012a; Masek et al. 2006). Forest changes between consecutive epochs from 1980s to 2000s (e.g. 1980s-1990s) were then mapped using the Training Data Automation-Support Vector Machine (TDA-SVM) algorithm (Huang et al. 2008b), which trains an SVM-based classification on a sample of “forest” pixels automatically identified using an a priori dark-vegetation heuristic rule. Clouds and their shadows were identified based on visible and thermal properties and masked from the dataset (Huang et al. 2010b). Landsat MSS image was classified using SVM based on forest and non-forest training samples collected through visual interpretation. Post classification change detection method was applied to map forest cover change between 1970s and 1980s. Pixels identified as cloud or cloud-shadow in any epoch were excluded from further analysis.

3.3.2 Corona forest cover mapping

Geometric correction

Geolocation information distributed with Corona images includes only approximate coordinates of image corners, and preliminary inspection revealed that images selected in this study had spatial errors of several kilometers. Corona images

were therefore registered to orthorectified GLS Landsat images. For each site, the Corona image was coregistered to the GLS 2000 image using road intersections and other stable ground features as ground control points (GCP). These GCPs (totaling 19 in the VM site and 8 in the MTP site) were manually selected through careful inspection of the Corona and Landsat images. In addition, 15 GCPs in the VM site and 10 GCPs in the MTP site were identified and used for independent assessment of registration accuracy. Several polynomial functions for coregistering the Corona images to Landsat data were tested, and the one with lowest RMSE was selected to correct the Corona image.

Texture-based classification

Panchromatic Corona images have limited spectral information for forest cover classification. However, their high spatial resolution allows calculation of texture measures at various spatial resolutions. Many of these texture measures were found useful to visually discriminate different land cover types land cover discrimination (Figure 3.1).

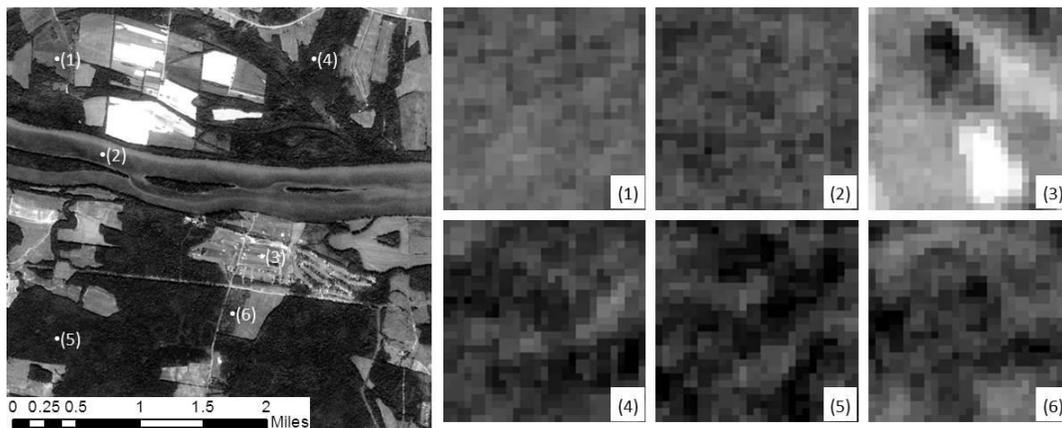


Figure 3.1 Corona image, aggregated to 27.5m resolution, in the left window has multiple land cover types, including forest, cropland, water body and scattered build-ups. Zoom-in windows 1, 2 and 3

show non-forest land cover, window 4 and 5 show dense forest cover and scatter forest cover is presented in window 6.

To evaluate the utility of texture measures for separating forest and non-forest in Corona data, I evaluated both occurrence (first-order) and co-occurrence (second-order) texture measures. Five occurrence textures were selected due to their effectiveness in land cover classification (Anys et al. 1994). Seven relatively uncorrelated and widely used co-occurrence textures (Anys et al. 1994) out of fourteen statistical features proposed by Haralick et al. (1973) were also selected. Occurrence metrics tested included data range (DR), variance (OC-VAR), mean (MN), entropy (OC-ENT) and skewness (SKE):

$$DR = GL_{\max} - GL_{\min} \quad \text{Eq. 3.1}$$

$$MN = \sum_{i=0}^{L-1} i \times P(i) \quad \text{Eq. 3.2}$$

$$OC - VAR = \sum_{i=0}^{L-1} (i - MN)^2 \quad \text{Eq. 3.3}$$

$$OC - ENT = - \sum_{i=0}^{L-1} P(i) \times \log_2 [P(i)] \quad \text{Eq. 3.4}$$

$$SKE = \frac{1}{(OC - VAR)^{3/2}} \times \left[\sum_{i=0}^{L-1} (i - MN)^3 \times P(i) \right] \quad \text{Eq. 3.5}$$

where L is the maximum grey level (GL), $P(i)$ is the frequency of pixels where $GL=i$ in a window, and “max” and “min” are the maximum and minimum value within a window. Co-occurrence texture measures included variance (CO-VAR), homogeneity (HOM), contrast (CON), dissimilarity (DIS), entropy (CO-ENT), second moment (SM), and correlation (COR):

$$CO - VAR = \sigma_x + \sigma_y \quad \text{Eq. 3.6}$$

$$CO - ENT = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} P(i, j, d, \theta) \times \log_2 (P(i, j, d, \theta)) \quad \text{Eq. 3.7}$$

$$CON = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} (i - j)^2 \times P(i, j, d, \theta) \quad \text{Eq. 3.8}$$

$$COR = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} [(i - \mu_x) \times (j - \mu_y) \times P(i, j, d, \theta)] / (\sigma_x \times \sigma_y) \quad \text{Eq. 3.9}$$

$$HOM = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} P(i, j, d, \theta) / (1 + (i - j)^2) \quad \text{Eq. 3.10}$$

$$DIS = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} (i - j) \times P(i, j, d, \theta) \quad \text{Eq. 3.11}$$

$$SM = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} P(i, j, d, \theta) \quad \text{Eq. 3.12}$$

where $P(i, j, d, \theta)$ is the frequency of co-occurrence between pixels where $GL=i$ and those where $GL=j$ given distance between the two pixels is d and direction of displacement is the angle θ (Anys et al. 1994). σ_x and σ_y are the standard deviation of lines and columns of the Grey Level Co-occurrence Matrix (GLCM), and μ_x and μ_y are the means of lines and columns of the same matrix.

To evaluate the impact of window size on class separability, the texture measures were calculated using window sizes (n) of 5×5-pixel, 7×7-pixel, 9×9-pixel, 11×11-pixel, 13×13-pixel, and 15×15-pixel at the original resolution (r) of the Corona images. Each texture image was resampled using the same window size with the nearest neighbor (NN) resampling method to produce a texture image at spatial resolution of $r*n$.

Table 3.3 Texture measures used in this chapter and their sources of literature.

References	Texture measurements
Haralick et al. (1973)	Angular second-moment (ASM), CON, COR, CO-VAR, inverse different moment, sum average, sum CO-VAR, sum CO-ENT, CO-ENT, difference variance, difference CO-ENT, information measure of COR, max COR.
Kushwaha et al. (1994)	ASM, ENT, inverse difference moment
Anys et al. (1994)	First order: MN, ENT, standard deviation, SKE; Second order: absolute value, CO-ENT, CON, COR, CO-VAR, cluster prominence; Third order: absolute value, OC-ENT, CON, small number emphasis, depth emphasis.
Hudak and Wessman (1998)	Standard deviation
Shaban and Dikshit (2001)	GLCM, grey level difference histogram (GLDH), sum and difference histogram (SADH)
Clausi (2002)	CON, COR, C-ENT.
Kim et al. (2011)	ASM, CONT, COR, DIS, CO-ENT, HOM, MN, CO-VAR.
This study	Occurrence texture: DR, MN, OC-VAR, OC-ENT, SKE; Co-occurrence texture: CO-VAR, HOM, CON, DIS, CO-ENT, SM, COR. Texture combination 1(COMB1): HOM + CO-ENT + COR + MN Texture combination 2(COMB2): DIS + SM + CO-VAR + MN

SVM classification of the Corona data

Reflectance and texture metrics were used to derive forest/non-forest classifications using Support Vector Machines (SVM) (Chan et al. 2001; Huang et al. 2002; Pal and Mather 2005), an advanced machine learning algorithm designed to locate an optimal multivariate boundary between classes. Distributed across each study area, training data included 4641 forest and 4618 non-forest pixels in the VM study area and 2272 forest and 3580 non-forest pixels in the MTP study area (Table 3.4). The non-forest class comprised bare soil, herbaceous vegetation, water, and

other cover features, and so a larger portion of non-forest pixels were sampled than their true proportion in the MTP study area in order to characterize the range of variability within this complex class. Training pixels were labeled through visual interpretation of the Corona images at their native resolution by an experienced image analyst. Because SVM requires training samples located near the discrimination boundary between two classes in the feature space (Huang et al. 2002), some mixed pixels were deliberately included in the training sample. Edge pixels were included to represent high texture values near edges. The radial basis function (RBF) kernel was used because it has been found to be robust for various classification problems (Huang et al. 2002). Optimal values for the cost parameter c and the RBF kernel parameter γ were selected through five-fold cross validation following the procedure in Chang and Lin (2011). I first varied the parameter values at coarse steps and, once the approximate ranges of the optimal parameter values were determined, finer steps were used to optimize parameter values within the approximate ranges.

Table 3.4 Number of training and validating pixels used for forest and non-forest classification using Corona data in each study site

Study site	Training pixels		Validating pixels	
	Forest	Non-forest	Forest	Non-forest
VM	4641	4618	1064	926
MTP	2272	3580	1313	184

To evaluate the utility of various texture measures for forest/non-forest classification, I ran the SVM with different combinations of texture measures together with the panchromatic brightness values (i.e., MN) (Table 3.5). For each set of texture inputs, the SVM parameters that yielded the highest cross-validation accuracy for a study area were chosen and used to classify the entire image subset for that area.

Since our goal was to use these classifications together with results derived from Landsat to estimate forest change, all classifications derived using the Corona data were aggregated to 30-m resolution using a majority rule.

Table 3.5 Texture features used in the classifications, windows sizes to derive textures, scales at which classifications and corresponding accuracy assessments were carried out at the VM study area. *Note: at MTP study area, window sizes are 5x5, 7x7, 9x9, 11x11, and classification resolution is 13.7m, 19.04m, 24.66m and 30.14m, other parameters are the same as the VM study site.

Texture features	Window size*	Classification resolution*	Accuracy assessment resolution
MN			
DR + MN			
OC-ENT + MN			
SKE + MN			
OC-VAR + MN	5×5,	9.14m,	
CON + MN	7×7,	12.80m,	
COR + MN	9×9,	16.50m,	
DIS + MN	11×11,	20.12m,	30 m
CO-ENT + MN	13×13,	23.77m,	
HOM + MN	15×15-pixel	27.43m	
SM + MN			
CO-VAR + MN			
ALL TEXTURES			
HOM + CO-ENT + COR + MN			
DIS + SM + CO-VAR + MN			

3.3.3 Accuracy assessment of classification result

Forest cover and change classifications were evaluated using overall accuracy, kappa coefficient, and class-specific user's and producer's accuracies derived from a confusion matrix (Congalton 1991; Stehman and Czaplewski 1998). Reference data were collected by stratified random sampling with strata defined by land cover classes in the output forest cover and change maps (i.e. forest, non-forest, forest gain, and

forest loss). This approach is commonly incorporated in sampling design for accuracy assessment of global to regional land cover maps (Olofsson et al. 2012). Stratified random sampling is a probability-based sampling method, which enhances the precision of accuracy estimates for minor classes. It is thus advantageous over simple random sampling in reducing the standard error for estimating overall accuracy (Stehman 1999; Stehman and Czaplewski 1998). For each selected pixel, the true forest-change type was determined by visual comparison of the Landsat series against high-resolution images from Google Earth, similar to the method described by Feng et al. (2012b). A summary of the reference sample for evaluating the Landsat-based cover and change maps is provided in Table 3.6.

Table 3.6 Number of validation pixels for assessing Landsat-based FCC (1980s-2000s) products at each study area

Study site	1970s FC		1980s-1990s FCC				1990s-2000s FCC			
	Forest	Non-forest	Persistent Forest	Forest loss	Forest gain	Persistent non-forest	Persistent Forest	Forest loss	Forest gain	Persistent non-forest
VM	922	596	977	72	59	874	853	183	81	865
MTP	--	--	1436	67	6	433	1430	11	1	500

Reference samples for evaluating the Corona classifications were also selected randomly within forest and non-forest stratum in both study areas. For the classification derived from Corona data, each sample pixel was labeled as forest or non-forest at 30-m resolution based on visual analysis of the Corona images at their native resolutions. The number of reference samples for each class in each study area is listed in Table 3.4.

3.3.4 Forest cover change rate calculating

The Corona-based, 30-m forest/non-forest classifications with highest accuracies were used together with the Landsat-based FCC products to calculate forest cover and change rates between the 1960s and 2000s for the two study areas. Gross forest loss and gain rate were converted to average annual change rate by dividing the rate of gross loss or gain to the total land area by the time difference (in years) of each pair of images.

$$\text{Average annual loss rate (\%)} = \text{total loss area} / \text{total land area} / \text{year difference} \\ * 100 \text{ Eq. 3.13}$$

$$\text{Average annual gain rate (\%)} = \text{total gain area} / \text{total land area} / \text{year} \\ \text{difference} * 100 \text{ Eq. 3.14}$$

3.4 Results

3.4.1 Geometric Correction of Corona Data

For both the VM and MTP sites, second-order polynomial transformations yielded the smallest root mean square errors (RMSE) for both the control and check GCPs. Corona images for both study areas were coregistered within one half pixel (15 m) of the Landsat data when measured using the control points (Table 3.7). Although the RMSE of check points were higher than those of control points, suggesting some model over-fitting. Control-point errors were still less than one 30-m pixel. The geolocation accuracy of corrected Corona image was comparable to the Landsat images themselves (Tucker et al. 2004), so few spurious change detection would be caused due to misregistration. The greater co-registration errors in the MTP site were

likely due to lack of road intersections and other stable ground features that could be used as GCPs. As a result, river banks and some small water bodies were selected as GCPs, some of which could have moved during the period between the Corona and Landsat data acquisitions.

Table 3.7 RMSE of manual image geometric correction for Corona data at two study areas

Study area		VM		MTP	
		Control points	Check points	Control points	Check points
Number of points		19	15	8	10
RMSE (meters)	x	8.08	15.29	11.38	20.33
	y	6.05	10.28	4.58	21.18
	total	10.09	18.24	12.29	29.35

3.4.2 Effectiveness of Textures for Corona Forest/non-forest Classification

Accuracies of the Corona-based forest/non-forest classifications derived using various combinations of texture measures calculated using different window sizes are shown in Figure 3.2. The overall accuracies in the VM study area ranged from 90.5% to 95%, with Kappa coefficients between 0.81 and 0.90. The maximum overall accuracy and Kappa statistic were approximately 95% and 0.9, respectively. Except for the classification derived using mean reflectance (MN) alone, the overall accuracies and Kappa coefficients in the MTP ranged from 94.5% to 96.5% and from 0.75 to 0.82 respectively.

Some of the texture measures may not be suitable for forest/non-forest classification using Corona data. In the MTP site, accuracies derived using MN alone within window sizes > 5 pixels were substantially lower than those derived using texture measures. MN also yielded lower accuracies when calculated using window

sizes of 5×5 and 9×9 pixels. In the VM site, occurrence (OC) and co-occurrence (CO-OC) textures calculated using a window size of 15 pixels also yielded markedly lower accuracies. Accuracies derived using the other texture measures differed by up to 4% (or 0.04), but no individual texture measure had consistently better accuracies than the others. Interestingly, and likely due to model over-fitting, use of all texture measures (ALL) did not yield the best accuracies. However, in the VM site, the combination of HOM, CO-ENT, COR and MEAN (COMB2) calculated using window sizes of 9 pixels or larger seemed to have slightly better accuracies than all other texture combinations, while in the MTP site, the combination of DIS, SM, CO-VAR, and MEAN (COMB1) had marginally better accuracies.

Performances of different textures calculated at the same window size usually do not vary much but the impact of varying window size on accuracies had obvious patterns (Figure 3.2). In the VM site, windows of 7×7 to 9×9 pixels (or $\sim 13 \times 13\text{m}$ to $\sim 16 \times 16\text{m}$) yielded best accuracies for most texture measures, but accuracy decreased with window size. In the MTP site, the window size of 11×11 pixels (or $\sim 30\text{m} \times 30\text{m}$) yielded the best accuracies for most texture measures. Although land cover in the VM study area shows more heterogeneous pattern than that in the MTP site, patch size of forest land in eastern U.S. is much smaller than that in the undeveloped Amazon rainforest. In the VM site, small forest patches result in more mixed pixels, which can be better captured by textures calculated from relatively small window sizes. Overestimation of forest was caused as window size increased to 15×15 pixels. On the other hand, peppered non-forest pixels were mapped as random speckle in homogeneous forests in the MTP site when small window sizes

were used. The use of windows as large as 11×11 pixels avoided the misclassification of non-forested pixels caused by poor image quality in this site.

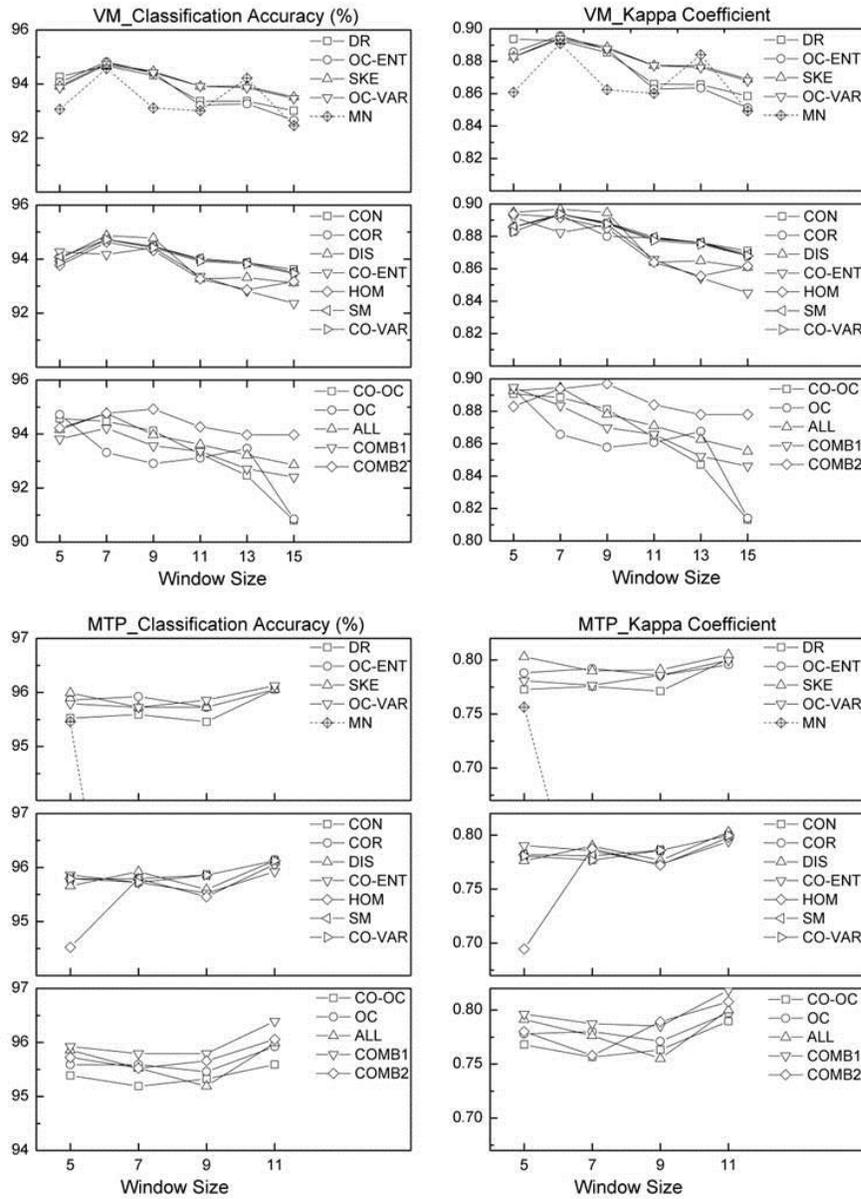


Figure 3.2 Classification accuracies and kappa coefficients of Corona forest/non-forest classifications using single and multiple textures in the VM and MTP study areas. Abbreviations are defined in Table 3.5. For the MTP study area, the value of classification using MN fell below the range of y-axis, kappa coefficients were 0.52, 0.57 and 0.46 at window sizes of 7×7 to 11×11 , and classification accuracies were 88.0%, 89.6% and 85.1% correspondingly.

3.4.3 Accuracy of Landsat Forest Cover Change Map

Overall accuracies were >91% for both the 1980s-1990s and 1990s-2000s periods in the VM study area, and were 96% and 87% for the two periods in the MTP area. Both persistent forest and persistent non-forest had user's accuracies >90% in all periods and producer's accuracies >90% in the majority of periods. The forest-loss class had slightly lower accuracies, but its user's and producer's accuracies were >70% in both sites during both periods. The forest-gain class had lower user's accuracies in the MTP site (62.5% and 50%), likely because the coverage of forest gains was very low, so only a small number of reference points were randomly sampled for estimating the accuracy for this class. Commission error increased significantly when the forest class was overestimated in time two; this led to reduced user's accuracy especially in forest gains. Overall accuracy of GLS1975 image classification achieved 94% with overall kappa coefficient of 0.89.

3.4.4 Forest Cover Change Rate

The highest-accuracy of the 30-m resolution map of forest cover based on Corona images was obtained by using the second texture combination (COMB2), calculated with window size of 9 for the VM study area and using the first texture combination (COMB1), calculated with window size of 11 for the MTP study area. Although both study areas lost substantial forest cover between the 1960s and 2000s, they had different forest change histories (Figure 3.3). Located in the suburbs of Washington, DC and the Chesapeake Bay region, the VM study area comprised areas in Fairfax and Loudoun Counties, Virginia and Montgomery County, Maryland (Figure 3.4). The annual forest loss rate during mid-1980s doubled to 2%. Most of the

loss happened before the 1980s and after the 1990s, especially in Loudoun County in Virginia. Annual forest gain rates were ~1% during the 1960s and the 1970s, then increased to ~2%—exceeding forest loss rate during the 1980s—and dropped afterwards. Most trees appear to have been planted around residential areas. As a result, forest cover in the VM study area slightly increased around late 1980s and then continued decreasing.

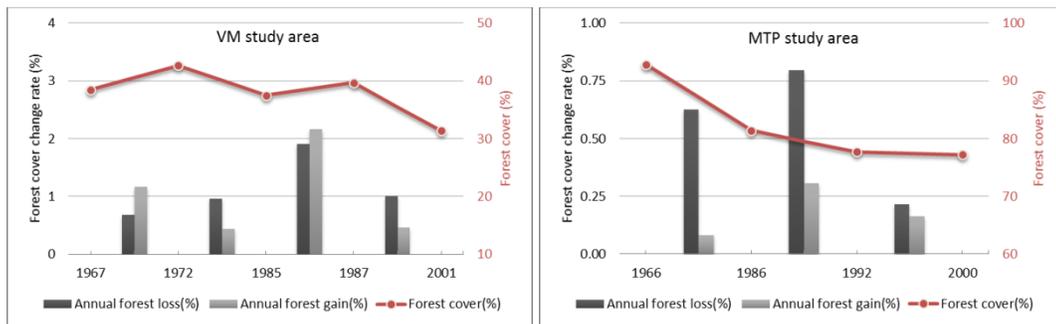


Figure 3.3 Forest cover change rates during three epochs for two study areas, forest loss (or gain) percentage = forest loss (or gain) area/forest area of beginning year * 100.

In contrast, the rate of annual forest loss in the MTP study area was relatively low (~0.6-0.76%) from the 1960s to the 1990s and dropped to ~0.2% during the period of 1992-2000. However, the patch size of cleared forest was much larger than in the VM study area (Figure 3.5), which was mainly cleared for cattle ranching and mechanized agriculture. The rate of forest gain was quite low (~0.1-0.26%) and was largely due to trees growing in abandoned agricultural land. Decrease of forest cover in the MTP study area continued from the 1960s to the early 1990s, driven by logging activities and gradually ceased till 2000.

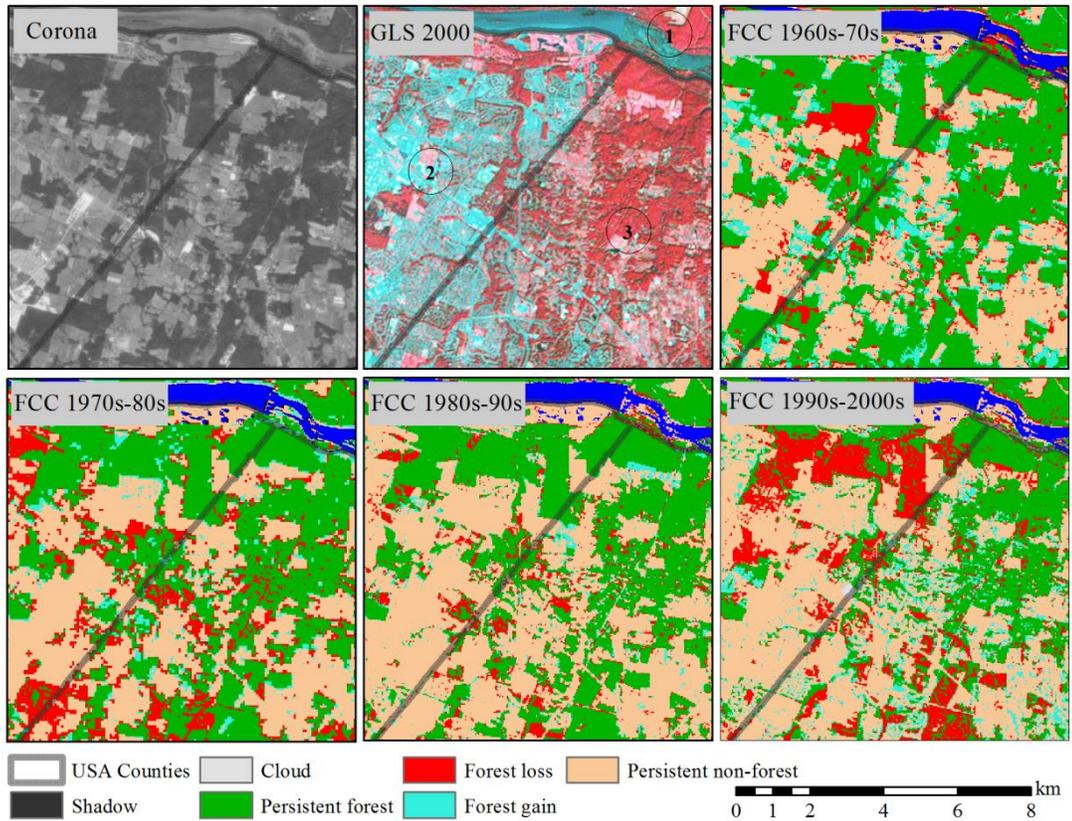


Figure 3.4 Georeferenced Corona, GLS 2000 images with band 4, 3, 2 in color R, G, B, and sequential forest cover change maps of each two epochs in the VM study area. (1) is Montgomery County in Maryland, (2) and (3) is Loudoun and Fairfax County in Virginia separately.

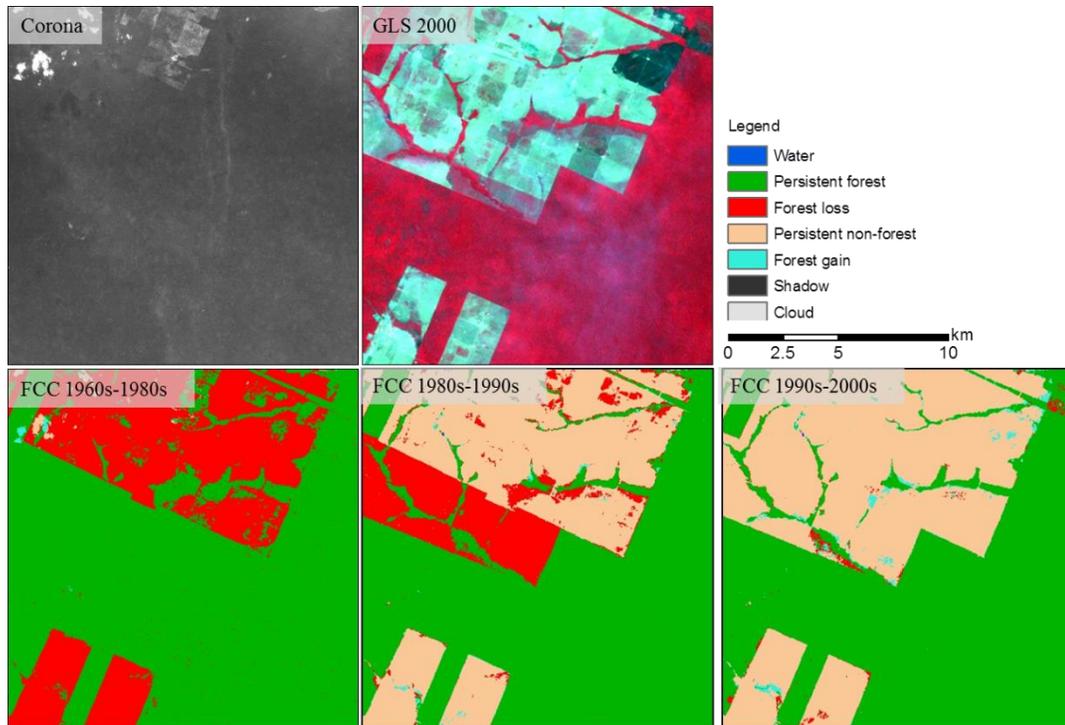


Figure 3.5 Georeferenced Corona, GLS 2000 images with band 4, 3, 2 in color R, G, B, and sequential forest cover change maps of each two epochs in MTP study area.

3.5 *Discussion*

In this study, Corona images were used with Landsat images to monitor forest cover and its changes in a suburban study area and a tropical forest area. Co-registration against terrain-corrected Landsat images yielded high georegistration accuracy of Corona images using manually selected GCPs and polynomial transformation functions. The co-registered Corona images were deemed to have adequate geolocation accuracy for comparison against Landsat images for mapping forest change without introducing excessive spurious changes (Townshend et al. 1992). However, more rapid or automated georeferencing methods, such as the Automated Registration and Orthorectification Package (AROP) (Gao et al. 2009), will be required to deal with large volumes of images necessary for studying larger regions. Any such method must be capable of overcoming spectral and resolution

differences between Corona and Landsat, as well as real differences in land cover accumulated over long times.

The classification method I developed here has improved in processing efficiency compared to visual interpretation and/or manual delineation of land cover used in previous studies (Bindschadler and Vornberger 1998; Challis et al. 2002). The use of texture measures increased the accuracy of classification compared to only using image grey-scale values alone. The combination of co-occurrence textures representing different spatial patterns with grey-scale value is recommended for studies in the future. While, different forest classes can hardly be separated so far using only texture information. Additionally, although the best window size for extraction of textural measurements is usually affected by image spatial resolution, vegetation structure, and forest patch sizes in landscape (Lu et al. 2008), small window sizes (e.g., 9×9 pixels) are recommended for both complex and homogeneous landscapes. Beyond the texture measurements I used for separating forest and non-forest class, automatic tree crown identification algorithm proposed by Palace et al. (2008) could potentially be applied to Corona image, given its very high spatial resolution and panoramic, to estimate tree crown width and percent tree cover in 1960s. Given the availability of adequate, representative training data, effective classification features and improvements in automated image registration, forest classification and change detection using Corona image can be automated for large areas around the globe.

3.6 Conclusions

The Landsat satellites have generated a primary source of information for studying changes in Earth's land surface that now spans four decades since the 1970s. Using data acquired from Corona satellites, which have spatial resolutions better than Landsat data and are available for most of the land areas of the globe, this record can be extended a decade further, to the 1960s. I demonstrated the feasibility of paring Corona with Landsat images for mapping forest cover changes between the 1960s and 2000s through case studies conducted in two areas, Virginia-Maryland (VM) in the US and Mato Grosso-Tocantins-Pará (MTP) in Brazil. For each area, the Corona image was accurately co-registered to an orthorectified Landsat image. Combination of metrics to represent multiple aspects of spatial texture achieved classification accuracies of ~95%. Forest changes during the Landsat era were mapped with overall accuracies around 90%. Different forest-cover change rates and trends were observed between the two study areas, with 18.9% and 16.8% net forest loss in the VM and MTP study area respectively from 1960s to 2000s. Results based on Corona data indicate that forest area in a suburban study area of Virginia and Maryland increased from the 1960s to the 1970s, before being converted to residential area. In contrast, severe forest loss in the Amazon region started from scattered small forest land cleared during 1960s and 1970s and eventually expanded thereafter.

These results demonstrate the potential of Corona data in land cover and change studies. However, using Corona imagery for land cover change studies over larger areas will require more automated georegistration and classification methods. Georeferencing Corona images must be automated to avoid overwhelming human

involvement, especially in the identification of ground control points. Automatic training data selection methods, similar to others used for Landsat-based classification, also need to be further developed for regional to global retrievals.

Chapter 4: Towards automated registration of historical satellite images for land cover land use change applications²

4.1 Introduction

4.1.1 Background

Concluded from Chapter 3, the biggest challenge remains in the accurate automated georegistration of the Corona image, mainly due to the following aspects. First, accurate satellite orbit information is currently not available; besides the satellite gesture, including satellite's location and height, and sensor's roll, pitch and yaw, was not as well controlled in the 1960s as for recent satellites. As a result, it is not possible to get an accurate estimate of the geolocation of each Corona scene. The prior knowledge on geolocation information of Corona data provided by the United States Geological Survey (USGS) is usually with errors ranging from several to tens of kilometers. Second, each Corona scene was originally recorded on a film and then digitalized into four overlapping subset images. Each pair of adjacent subsets has a slight offset in their image locations along-track and also differs in spatial resolutions due to the panoramic projection of the Corona data. In addition, there is no radiometric calibration parameter available, and the signal-to-noise ratio (SNR) is predictably low. The poor radiometric quality of the Corona image makes it difficult to identify image feature based on spectral characteristics. Third, cumulative land cover changes caused by urbanization, agricultural expansion, forest disturbances, or

² This chapter has been prepared for submission to the journal of *Remote Sensing*.

climate change since the 1960s bring a lot of uncertainties in matching Corona image with other satellite images obtained at least a decade later. Traditional image matching approaches relying on identification of consistent land features are usually unsuccessful for registering the Corona image against other satellite images. Last, there is significant geometric distortion due to the panoramic design of the camera system and low-orbit feature of satellite in the Corona program (e.g., 150- and 180-km for KH-4A and KH-4B satellites separately) (Galiatsatos 2009a). Low orbit satellite missions and large field-of-view (FOV) resulted in large view angles for the side looking part of image, which led to significant surface anisotropic effects (also different effects over different land cover types). All these factors make it difficult to apply the existing approaches for the automated georegistration of the Corona images.

To partially solve the above-mentioned issues, the existing method for georegistering high/medium resolution image are mainly based on two elements, 1) ground control points or tie points collected from either field measurements or reference image, and 2) the transformation function, such as polynomial function or mathematical collinearity function (Richards and Jia 2006). Due to the large amount of time consumed by collecting ground points and the inaccessibility of the remote region, control points are usually collected from reference data such as the georeferenced remote sensing image or topographic map, which is usually known as image matching procedure. Correlation, describing the greyscale similarity of two spatial windows in target and reference images, has been widely used to search for the tie points (Gao et al. 2009). This method is more suitable for images acquired from the same platform or multi-sensor images with similar characteristics such as

viewing angle, band setting, and so on. Preliminary registration of the target image is also required in order to limit the searching distance as well as to avoid generating pseudo points. In addition, significant LCLUC may also prevent the application of methods that are based on grey scale matching for the registration of historical images.

To overcome these limitations as well as to reduce the human effort involved in the georegistration, I propose a semi-automated approach to correct the Corona images. First, each Corona scene that consists of four overlapping subset images is mosaicked into one whole image. Second, each mosaic Corona image is resized to 30 m based on the nominal coverage provided by USGS, and then manually registered to a Landsat reference image with four control points. The purpose of this step is to constrain the initial registration error up to ~4.5 km. The third step is the precision registration, which relies on tie points selected manually or from the state-of-art feature extraction and matching algorithms to achieve a high level of registration accuracy by using a transformation function (e.g., polynomial, collinear functions) designed for Corona.

4.1.2 Objectives

The objectives of this chapter include: 1) evaluating the use of collinear function and Landsat reference data to register Corona image and 2) exploring the possibility of semi-automated registration of Corona data based on a feature matching algorithm. The resultant registered Corona data are expected to satisfy land cover change mapping between Corona and Landsat data at a spatial scale not coarser than 250 m. This chapter presents the first attempt towards automated georegistration of the Corona images with Landsat data as references. Targeting at the georegistration of

the Corona image for the large area change detection application, this chapter proposed methods for preliminary processing, manual/automated tie point selection and matching, precision georegistration based on a transformation function (including the polynomial function and collinear function), and error estimation, respectively in the following sections.

4.2 *Data*

4.2.1 Corona data

Four Corona images, located in North America, South America, and Asia, were selected to demonstrate the capability of the registration method across different landscapes and regions with different types of land cover and change (Table 4.1). Images in Maryland, USA (VM, hereafter), and Mato Grosso do Sul, Brazil (MGdS, hereafter), are characterized by flat terrain and significant land cover changes caused by urban expansion and tropical forest cutting, respectively. The image in St. Louis, USA (SL, hereafter), has very rich features comprised by regular-shaped crop field, while changes were also observed in crop fields and in rapidly developed suburb areas since the 1960s. The most complicated terrain among the four study regions happened in Sichuan, China (SC, hereafter), where elevation varied from 1800 m to 5500 m and was covered by temperate and sub-tropical forests.

4.2.2 Landsat and elevation data

Global Land Surveys (GLS) data (https://landsat.usgs.gov/science_GLS.php) was used as references to register Corona images, considering the potential for generating a change product by combining the two datasets. Five epochs of the GLS

dataset, 1975, 1990, 2000, 2005, and 2010, are featured with high geodetic accuracy against each other, and quality of GLS2000 and afterwards is better than the previous two collections (Gutman et al. 2013). This dataset has been widely used to generate the finest national or global land cover and change products (Hansen et al. 2013; Kim et al. 2014; Sexton et al. 2013a). GLS2000 data was chosen as a result of its global data availability and mitigation of land cover change impact (Gutman et al. 2013). More important, the RMSE of GLS2000 is less than one Landsat pixel, *i.e.*, 30 m for Landsat Thematic Mapper (TM) (Gutman et al. 2013).

Table 4.1 Basic information of the three Corona images used in this study.

	Manual method		Automatic method	
Image ID	DS1101- 2157DF026	DS1021- 2093DF063	DS1035- 1046DF018	DS1105- 2248DF067
Mission No.	KH-4B	KH-4A	KH-4A	KH-4B
Acquisition date	9/25/1967	4/24/1965	9/23/1966	11/19/1968
Covered region	Virginia/Maryland, USA	Mato Grosso do Sul, Brazil	St. Louis, USA	Sichuan, China
Abbreviation	VM	MGdS	SL	SC
Spatial resolution	6 ft. (~1.83 m)	9 ft. (~2.74 m)	9 ft. (~2.74m)	6 ft. (~1.83 m)
Camera	Forward looking	Forward looking	Forward looking	Forward looking

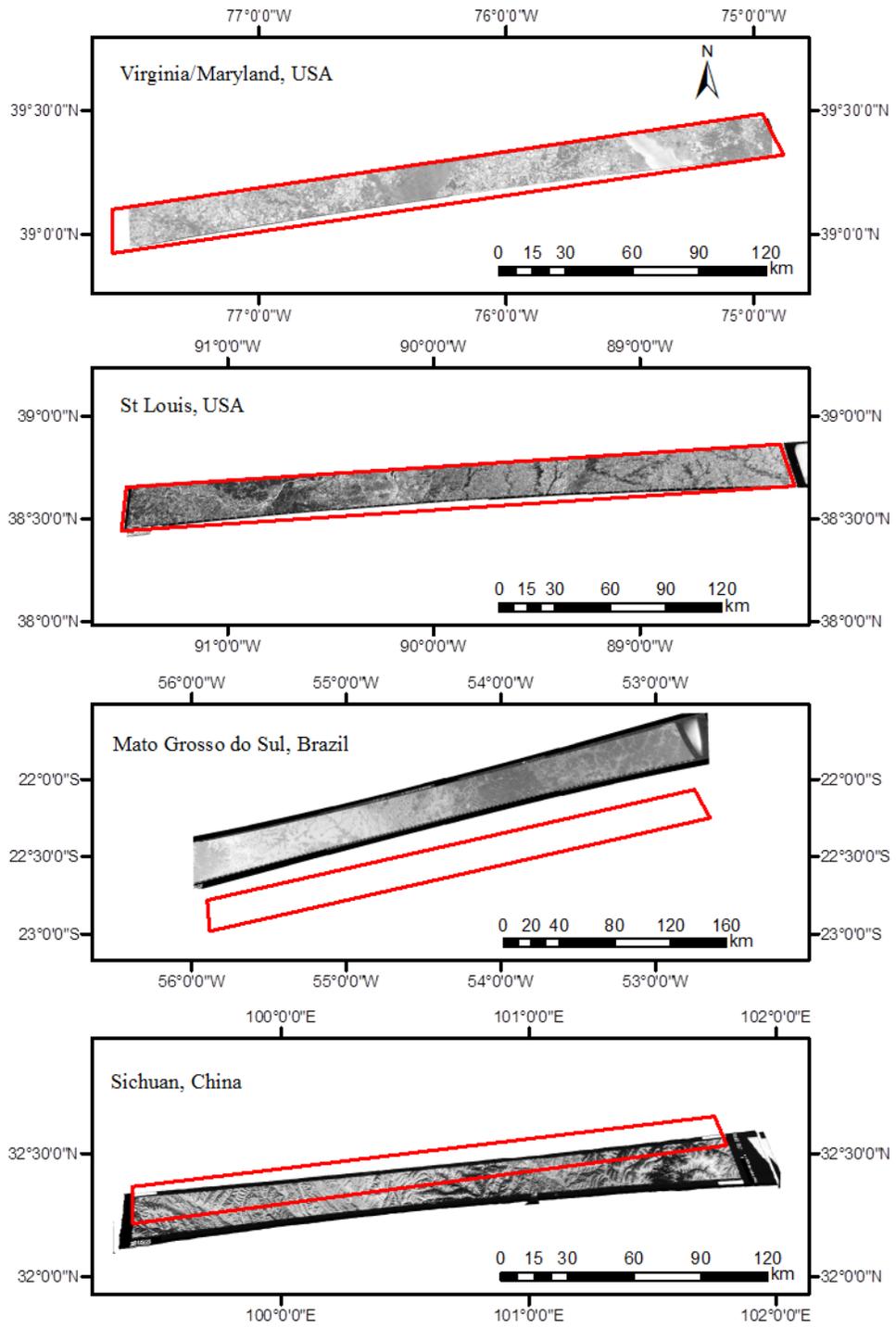


Figure 4.1 The actual Corona coverage (b/w image) overlaid by the estimated footprint provided by USGS (red frame, downloaded from <http://earthexplorer.usgs.gov/>).

As the Corona greyscale image covers only the visible spectrum, only the first three bands of Landsat TM images are potentially suitable for image matching. To find the best band for feature matching, pure samples of water, forest, crop, and urban areas were collected from the collocated Corona and Landsat images and compared their sensitivities to the surface reflectivity. Experiments show that Landsat visible bands and its mean have higher correlations with Corona than SWI and pan bands. As a result, band 3 (red band) of the Landsat TM image is selected in searching for tie points against Corona image. Elevation information is derived from the 1-arc second data (~30 m) from SRTM data (USGS 2004).

4.3 *Methodology*

4.3.1 Data preparation (1): Automated mosaicking of Corona data

Each Corona image was scanned by USGS using a Leica scanner that is commonly used for digitizing a 241×241 mm aerial film size. As a result, the approximately 750mm length of each declassified Corona film was divided into four subsets. There is an overlap among each adjacent pair of subsets, the size of which varies depending on the magnitude of tilting at the head of the image or along the edge (Personal communication with USGS personnel). In addition, there is a difference in geolocation for the overlapping area between adjacent subsets. To remove such overlaps and to reconstruct the complete Corona scene, an automated, feature-based panoramic image stitching algorithm was applied (Brown and Lowe 2007). After this step, a complete Corona scene is generated as shown in Figure 4.2, however, the Corona image still lacks of geolocation information.

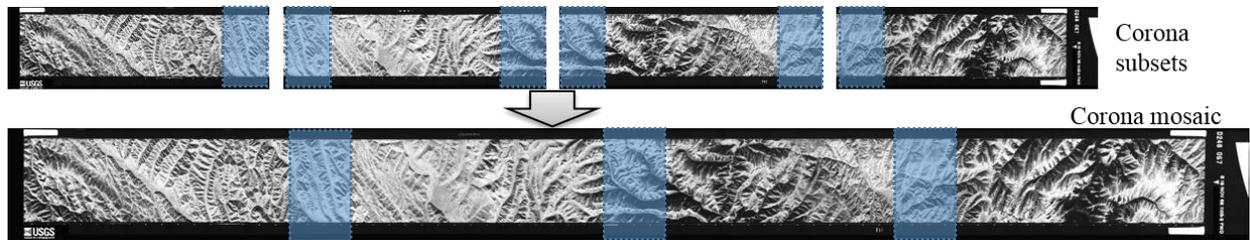


Figure 4.2 Image mosaicking for scene DS1105-2248DF067. Four subset images are shown in the upper panel and the final mosaic image is shown in the lower panel. Approximate overlapping areas are shaded in blue in the subsets and mosaic image.

4.3.2 Data preparation (2): Preliminary registration of Corona data

The purpose of this step is to assign the Corona data a rough geolocation with maximum error of 150 Landsat pixels at 30 m resolution (~4500 m), which will allow an accurate image matching against referenced Landsat data. Although being lack of accurate geolocation information, the initial coordinates provided by USGS can help locate the Corona image and roughly link it to Landsat images. The image tilting angle (t) can be calculated from the nominal coordinates of four corners, which can then be used to rotate the Corona mosaic and to reproject it to a Universal Transverse Mercator (UTM) projection. I did not assign the coordinates of four corners directly to the Corona image in order to minimize the image signal loss due to multiple transformation steps. The rotated Corona image was then resized to 30m, the scaling factor (SF), of which was decided by the best resolution of its mission (e.g., SF=30m/1.8m used for the image of KH-4B and SF=30m/2.74m for KH-4A). This rotation and rescaling processing will not be used directly in the following registration, but serves as a reference for us to identify the preliminary tie points between Corona and Landsat data. As a result, the reprojected Corona data often intersects with 2 to 4 Landsat scenes due to its rectangular shape and wide coverage.

To satisfy the needs for preliminary registration, the overlapped Landsat scenes were mosaicked into one image.

Preliminary registration of the Corona image to GLS2000 data was carried out based on four manually selected tie points at the corners of the Corona image, in order to constrain the initial geolocation accuracy to <150 Landsat pixels (~4500 m) in both along-track and cross-track directions. Then, the intersected area to the preliminarily registered Corona data was extracted from the reference GLS2000 mosaic. Considering that the preliminarily registered Corona data still carries a geolocation error of up to 150 Landsat pixels, to maintain a reasonable image-matching accuracy only a subset of the corresponding Landsat mosaic image was extracted with a 150-pixel (4500 m) buffer zone to the range of preliminarily registered Corona data. The preliminarily registered Corona data and the Landsat subset with a buffer zone were used together as input data for tie point selection and georegistration. Examples of the Corona data after preliminary registration and the referenced Landsat image are shown in Figure 4.3. To maintain a reasonable accuracy in both tie point selection and precision georegistration in the following section, every Corona scene was equally divided into three subsets with the same dimensions and the preliminary registration process was carried out for each subset individually.

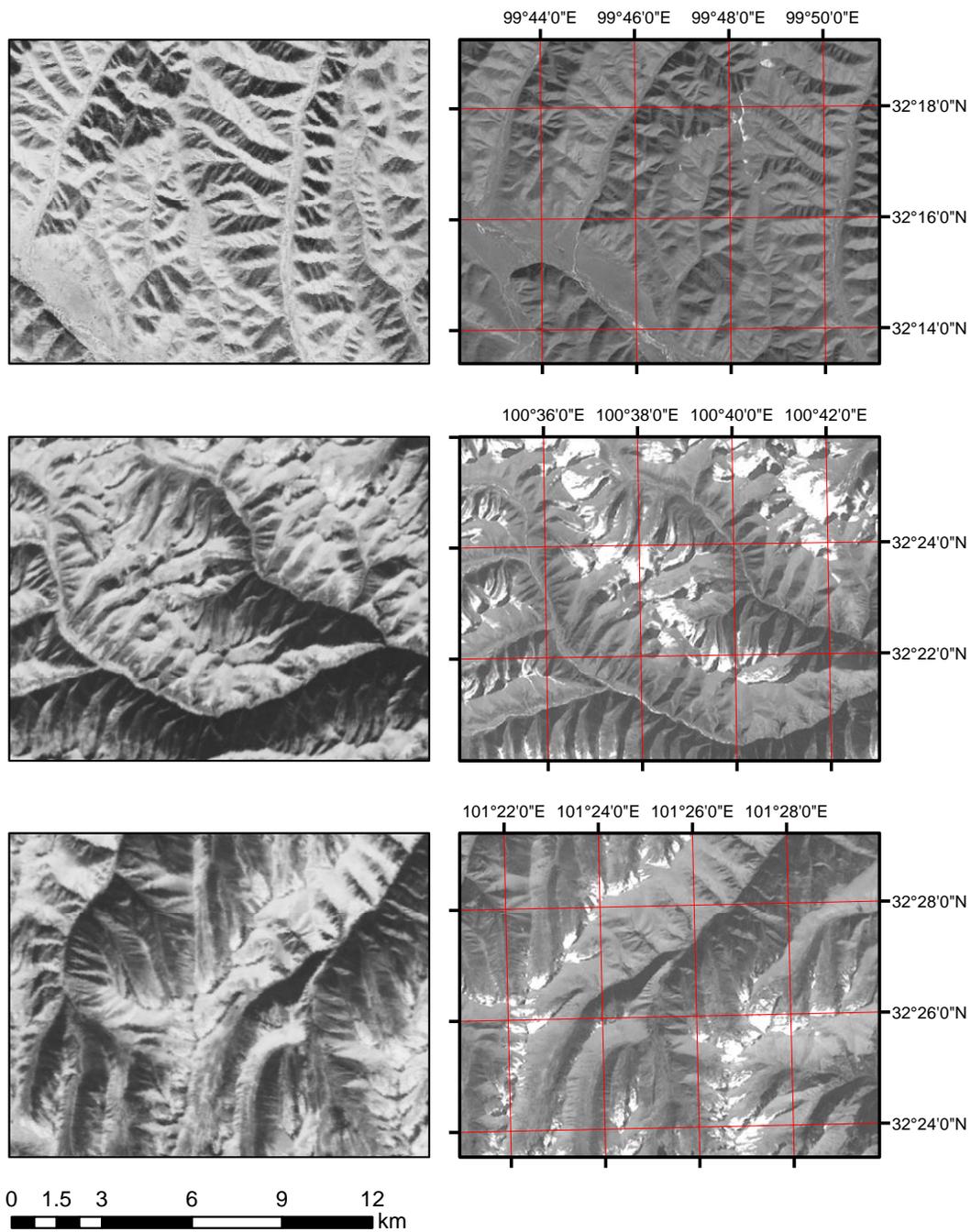


Figure 4.3 Preliminary registered Corona (left) and reference Landsat (right) images in Sichuan. Each pair of windows (windows in the same row) represents the same range according to the geoinformation of Corona and Landsat images.

4.3.3 Transformation model

Polynomial function has been widely used in correcting image distortion considering its simplicity in application. However, lower degree polynomial function (e.g., 2nd order) is not efficient to describe the complex distortion embedded in an entire Corona scene, especially over rugged terrains. The solution for higher degree polynomial function (e.g., 3rd to 5th order) may sometimes be an alternative but is usually unstable, although it can result in a small fitting error (Galiatsatos et al. 2008) if the polynomial coefficients can be well constrained.

The 2nd-order polynomial function was chosen as one of the two transformation models in this study, which is described as:

$$x = a_0 + a_1X + a_2Y + a_3XY + a_4X^2 + a_5Y^2 \quad \text{Eq. 4.1}$$

$$y = b_0 + b_1X + b_2Y + b_3XY + b_4X^2 + b_5Y^2 \quad \text{Eq. 4.2}$$

Where (x, y) are the image coordinates of target image, (X, Y) are the ground coordinates derived from the coordinates of control points on reference image, a_i and b_i are the regression coefficients.

Usually, the collinearity equations are used for modeling the relationship among the center of the camera, the coordinates on the perspective plane (image), and the ground coordinates with the use of a digital elevation model.

$$x = -f \frac{a_1(X - X_0) + a_2(Y - Y_0) + a_3(Z - Z_0)}{a_7(X - X_0) + a_8(Y - Y_0) + a_9(Z - Z_0)} \quad \text{Eq. 4.3}$$

$$y = -f \frac{a_4(X - X_0) + a_5(Y - Y_0) + a_6(Z - Z_0)}{a_7(X - X_0) + a_8(Y - Y_0) + a_9(Z - Z_0)} \quad \text{Eq. 4.4}$$

Where (x, y) are the image coordinates, (X, Y, Z) are the ground coordinates, (X_0, Y_0, Z_0) are the ground coordinates of the camera, f is the focal length of Corona camera, and a_1 to a_9 are coefficients given by the following rotation matrix:

$$\begin{pmatrix} a_1 & a_2 & a_3 \\ a_4 & a_5 & a_6 \\ a_7 & a_8 & a_9 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos \omega & \sin \omega \\ 0 & -\sin \omega & \cos \omega \end{pmatrix} \begin{pmatrix} \cos \varphi & 0 & -\sin \varphi \\ 0 & 1 & 0 \\ \sin \varphi & 0 & \cos \varphi \end{pmatrix} \begin{pmatrix} \cos \kappa & -\sin \kappa & 0 \\ \sin \kappa & \cos \kappa & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \text{Eq. 4.5}$$

In Eq. 4.5, ω , φ , κ are the rotation angles of the camera. However, there is no native collinearity equation for the panoramic camera on board of Corona satellites. To account for the distortions caused by the panoramic projection, satellite flight velocity, film scanning time, and non-nadir pointing camera, the above equations (Eqs. 4.3-4.5) need to be revised to model the Corona data by adding several correction terms, which is used to convert the Corona image coordinates from the panoramic image to the perspective plane. A set of modified collinear equations were proposed by Sohn et al. (2004):

To convert the image coordinates in the panoramic film (x_p, y_p) to the corresponding coordinates in the perspective plane (x_f, y_f) for a particular pixel, the pixel's scan angle α and camera's focal length f are used here:

$$x_p = f \alpha \quad \text{Eq. 4.6}$$

$$x_f = f \tan \alpha \quad \text{Eq. 4.7}$$

$$y_f = y_p \sec \alpha \quad \text{Eq. 4.8}$$

The image coordinates in the perspective plane (x_f, y_f) are usually used in Eqs. 4.3 and 4.4 for frame cameras. For Corona data, two types of image distortion need to be corrected following the equations proposed by Sohn et al. (2004).

One type of distortion is the scan positional distortion caused by the moving of the camera's position during the scan, which is corrected using the following equations:

$$y_s = \frac{f}{H} V t \cos \alpha \quad \text{Eq. 4.9}$$

$$t = \frac{\alpha}{\delta} \quad \text{Eq. 4.10}$$

$$y_s = \frac{V f \alpha \cos \alpha}{H \delta} \quad \text{Eq. 4.11}$$

where y_s is the scan positional distortion component for a given scan angle α , V is the velocity of the platform, t is the scan time of the camera, δ is the angular velocity of the camera scan arm, and H is the flight height.

The other distortion is caused by the image movement during the exposure time and can be compensated with the following equations:

$$v = \frac{fV \cos \alpha}{H} \quad \text{Eq. 4.12}$$

$$y_{im} = -\frac{fV}{H \delta} \sin \alpha \quad \text{Eq. 4.13}$$

where v is the velocity of the image and y_{im} is the corresponding distortion component.

Taking into account of these two types of distortion, the corrected image coordinates in the perspective plane (x'_f, y'_f) can now be expressed as:

$$x'_f = x_f = f \tan \alpha \quad \text{Eq. 4.14}$$

$$y'_f = y_f + (y_s + y_{im}) \sec \alpha = \frac{y_p}{\cos \alpha} + \frac{Vf}{H \delta} (\alpha - \tan \alpha) \quad \text{Eq. 4.15}$$

Detailed descriptions of the derivation for the distortion correction can be found in Sohn et al. (2004).

The collinear function, by combining Eqs. 4.3, 4.4, 4.5, 4.14, and 4.15, was chosen as another transformation model to register the Corona data with Landsat in comparison against polynomial function, with the help of a digital elevation model.

Cost function is formulated as the least-squares of differences between measured image coordinates and simulated values from transformation function (either the 2nd-order polynomial function or collinear function):

$$F = \sqrt{\frac{\sum_{i=1}^n \left[(x_i - \bar{x}_i)^2 + (y_i - \bar{y}_i)^2 \right]}{n}} \quad \text{Eq. 4.16}$$

Where (x_i, y_i) and (\bar{x}_i, \bar{y}_i) correspond to the measured and estimated image coordinates of the i th $(i \in [1, n])$ selected tie points on the Corona image plane.

There are seven unknown parameters to solve the modified collinear function, including three for exterior orientation, three for interior orientation, and one for satellite velocity (the ratio between V and δ in Eq. 4.15). Initial values of exterior position elements were determined based on the center coordinates of the preliminary

registered Corona data and the nominal flight height. Gesture elements were initialized to be zero. Initial values of interior elements were assigned based on approximate image center identified by visual interpretation. It is difficult to accurately estimate the initial values of the unknown variables since the orbit control was poor during KH missions and the orbit varied among different missions. To mitigate the risk of finding only a local minimum of the cost function, the modified collinear function was built for each of the three subsets, respectively. In other words, for each Corona subsets one set of unknown parameters of the cost function is assumed and retrieved, which is to reduce the complexity of collinear equations for Corona (Sohn et al. 2004).

4.3.4 Tie points extraction

Two types of tie point extraction of Corona and Landsat images were tested in this study for the georegistration of Corona data. Manual tie point selection was chosen to examine the accuracy of the proposed method based on collinear equation transformation function with Landsat data as references. Automatic tie point (feature) extraction and refinement method was investigated in order to explore the applicability of large scale automated registration between Corona and Landsat data.

Manual tie points selection

Two Corona images, located in Maryland, USA and Mato Grosso do Sul, Brazil, were used to test the georegistration accuracy for the two transformation models with manually selected tie points between Corona and Landsat data. Elevation values for tie points were extracted from SRTM. Tie points were designed to evenly locate across whole Corona scenes.

The unknown variables in both the collinear function and polynomial function of the Corona camera were retrieved using the collected tie points, after which the registration of Corona image can be done using the forward calculation. Two experiments were carried out to assess the retrieving accuracy of the exterior elements. First, the whole scene was treated as one model (*i.e.*, method 1). All selected tie points were included in the retrieval process. Second, the whole scene was divided into three parts with the same dimension (*i.e.*, method 2), each of which has its own model and the exterior elements for each part were retrieved separately. The reason of doing this is that it takes 0.5 second for the Corona camera to scan the whole area. During the scan period, the satellite can actually fly a relatively long distance so that it cannot be assumed the camera position remains the same and, hence, the flight speed was unknown. The result will demonstrate the accuracy of revised collinear function in the registration of Corona image using reliable tie points collected from reference Landsat and elevation value derived from SRTM data. Through the comparison, it is also possible to find out which method can achieve higher accuracies and whether the image partition can improve the results.

Automated tie point extraction and refinement

Having the transformation models tested with tie points selected manually, the question remains whether tie points can be generated automatically and whether an automated tie point selection allows the accuracy of Corona registration at a large spatial domain.

Due to the geometric and radiometric characteristics of Corona camera, our experiments on testing the traditional image matching approach, such as Automated

Registration Orthorectification Package (AROP) (Gao et al. 2009), failed to find enough tie points between the Corona and reference image. Thus, a more sophisticated approach is needed, which would suffer less from the image distortion (scaling, offsetting, transforming, topographic effects, etc.) and difference in spectral response functions between Corona data and reference image. The Scale Invariant Feature Transform (SIFT) (Lowe 1999, 2004), an efficient feature selection algorithm characterized with invariant feature to spatial scale and image rotation, is used here to minimize the above-mentioned effects of image distortion in image matching. The SIFT feature is described by the SIFT descriptors consisting of gradient information in the detected regions. The selected feature points are partially invariant to change in illumination, camera projection, occlusion, clutter, or noise (Lowe 1999, 2004) enabled by a four-step process in the tie point (feature) selection, which consists of scale-space extrema detection, feature localization, orientation assignment, and feature descriptor calculation (Lowe 2004).

The SIFT descriptor has been demonstrated to be the most robust local invariant feature descriptor for images with different geometrical properties (Mikolajczyk and Schmid 2004), which allows the feature to be matched between multiple images acquired by different satellites or at different time periods assuming there are no surface changes. The SIFT algorithm has been applied to the registration of both high and medium-resolution images such as QuickBird and HuanJing (HJ) images, resulting in georegistration errors of less than one satellite pixel (Wang et al. 2012).

To mitigate the errors in the preliminary registration and improve tie point (feature) matching accuracy, the entire Corona scene was evenly partitioned along the long side into three subsets and the SIFT algorithm was applied to each Corona subset and buffered Landsat subset separately to generate their own feature point set. The SIFT matching process was then carried out over the feature point sets of target Corona and reference Landsat subsets to search for tie point pairs which satisfied the predefined requirement on similarity measurement. However, the collected SIFT matches may sometimes contain pseudo matches or do not follow an even distribution. Two key steps were carried out to mitigate the impacts. First, the pair of SIFT matches with a geolocation difference of more than 150 30m pixels are considered as pseudo matches and thus excluded. Second, a 2nd-order polynomial transformation function is estimated between the image coordinates of SIFT points in the Corona image and Landsat image. A fitting residual is calculated for each of the SIFT matches. The matched SIFT tie points were considered as “pseudo” matches if their residuals calculated from polynomial functions exceed a certain threshold. Based on our experiments over multiple scenes, the threshold was set to six pixels at 30m resolution. The searching will be iterated only until the defined threshold was achieved. The coordinates of refined SIFT features on the Corona and Landsat images were recorded and used to solve the transformation models and estimate error in the following step. The whole process is illustrated in Figure 4.4.

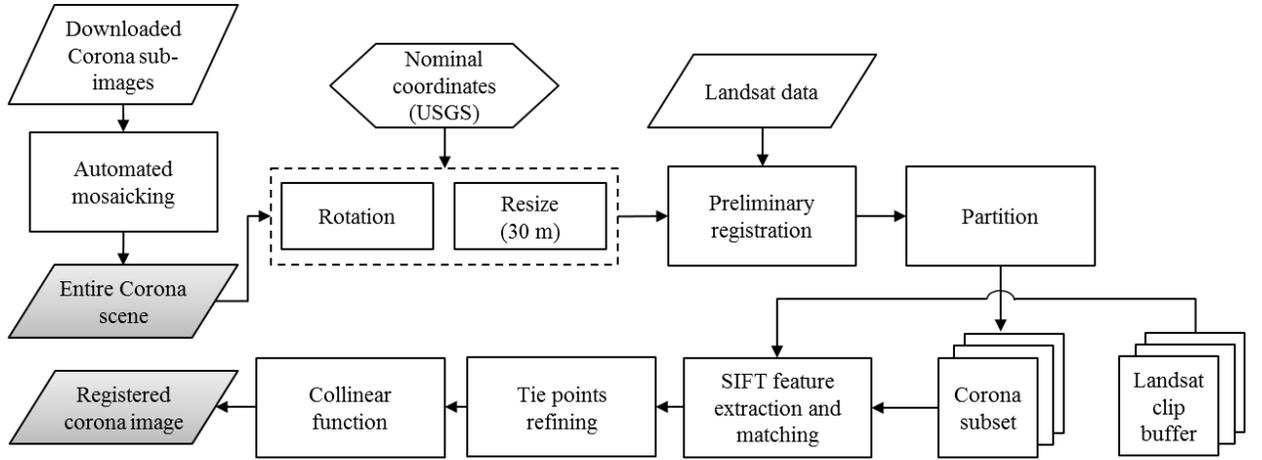


Figure 4.4 Method of the semi-automated registration of Corona image with Landsat image as reference. SIFT: the Scale Invariant Feature Transform. USGS: the United States Geological Survey.

4.3.5 Error estimation

Two types of error estimation were carried out: one is to evaluate the fitting accuracy of transformation model with matched SIFT tie points and the other is the independent accuracy assessment with manually selected tie points from the corrected Corona and Landsat data. RMSE along two dimensions ($RMSE_x$ and $RMSE_y$) and total $RMSE_{xy}$ are calculated using the following equations:

$$RMSE_x = \sqrt{\frac{\sum_{i=1}^n (X_i^C - X_i^L)^2}{n}} \quad \text{Eq. 4.17}$$

$$RMSE_y = \sqrt{\frac{\sum_{i=1}^n (Y_i^C - Y_i^L)^2}{n}} \quad \text{Eq. 4.18}$$

$$RMSE_{xy} = \sqrt{RMSE_x^2 + RMSE_y^2} \quad \text{Eq. 4.19}$$

Where (X_i^C, Y_i^C) and (X_i^L, Y_i^L) represent the geographic coordinates of the i^{th} ($i = 1, \dots, n$) selected feature center or tie point from Corona and Landsat image, respectively.

For the accuracy assessment using manually selected tie points, the same set of tie points were selected for preliminary registered Corona data, Corona data registered with 2nd-order polynomial function, and Corona data registered with collinear function.

4.4 *Results*

4.4.1 Manual point selection and collinear function accuracy

There are 87 and 56 tie points selected for MGdS scene and VM scene, respectively, which are located across the whole Corona scenes on purpose.

For method 1, which treated the whole Corona scene as one model, results showed large RMSEs in both x (247.7 m) and y (139.0 m) directions for the MGdS case (Table 4.2). The VM case showed better results with lower RMSEs in x (177.9 m) and y (57.8 m), which however are still too large to meet the accuracy requirement for land cover change detection at a medium to coarse resolution particularly along the x direction. This is probably due to the fact that the change in image center (satellite position) during the scanning period was not considered. Since the x dimension approximates the camera scan direction, larger uncertainties happened in the x direction (RMSE_x) due to failing to consider the change in scan center.

In method 2, the whole scene was divided into three parts with the same size, each of which has its own collinear model and the exterior elements for each part were retrieved separately. In this experiment, the accuracies of georegistrations improved much, where the total RMSEs dropped to 40~53 m in both x and y directions (Table 4.2). This demonstrated the high accuracy of using manually

selected tie points from reference Landsat and SRTM data in the correction of Corona image. It also indicates that multiple sets of parameters of collinear equations are needed for different Corona subsets of a complete scene and accurate geolocation could be achieved individually. In other words, dividing the Corona image into three subsets can help reduce the errors resulted from the change in satellite position during the scanning period. It is noteworthy that the accuracy for the central subset (RMSE less than 20 m) is better than the subsets on two sides (RMSE less than 70 m) likely because the image distortion is larger at the edge.

Table 4.2 RMSE for collinear transformation test using manually selected tie points

	Method 1	Method 2			
MGdS	Total	Total	West	Center	East
RMSE_X (m)	247.7	39.9	48.7	18.9	41.4
RMSE_Y (m)	139.0	44.3	58.5	16.5	42.0
Number of tie points	87	87	31	23	33
VM	Total	Total	West	Center	East
RMSE_X (m)	177.9	52.9	47.9	11.3	67.5
RMSE_Y (m)	57.8	43.9	40.2	14.6	55.3
Number of tie points	56	56	26	9	21

4.4.2 Automated tie point (feature) extraction and refinement

Another two Corona scenes were chosen here to evaluate the effectiveness of SIFT feature matching based automated procedure of georegistration for the Corona data with Landsat data as a reference. One scene is located at a mountainous region in Sichuan, China and the other is located at a suburban-agriculture region in St. Louis, USA.

After preprocessing, the geolocation difference between the preliminarily registered Corona and reference Landsat images was constrained up to 4500 meters

(*i.e.*, ~150 Landsat 30m pixels). The SIFT algorithm found features from each individual image and successfully matched pairs of feature points with the largest similarity. For the three subsets of the mountainous scene in Sichuan (DS1105-2248DF067), SIFT algorithm exported 677, 625, and 2000 “qualified” matching points individually from the west-to-east subset. The refinement step reduced the matching points to 183, 279, and 928 separately which are still adequate to solve the collinear function.

Examples in Figure 4.5 demonstrated the distribution of refined SIFT matches in a zoomed-in view. Different from the feature points usually selected by experts through visual interpretation, SIFT did not tend to pick features such as river turning points, mountain ridge or valley; instead, the algorithm selected features at multiple scales and finalized the location at the center of a multi-scale feature. Although the feature is not as “obvious” as manually selected features, the accuracy is usually high, as it reports the location of feature at sub-pixel scale, and the number of feature points is ensured.

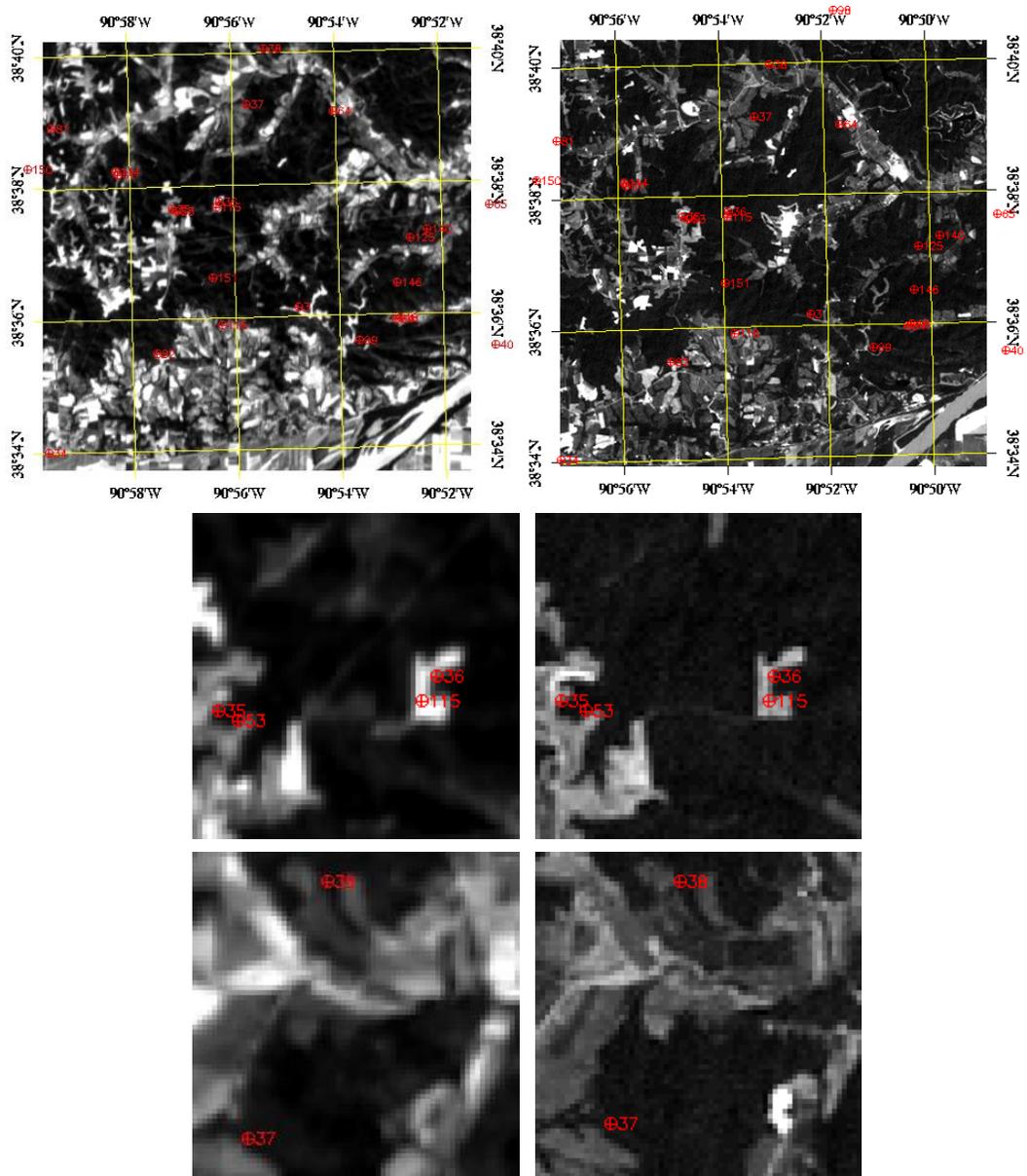


Figure 4.5 Refined matching points selected by SIFT shown on Corona (left) and referenced Landsat (right) image for Sichuan.

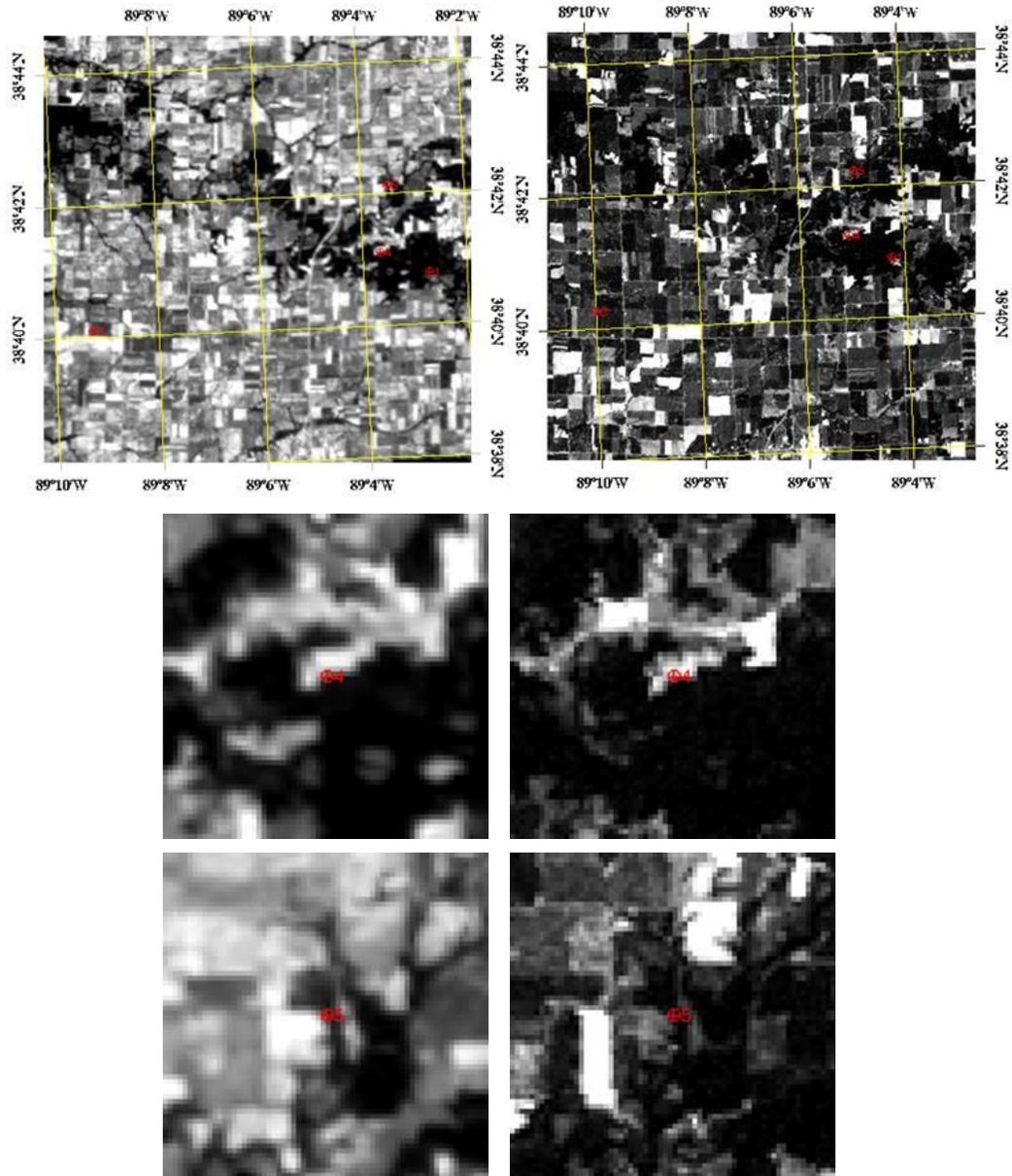


Figure 4.6 Refined matching points selected by SIFT shown on Corona (left) and referenced Landsat (right) image for St. Louis.

For the three subsets of the scene in St. Louis (DS1035-1046DF018), SIFT algorithm exported 198, 53, 39 matched features from west to east, which dropped to 151, 13, and 13, respectively, after the refinement (Figure 4.6). This significant reduction in matching points was due to the removal of a large number of “pseudo” matches resulting from similar image patterns and land cover changes, particularly at

the east part of the scene where was covered by agriculture lands. The refined number of SIFT matches remain adequate to solve the collinear function.

Table 4.3 Fitting accuracy of polynomial and collinear function separately based on refined SIFT matching points.

Sichuan		2 nd order polynomial			Collinear function		
Subset	Number of points	RMSE_X (m/pixel)	RMSE_Y (m/pixel)	RMSE_XY (m/pixel)	RMSE_X (m/pixel)	RMSE_Y (m/pixel)	RMSE_XY (m/pixel)
1	183	34.2 (1.14)	53.7 (1.79)	63.7 (2.12)	26.64 (0.89)	33.85 (1.13)	43.08 (1.44)
2	279	45.6 (1.52)	63.9 (2.13)	78.5 (2.62)	30.81 (1.03)	36.04 (1.20)	47.41 (1.58)
3	928	49.5 (1.65)	75.3 (2.51)	90.1 (3.00)	41.76 (1.39)	38.18 (1.27)	56.58 (1.89)
St. Louis		2 nd order polynomial			Collinear function		
Subset	Number of points	RMSE_X (m/pixel)	RMSE_Y (m/pixel)	RMSE_XY (m/pixel)	RMSE_X (m/pixel)	RMSE_Y (m/pixel)	RMSE_XY (m/pixel)
1	151	50.9 (1.70)	38.13 (1.27)	63.59 (2.12)	46.0 (1.53)	54.5 (1.82)	71.3 (2.38)
2	13	15.1 (0.50)	25.6 (0.85)	29.7 (0.99)	43.3 (1.44)	61.0 (2.03)	74.8 (2.49)
3	13	20.2 (0.67)	15.3 (0.51)	25.3 (0.84)	41.4 (1.38)	106.2 (3.54)	114.0 (3.80)

4.4.3 Comparison of model fitting accuracies

Based on the refined SIFT matching points, the overall error of collinear function is lower than 2nd order polynomial function by 30-40% with same set of match points for Sichuan (as shown in Table 4.3). Both the RMSEs in x and y directions have been reduced by using collinear function for all three subsets in the two regions. The test Corona scene in Sichuan used here covers a region with

complex topographic relief, which increased the difficulty in image registration. In this circumstance, collinear function outperformed the polynomial function by correcting the image distortion caused by terrain as it accounts for the elevation of terrain. Although higher order of polynomial function is also capable of modeling complex distortion (Galiatsatos et al. 2008), they are a lot more unstable than the collinear function, and thus not suitable for large scale application.

On the contrary, the 2nd order polynomial function has lower RMSEs than the collinear function in two out subsets in St. Louis. It noteworthy that the numbers of selected tie points are only 13 in those two subsets, which is likely the reason why the retrieved polynomial equations tend to have lower fitting RMSEs. For the first subset in the scene, the fitting accuracies are much closer between the two transformation models when the sample size was 151.

4.4.4 Accuracy assessment

The accuracy of SIFT-based georegistration with different transformation functions was further examined based on manually selected tie points at the two regions, Sichuan and St. Louis. Manually selected tie points distributed across the whole scene (Figure 4.7), allowing a reliable accuracy assessment of the transformation models.

Table 4.4 shows the accuracy assessment for the two cases with two transformation models. Although the modeling accuracy of collinear function did not show much improvement over the polynomial function, results from manually selected tie points showed better accuracy from using collinear function than those

using 2nd-order polynomial functions in both cases. The collinear function method could serve the requirement on accuracy for the land cover change mapping purpose with the maximum RMSE being around 100 m.

For the Sichuan case, RMSEs for the results using polynomial function are in general two to three times larger than those using collinear function. This is not surprising because topography effects are well-accounted for in the collinear function with the help of elevation data.

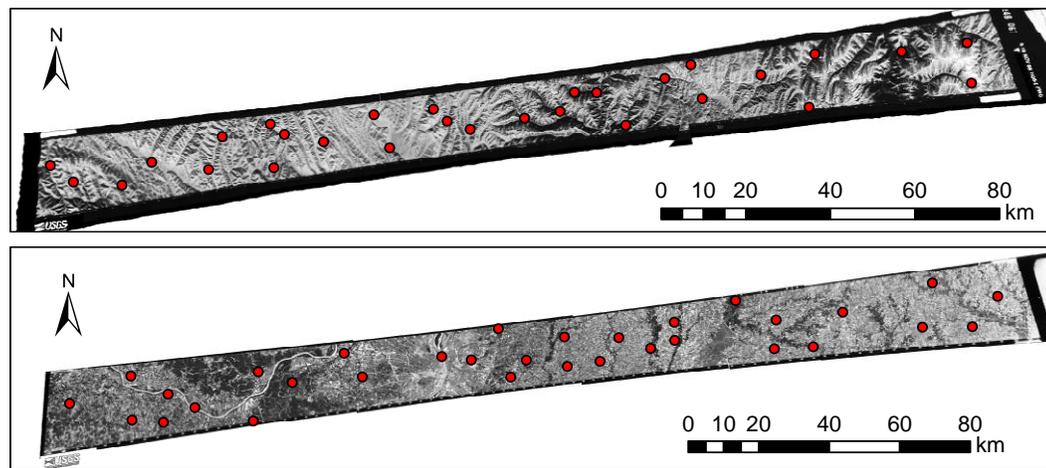


Figure 4.7 The distribution of manually selected validation points in Sichuan (upper) and St. Louis (lower) overlaid on a registered Corona image.

Unlike the fitting RMSEs for the St. Louis case, manual accuracy assessment showed much smaller RMSEs for the results using collinear function, which are also similar in magnitude compared with the fitting RMSEs for the case using collinear function. This demonstrates that the use of collinear function over polynomial function can still generate reasonably accurate results when the number of automatically selected tie points is not high. Examples of registered Corona image after preliminary registration, and precise registrations using 2nd order polynomial function and collinear function based on refined SIFT matching points are shown in

Figure 4.8 and Figure 4.9 for Sichuan and St. Louis separately. Significant improvements of precise registration over preliminary registration were observed and the collinear function was preferable to the polynomial function especially at regions with complex terrain.

Table 4.4 Accuracy assessment using manually selected validation points (vpts) for two transform models.

Sichuan		2 nd order polynomial			Collinear function		
Subset	Number of vpts	RMSE_X	RMSE_Y	RMSE_XY	RMSE_X	RMSE_Y	RMSE_XY
1	11	141.5 (4.72)	82.1 (2.74)	163.6 (5.45)	49.7 (1.66)	32.8 (1.09)	59.5 (1.98)
2	10	145.1 (4.84)	115.1 (3.84)	185.2 (6.17)	60.7 (2.02)	54.5 (1.82)	81.5 (2.72)
3	8	270.6 (9.02)	125.9 (4.20)	298.5 (9.95)	78.2 (2.61)	70.5 (2.35)	105.2 (3.51)
Total	29	187.2 (6.24)	107.2 (3.57)	215.7 (7.19)	62.4 (2.08)	52.9 (1.76)	81.8 (2.73)
St. Louis		2 nd order polynomial			Collinear function		
Subset	Number of vpts	RMSE_X	RMSE_Y	RMSE_XY	RMSE_X	RMSE_Y	RMSE_XY
1	11	182.6 (6.09)	114.5 (3.82)	215.5 (7.18)	37.5 (1.25)	48.5 (1.62)	61.4 (2.04)
2	12	851.2 (28.37)	550.2 (18.34)	1013.5 (33.78)	78.4 (2.61)	81.2 (2.71)	103.0 (3.43)
3	9	1326.4 (44.21)	1299.7 (43.32)	1857.0 (61.90)	58.0 (1.93)	87.7 (2.9.2)	105.1 (3.50)
Total	32	882.0 (29.40)	770.1 (25.67)	1170.9 (39.03)	57.9 (1.93)	70.9 (2.36)	91.6 (3.05)

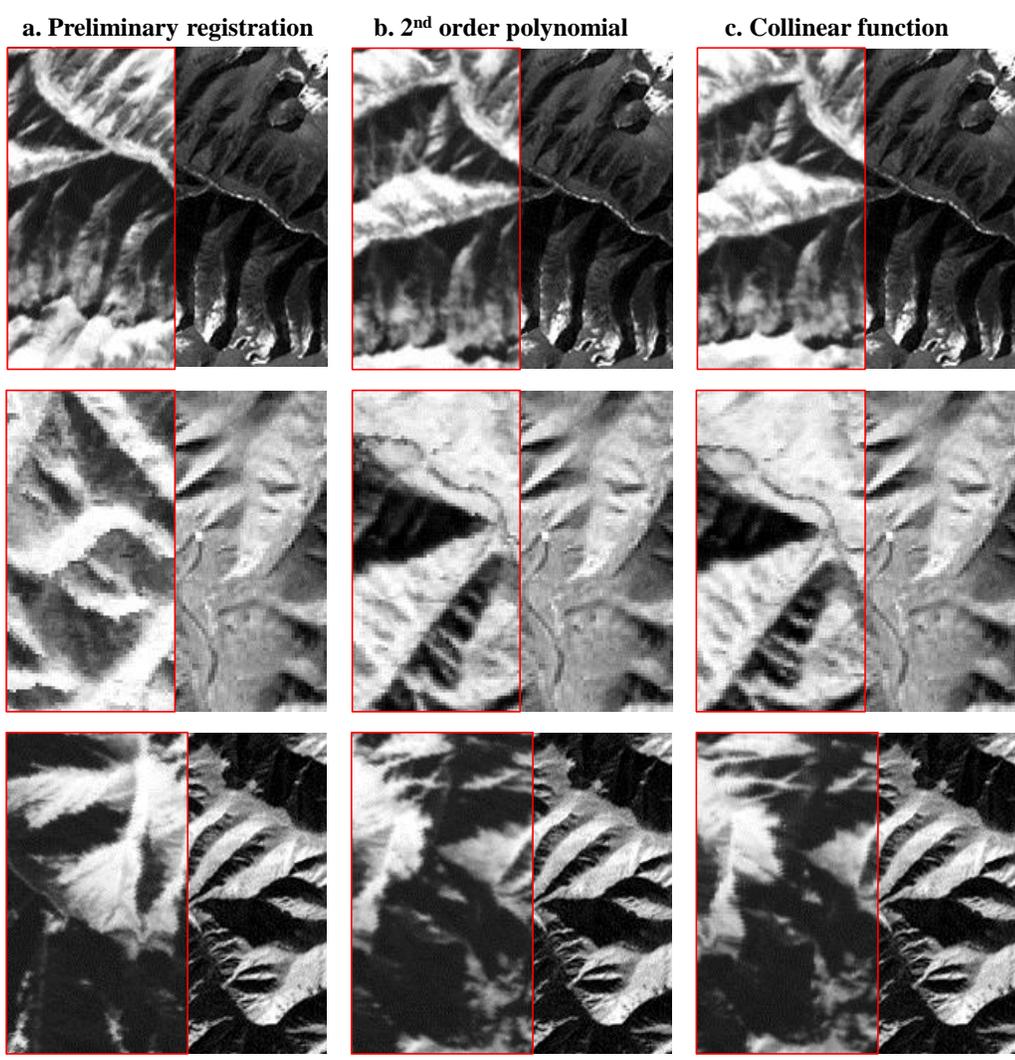


Figure 4.8 Corrected Corona images (red framed) after a) preliminary registration, and registration using b) 2nd order polynomial function and c) collinear function based on the refined SIFT matching points, overlaid on Landsat reference image in Sichuan. Each window is composed by 200-by-200 pixels at 30m resolution.

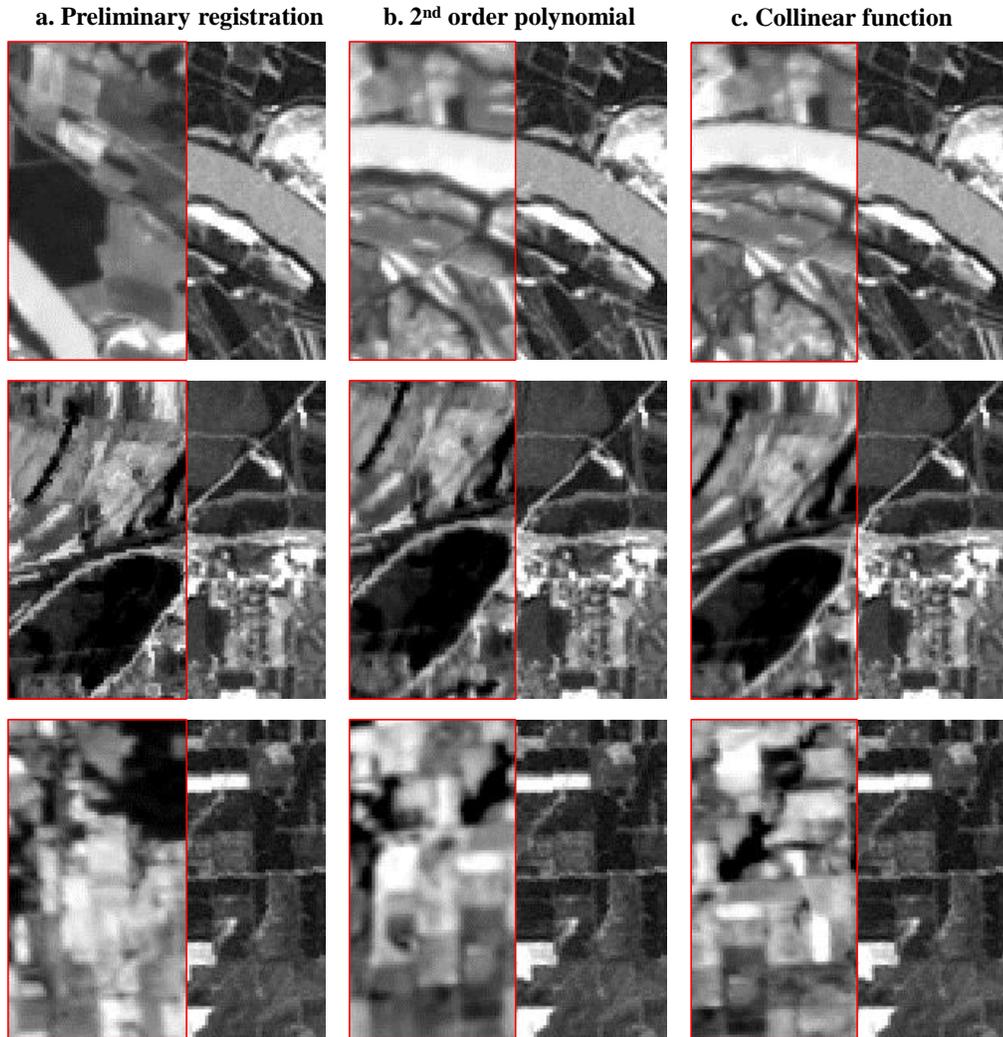


Figure 4.9 Corrected Corona images (red framed) after a) preliminary registration, and registration using b) 2nd order polynomial function and c) collinear function based on the refined SIFT matching points, overlaid on Landsat reference image in St. Louis. Each window is composed by 100-by-100 pixels at 30m resolution.

4.5 *Discussion and conclusions*

In this chapter, I explored towards automated georegistering of the historical Corona data by presenting a pioneer study of a semi-automated method with Landsat data as a reference. It consists of three steps, namely, preliminary registration, feature points extraction and matching, and precise registration. The major findings of this chapter include:

- 1) Initial registration is critical in the Corona registration, as it would significantly affect the SIFT matching accuracy and efficiency. Our experiments show that the errors from the initial registration constrained within 150 Landsat pixels at 30 m resolution would be proper for the following steps. It can be achieved with limited human intervention by locating the four image corners.
- 2) Collinear function has been proved to work for Corona data with tie points selected from Landsat data. However, it is noteworthy that due to the image distortion led by panoramic scanning and unknown flight velocity of Corona cameras, an assumption of one set of collinear function for each Corona scene would generate large uncertainties. To mitigate this problem, it can achieve reasonable georegistration accuracies by dividing the Corona scene into three subsets and assigning one set of collinear function for each subset.
- 3) The SIFT tie point (feature) extraction and matching algorithm is demonstrated to generate not only abundant but also accurate matching points. Especially, SIFT features are extracted at subpixel accuracy, which would improve the precise registration at a later stage. To account for the errors in the preliminary registration, a buffer zone, e.g., 150 Landsat pixels used in this study, should be considered in the tie points searching in order to remove potential pseudo matches. One limitation of the tie points selection algorithm is that due to the poor radiometric performance of Corona data and land cover changes that occurred between the acquisition times of Corona and Landsat, SIFT feature matching often failed when the study region does not have very

clear features and/or dramatic land cover change occurred. In addition, an iterative procedure was carried out to remove the potential pseudo tie points from the SIFT algorithm, which worked well for the selected Corona scenes. However, further efforts are still required to demonstrate the applicability of this approach on other Corona scenes.

Although the total RMSE of around 50 meters does not meet the requirement of pixel-to-pixel mapping at Landsat TM resolution (30 m), pixel level mapping at MSS level (60 m) or at a slightly coarser resolution could be fulfilled using the registered Corona image. Change mapping could still be carried out if the corresponding Landsat TM image is aggregated to a coarser resolution (say, 60 to 240 m) and the RMSE is constrained to sub-pixel level (Townshend et al. 1992), in order to detect the long-term dynamics on Earth surface (as shown in Figure 4.10 and Figure 4.11).

The results demonstrated that collinear function could solve the orthorectification problem of historical Corona image and the automated SIFT point selection method could successfully select adequate matching points to solve the function. There still are unsolved issues hindering a full automation of the Corona image registration and further the application over large area.

First, “pseudo” matches due to land cover change have prevented a more adequate number of matches, which can otherwise ensure a more stable fitting of transformation models. The current choice of GLS as a reference may suffer more from land cover changes than using the MSS data with a smaller temporal gap between Corona and Landsat acquisition times. However, the current MSS data also

suffers from relatively low georegistration accuracy and poor image quality (e.g., low signal-to-noise ratio) compared with TM and ETM+ data. Once an improved MSS dataset is provided, an improvement in Corona registration may also be achieved.

Second, regarding the shortage of SIFT technique, several improvement methods (Han et al. 2012; Kupfer et al. 2015; Li et al. 2009; Yu et al. 2008), including reducing false matches and enhancing a spatially even distribution of SIFT matches, have been recently proposed and could be tested for the Corona image registration. Further efforts will be made to incorporate these methods into the current Corona registration processing chain and to improve the accuracy.

Last, although our target of Corona registration does not lie on the transformation model parameters, an improvement in constraining the model parameters would be of help in reducing the final registration error if the Corona flight status (location and gesture) is better understood and considered into the data processing.

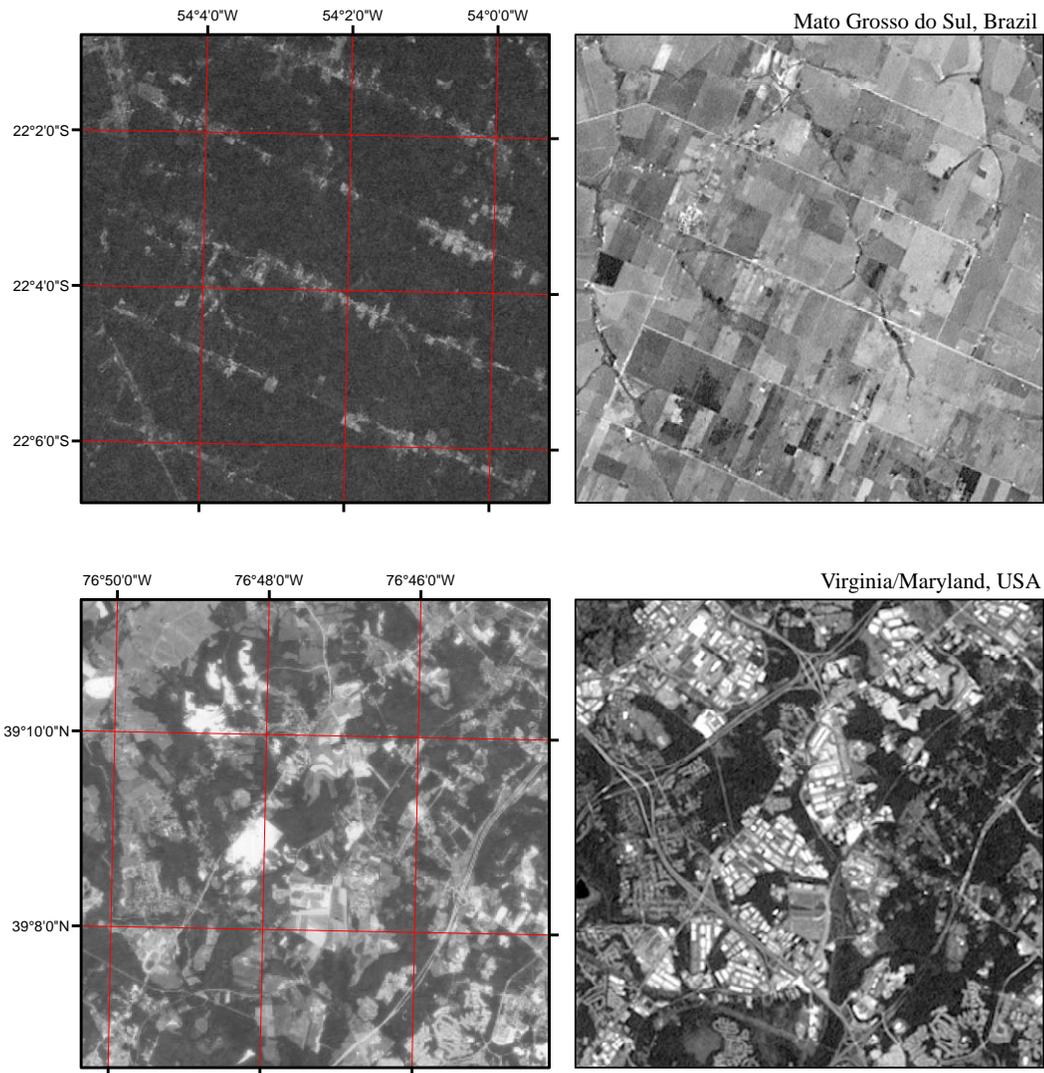


Figure 4.10 Example of land cover change between 1960s and circa 2000 in MGdS and VM study areas captured by Corona (left) and Landsat (right) data.

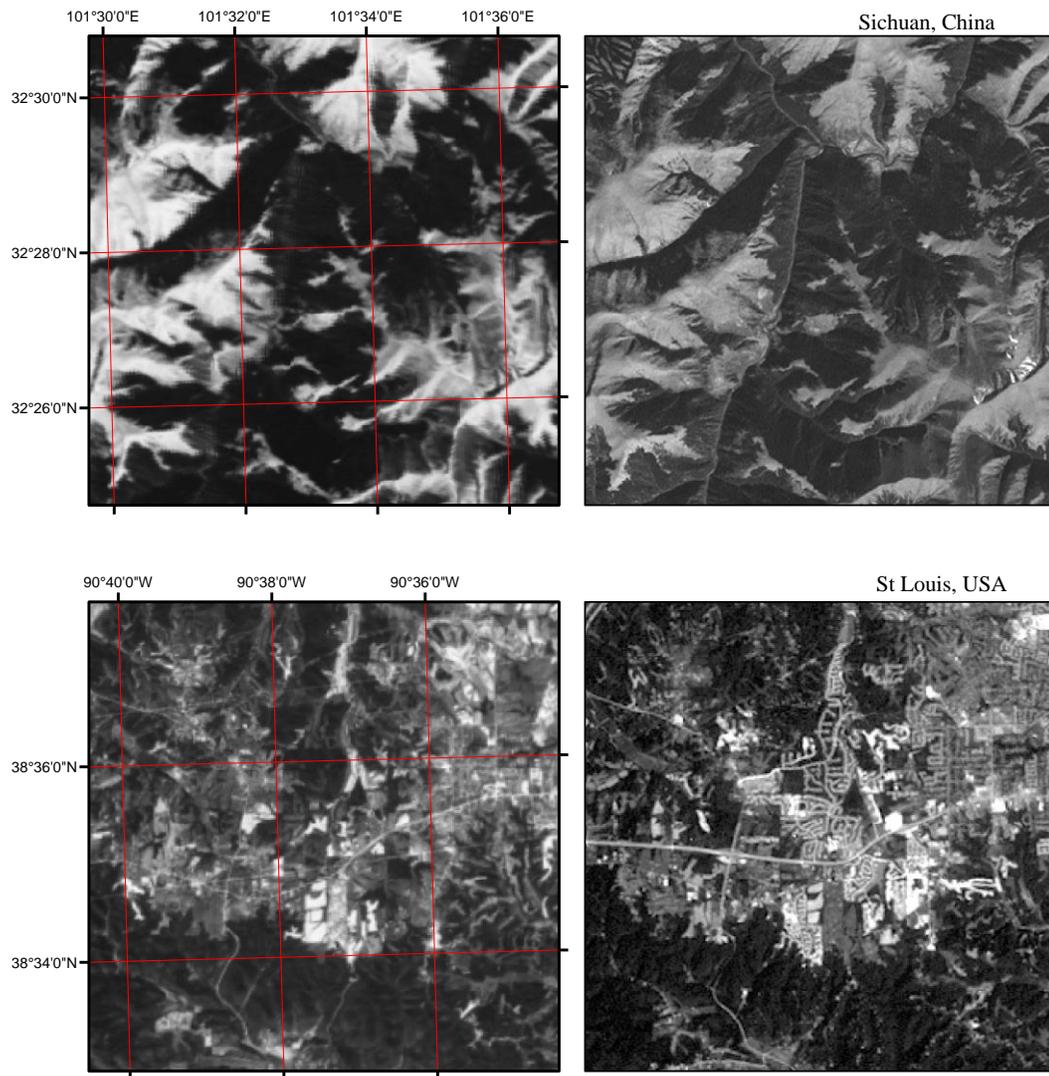


Figure 4.11 Example of land cover change between 1960s and circa 2000 in SC and SL study areas captured by Corona (left) and Landsat (right) data.

Chapter 5: Remote sensing-based estimation of forest cover and change since the 1960s in Sichuan, China³

5.1 *Introduction*

China has undergone frequent social and economic changes which at times led it to fast economic development and natural resource consumption since the founding of the country in 1949. Four of the most important forest-related policies are as follows: 1) the Great Leap Forward (1958-1963), known as the second Five Year Plan aiming to modernize the Chinese economy through industrialization and collectivization, 2) the Cultural Revolution (1966-1976), known as a ten-year revolution throughout the entire country with chaotic political and economic policies, 3) the Reform and Opening (1978-now), resulting in greatly increased income, prosperous urban and rural industry, but also polluted environment and generated huge wealth inequality, 4) the National Greening Programs (1998-now), two of the biggest programs known as the Natural Forest Conservation Program (NFCP), with goals of reducing timber harvesting and increasing capacity in natural forest, and the Grain to Green Program (GTGP) aiming to restore forest and grassland cover especially on steep slopes from cropland (Liu et al. 2008). For the half-century after 1949, forest sector has been playing role of providing timber products and fuel woods to support the nation's development and population growth, therefore, vast area of natural forests were lost (Albers et al. 1998). Hit by the devastating flooding in 1998,

³ This chapter has been prepared for submission to the journal of *International Journal of Digital Earth*.

China set the logging bans and afforestation as the primary focus for the implementation of conservation policies and aimed at mitigating environment degradation, and conserving natural resources (Liu et al. 2008).

However, there are controversies on the trend of resultant forest cover change due to changes in policy in China. On one side, some studies depicting forests in China as carbon pool by citing the official statistics which reported a continuous increase in forest cover from 12.7% to 20.36% during 1976 and 2008 (Fang et al. 2001; FAO and JRC 2012; Nabuurs et al. 2007). On the other side, such increase trend is doubted by researchers who criticized that the natural forests in China experienced irreversible destruction (Smil 1984). As a result, a spatially and temporally comprehensive estimation on the change of forest cover is desirable, in order to evaluate the effectiveness of long-term forest policies (Vina et al. 2016).

As discussed in Chapter 2, remote sensing is an advanced technology in land use and land cover change mapping equipped with various spatial resolutions and temporally continuous coverages. The combination of Landsat and Corona data provides continuous observation of the Earth's surface from the 1960s until now, which enables consistent forest cover mapping and provides spatial explicit distribution of where forest cover change have happened.

The primary objective of this chapter is to assess forest change in Sichuan province over four decades (1960s-2005) and examine potential connections between the observed change and social economic and policy changes. This is achieved by quantifying forest cover in the 1960s using Corona data and analyzing the derived results with existing forest cover change products derived from Landsat for 1975 and

afterwards. The satellite-based results are compared with forest cover estimates reported by national and provincial agencies. More details of the study region and data are described in Section 5.2. Sampling design and the accuracy assessment of forest cover change products are presented in Section 5.3. The estimated forest cover trend is presented in Section 5.4 and the observed change on forest cover was linked to concurrent forest and socioeconomic policies from the 1960s to the 2000s, which is discussed in Section 5.5.

5.2 *Study area and data*

5.2.1 Study area

Sichuan province, located in Southwest China, has a terrestrial area of $5.66 \times 10^5 \text{ km}^2$ before and $4.84 \times 10^5 \text{ km}^2$ and after the autonomy of Chongqing city in 1997. Besides Yunnan and Tibet, Sichuan is another major part of the second largest forest of the nation, Southwest forest (Huang et al. 2008a) which accounts for 43.9% of timber reserves (Wang et al. 2004) of China and originally had tremendous primary forest. Adjoining the Qinghai-Tibetan Plateau forest in Sichuan has terrain characterized by high spatial diversity and high sensitivity to climate change. Before the 1950s, most forests in China were naturally regenerated and there were large areas of old-growth forest (Zhang et al. 2000a). Average carbon density of forest is $38.04 \text{ MgC} \cdot \text{hm}^{-2}$. In general, forest carbon density decreases from subalpine coniferous forest in western Sichuan ($30 \sim 40 \text{ MgC} \cdot \text{hm}^{-2}$) to the montane evergreen broadleaf forest located in the east and the edge of Sichuan Basin ($20 \sim 30 \text{ MgC} \cdot \text{hm}^{-2}$), and increases along altitude and slopes (Huang et al. 2009a).

Forests in Sichuan are currently under a mixed management system of state-run and collective-run forest. Forest disturbance is mainly driven by long-term, complex human activities, especially at low-altitude regions and the Sichuan Basin. Thus, primary and undisturbed forests are mainly distributed in the high-elevated, steep-sloped regions ranging from mid-west to mid-north of Sichuan. With the great potential of forest resources, Sichuan has been selected as the pioneer region for several forestry programs such as Natural Forest Conservation Program (NFCP), Grain to Green Program (GTGP), and Yangtze River Shelter Forest Program (Liu et al. 2008). However, the complex terrain hinders the field inventory of forest resources, thus making the use of remote sensing technology necessary in forest cover and change estimates. Sichuan provides major habitats for many wildlife species such as the giant panda. Liu et al. (2001) found that suitable habitat for the giant panda was reduced by around 5100 ha before the establishment of the Sichuan Wolong Nature Reserve in 1975 and further reduced by 13% (~7300 ha) from its establishment until 1997 (Liu et al. 2001).

5.2.2 Data

Corona data

The characteristics of Corona data are described in Chapter 2. In this chapter, the number and location of Corona data are determined by the sampling strategy deliberated in Section 5.3. Corona data acquired by KH-4A and -4B mission are used when available.

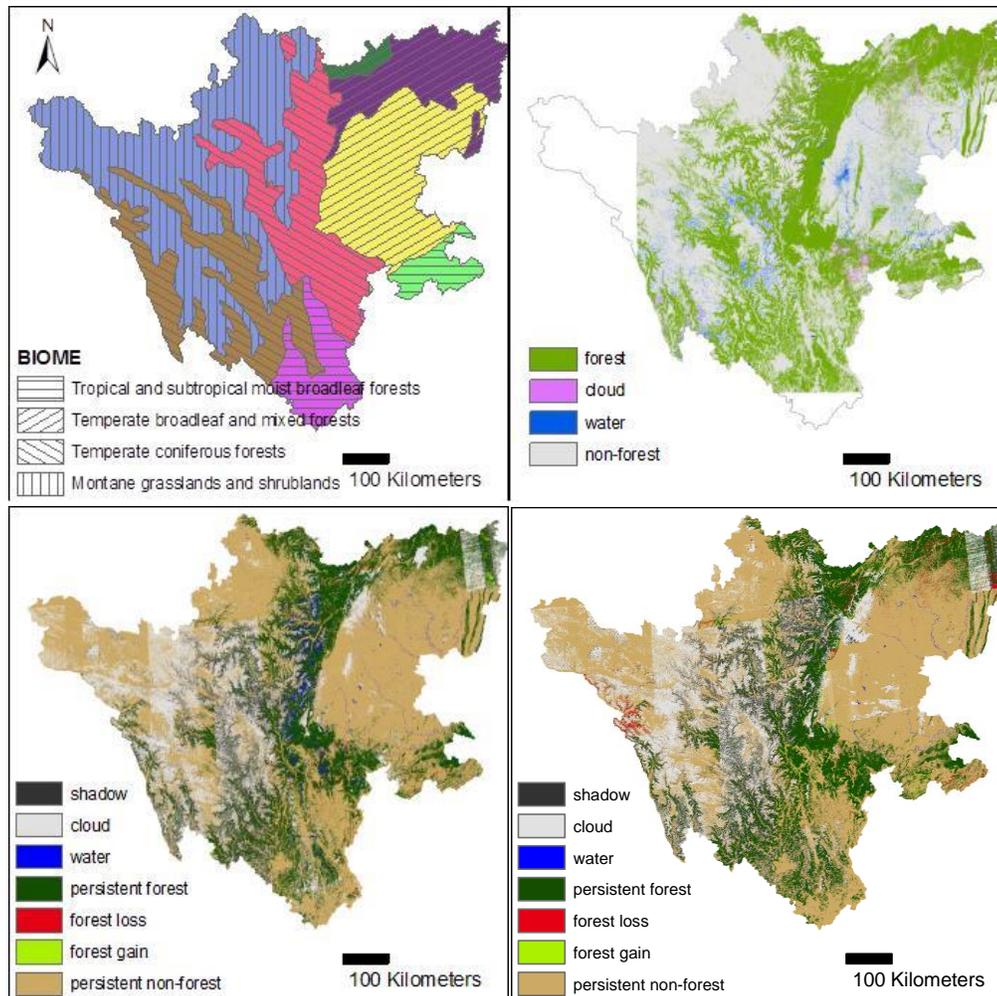


Figure 5.1 Biome map (upper-left), downloaded from the World Wildlife Fund (<http://www.worldwildlife.org/publications/terrestrial-ecoregions-of-the-world>), and GLCF-GFCC maps for Sichuan, including forest cover map in the 1975 (upper-right), and change map from 1990 to 2000 (lower-left), and from 2000 to 2005 (lower-right), downloaded from Global Land Cover Facility (<http://landcover.org/data/landsatFCC/>).

GLCF-GFCC maps for epoch 1975 and after

Among different products, the global forest cover change map produced by Global Land Cover Facility, named the Global Forest Cover Change (GFCC) map, is selected for analyzing the forest cover change trend in Sichuan province. The forest cover map in 1975 and change map from 1990 to 2000 and 2000 to 2005 are presented in Figure 5.1, along with the ecoregion map of Sichuan. The regional

accuracy of the GLCF-GFCC map needs to be evaluated before applying it to regional analysis.

Forest statistical data

Seven national forest inventories (NFI) have been implemented by the Chinese government from 1973 to 2008 (i.e. 1st NFI: 1973-1976, 2nd NFI: 1977-1981, 3rd NFI: 1984-1988, 4th NFI: 1989-1993, 5th NFI: 1994-1998, 6th NFI: 1999-2003, and 7th NFI: 2004-2008) and forest statistics were derived from field inventory (SFA 2005, 2009; Thematic Database for Human-Earth System). This dataset was used to investigate the differences in forest cover estimates obtained from remote sensing data. NFI reports were downloaded from the Chinese Forest Science Data Center (CFSDC, <http://www.cfsdc.org/>). Sichuan statistical yearbooks were downloaded from the Sichuan Provincial Bureau of Statistics (<http://www.sc.stats.gov.cn/>). Forest records for 1949 and 1950-62 were acquired from the Thematic Database for Human-earth System (TDHS, <http://www.data.ac.cn/index.asp>). However, it is noteworthy that the technical details, especially of TDHS statistics, are not available yet.

5.3 Methods

5.3.1 Sampling strategy for Corona data

Although wall-to-wall mapping has been widely implemented over large regions for many satellite data, such as Landsat or MODIS, there are three major challenges hindering the wall-to-wall mapping of Corona data. The first issue is the time and cost of Corona data processing. A full coverage for Sichuan requires more than 200 Corona images (Figure 5.2). Although the developed semi-automated

registration method could significantly reduce processing cost, further efforts such as initial registration and refining pseudo matches are still needed to process large volume of images. Second, clouds are key sources of contamination for the Corona image. The gaps caused by clouds and shadows could hardly be filled by other images acquired within a short period of time, because Corona data do not have a regular revisit period. Lastly, free Corona images cannot provide a complete coverage of the study region (Figure 5.2), and additional images need to be purchased in order to achieve a wall-to-wall mapping. Considering these limitations, a sampling method is applied to the Corona data to estimate forest cover in the 1960s.

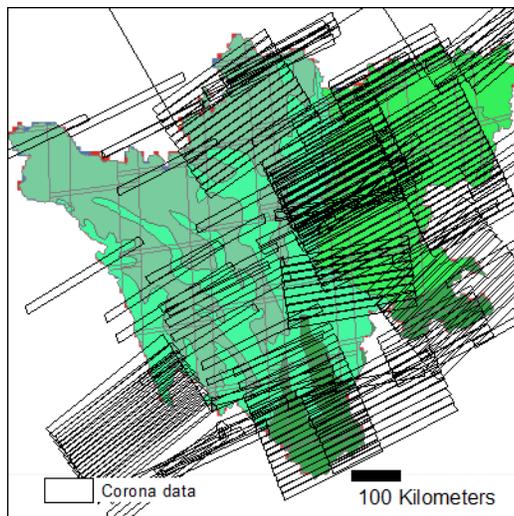


Figure 5.2 Footprint of Corona data for 1960s in Sichuan that are freely available for download, inquired from EarthExplorer website (<http://earthexplorer.usgs.gov/>).

A sampling method is commonly adopted for land cover and change estimation when data availability is limited (DeFries et al. 2007; Hansen et al. 2008). Among different sampling strategies, the stratified random sampling (SRS) method is favored over simple random sampling or systematic sampling because of its advantages in reducing uncertainty and maintaining the accuracy of the estimate

(Broich et al. 2009). The major issue with systematic sampling is that it cannot provide an unbiased estimate, and the issue with random sampling is that forest loss usually does not follow random distribution. Information of the forest cover and change in the 1960s, especially with explicit spatial distribution, is needed to define the stratum for the SRS method. However, such data is not available over many regions including Sichuan.

Due to the constraints on data availability, a forest cover change (FCC) probability map is produced using an experiential model. Assuming that forest cover and change after the 1970s along with other factors such as land cover types provides a good indication to forest change from the 1960s to the 1970s, a set of independent variables were used to predict the change probability between the 1960s and 1970s ($FCC_{60s} \%$), including persistent forest ($ff \%$) and non-forest ($nn \%$), forest loss ($fn \%$) and gain ($nf \%$) between 2000 and 2005, DEM and fractions of land use types (including crop, forest, grass, water, buildup, unused land). The model was trained in south Sichuan where forest loss was reported to occur during the 1960s (Chinese Academy of Forestry 2016), and was then applied to the entire province. The model was trained and applied at block level with a size of $10 \times 10 \text{ km}^2$. The reason for selecting this block size is mainly because it is close to the width of one Corona scene. Input images, including GFCC 2000-2005 at 30m resolution and land use data in 2000 at 1km resolution produced by the Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Science were aggregated to 10 km to derive the fractional value. DEM data at 90m resolution (USGS 2004) was also

aggregated to 10 km resolution using a spatial averaging method. The derived prediction model is as below:

$$\begin{aligned}
 FCC_{60s} \% = & 1.5 + 0.14 \times ff \% + 3.41 \times fn \% + 1.75 \times nf \% + 0.17 \times nn \% \\
 & - 0 \times DEM - 0.014 \times crop \% - 0.014 \times f \% - 0.016 \times grass \% \\
 & - 0.011 \times water \% - 0.012 \times buildup \% - 0.136 \times unused \%
 \end{aligned}
 \tag{Eq. 5.1}$$

By applying Eq. 5.1 to the entire study region, the FCC probability was predicted for all 10×10 km² blocks in Sichuan. Three change strata including low, medium and high change to represent different degrees of forest cover change, plus a no-change stratum that was mainly located in regions without forest cover or regions with no valid data due to persistent cloud and shadow contamination. Thresholds used for separating different strata were derived following the Dalenius-Hodges rule (Cochran 1977) and are listed in Table 5.1. First of all, 10 blocks were selected from the no change stratum. For the three change strata, 50 blocks in total were sampled, considering the factors such as processing time and image availability. In total, 60 blocks were sampled for the change and no-change strata in Sichuan, representing a sampling ratio of 1.24% to the total blocks.

Furthermore, an optimal allocation method, called Neyman allocation method (Cochran 1977), was adopted to decide the number of sample blocks to be assigned to each change stratum. The number of sample blocks for each stratum is listed in Table 5.1. The distributions of strata and sample blocks are shown in Figure 5.3.

Table 5.1 The summary of stratified sampling design, including the boundary, size of total population, size of sampled blocks and the ratio of sampling for each stratum.

Stratum (threshold of FCC%)	Percent of area	No. of blocks (N_h)	No. of sampled blocks (n_h)	Percent sampled	Percent of total sampled bocks
No change (<0.078%)	22.64%	1096	10	0.91%	16.67%
Low change (0.078-24%)	23.60%	1143	13	1.14%	21.67%
Medium change (24.01-40%)	41.84%	2026	17	0.844%	28.33%
High change (40.01-99.08%)	11.92%	577	20	3.47%	33.33%
Total	100%	4842	60	1.24%	100%

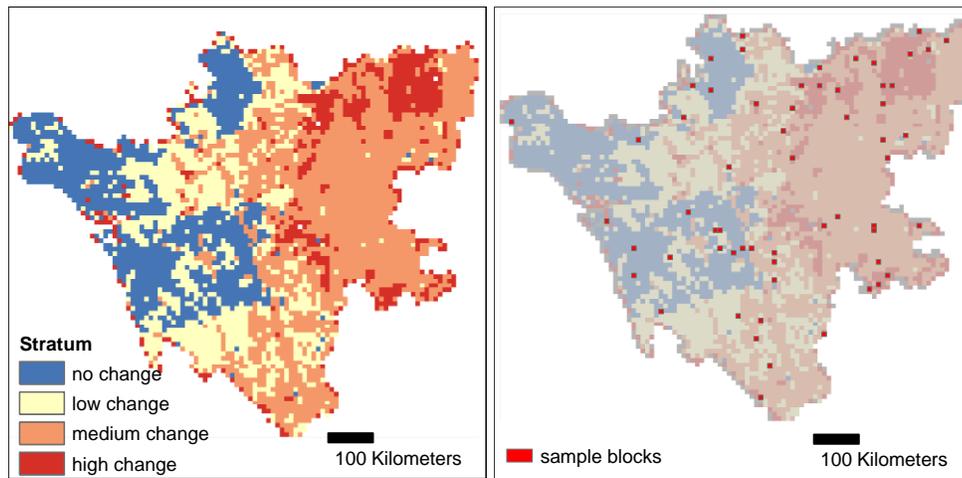


Figure 5.3 Strata map (left) and the location of selected sample blocks (right).

5.3.2 Block level forest cover estimate from Corona

According to the nominal coverage of the Corona image, 60 Corona images were selected to cover all sample blocks. Since the block size is $10 \times 10 \text{ km}^2$, one block can be covered by a Corona subset rather than the entire image. Considering the fact that one Corona scene is distributed in four separate files (subsets), only one out of four subsets is therefore needed and selected. Thus, image registration and classification were applied only to the selected subset, to derive the forest cover for the overlapped sample block.

Corona image registration

Because the nominal coverage file and the Corona image are distributed separately, an additional step is required to add preliminary geolocation to the Corona image. For each selected subset, up to four tie points were manually selected to link the Corona image to the reference Landsat data. Here, GLS 1990 was used as reference data considering its smaller difference on acquisition time than GLS 2000 and 2005 with the Corona data, and higher geolocation accuracy than GLS1975 (Gutman et al. 2013). Following the preliminary registration, the SIFT method was applied to automatically select tie points among Corona and Landsat images. After the precise registration, 32 out of 60 sampled Corona subsets were successfully registered to Landsat images with RMSE up to six Landsat pixels (30m). For the remaining 28 images, manual precise registration method was adopted, and 14 images were successfully processed. The average RMSE of manual registration is 2.8 Landsat pixels. The failure in registration of the remaining 14 images is mainly caused by the poor quality of the Corona image due to a film exposure problem or cloud contamination. Limited by image availability, substitute images could not be found for these sample blocks. However, since the 14 failed images appeared to be randomly distributed in Sichuan, the 46 registered Corona subsets still follow a random distribution and it keeps the predefined SRS strategy valid. As a result, 46 Corona image subsets were used towards the next step.

Another issue appeared after I finished the image registration. Since the coverage of the Corona image provided by USGS was an approximate range, it was observed that there were common shifts between the nominal coverage and the actual

location with a degree of several kilometers. Thus, the previously selected sample blocks probably could not be covered by the registered Corona data, and forest cover could not be estimated on these blocks. To solve this problem, blocks covered by the successfully registered Corona subsets composed a new set of samples, facilitating the predefined sampling strategy. This set of sample blocks was utilized and the inclusion probability of each sample block within each stratum was recalculated. The actual number of sample blocks and the sampling percentage is summarized in Table 5.2.

Table 5.2 The size of sample blocks and the sampling ratio (%) each stratum by using all blocks covered by registered Corona image. n_h and N_h are denoted the sample size and size of total populations for stratum h .

	n_h / N_h	Sample%
Stratum 1: no change	85/1096	7.8%
Stratum 2: low change	60/1143	5.2%
Stratum 3: medium change	113/2026	5.6%
Stratum 4: high change	74/577	12.8%
Total	332/4842	6.9%

Corona image classification

Concluding from Chapter 3, a combination of texture features, including dissimilarity, second moment, co-occurrence variance and mean, is suitable for forest/non-forest classification. In addition, a window size of 7-by-7 pixels is recommended for classification in complex landscapes such as Sichuan. So, the four textures were calculated at the original resolution of the Corona image and aggregated using a scaling factor of 7-by-7 pixels. Image classification was implemented at this

coarser resolution with SVM classifier and further aggregated to 30 m. Georegistration transformation function was then applied to these Corona classification images to derive the forest cover maps for each sample block. Since the average RMSE of Corona image registration is larger than an MSS pixel (~60 m), block-level forest cover estimate is more appropriate than a pixel-to-pixel change map. As a result, the fraction of forest cover for each 10-by-10 km² sample block was derived.

5.3.3 Region level forest cover estimate for the 1960s

By adopting the equations for calculating the regional estimate from samples (Stehman 2014), forest cover rate and standard error over the entire Sichuan province can be inferred from a sample-based estimation of forest cover percentage. For each stratum h , a simple random sample of n_h blocks was selected from all blocks N_h . N is the total number of blocks in Sichuan. The regional averaged forest cover rate \hat{Y} was calculated using the following equation:

$$\hat{Y} = \sum_{h=1}^H \frac{N_h \bar{y}_h}{N} \quad \text{Eq. 5.2}$$

$$\bar{y}_h = \sum_{u \in h} y_u / n_h \quad \text{Eq. 5.3}$$

where \bar{y}_h is the average forest cover of all selected sample blocks in stratum h . y_u is the forest cover for each sample block u ($u \in h$) derived from selected Corona classification maps. H is the number of strata and N is the total number of blocks in Sichuan. The variance \hat{V} of the estimated forest cover was calculated following the equations:

$$\hat{V}(\hat{Y}) = \left(\frac{1}{N^2}\right) \sum_{h=1}^H \frac{N_h^2 \left(1 - \frac{n_h}{N_h}\right) s_{yh}^2}{n_h} \text{ Eq. 5.4}$$

$$s_{yh}^2 = \sum_{u \in h} (y_u - \bar{y}_h)^2 / (n_h - 1) \text{ Eq. 5.5}$$

where s_{yh}^2 is the sample variance of stratum h . The standard error SE of estimated regional forest cover \hat{Y} is the square root of the estimated variance \hat{V} .

5.3.4 Estimation of forest cover and change for 1975 and after

Since the GFCC product for 1975 used in this study is a pre-released version, its accuracy has not been evaluated either at a regional or global scale. Accuracies of GFCC at the other three epochs have only been conducted at the global scale (Feng et al. 2016). In order to utilize this dataset, accuracy assessment needs to be conducted particularly for Sichuan. Four WRS2 path/rows were randomly selected from the total of 24 path/rows (Figure 5.4), representing the four biomes of the region. A stratified random sampling algorithm with proportional allocation was adopted with the forest and non-forest classes as two strata (Olofsson et al. 2014). A sampling percentage of 0.003% was defined, considering the time and efforts needed in reference class labeling. Thus, 384 forest pixels and 572 non-forest pixels in total were randomly sampled from the four forest/non-forest maps in 1975. Due to the lack of reference data for early 1970s, it is unavoidable to use ancillary information to make the best estimation of the forest/non-forest class for the 1970s. Each sample pixel was examined by visually interpreting the MSS imagery in the 1970s and high-resolution imagery from Google Earth (www.earth.google.com). Forest cover in the 1960s was checked to determine if the location was covered by a registered Corona image.

Knowledge on the temporal change derived from other GFCC epochs, along with the spatial context of the pixel, was used to label each sample pixel. One assumption for the labeling is that from the 1970s to the 1990s, forests are more likely to be disturbed than regrowth. Therefore, a pixel is likely to be forest in the 1970s if it exhibits vegetation characteristics in the MSS image and is classified as forest in the 1990 and afterwards. On the contrary, if a pixel is classified as non-forest after 1990, it may be a forest pixel depending on the spectral characteristics in the MSS image since forest loss could occur during the 1970s and 1990. The set of reference pixels for 1975 was reserved for assessing the accuracy of the forest cover maps after 1990.

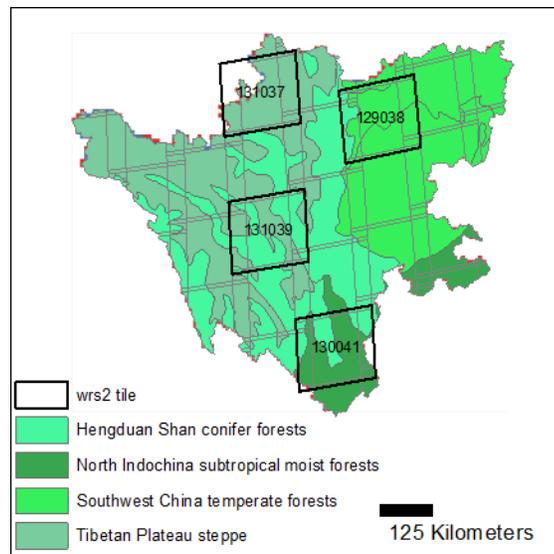


Figure 5.4 Path/row selected for the accuracy assessment of GLCF-GFCC product in Sichuan, displayed on top of the biome map, downloaded from the World Wildlife Fund (<http://www.worldwildlife.org/publications/terrestrial-ecoregions-of-the-world>).

The common form of error matrix in terms of sample counts is illustrated in Table 5.3, accompanied by unbiased estimator of the proportion of area in cell i,j of the error matrix:

$$\hat{p}_{ij} = W_i \frac{n_{ij}}{n_{i\cdot}} \text{ Eq. 5.6}$$

where w_i is the proportion of area mapped as class i . Estimates of accuracies, including user's accuracy (U or UA), producer's accuracy (P or PA), and overall accuracy (O or OA), based on the stratified random sampling design can be obtained by applying the following equations:

$$\hat{U}_i = \frac{\hat{p}_{ii}}{\hat{p}_{i\cdot}} \text{ Eq. 5.7}$$

$$\hat{P}_j = \frac{\hat{p}_{jj}}{\hat{p}_{\cdot j}} \text{ Eq. 5.8}$$

$$\hat{O} = \sum_{j=1}^q \hat{p}_{jj} \text{ Eq. 5.9}$$

The unbiased estimates of forest cover and standard error were then calculated using the following equations (Olofsson et al. 2013):

$$\hat{p}_{\cdot 1} = \sum_{i=1}^2 W_i \frac{n_{i1}}{n_{i\cdot}} \text{ Eq. 5.10}$$

$$S(\hat{p}_{\cdot 1}) = \sqrt{\sum_{i=1}^2 W_i^2 \frac{\frac{n_{ij}}{n_{i\cdot}} \left(1 - \frac{n_{ij}}{n_{i\cdot}}\right)}{n_{i\cdot} - 1}} \text{ Eq. 5.11}$$

Table 5.3 Error matrix of sample counts, n_{ij} and estimated area proportion, \hat{p}_{ij} . Map classes are the rows while reference classes are the columns.

Class	Forest	Non-forest	Total
Forest	$n_{11} (\hat{p}_{11})$	$n_{12} (\hat{p}_{12})$	$n_{1\cdot} (\hat{p}_{1\cdot})$
Non-forest	$n_{21} (\hat{p}_{21})$	$n_{22} (\hat{p}_{22})$	$n_{2\cdot} (\hat{p}_{2\cdot})$
Total	$n_{\cdot 1} (\hat{p}_{\cdot 1})$	$n_{\cdot 2} (\hat{p}_{\cdot 2})$	$n (1)$

Table 5.4 Accuracy of forest class for GLCF forest cover product at four epochs in Sichuan. The biome where each Landsat scene located is also shown. Accuracy measurements include overall accuracy (*OA*), producer's accuracy (*PA*), and user's accuracy (*UA*) of the forest class.

<i>Forest in Sichuan</i>		1975	1990	2000	2005
p131r039 Coniferous forest	<i>UA</i>	89.72	96.67	100.00	93.33
	<i>PA</i>	92.27	91.44	95.92	91.57
	<i>OA</i>	92.84	97.15	99.04	96.59
p129r038 Temperate forest	<i>UA</i>	92.14	91.40	95.95	93.24
	<i>PA</i>	91.56	94.87	79.96	82.88
	<i>OA</i>	91.64	94.32	89.20	90.13
p130r041 Subtropical forest	<i>UA</i>	85.34	92.54	98.55	98.55
	<i>PA</i>	89.97	83.64	83.33	87.12
	<i>OA</i>	87.22	89.49	91.05	93.28
p131r037 Mountain steppe	<i>UA</i>	66.67	88.24	94.44	94.12
	<i>PA</i>	80.68	60.02	72.47	55.28
	<i>OA</i>	95.24	97.65	98.54	97.29
Total	<i>UA</i>	88.02	92.27	97.40	95.26
	<i>PA</i>	90.91	88.63	83.13	84.64
	<i>OA</i>	91.52	94.34	93.69	93.86

5.4 *Results*

5.4.1 Accuracies of forest cover maps in 1975, 1990, 2000, and 2005

Accuracy of forest class in 1975 was consistently high across all biomes in Sichuan, with overall accuracy (*OA*) being 91.5% (*SE* = 1.57%) (Table 5.4).

Producer's accuracy was >90% and user's accuracy was 88% for forest class.

Therefore, commission errors ($CE = 1 - UA$) and omission error ($OE = 1 - PA$) were ~10%. The highest accuracies were found in conifer forests (p131r039), temperate forests (p129r038), followed by subtropical moist forests (p130r041) and mountain steppes (p131r037) – which respectively had producer's accuracy (*PA*) of 92.27%, 91.56%, 89.97% and 80.68%, and user's accuracy (*UA*) of 89.72%, 92.14%, 85.34%,

and 66.67% for the forest class. The omission errors were all < 20%. The accuracies for epoch 1990 and afterward are consistent with the map of 1975, with maps' overall accuracy of ~94% for all. User's accuracy for forest class is consistently over 90% and producer's accuracy is slightly lower ranging from 83.13% in 2000 to 88.63% in 1990.

Table 5.5 Forest cover (%) from 1960s to 2005 estimated from remote sensing data and statistics of forest inventory. The standard error associated with forest cover is reported in parentheses.

Epoch	RS estimate d forest cover	NFI stats			Yearbook		TDHS	
		forest cover	forest land use	forestla nd	forest cover	forestlan d	forest land use	forestlan d
1949	-	-	-	-	-	-	-	16.92
1960s	45.19 (1.62)	-	-	-	-	-	36.31	12.89
Circa 1975	38.98 (2.06)	13.3	36.04	13.26	-	-	-	-
Circa 1990	28.91 (2.07)	20.37	47.21	20.37	19.17	-	-	-
Circa 2000	28.19 (2.27)	30.27	46.84	25.51	39.7	24.23	-	-
Circa 2005	27.87 (2.14)	34.31	47.79	27.15	28.98	27.13	-	-

5.4.2 Estimates of forest cover from the 1960s to 2005

Table 5.5 reports the estimates of forest cover rate for Sichuan from the 1960s to 2005, as well as the measurements derived from multiple statistical reports. An estimated forest cover of $45.19 \pm 3.11\%$ in the 1960s was derived at the 95% confidence level from the sampled Corona classification maps. Using the GFCC product circa 1975 yielded an estimated forest cover of $38.98 \pm 4.04\%$ at the 95% confidence level. For the following epochs, GFCC products also yielded area-

adjusted estimates of forest cover of $28.20 \pm 4.06\%$, $27.73 \pm 4.45\%$ and $27.27 \pm 4.19\%$ at the 95% confidence level in 1990, 2000 and 2005 respectively. Examples in Figure 5.5 further illustrate three examples of changes of Sichuan's forest cover over a 40-year period observed from Corona, MSS, TM and ETM+ images, including (a) deforestation due to mining, (b) deforestation and regrowth, and (c) conversion of forest to cropland.

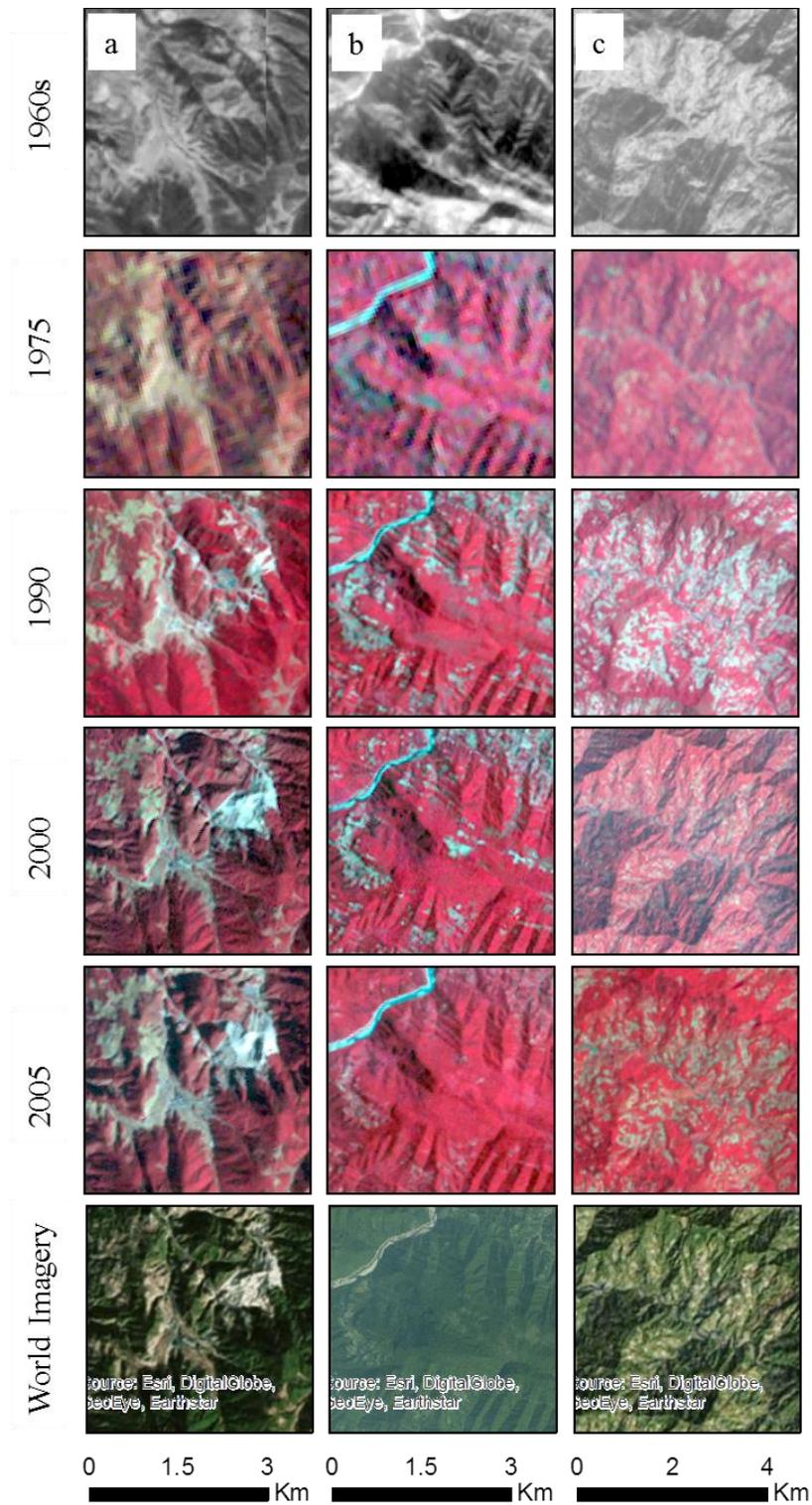


Figure 5.5 Examples demonstrate forest cover dynamics over five epochs in Sichuan, including a) forest cover loss due to coal mining; b) forest logging and afforestation; c) conversion from forest to agricultural land. ESRI World Imagery maps at very high resolution are shown as references (the last row).

5.4.3 Change in forest cover from the 1960s to 2005

Continuous decrease in forest cover during the 40-year period was observed from the remote sensing-based estimates. By adopting Eqs. 5.13-14, it is estimated that forest cover has totally decreased by 38% from the 1960s to 2005. 31% of the total forest loss was due to the loss happened in the first decade (circa 1965-1975), and 58% of the total loss happened between 1975 and 1990. Since 1990, forest cover has become relatively stable with a much lower decrease rate of 0.24% per year until 2005.

$$FCC_{total}(\%) = \frac{|FC_{t_1} - FC_{t_2}|}{FC_{t_1}} \times 100 \quad \text{Eq. 5.12}$$

$$FCC_{annual}(\%) = \frac{|FC_{t_1} - FC_{t_2}|}{FC_{t_1} \times (t_2 - t_1)} \times 100 \quad \text{Eq. 5.13}$$

In comparison to the forest cover estimated from satellite data, forest cover obtained from various statistical data source presents different trends over the 40 years. Only TDHS stats contain forest cover before the 1960s, reporting rough estimates of 16.92% of forestland in 1949 and 36.31% of forest land use for the early 1960s. The percent of forest cover reported by NFI stats is 13.3% in the 1970s, which was increased to 20.37% until circa 1990. The 6th NFI conducted around 2000 shows that forest cover is 30.27%, which was increased to 34.31% in circa 2005 according to the 7th NFI statistics. Overall, forest cover increased by 158% from the 1970s to 2005, with 66% of the total increase occurred after 1990. The Sichuan statistical yearbook reports forest cover from after 1990, describing a pattern with the highest

forest cover of 39.7% in 2000, increased from 19.17% in 1990 and decreased to 28.98% in 2005. Definitions of different variables are provided in Figure 5.7.

5.5 *Discussion*

5.5.1 Differences between satellite- and NFI-based forest cover estimates

The difference between the remote sensing- and NFI-based forest cover change estimates for Sichuan mainly existed in the 1960s and circa 1975 and was reduced after 1990 (Table 5.5). Similar differences were also identified in the other provinces in China between 2000 and 2005 (Figure 5.6). For 2000, NFI reported a higher forest cover for provinces especially in north China dominated by sparse forest cover. In provinces with medium forest cover, the two data sets have similar estimates. For provinces in the south with dense forest cover such as Guangdong and Fujian, NFI did not report as much forest as remote sensing estimates. Figure 5.6 also shows that almost all provinces experienced big increases in forest cover according to two NFI reports, however, net losses were observed from the GFCC maps. The trend of forest decline was similar to what was reported by Hansen et al., (2010), which stated that the net forest increase reported by China's government could not yet be detected using remote sensing observations.

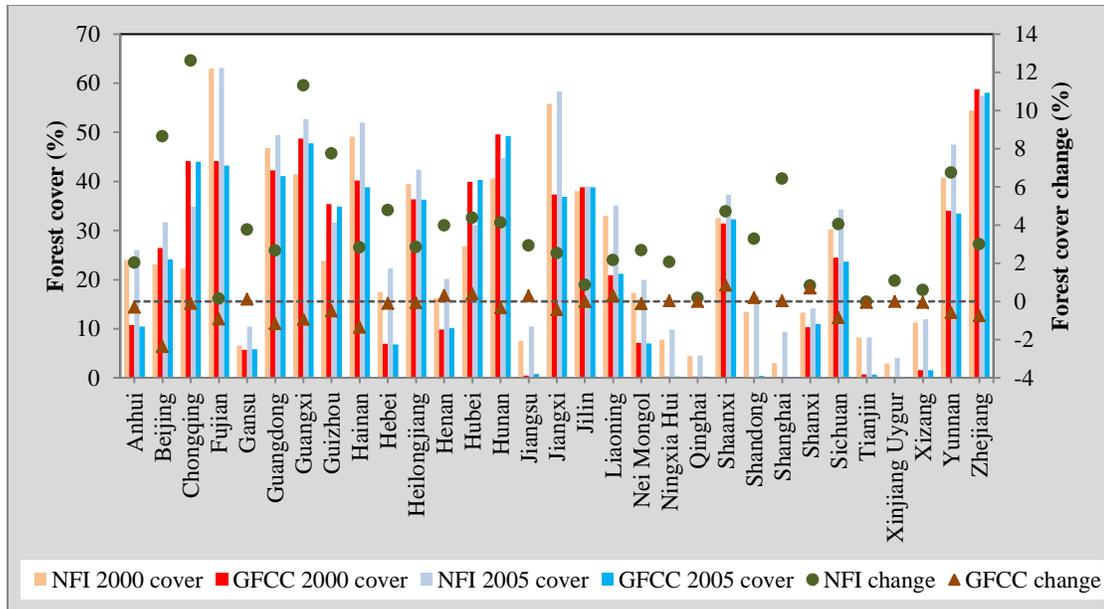


Figure 5.6 Comparison of forest cover rates for 2000 and 2005 and change rates during the five years derived from GLCF-GFCC maps and 6th and 7th NFI statistics respectively (Song et al. 2014).

5.5.2 Sources of difference in forest cover estimates

The difference on forest cover estimates from the remote sensing and forest inventory could be caused by two factors. The first factor is the mismatch of definition. In the remote sensing-based estimate, forest cover is defined as parcels with area > 1 ha and with tree cover > 30%, measuring the biophysical properties of forests. Forest cover in NFI statistics, what the remote sensing estimate was compared, is defined as the fraction of forestland, further including “special purpose” scrubs in area with annual rainfall less than 400 mm since 2003 (SFA 2003). Forestland in NFI is defined as parcels > 0.067 ha in area and comprising a canopy cover > 30%, excluding scrubland or urban forest. The definition was changed in 1994 to a minimum canopy cover of 30% reduced to 20% (Zhang and Song 2006). In addition to forestland, forest land use in NFI also includes sparse forest, afforested

land, post-disturbance forest land, and other land used for forest purposes (Figure 5.7).

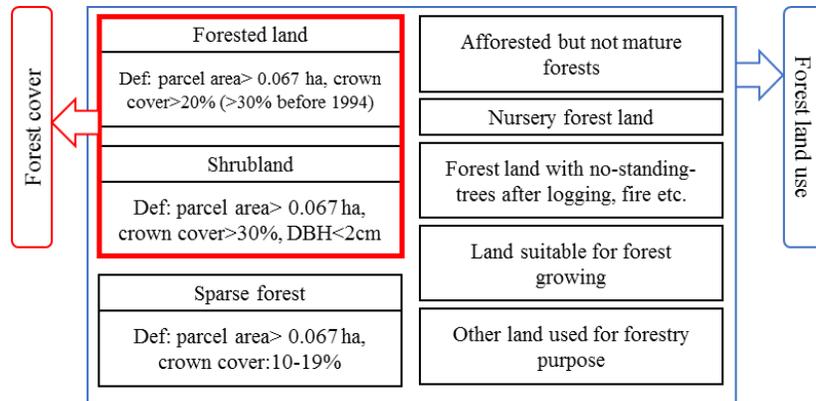


Figure 5.7 Definitions of major forest land types and key variables in forest inventory and statistics.

“Forest cover” is used for comparison with remote sensing-based estimate of forest cover.

Second, the inconsistency in method adopted in different NFIs brings more difficulties to understand the statistics. For instance, Sichuan’s first inventory in 1976 was mainly carried out by people from the non-forest sector, and was based on county-level inventory. A standardized inventory procedure was established since the 3rd NFI in the 1980s, and remote sensing technology was first introduced in NFI by 2000 (Zhang and Song 2006).

Regarding accuracy of the two data sources, consistent definition and algorithm were adopted in remote sensing-based estimates and the accuracy and uncertainty were quantified; however, uncertainty in statistics was not reported. Therefore, remote sensing-based estimates are more reliable for analyzing the long-term dynamics, especially when the field data for deriving statistics is not available.

5.5.3 Relationships of long-term forest cover change to China’s forest policies

Forest policies have multiple effects on the change of forest area and the management intensity, through the impacts of creating secure or insecure tenure system, changing incentives for forest production, affecting timber price and taxes, or reforestation/afforestation programs (Yin and Newman 1997; Zhang et al. 2000b).

Forests in China have significant spatial heterogeneity driven by both natural conditions and various socioeconomic policies (Yin and Newman 1997). Since the founding of the current government in 1949, China has undergone many social and economic changes (Table 5.6), which at times led it to fast economic development and natural resource consumption. Using the estimates of forest cover change over four decades from the 1960s in Sichuan, it is able to connect the change magnitude to the forest policies over different time periods.

Table 5.6 List of key forest and relevant policies in China from late 1950s to 2000s.

Years	Period	Key Forest Policies
1958-1962	The Great Leap Forward	Establishment of people’s commune (1958) The Second Five Year Plan (1958-1962)
1963-1977	Cultural Revolution	“Who plants, who owns” regulation (1964) National Forestry Development Plan (1971)
1978-1998	Economic Reform	Green Great Wall program (1978) Household production responsibility system (1977 – 1991) The forest ownership policy (1981) Forestry Law (1985) Forestry Reform (1992-1998)
1998-2010	National afforestation and deepened reform	Natural forest conservation program (1998-2050) Grain to green program (GTGP) (1999-) Deepened forestry reform (2003-)

Tenure change and underpriced logs before 1978

Forest cover rate in Sichuan province decreased by 13.7% from $45.19 \pm 3.11\%$ in the 1960s to $38.98 \pm 4.04\%$ in the mid-1970s. This falling trend during this period could mainly be explained by the tenure system of forest land and the timber output from forests.

From a tenure perspective, forest degradation was mainly due to the lack of confidence in tenure security caused by frequent change in forest policies in the 1960s (Liu 2001). Establishment of people's communes (1958) caused the entire nation get involved in agriculture and steel-making activities by collecting fuel from trees. This could have caused the first severe destruction to natural forestry since the founding of P.R. China (Liu 2001). In the adjustment of national economy (1961-1964) period, after the strike of forest loss during the previous period, central government carried out series of policies to recover the destroyed forests. Afforestation-based forest policy was implemented in 1964 including "who plants, who owns" regulation and providing afforestation funding. Forest ownership and management were devolved from commune to production team or production brigade, and scattered trees were assigned back to households (Liu 2001). This short recovery period may not have caused significant changes on forest cover, but at least it could cease the severe logging in late 1950s. Another reversal of policy on the house-hold ownership of forest happened in the Cultural Revolution period. In terms of forestry, forest management institutions were dissolved and experts and technicians in forestry institutions were laid off. Exploiting of forest resources was recorded and rampant timber cutting happened around the nation, at the same time, afforestation campaigns

were highly inefficient, with nominal surviving rate of only 20% (Wang et al. 2004). Until 1971, adjustments policies were carried out by drafting the National Forestry Development Plan. Rapid-growth forest planting strategy was promoted in southern provinces, and national forestry farms were established especially in northern China (Liu 2001).

From a financial perspective, the forest sector in China mainly supplied underpriced timber products to support the national economic development during the three decades before 1978 (Wang et al. 2004). Deforestation in Sichuan as a whole began in the 1950s, and in its major forested regions such as Aba Prefecture in the 1960s, when it supplied about 84% of timber output of the province, and Garzê Tibetan Autonomous Prefecture (Chinese Academy of Forestry 2016; Hayes 2013). Another example is Baoxing County, Sichuan, where the timber production was 100,000 m³ in 1960s but increased by 16,000 m³ per year during the 1970s-80s (SFA 2014).

Demand for timber during 1978 to 1998

A continuous deforestation trend from the mid-1970s to the late 1990s was observed according to the remote sensing-based estimates, composed by a sharp drop from $38.98 \pm 4.04\%$ to 28.20% during the first fifteen years and a further mild decrease to 27.73% during the next decade. The loss of one-third of forest cover during the two decades before 2000 is possibly relevant to the high demand for timber products during the country's economic reform and development.

Economic reforms were carried out around the country since 1978, followed by a reform to forestry sector in 1981 that changed the ownership of forests. In 1985,

the Forestry Law was finalized, which opened the timber market to meet the demand of timber products during the period of rapid economic development. State-owned forest companies became more autonomous with increasing controls on timber production; on the other hand, collectively-owned woodland was distributed to peasants, with entitlements to harvest of the forest land (Wang et al. 2004). To meet the demand for timber during the economic reform in the late 1970s and 1980s, annual timber harvest increased from 20 million m³ in 1950s to 63 million m³ in the 1990s. Total volume of forest harvest from the 1970s to the late 1980s was higher than the total volume of forest growth (Zhang and Song 2006), indicating the decrease of natural (mature) forest coverage and increase of newly planted young trees. Although the large-scale plantation-style forest increased the total forest cover in NFI statistics, extensive cutting of forests caused the decline of natural forests in 1995 to 30% of the total forest area in 1950s (Zhang et al. 2000a). The undetectable young trees made the remote sensing estimates show a trend of forest loss from 1975 to 2000. In Sichuan, forest destruction is most severe along rivers where logged timber can be floated downstream to the Sichuan basin, and along major highways for transportation purpose. Forest cover has also been estimated to decrease from 30% in 1950s to 14% in 1980s (Clarke 1998) in west Sichuan, which hosts most of the conifer forests.

Afforestation since 1998

The drastic demise on forest cover ceased until the late 1990s and the forest cover in Sichuan stabilized at around 27% after 2000. This sudden turn around on the deforestation trend lasting for three decades could be attributed to the national

afforestation programs and conservation activities such as logging bans in the important forest regions such as Sichuan.

Alarmed by the massive floods in China happened in 1998, six key forest conservation programs were launched to recover the environment, encompassing 97% of the nation. Further reform happened to the forest sector, which was aimed at establishing mature timber market in collective-run forests region, and increasing ecological values of forests in state-run forests. The Natural Forest Conservation Program (NFCP) that was carried out in Sichuan since 1998 was characterized with a strict logging ban in natural forest and incentive afforestation. The duty of most employment in the forest sector has shifted from logging to tree planting (Liu et al. 2008). Afforestation accounted 2.34 million ha in area in Sichuan from 1989 to 2000. Wildlife habitats, such as the panda habitat, has been recovering since the NFCP (Vina et al. 2007). Area converted to forest was 90,700 ha from cropland, and was 72,100 ha from barren hills by the Grain to Green Program (GTGP) between 1998 and 2003 (Liu et al. 2008). However, remote sensing data could not yet provide observational evidence for the continuous net increase in forest cover as reported by NFI (Hansen et al. 2010). It may take more than five years for the large-scale of the afforestation area planted in late 1990s to be observable by remote sensing data.

Although remote sensing provides continuous observations throughout the past several decades, the forest cover and change maps do not carry sufficiently fine temporal information to address issues in the change analysis compared with forest policies, primarily due to issues in data processing such as data availability and gap

filling. Thus, there is still a gap to link long-term policies with remote sensing estimates of forest cover change.

5.6 Conclusions

In this chapter, forest cover in the 1960s was quantified for the first time using satellite data, extending the Landsat record of forest cover to the 1960s in Sichuan, China. A sampling method facilitates the application of the Corona image to a regional estimate by reducing the processing time and quantifying the uncertainty. By using multi-temporal Landsat-based forest cover change products, a continuous forest cover change trend was estimated for Sichuan province. Continuous forest loss was identified with forest cover reduced by 38% from the 1960s to 2005 in Sichuan, contradicting with the assertion of continuous forest increase in NFI statistics. The remote sensing based estimations are more transparent and consistent than the government's statistics by providing the spatially explicit distribution of changes and uncertainty estimation. It is also more consistent in the forest definition and approach used for mapping.

Supplementing the forest statistics, the remote sensing observed decrease in forest coverage from the 1970s to 1990 reveals the actual impact of the concurrent forest and socioeconomic policies. Timber consumption during economic reform and the transition on forest tenure during forestry reform intensified the forest cutting during the 1970s and the 1980s, and resulted in the sharp drop of Sichuan's forest cover circa 1990. Since 1998, the Natural Forest Conservation Program and the Grain to Green Program have played important roles in conserving forest resources in Sichuan and improving the natural habitats of wildlife. Although the large-scale

afforested area is not observable in remote sensing images and an increase in forest coverage has not been detected yet, stable forest coverage after 1990 can demonstrate the positive impact of these forestry programs.

Chapter 6: Improved modeling of the size-frequency distribution of forest disturbances based on Landsat forest cover change product⁴

6.1 *Introduction*

By altering vegetation cover, structure, and composition, forest disturbance – including natural mortality, hazards, logging, and clearing for agriculture or urban expansion – impacts the water cycle, surface energy budget, carbon flux, forest structure, and habitability for biodiversity (Foley et al. 2005). Spatially continuous forest cover change data over large regions have not been used to improve the modeling of size and frequency of forest disturbances. Over large regions, disturbance regimes are characterized by intensity, frequency, and size of individual events, which can range from an individual fallen tree to large clearings of many square kilometers. The frequency distribution of disturbance area impacts the structure and composition of the remnant forest through effects of edge, area, and isolation (Haddad et al. 2015). Large disturbances tend to result in large gaps and isolated forest patches, reduce connectivity among them, and make it difficult for pollen, and seeds to be dispersed over the gaps. However, if the same amount of disturbance area is caused by small disturbance events, it will result in large numbers of small patches that create numerous forest edges. Tree mortality in tropical forests increases within the first 100 m of forest edge due to microclimate change, wind-

⁴ This chapter has been submitted to *Landscape Ecology* and is now under review.

throw, and competition with non-forest species (Puetz et al. 2014); and along with factors including disturbance severity, pre-disturbance state, and process of conversion, the spatial pattern of disturbances determines the subsequent trajectory of regrowth and succession (Asner et al. 2013; Frohking et al. 2009). Recovery of large disturbance areas is usually slower and more variable due to longer distances for seed dispersal than over small disturbance patches (Shure et al. 2006; Turner et al. 1998). Disturbance patch size also affects light availability across the resulting gaps, and hence may affect post-disturbance species diversity due to different levels of shade tolerance by different species (Denslow 1987; Dietze and Clark 2008).

Accurate characterization of disturbance regimes is needed to understand the impacts of disturbances (Asner et al. 2013; Malamud et al. 2005); specifically, models of disturbance size-frequency distribution provide a basis for prescribing past and future disturbance scenarios in simulations of climate change and land-use (Fisher et al. 2008; Hurtt et al. 2016; Moorcroft et al. 2001). To this end, great effort has recently been devoted to studying the frequency distribution of forest disturbances (Asner et al. 2013; Chambers et al. 2009; Di Vittorio et al. 2014; Espirito-Santo et al. 2014; Fisher et al. 2008). The frequency (y) of disturbances of a given size (x) is often modeled by a power law relationship:

$$y = cx^{-\alpha} \quad \text{Eq. 6.1}$$

Where x is the area of individual disturbance events, y is the frequency or probability of disturbance events of each event-size, c is a linear parameter, and α is an exponential scaling parameter (Asner et al. 2013; Chambers et al. 2009; Di Vittorio et al. 2014; Fisher et al. 2008). This model characterizes the disturbance

regime from common small disturbances (i.e., individual branch losses or treefalls) through exponential decreasing size to large disturbances such as timber harvest, forest-land conversion, landslides, wildfire, and hurricanes or other windthrow events. Average patch size decreases as α increases (Fisher et al. 2008); small α values (e.g., $\alpha = 1$) indicate frequent large disturbances, whereas large values (e.g., $\alpha = 2.5$) indicate dominance by small disturbances.

Di Vittorio et al. (2014) compared six methods for modeling the size-frequency distribution of forest mortality in the Central Amazon and concluded that a power law fit by ordinary least squares (OLS) or maximum-likelihood estimators (MLE) with binned data were the best methods to estimate the exponent. A variety of binning methods, such as binning at the original data scale (Chambers et al. 2009; Fisher et al. 2008; Lloyd et al. 2009), logarithmic binning (Di Vittorio et al. 2014; Milojevic 2010), and normalized logarithmic binning (Di Vittorio et al. 2014), have been applied to reduce the impact of heteroscedasticity on model fitting (Di Vittorio et al. 2014; Milojevic 2010). However, binning may increase the uncertainty in reconstructing the patch-size frequency distribution, especially for very large patches. Numerical frequency values have also been converted into cumulative probabilities for the same purpose (Clauaset et al. 2009; Di Vittorio et al. 2014). A truncated power-law method was sometimes introduced to fit the size frequency distribution of wildfires in order to avoid the overestimation of the frequency of large fires (Cumming 2001).

Existing methods for modeling the patch size-frequency distribution suffer from two major issues. First, the power-law model assumes that the regional driver of

disturbance is a single, scale-invariant process (Wu and Li 2006). This assumption yields large uncertainties when multiple processes are in action – as evidenced by non-constant variance, or heteroscedasticity, over the range of disturbance-patch sizes. Empirical distributions typically exhibit increasing variation with patch size (Figure 6.1), and so the power law tends to poorly represent large, rare disturbances, especially when multiple disturbance processes are in action. Second, although basing the power law relationship on binned data improves representation of regional trends, it disregards potentially informative variation contained in the non-constant variance around the model. Plotting the patch sizes at each frequency level shows that this variation may follow a gamma distribution, especially at low-frequency levels (Figure 6.1-b, -c, and -d). Our observation of distributions similar to Figure 6.1-a over all study regions led us to develop a method to characterize the “noisy” tail in lieu of smoothing out the information previously treated as statistical noise.

Improvements to existing methods should describe the underlying relationship between size and frequency while also representing very large patches at low frequency levels. In addition, there is a lack of accounting for uncertainty when using the single power law model in applications of ecosystem modeling, such as disturbance area estimation. Hence, it is necessary to document the uncertainty of the existing methods and develop new methods to characterize the size distribution for ecosystem models in order to reconstruct the disturbance pattern.

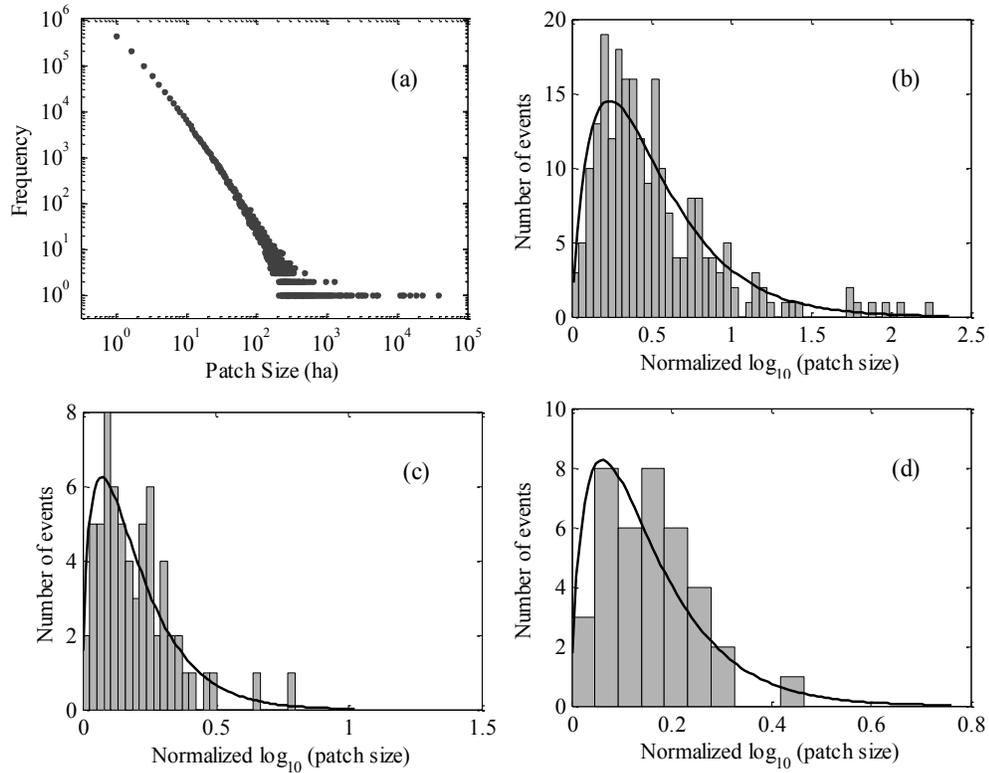


Figure 6.1 (a) Size-frequency distribution of forest disturbances in China based on Global Land Cover Facility 2000-2005 forest cover change data. b, c, and d show the horizontal slices through part (a) at frequency level of 1 ($=10^0$), 2 ($\approx 10^{0.3}$), and 3 ($\approx 10^{0.48}$) respectively. Number of bins are 50, 30 and 10 for (b), (c) and (d), respectively, but some bins include zero event. The histograms are fitted using gamma distribution (black curves in b, c and d). Note that all dots in (a) are separated when zoomed all the way into a resolution where very slight differences in patch size could be resolved.

In this chapter, I present a hierarchical method to describe the disturbance size-frequency distribution. A major goal of this method is to explicitly model the disturbance patches at each frequency level using a gamma function, further to substantially improve the characterization of the “heavy-tails” typical of disturbance size-frequency distributions in many regions (Malamud et al. 2005; Song et al. 2001). This method was applied to forest cover loss mapped over China and was compared against two existing size-distribution modeling methods.

6.2 *Data and study region*

6.2.1 Data

The GLCF-GFCC 2000-2005 (GLCF and GSFC 2014) was used to calculate disturbance size and frequency and to prototype the proposed model. The forest-loss class represents both natural and anthropogenic forest disturbances that occurred after the national forest conservation program started in circa 2000 (Song et al. 2014). Disturbance patches are discrete areas composed of connected forest-loss pixels, the area of which was used to estimate disturbance size. The forest-cover change maps are produced with a minimum mapping unit of 1 ha, which is thus the minimum disturbance size used for modeling. Frequency was defined as the number of these disturbances of a given size happened during the five-year period. In contrast to disturbances < 1 ha monitored from field plots, disturbances > 1 ha are more likely driven by human activities such as logging, and natural events like landslides, fire, and insects and have been mapped at regional, continental and global scales through satellite images (Espirito-Santo et al. 2014; Hansen et al. 2013; Kim et al. 2014; Masek et al. 2013). The size and frequency of large disturbances were related to environmental and policy factors varying across ecoregions and administrative units.

6.2.2 Study region

China's national forest inventory reported steady growth of forest area, from 12.7% in 1976 to 18.2% in 2003 (SFA 2005). However, recent satellite-based analysis revealed 3.32% of gross loss in forest cover happened in China between 2000 and 2005, and that net forest loss dominated most provinces over the period

(Song et al. 2014). Following extensive flooding in 1998, forest policies across the country were strengthened to improve forest ecosystem services and promote afforestation (Xu et al. 2006). Although plantations and natural regrowth led to increases in densely forested provinces, gross gains were accompanied by losses due to human-induced or natural disturbances (Song et al. 2014). Characterizing forest disturbances after 2000 is important for evaluating the effectiveness of these new policies, implementing forestry programs, and further improving forest management. According to the 6th National Forestry Inventory (SFA 2005), fourteen provinces with >30% forest cover were therefore selected 1) to examine the validity of the proposed method and 2) to compare the forest disturbance size-frequency distributions across regions. In addition, 27 ecoregions within three major woody biomes were selected to investigate the variation in disturbance patterns across different vegetation types (Figure 6.2).

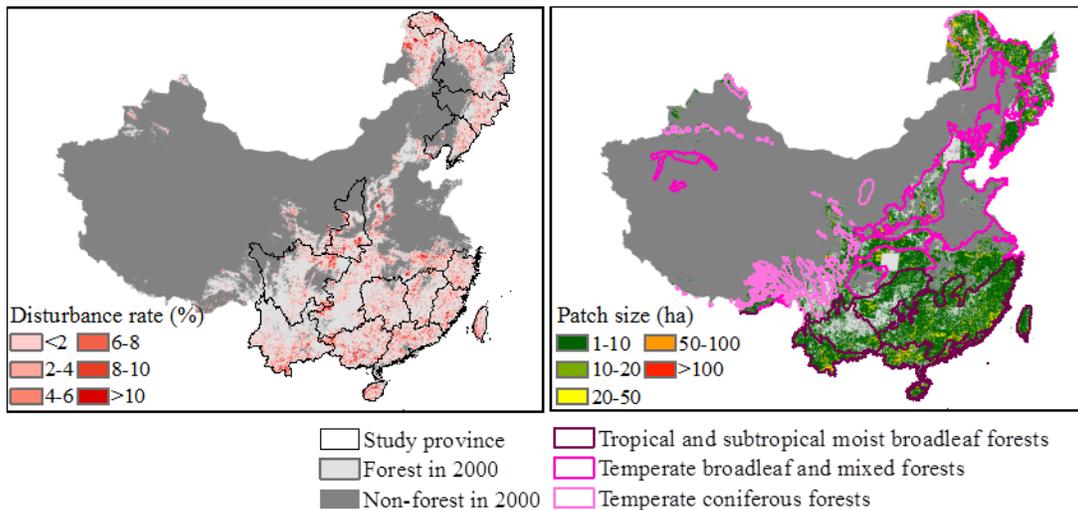


Figure 6.2 Forest disturbance rate and patch size map for China at 5-km resolution. Administrative boundary map is produced by National Geomatics Center of China, and terrestrial ecoregion map is distributed by World Wildlife Foundation.

6.3 Methodologies

In this section, the model, of the trend and then of the variation, was first described in Section 6.3.1, followed by the details of the fitting process. I then presented the methods for model assessment and sensitivity analysis in Section 6.3.2 and 6.3.3, elaborated by the metrics used to compare different methods. The workflow on how to utilize the proposed model in ecosystem modeling was then presented in Section 6.3.4. The section is ended by presenting the metrics used for regional analysis.

6.3.1 Modeling process

In this method, a relationship is first established between patch frequency and the median patch size at each frequency level. The distribution of patch sizes at each frequency level is then modeled following a gamma distribution. Finally, relationships between the parameters of the gamma distribution and patch frequency are established.

Modeling the trend in the size-frequency distribution – I used the original, non-binned, disturbance-size permutations in units of hectares (ha) as x and their frequencies as y for modeling. The frequency of disturbance-patch area decreases monotonically in the log-log domain. Using the original, non-binned data can bias estimates of the exponent towards that of medium to large disturbance patches (e.g. 10^2 - 10^3 in Figure 6.4), which are clustered and thus dominate fit. To avoid this, I extracted the medians (x_{med}) of all patch sizes at each frequency level (y) and then fit y' using x'_{med} using OLS:

$$y' = -a_1 x'_{med} + b_1 (a_1 > 0) \quad \text{Eq. 6.2}$$

where y' is the frequency after logarithmic transformation and x'_{med} is the median patch size at each frequency level after logarithmic transformation. A second-order (quadratic) polynomial was also tested, but the first-order polynomial performed sufficiently and so was retained. The coefficient a_1 in Eq. 6.2 is equivalent to the exponent α in Eq. 6.1, and b_1 is the predicted theoretical frequency after logarithmic transformation when the disturbance patch size approaches zero in a given landscape. For regions with similar values of b_1 , a large a_1 value indicates a disturbance pattern dominated by small patches and a small a_1 indicates the dominance of large disturbance patches.

Modeling the variation in the size-frequency distribution – The variation of patch sizes at each frequency level (Figure 6.1-b, -c and -d) follows the gamma distribution (Hahn and Shapiro 1994):

$$y = f(x | g_{shp}, g_{sc})$$

$$= \frac{1}{g_{sc}^{g_{shp}} \Gamma(g_{shp})} x^{g_{shp}-1} e^{-\frac{x}{g_{sc}}} \quad \text{Eq. 6.3}$$

where y is frequency, x is patch area, and two coefficients describe shape (g_{shp}) and scale (g_{sc}). A smaller shape and a larger slope describe a narrower distribution, both of which indicate a lower likelihood of large patches at a high frequency levels. The coefficients g_{shp} and g_{sc} were found to have a linear relationship with log-transformed frequency (y') and thus may be predicted using the following relationships:

$$g_{shp} = -a_2 y' + b_2 \quad \text{Eq. 6.4}$$

$$g_{sc} = -a_3 y' + b_3 \quad \text{Eq. 6.5}$$

where a_3 (or a_2) describes the change of g_{sc} (or g_{shp}) with frequency.

Regarding regions with the same parameters of Eq. 6.2 (i.e., a_1 and b_1), a large a_3 delineates a pattern with a more scattered tail at a low frequency level as a result of the occurrences of large disturbances than does a small a_3 . When the shape parameter is large, the gamma distribution approximates a normal distribution, so a smaller a_2 indicates a more significant characteristic of gamma distribution at a lower frequency level. In regions with similar estimates of a_1 and b_1 but a large a_3 , a higher a_2 value indicates more frequent occurrences of both small and very large disturbance patches.

The number of disturbance-patch sizes \hat{x}_{num} at each frequency level to be randomly generated from the gamma distribution after logarithmic conversion is predicted by:

$$\hat{x}_{num} = a_4 y' + b_4 \quad \text{Eq. 6.6}$$

Since the patch-size value retrieved from the gamma distribution naturally has a minimum value of zero, it must be adjusted to the observed range of values. The minimum value of all patch sizes at each frequency level was also found to be linearly related with the medians (x_{med}) and can be modeled as:

$$x'_{min} = a_5 x'_{med} + b_5 \quad \text{Eq. 6.7}$$

where x'_{min} is the lower bound of patch size after logarithmic transformation and x'_{med} is as denoted above. By adding the predicted minimum value to the

simulated size values from gamma distribution, the variability of patch size at each frequency can be reconstructed.

6.3.2 Model validation

In order to test the fit of the hierarchical model and compare to existing methods, I used (1) the hierarchical model to derive the best fit of patch-size frequency distribution, and fitted a power-law distribution using (2) a discrete MLE method (Clauset et al. 2009) on the original patch-size data permutations, and (3) a OLS method to the log10-log10 transformed patch-size values. The minimum patch size was set to the minimum mapping unit of 1 ha for all three methods, implying that all observed disturbances were preserved in the parameter fitting. After modeling the size-frequency distribution, the frequencies ($y_i (i = 1, \dots, k)$) of given patch-size values $x_i (i = 1, \dots, k)$ were estimated using the three methods individually. I measured the performance of each method by comparing the difference between the observed and simulated disturbance area, which was calculated as:

$$Area_{diff} (\%) = \frac{\sum_{i=1}^k x_i \cdot y_i}{\sum_{i=1}^k x_i \cdot y_{i0}} \times 100 \quad \text{Eq. 6.8}$$

where y_{i0} is the actual frequency of patch x_i derived from map.

In addition to comparing the total area estimation, I also evaluated the performance of the three methods at different patch size and frequency intervals, to avoid the case that a large patch is missed and compensated by an overestimation of small patches. Patch size was divided into five ranges: 10^0 - 10^1 , 10^1 - 10^2 , 10^2 - 10^3 , 10^3 - 10^4 , and 10^4 - 10^5 ha, and the total frequencies of all patch sizes falling in each size

interval were calculated. Frequency was likewise divided into six ranges: 10^0 - 10^1 , 10^1 - 10^2 , 10^2 - 10^3 , 10^3 - 10^4 , 10^4 - 10^5 , and 10^5 - 10^6 . The total area of patches within each frequency interval was summed. This interval-based analysis would provide insight into where the each method performs well or has large uncertainties.

6.3.3 Sensitivity analysis of model parameters and sampling intensities

The impacts of model parameters on disturbance patterns were examined through simulations of disturbance landscapes under various parameter combinations. Two pairs of parameters were examined separately, the power-law coefficients (a_1 and b_1), and the gamma-shape coefficient (a_2 and b_2), considering their important roles in describing the trend and variation of size frequency distribution. a_1 was increased from 1.9 to 2.3 at 0.1 intervals, while b_1 was increased from 3.2 to 4.2 at 0.2 intervals. a_2 was varied between 0.6 and 1.2 with increment of 0.2, and b_2 was increased in the range between 0.5 and 2.5 at 0.5 intervals. For the other parameters in the simulations, I used the model fitted parameters for Zhejiang province (Table 6.1). For each simulation, the percentage of disturbance area accounted for was then calculated by different disturbance sizes ranging from 10^0 to 10^3 , the change of which was used as a sensitivity measurement.

Because sampling is required to estimate the disturbance pattern of a given landscape when complete maps are not available (Asner et al. 2013; Di Vittorio et al. 2014), I also examined the sensitivity of the hierarchical model to sampling intensity by sub-sampling the original data. By increasing the sampling intensities from 10% to 90%, random sampling was repeated 50 times, and the variance of the estimated total

disturbance area was calculated as an indicator of sensitivity; large variance indicated sensitive model, while small variance indicated model robustness at that sampling intensity level.

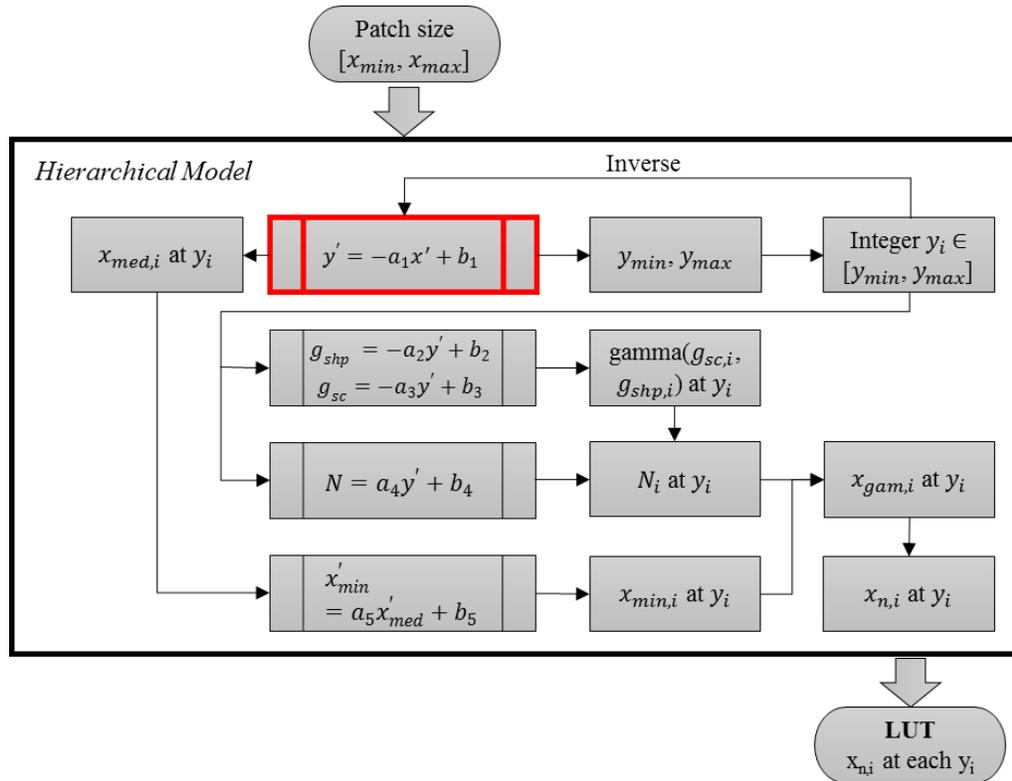


Figure 6.3 Flowchart for building a look up table for the patch size $x_{n,i}$ and frequency y_i using the hierarchical model. The range of patch size $[x_{min}, x_{max}]$ is needed as input (for the red box) to initialize the procedure and the output is a LUT. Variables with superscripts are after logarithmic transformation and the ones without superscripts are of their original value.

6.3.4 Model utilization

In order to utilize the hierarchical model to predict the frequency of disturbance patches, a look up table (LUT) method is recommended. As demonstrated in Figure 6.3, the range $[x_{min}, x_{max}]$ of disturbance patch size values of a given landscape is the only input required for a prescribed model with known model

parameters (a_i, b_i) . Following the procedure described in Section 6.3.1, the patch sizes that fall in the given range along with the corresponding frequencies will be output in the fashion of a LUT. Therefore, the frequency of any desired patch size within the range could be derived by looking up for the closest patch size value.

6.3.5 Regional disturbance analysis

Although the hierarchical model is defined by ten parameters $(a_1, \dots, a_5, b_1, \dots, b_5)$, a_1 (equivalent to the power-law exponent α) indicates the dominant disturbance size and defines the overall trend of the size-frequency distribution, while the others describe the variation. The power-law exponent has been often used to characterize the disturbance regime, such as fire regime, by describing the ratio of large to small fires in a given landscape (Malamud et al. 2005). For comparison analysis across provinces and ecoregions, I thus used the model retrieved α (i.e. a_1) to compare the trends of forest disturbance distributions across study regions. For modeling purpose, all ten parameters are needed to prescribe the hierarchical model.

6.4 *Results*

6.4.1 Model assessment

From each of the three models, the simulated size-frequency distribution of forest disturbances in China from 2000 to 2005 is plotted on top of the observed distribution in Figure 6.4. The hierarchical method estimated a power-law exponent of 2.28 for China, which was slightly lower than the MLE derived exponent of 2.47.

In contrast, the OLS approach underestimated the exponent to be 1.61. The power law fit by MLE simulated data with a bias toward small patches, and the one fit by OLS simulated data underestimated the frequency of small patches. Compared to the two existing models, the hierarchical model estimated patch-size frequencies more accurately over the range of patch sizes, as well as their variation at low frequencies (Figure 6.4).

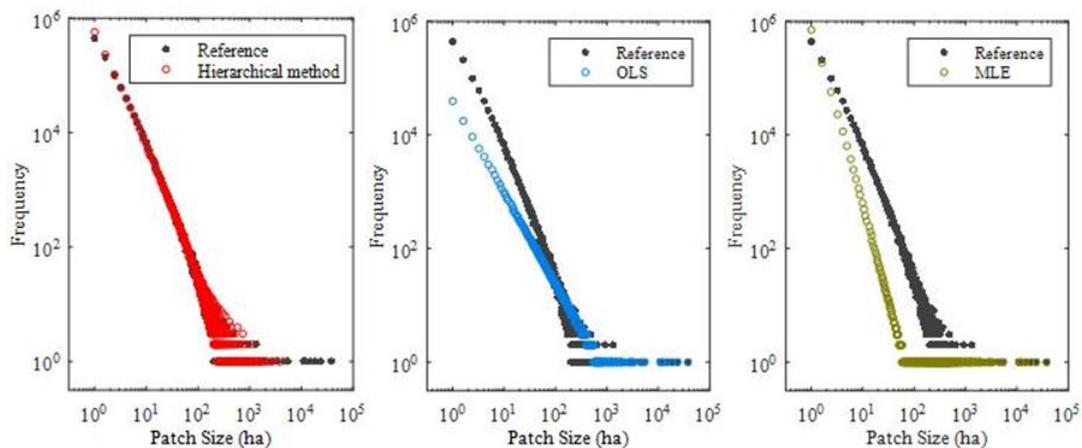


Figure 6.4 Comparison of forest-disturbance patterns in China from 2000-2005 simulated using the hierarchical method (left), simple ordinary least squares (OLS), (center), and maximum-likelihood estimation (MLE) (right). Simulated frequencies less than one were forced to one, and those larger than one were rounded to the closest integer value.

Regional disturbance-area estimates

The hierarchical method estimated regional disturbance area more accurately than the simple MLE and OLS fitted power laws across the provinces (Figure 6.5).

The MLE and OLS approaches consistently underestimated total disturbance area by 20-40%, due to their misrepresentation of the frequency of large patches, while the hierarchical method estimated area within $\pm 5\%$ against the reference. Large errors of -10 and -15% for Heilongjiang and Sichuan provinces, respectively, were probably due to the occurrence of very large patches located beyond the assumed range of the

gamma distribution. This result indicates that when total disturbance area or disturbance rate is the only known knowledge for a landscape where experienced multiple disturbance types, our hierarchical method has highest accuracy in reconstructing the size-frequency distribution, meanwhile, exports the closest approximation of total area.

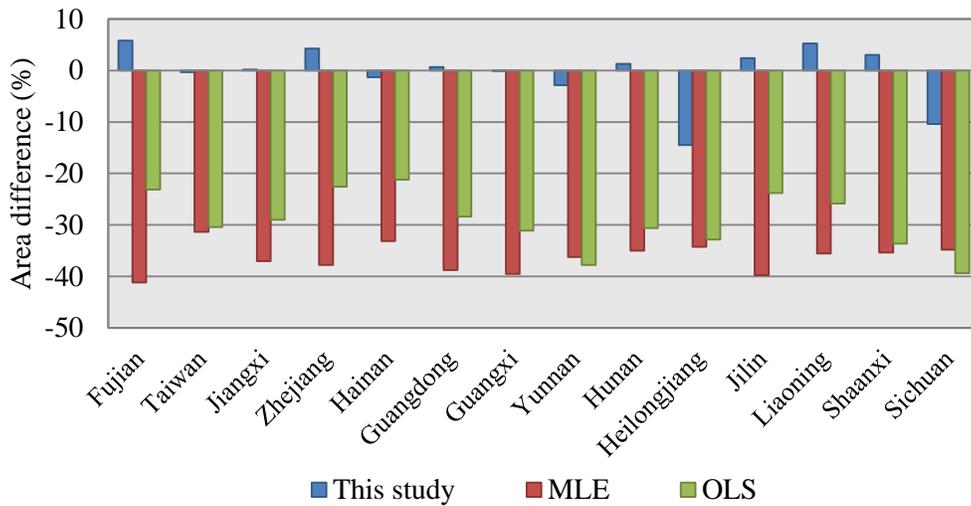


Figure 6.5 The difference between the estimated total disturbance area using different modeling methods separately and the true total area for selected provinces (or regions) in China.

Interval area estimate

The total disturbance area accounted for by patch sizes with frequencies in each interval (10^0 - 10^1 ha, 10^1 - 10^2 ha, 10^2 - 10^3 ha, 10^3 - 10^4 ha, 10^4 - 10^5 ha, and 10^5 - 10^6 ha) was calculated for selected provinces (Figure 6.6). The hierarchical method is more accurately and consistently than both the MLE and OLS fitted power laws. The hierarchical method mostly coincided with the actual area of disturbances in the 10^1 - 10^2 ha, 10^2 - 10^3 ha, and 10^3 - 10^4 ha intervals, indicating that, at moderate patch-size frequencies, the gamma distribution was suitable for accurately estimating patch area and frequency. At intervals of 10^4 - 10^5 ha and larger, where most of the large patches

occurred, estimates of total area occasionally deviated from the actual values by -40% – indicating that, although the gamma distribution was good at simulating the frequency distribution over a wide range of patch sizes, it sometimes could not simulate the extreme tails of some distributions (e.g. Figure 6.1-b). This might cause underestimation at the 10^0 - 10^1 frequency level. The 10^4 - 10^5 level was another interval where the maximum difference between the predicted and actual area could reach -25%. The major reason was error in the frequency estimate, which will be discussed in the following section. However, the simple MLE and OLS methods both performed worse than the hierarchical method. The OLS significantly underestimated the total area in the 10^2 - 10^3 ha level and above due to its monotonic decreasing trend. MLE also underestimated the disturbance area at middle frequency levels while the accuracy was improved at the 10^0 - 10^1 interval, which benefited from minimum x value being set to the actual minimum patch size value of 1 ha in MLE method.

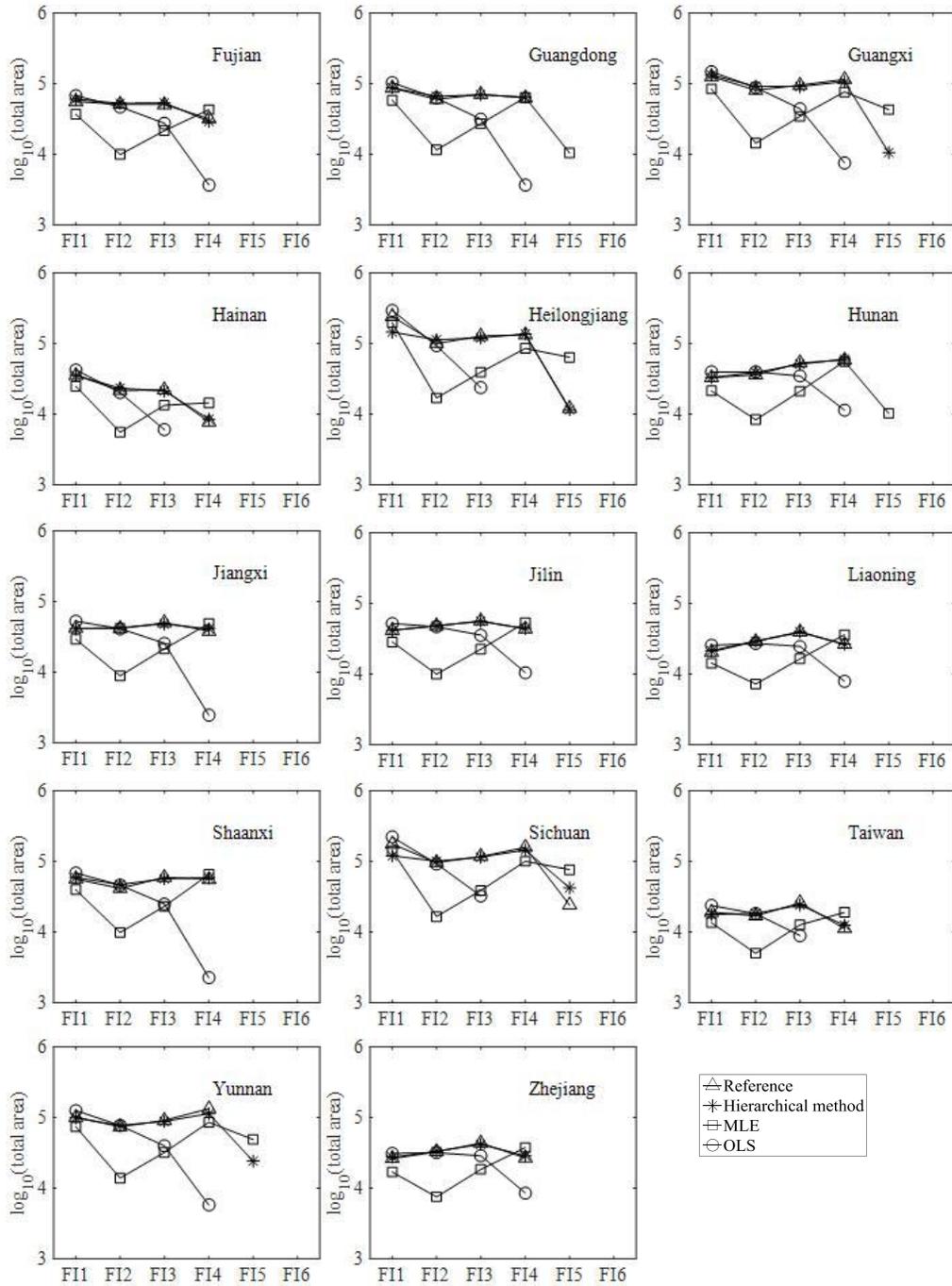


Figure 6.6 Model accuracy assessment measured by total area estimate at different frequency intervals, F11 (10^0 - 10^1), F12 (10^1 - 10^2), F13 (10^2 - 10^3), F14 (10^3 - 10^4), F15 (10^4 - 10^5), and F16 (10^5 - 10^6) for selected provinces (or regions).

Interval frequency estimate

The frequency of patches in each size interval, 10^0 - 10^1 , 10^1 - 10^2 , 10^2 - 10^3 , 10^3 - 10^4 , and 10^4 - 10^5 ha as defined above, was calculated for selected provinces (Figure 6.7). The hierarchical method resulted in the smallest deviations from the reference data, although all three methods captured the overall trend. The hierarchical method were much more accurate at 10^0 - 10^1 , 10^1 - 10^2 , 10^2 - 10^3 ha size intervals, with maximum differences of +17%, -4%, and +9% respectively. OLS underestimated the frequency for the 10^0 - 10^1 and 10^1 - 10^2 levels by up to 90% and 75%, respectively, while MLE underestimated the frequency for the 10^1 - 10^2 and 10^2 - 10^3 levels by 95% and 78%, respectively. Occasionally, both methods had lower errors than did the hierarchical method for the 10^3 - 10^4 level and above because of their frequency estimation for each input size value. The limitation of the hierarchical method at the 10^3 - 10^4 level and above was because the large patches were sometimes not adequately randomized by the gamma distribution.

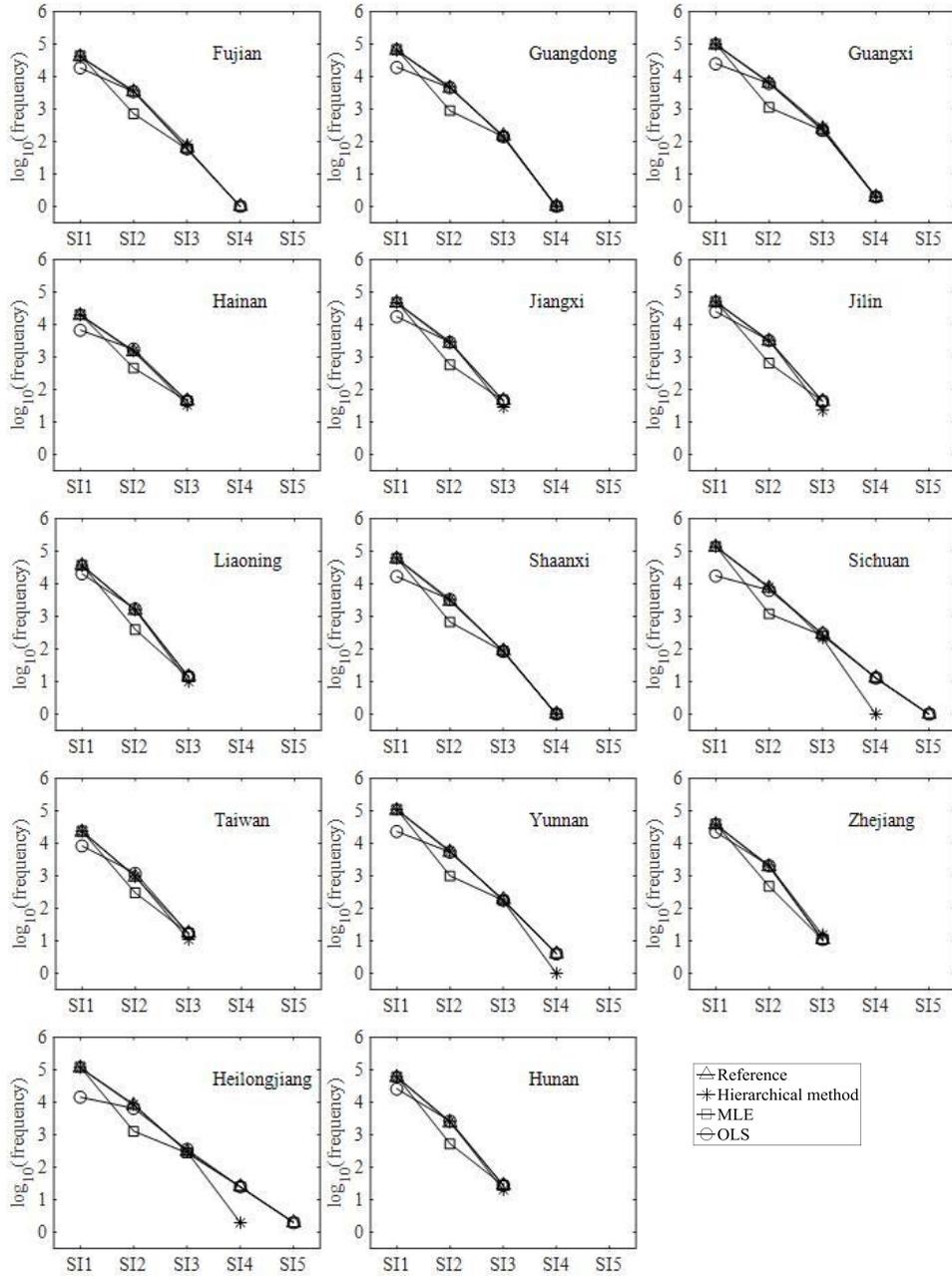


Figure 6.7 Model accuracy assessment measured by frequency estimate at different patch-size intervals: SI1 (10^0 - 10^1 ha), SI2 (10^1 - 10^2 ha), SI3 (10^2 - 10^3 ha), SI4 (10^3 - 10^4 ha), and SI5 (10^4 - 10^5 ha), for selected provinces (or regions).

Sensitivity analysis

Simulation results demonstrated that the power-law coefficients, a_1 and b_1 , and the gamma-shape coefficients, determined by a_2 and b_2 , had obvious impacts on

the size frequency distribution. When solely increasing b_1 , the percentage of the total disturbance accounted for by patches larger than $10^{0.5}$ ha increased rapidly from 0 to around 25%, indicating a landscape dominated by large disturbances. In contrast, when solely increasing a_1 , the dominant disturbance size decreased to 1 ha, and the landscape was fragmented with many small gaps (Figure 6.8).

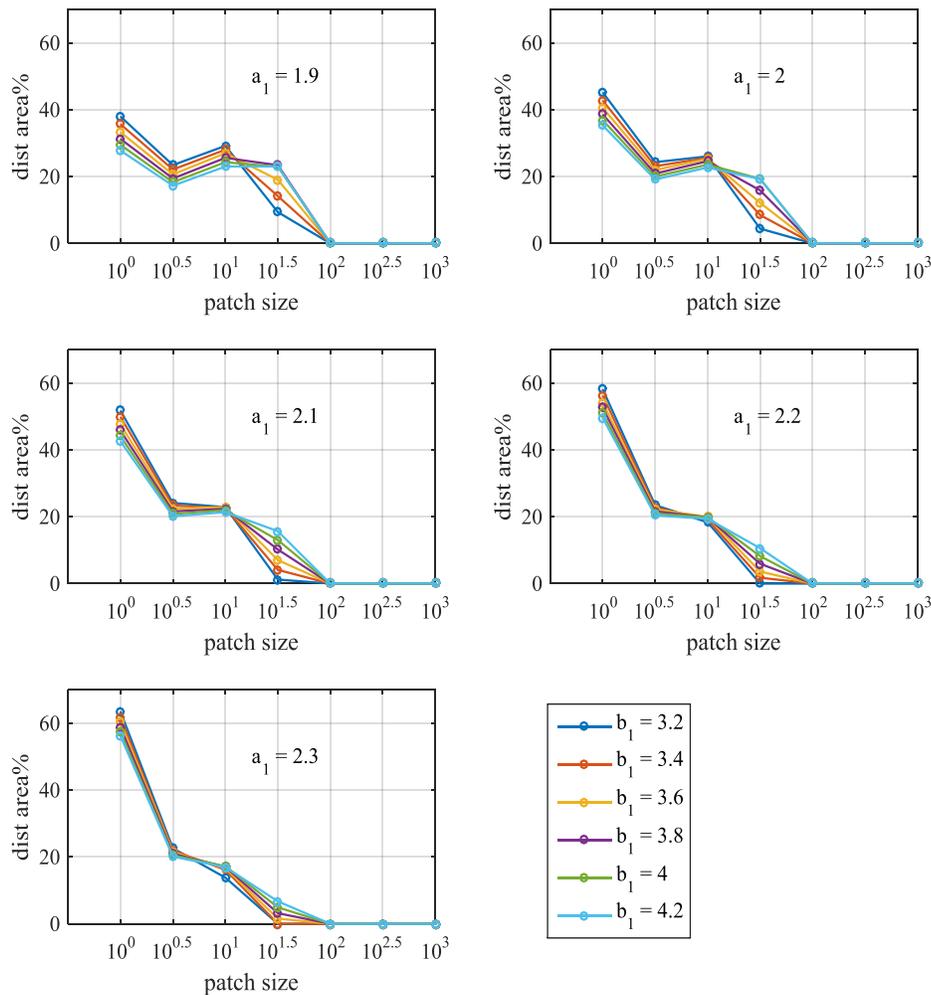


Figure 6.8 The percentage of total disturbance area accounted for by different size of patches, simulated by increasing two parameters of the size-frequency distribution, a_1 from 1.9 to 2.3 with an interval of 0.1 and b_1 from 3.2 to 4.2 with a step of 0.2 respectively.

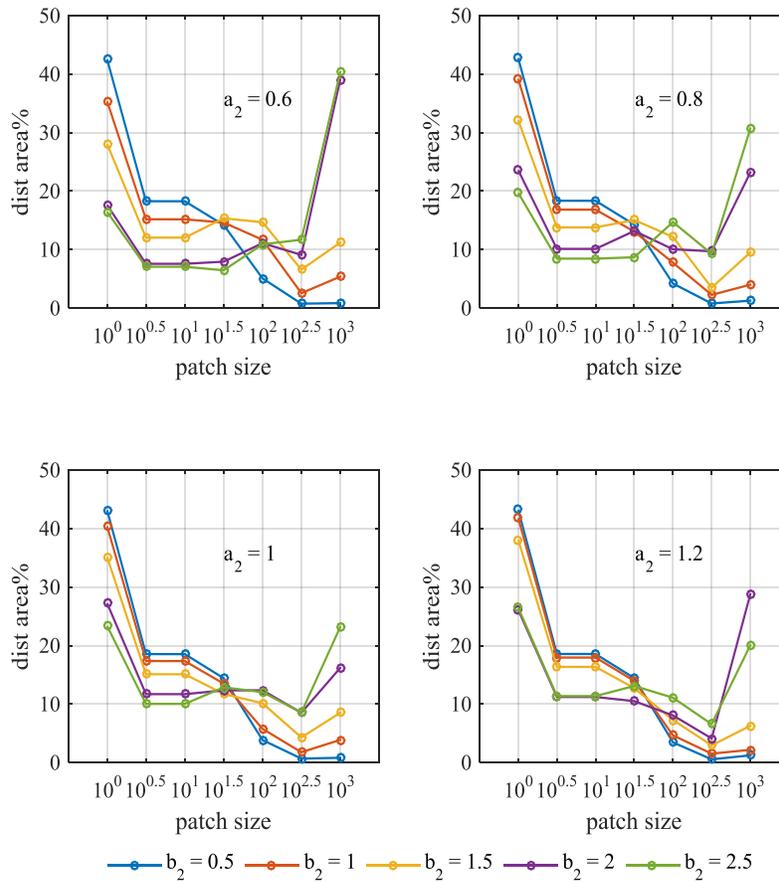


Figure 6.9 Simulation of the percentage of total disturbance area accounted for by patches when varying gamma distribution parameters in size-frequency modeling. The horizontal axis is disturbance and the vertical axis is the percentage of disturbance area accounted for by disturbances within each size interval.

On the other hand, the percentage of the total disturbance area accounted for by small disturbances (1 ha) decreased and was eventually replaced by large disturbances ($>10^3$ ha) when only increasing b_2 from 0.5 to 2.5 (Figure 6.9). The intercept referred to the shape at a frequency of 1, and both spread and likelihood of large patches increased with shape. In contrast, an increase of the absolute value of a_2 caused a slight drop in the percentage of the total area composed of larger patches and a consequent increase in the contribution of smaller patches. The larger the

absolute value of a_2 , the faster the decrease in the gamma-shape in relation to frequency.

Although the hierarchical method successfully modeled the size frequency distribution, it was established based on complete maps of the entire study region. However, when the proportion of sampling was around 10-20%, the hierarchical method resulted in errors of 30-50% in total disturbance area (Figure 6.10). The difference was reduced to <10% when the sampling proportion was >30%. This suggests that >30% of the disturbance area must be sampled in order to reconstruct a reliable size-frequency relationship using the hierarchical method.

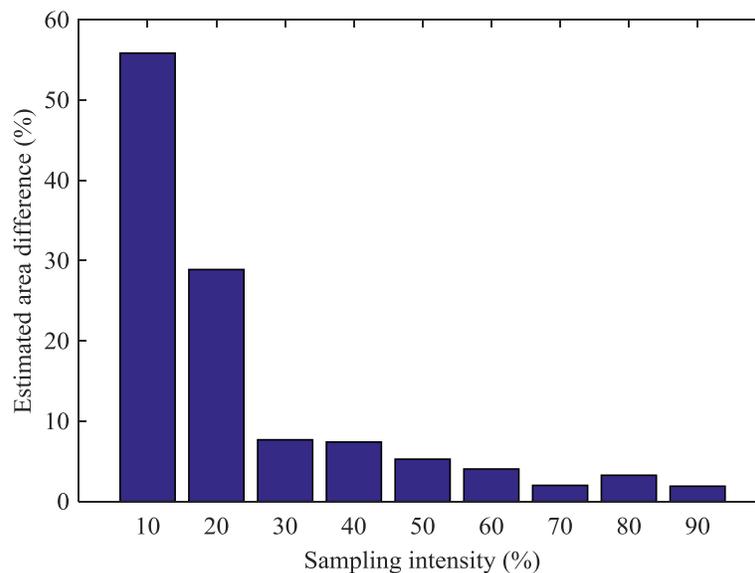


Figure 6.10 Differences in total estimated area using proposed, hierarchical method when increasing sampling proportion from 10 to 90% at 10% intervals.

6.4.2 Spatial pattern of forest disturbance in China

Disturbance pattern of China

In general, the forest disturbance regime in China from 2000 to 2005 was dominated by small events. Nationally, the slope of the log-log relationship between forest-disturbance area and frequency, α , was estimated to be 2.275 using the hierarchical method. Previous studies have suggested an α value of 2.0 to be the boundary between forest landscapes dominated by small ($\alpha > 2.0$) and large ($\alpha < 2.0$) disturbances (Asner et al. 2013; Fisher et al. 2008). Across the country, 10.99% of the total disturbed area was in sizes of 1 ha or less, while 40.10% came from 1- to 10-ha disturbances, 33.62% came from 10- to 100-ha disturbances, and 9.56% came from 100- to 1000-ha disturbances. Although large disturbances (>1000 ha) were rare, they comprised 5.73% of the total loss area.

Disturbance pattern across ecoregions

Across different forest types, Jian Nan subtropical evergreen forests in southern China experienced the most extensive loss in area (~0.7 million ha) from 2000 to 2005, followed by Manchurian mixed forests of the temperate forest region in the northeast of China with $> 500,000$ ha lost (Figure 6.11-a). The estimated power law exponent was > 2 for most ecoregions (Figure 6.11-b), indicating that small disturbances dominated the forest landscape and that large disturbances were relatively rare. However, the disturbance regimes of most ecoregions in central and western China, where vegetation cover is comparatively sparse, were dominated by large events ($\alpha < 2$). Tropical forest ecoregions in southern China generally had larger exponents ($\alpha > 2.2$) than temperate forests in the north ($\alpha = 2.0-2.2$), except the three subtropical evergreen forests near the southern boundary of China.

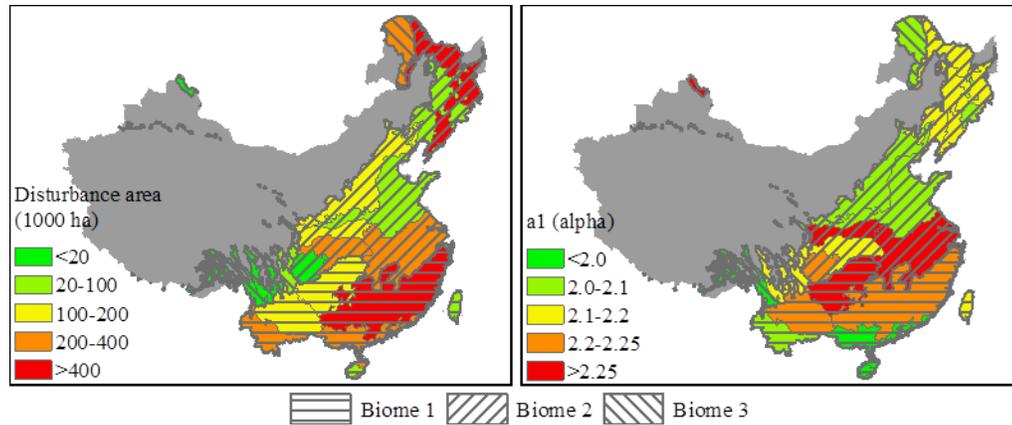


Figure 6.11 Forest disturbance areas (a) and the α estimates (b) for terrestrial forest ecoregions in China. Biome 1: tropical and subtropical moist broadleaf forests; Biome 2: temperate broadleaf and mixed forests; Biome 3: temperate coniferous forests.

Disturbance pattern across administrative provinces

Among the provinces, Heilongjiang and Sichuan experienced the highest rates of forest-cover loss between 2000 and 2005, followed by Guangxi, Yunnan, Guangdong, and Fujian provinces in the south of China (Table 6.1, Figure 6.12-a). Other provinces with dense forest cover (e.g., Jilin, Zhejiang, and Liaoning) each lost approximately 0.1 million ha of forest cover as well. However, α did not change with the variation in forest loss rates and ranged from 1.8 to 2.5 (Figure 6.12-b). Three northern provinces, Heilongjiang, Jilin, and Liaoning, had α values from 2.1 to 2.3, generally smaller than that of the central and southern provinces e.g. Shaanxi, Sichuan, and Yunnan with α values of 2.2-2.3, and Hunan with α of 2.45. Zhejiang and Taiwan also had large α estimates of > 2.2 . Guangdong, Fujian and Hainan provinces had the lowest α estimates from 1.98 to 2.07.

Table 6.1 Model parameters of size-frequency distribution for selected provinces in China. They are listed for reference of the value range of model parameters at regional scale.

Province	a_1	b_1	a_2	b_2	a_3	b_3	a_4	b_4	a_5	b_5
Fujian	1.98	3.59	0.96	1.89	0.07	0.18	-0.94	1.92	0.71	0.27
Guangdong	2.07	3.79	1.22	2.27	0.06	0.17	-1.10	2.15	0.63	0.37
Guangxi	2.09	3.96	1.24	2.45	0.05	0.17	-1.02	2.14	0.66	0.35
Hainan	1.99	3.24	1.15	2.05	0.06	0.18	-1.00	1.91	0.65	0.28
Heilongjiang	2.08	4.05	1.07	2.17	0.07	0.18	-1.09	2.25	0.65	0.38
Hunan	2.29	3.83	1.08	2.01	0.04	0.15	-0.97	1.93	0.66	0.32
Jiangxi	2.13	3.68	0.96	1.89	0.05	0.15	-0.96	1.93	0.71	0.27
Jilin	2.11	3.71	0.80	1.77	0.05	0.13	-0.79	1.77	0.78	0.19
Liaoning	2.22	3.62	0.79	1.59	0.06	0.16	-0.89	1.78	0.70	0.26
Shaanxi	2.19	3.80	1.22	2.24	0.09	0.20	-1.10	2.07	0.57	0.44
Sichuan	2.19	4.15	0.82	1.82	0.07	0.18	-0.95	2.05	0.70	0.31
Taiwan	2.24	3.41	0.67	1.36	0.10	0.21	-0.94	1.78	0.70	0.24
Yunnan	2.21	4.04	1.26	2.42	0.07	0.16	-1.05	2.14	0.63	0.39
Zhejiang	2.17	3.62	1.03	1.94	0.05	0.15	-0.92	1.82	0.67	0.30

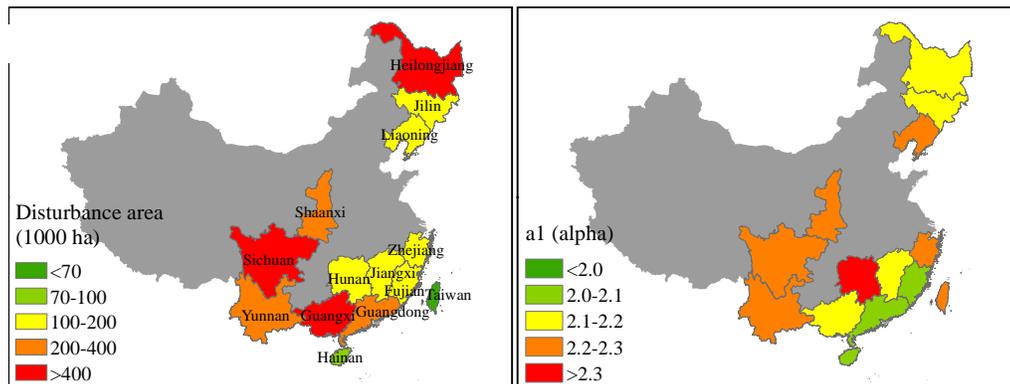


Figure 6.12 Forest disturbance areas (left) and the estimated α (right) for selected provinces in China.

6.5 Discussion

The relationship between disturbance size and frequency is key to understanding the disturbance regime. Given the synoptic coverage of remote sensing products at large regions, models of the size-frequency distribution of forest

disturbances should describe the regional relationship while also representing very large patches at low frequency levels. Developed from the general method of fitting size-frequency distribution with a power-law function, I is able to interpret more of the information in the original data and quantitatively describe the distribution by developing a hierarchical model. The power law exponent was estimated by extracting the median value of sizes at each frequency level; and the distribution of patch sizes at lower frequencies, usually considered noise, was kept and modeled to supplement the simple power law function. The model is “simple”, since all model parameters are able to be derived from the frequency level, but also accurate in estimating the total disturbance area and frequency across the region.

The estimated exponent (α) of 2.275 indicates that between 2000 and 2005, forest in China was dominated by small disturbances with occasional occurrence of large disturbances. The higher exponent in the southern ecoregions revealed that larger disturbances dominated the north while smaller disturbances dominated the southern ecoregions, which could be related to environmental factors, such as topography, geology, and soils, and socioeconomic factors such as population and forest management types. It is speculated that the collective forest management and complex terrain in South China contributed to this disturbance pattern. The deciduous and coniferous forests managed by State forest farms in Northeast China are mainly for industrial timber production, which leads to the harvest over large areas of forest land. Forest fires are more frequent in the north than in the south (Lu et al. 2006), potentially contributing to the comparatively large disturbances in the north. Population density is another possible driving force of different disturbance patterns

in the south and northeast forests; the higher population density in the south requires more residential area in forested areas, causing small disturbances and reducing forest carbon storage (Wang et al. 2001). Among provinces, the different disturbance patterns are also caused by political factors, especially by the specific implementation of forestry policies and management activities by local governments and forestry bureaus.

Regarding the disturbance data, the estimates of model parameters are expected to be further improved if a dataset is available at improved spatial and temporal resolutions, and with detailed disturbance types. Partial samples of plot data can lead to biases toward biomass gains when the frequency of large-scale disturbances increases (Di Vittorio et al. 2014; Fisher et al. 2008), while disturbance maps covering entire regions can potentially avoid these errors and uncertainties. High-resolution, continental and global maps of forest cover and change (Hansen et al. 2013; Huang et al. 2010a; Kennedy et al. 2010; Kim et al. 2014; Sexton et al. 2013a) enable quantitative understanding of forest disturbances over large areas (Townshend et al. 2012). Multi-temporal satellite remote sensing can detect losses of tree canopy cover that result in significant deviations of the canopy's reflectance over time. Further efforts should thus be devoted to assess alternative forest disturbance data for the size-frequency distribution modeling and to investigate the driving forces of different disturbances distributions. For example, the model could be extrapolated to estimate the frequency of patch sizes $< 1\text{ha}$ if including multiple data sources, such as plot data and LiDAR (Asner et al. 2013; Espirito-Santo et al. 2014). For regional studies, disturbance data with higher temporal resolution, such as yearly data, should

be collected to investigate the size frequency distribution and its dynamic through time.

Overall, the modeled patch-size frequency distribution is useful for studying forest disturbance patterns over large regions to investigate the impacts of environmental and forest policy on forest disturbances. The application of the size frequency distribution model in ecosystem modeling will be greatly enhanced by having individual size-frequency models for different disturbance types, such as insects, fire, and logging.

6.6 *Conclusions*

Variation around the widely used simple power-law relationship can be modeled and interpreted to better characterize the size-frequency distribution of forest disturbances. A hierarchical method incorporating five linear relationships to describe the trend and variation of disturbance sizes improved fit and interpretation of the power law and outperformed current methods in reconstructing forest disturbance patterns, especially at low frequency levels. All model parameters were able to be derived from a single variable, the frequency level. The hierarchical model significantly improved representation of the size-frequency distribution and accurately estimates total area and frequency of disturbances. Applying the hierarchical method across China and a selection of provinces therein revealed different disturbance patterns that coincided with spatial variation in forest management and natural disturbance patterns across the regions. In general, large disturbances (>10 ha) were rare, and small disturbances (<10 ha) dominated the forest landscape. Forest disturbances in the southern provinces were typically smaller than

in the north – except Guangdong, Fujian and Hainan provinces, which were disproportionately dominated by large disturbance patches. This hierarchical approach can be applied to maps of any land-cover or disturbance type and at any time period to fulfill the purpose of better characterizing and understanding size-frequency characteristics of ecosystem disturbance regimes.

Chapter 7: Conclusion

The studies presented in this dissertation addressed a range of issues towards improving forest cover change characterization. Original contributions include demonstrating the usefulness of Corona data for extending satellite-based forest cover change mapping back to pre-Landsat years in the 1960s, developing a semi-automated registration method to achieve efficient geometric registration of Corona data for large area forest cover assessment, quantifying forest cover change in Sichuan, China from the 1960s to 2005, and developing a new patch size-frequency method for improving the representation of in ecosystem and other spatially explicit models forest disturbance. The developed approaches provide potential solutions to quantify the long-term forest cover change, and hence have significant impact on understanding the interaction between the changes in ecosystem and dynamics in social environment. This final chapter briefly summarizes the key findings from each chapter, presents the significance and implications of the findings, and suggests several essential aspects that could be explored in future research.

7.1 Major findings of the dissertation

After briefly reviewing the Landsat-based forest cover change mapping and proposing to extend forest cover change record back to the 1960s using Corona data in Chapter 2, the issues hindering the usefulness of Corona data were addressed in Chapter 3. It was demonstrated that a combined use texture metrics and the support vector machine allowed forest mapping with accuracies of up to 95% using Corona data. Then, a semi-automated georegistration procedure, including a modified

collinear function and the automated SIFT feature selection and matching method, was developed in Chapter 4. The proposed method implemented a rapid registration with error of up to 100 m for Corona image. The proposed classification and registration methods enabled the forest/non-forest mapping for the 1960s and the Landsat-based forest cover record was successfully extended by a decade.

In Chapter 5, forest cover in the 1960s in Sichuan Province, a major forested province in China was estimated using the developed methods. Combined with a Landsat-based forest cover products, the results revealed that forest cover in Sichuan was reduced by 38% during the forty years from 45.19% in the 1960s to 28.91% by 1990 and then stayed relatively stable until 2005. Although the derived estimates contradicted the trend reported by the inventory data, they were shown to be more reliable and transparent than the government's forest statistics as a consistent definition and method was adopted and the uncertainty of the forest cover for each epoch was provided. This observed decline in forest cover over forty-year period provides evidence showing the impact of the changed forest policy during drastic socioeconomic transitions to forest. The turning point between sharp decreases before 1990 and the stable period after 1990 could possibly be explained by the decline in deforestation and the increase in afforested area during the afforestation and conservation programs since late 1990s in China, and likely reflected the transitions in forest policies from focuses on timber production to forest conservation.

Representation of the patch-size frequency distribution of forest disturbances was improved in Chapter 6 by developing a hierarchical method. The method was more accurate in representing both the major trend and patch size at different

frequency levels than the existing methods and had smaller errors in the estimation of total disturbance area as well. By applying the developed method across China, it was found that although the total area of forest loss in the southern ecoregions were similar to that in the north, the dominant disturbance patch in the southern forests was smaller than that in the north.

7.2 *Significance and implications of this dissertation*

The first significance of this dissertation is the development of solutions to extend satellite-based forest cover change mapping back to pre-Landsat years in the 1960s using the historical satellite data – Corona. Maps of historical forest cover provide critical baseline for monitoring the change in Earth's forest. The derived long-term (50 years) trend of forest change is necessary for assessing the impact of forest change, understanding the natural and anthropogenic causes of change, evaluating the effectiveness of conservation policies.

The second contribution of this research is that it advances the understanding of the trends of forest cover change since the founding of the current government of China in 1949. The results from Chapter 5 presented an overall opposite trends of forest change contradicting the statistics reported by inventory data, showing that 38% of forest was lost during the four decades since the 1960s. Besides the opposing estimates, my research is distinguished from the statistics as 1) it adopted a spatially and temporally consistent definition of forest, enabling a direct comparison of forest cover among epochs together with known uncertainties and 2) it used publicly available datasets to estimate the forest cover, enabling the expansion to other regions and spatial analysis. Thus, it is indicated that satellite data should be increasingly

used in evaluating the effectiveness of forest policies because the satellite-based mapping is advanced with the use of a consistent forest definition and a consistent method in monitoring the forest cover change over a long-term period. Enabled by the first significance, the fifty-year forest cover change map provides opportunity to understand the drivers of forest change and to evaluate the effectiveness of forest policies in various regions under different socioeconomic contexts around the world, as the understanding of the relationship between the long-term change of forest and the its driver is particularly hindered by the lack of accurate forest estimate in the early epochs.

The last contribution of this dissertation is that it further improved the characterization of forest cover change, in particular the patch size distribution of forest disturbance, for ecosystem and other spatially explicit models. With ever increasing advances in forest cover change maps derived from remote sensing imageries, the capacity to simulate vegetation dynamics using ecosystem models could be improved by accurate measurement of change area and representation of the geolocation of forest change at fine resolution. The patch-size distribution modeling method developed in the last research chapter inherited the advances of forest cover change maps derived from fine resolution satellite data. It improved 1) the static representation of forest disturbance by more accurately estimated the total disturbance area over a long-term period, and 2) the spatial representation of disturbance by allocating an appropriate amount of disturbances to each patch size level, hence the capacities of spatially-explicit simulation of the vegetation dynamics and the resultant biodiversity and carbon dynamics.

7.3 *Implications for future research*

A long-term forest cover change record is desired to cover the entire land surface, while both opportunities and difficulties rely on the successful processing of Corona data. Enabled by the global coverage and repeated acquisitions, application of Corona data on a continental or global scale is possible to establish the baseline of global forest cover in the 1960s. Moving forward from the registration and classification methods for Corona data developed in this dissertation, there are still challenges, particularly regarding image registration, that need to be overcome in order to achieve the goal of mapping global forest in the 1960s. There are three directions that could be explored. First, the image registration method needs to be further automated, especially to address issues in a pseudo match of tie points, and in an uneven distribution of tie points. More accurate information on flight location and gesture is also very important for accurate registration. Second, the reference image at an earlier date, such as the Landsat Multispectral Scanner (MSS)-based GLS1975 dataset, can be used, to avoid a pseudo match of tie points caused by land cover change. Third, a sub-pixel level registration accuracy should be obtained in order to derive a pixel-to-pixel change mapping at Landsat resolution. Without accurate sub-pixel level registration accuracy, the pixel-to-pixel change detection is only possible at a resolution of hundreds of meters.

The remote sensing-based estimate provides a more reliable and transparent quantification of forest cover change in China than the government-reported forest statistics, as a consistent forest definition was adopted and uncertainty measurements were provided. However, the value of the continuous forest inventory from the 1970s

is not denied. Forest inventory often provides the measurements of multiple variables, such as tree height, volume, age, which are collected at plot scale, complementing remote sensing. The difference between the remote sensing-based estimates and statistics could be explained with the availability of plot data and a more comprehensive estimate of long-term changes could be derived (Tomppo et al. 2008). The inclusion of plot data could also make the remote sensing-based estimate more comparable with the government report by capturing the changes, such as the gain in immature forest, which are beyond the biophysical definition of forest increase (Sexton et al. 2016). Although the strict data policy in China poses a big challenge in accessing these plot data, it is worth to put more effort in accessing the plot data.

Another important application of quantifying the historical baseline of forest cover and the long-term change is to understand the cause of the changes. Forest change can be caused by both socioeconomic factors, such as forest policies and economic growth, as well as the interaction with biophysical factors such as wildfires, droughts and floods. In this dissertation, Sichuan province was subject to the analysis of forest cover change from the 1960s to 2005, and the effects of policies were qualitatively evaluated. Inclusion of more provinces will enable the evaluation of the effectiveness of forest policies and the spatial variations. A quantitative approach linking the forest cover change rate and the socioeconomic factors is particularly desired, which brings another challenge of collecting long-term socioeconomic data, such as policies and their status of enforcement, starting from the tumultuous period of the 1950s-60s. At a broader spatial scale, many possible changes happened during the 1960s around the world will be verified with substantial evidence from remote

sensing observations. The spatial pattern of forest loss and gain could also be mapped, which would reveal different forest cover change trends under various social and economic environments among countries after World War II.

When predicting future forest dynamics and its impact on carbon balance using ecosystem models, forest disturbance is mostly represented by rate. Quite a few models have adopted the simple power-law model to characterize the patch size-frequency distribution. This dissertation has demonstrated that using a power-law model could make the estimates of total disturbance area and total frequency deviated from the actual scenario. Thus, I proposed a hierarchical model to improve accuracy. Possible improvements to the proposed model include incorporating the disturbance attribution (Kennedy et al. 2015; Zhao et al. 2015) to separate the projection by disturbance types and specifying the patch size-frequency distribution for different time periods. Application of the hierarchical method in ecosystem models is straightforward and desirable, which will improve the allocation of disturbances with various patch sizes, and further enhance the prediction of forest fragmentation, forest edge effect, post-disturbance resistibility and recovery of forest and hence the carbon balance following the forest disturbances. This is also a future research direction to be explored based on the findings of this dissertation.

Bibliography

Albers, H.J., Rozelle, S.D., & Guo, L. (1998). China's forests under economic reform: Timber supplies, environmental protection, and rural resource access. *Contemporary Economic Policy*, 16, 22-33

Anys, H., Bannari, A., He, D.C., & Morin, D. (1994). Texture analysis for the mapping of urban areas using airborne MEIS-II images. In (pp. 231-245): Proceedings of the First International Airborne Remote Sensing Conference and Exhibition

Asner, G.P., Kellner, J.R., Kennedy-Bowdoin, T., Knapp, D.E., Anderson, C., & Martin, R.E. (2013). Forest Canopy Gap Distributions in the Southern Peruvian Amazon. *Plos One*, 8, 10

Beck, A., Philip, G., Abdulkarim, M., & Donoghue, D. (2007). Evaluation of Corona and Ikonos high resolution satellite imagery for archaeological prospection in western Syria. *Antiquity*, 81, 161-175

Belward, A.S. (1996). The IGBP-DIS global 1 km land cover data set "DISCover": Proposal and implementation plans. In, *Report of the Land Cover Working Group of IGBP-DIS*. Toulouse, France: IGBP-DIS Office

Bindschadler, R., & Vornberger, P. (1998). Changes in the West Antarctic ice sheet since 1963 from declassified satellite photography. *Science*, 279, 689-692

Bonan, G.B. (2008). Forests and climate change: Forcings, feedbacks, and the climate benefits of forests. *Science*, 320, 1444-1449

Brandt, J.S., Kuemmerle, T., Li, H., Ren, G., Zhu, J., & Radeloff, V.C. (2012). Using Landsat imagery to map forest change in southwest China in response to the national logging ban and ecotourism development. *Remote Sensing of Environment*, 121

Broich, M., Stehman, S.V., Hansen, M.C., Potapov, P., & Shimabukuro, Y.E. (2009). A comparison of sampling designs for estimating deforestation from Landsat imagery: A case study of the Brazilian Legal Amazon. *Remote Sensing of Environment*, 113, 2448-2454

Brown, M., & Lowe, D.G. (2007). Automatic panoramic image stitching using invariant features. *International Journal of Computer Vision*, 74

Casana, J., & Cothren, J. (2008). Stereo analysis, DEM extraction and orthorectification of CORONA satellite imagery: archaeological applications from the Near East. *Antiquity*, 82, 732-749

- Challis, K., Priestnall, G., Gardner, A., Henderson, J., & O'Hara, S. (2002). Corona Remotely-Sensed Imagery in Dryland Archaeology: The Islamic City of al-Raqqā, Syria. *Journal of Field Archaeology*, 29, 139-153
- Chambers, J.Q., Negron-Juarez, R.I., Hurtt, G.C., Marra, D.M., & Higuchi, N. (2009). Lack of intermediate-scale disturbance data prevents robust extrapolation of plot-level tree mortality rates for old-growth tropical forests. *Ecology Letters*, 12, E22-E25
- Chan, J.C.-W., Huang, C., & DeFries, R.S. (2001). Enhanced algorithm performance for land cover classification using bagging and boosting. *IEEE Transactions on Geoscience and Remote Sensing*, 39, 693-695
- Chang, C.-C., & Lin, C.-J. (2011). LIBSVM: A library for support vector machines. *ACM Trans. Intell. Syst. Technol.*, 2, 1-27
- Channan, S., Feng, M., Kim, D.H., Sexton, J.O., Song, X.P., Song, D.X., Noojipady, P., Collins, K., Anand, A., & Townshend, J.R. (2015). THE GLS plus AN ENHANCEMENT OF THE GLOBAL LAND SURVEY DATASETS. *Photogrammetric Engineering and Remote Sensing*, 81, 521-525
- Chaplin-Kramer, R., Ramler, I., Sharp, R., Haddad, N.M., Gerber, J.S., West, P.C., Mandle, L., Engstrom, P., Baccini, A., Sim, S., Mueller, C., & King, H. (2015). Degradation in carbon stocks near tropical forest edges. *Nature Communications*, 6
- Chen, J., Chen, J., Liao, A., Cao, X., Chen, L., Chen, X., He, C., Han, G., Peng, S., Lu, M., Zhang, W., Tong, X., & Mills, J. (2015). Global land cover mapping at 30 m resolution: A POK-based operational approach. *ISPRS Journal of Photogrammetry and Remote Sensing*, 103, 7-27
- Chinese Academy of Forestry (2016). Review of forest resources in Sichuan. In: <http://www.lknet.ac.cn/qz/scsl.htm> (access date: 2016.8.24)
- Clarke, G.E. (1998). Development, society and environment in Tibet. *Development, Society, and Environment in Tibet*, 1-45
- Clauset, A., Shalizi, C.R., & Newman, M.E.J. (2009). Power-Law Distributions in Empirical Data. *SIAM Review*, 51, 661-703
- Clausi, D.A. (2002). An analysis of co-occurrence texture statistics as a function of grey level quantization. *Canadian Journal of Remote Sensing*, 28, 45-62
- Cochran, W.G. (1977). *Sampling techniques*. (3rd ed.). New York: Wiley
- Congalton, R. (1991). A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of Environment*, 37, 35-46

- Coppin, P., Jonckheere, I., Nackaerts, K., Muys, B., & Lambin, E. (2004). Digital change detection methods in ecosystem monitoring: a review. *International Journal of Remote Sensing*, 25, 1565-1596
- Cumming, S.G. (2001). A parametric model of the fire-size distribution. *Canadian Journal of Forest Research*, 31, 1297-1303
- DeFries, R., Achard, F., Brown, S., Herold, M., Murdiyarso, D., Schlamadinger, B., & de Souza, C., Jr. (2007). Earth observations for estimating greenhouse gas emissions from deforestation in developing countries. *Environmental Science & Policy*, 10, 385-394
- Defries, R.S., Field, C.B., Fung, I., Justice, C.O., Los, S., Matson, P.A., Matthews, E., Mooney, H.A., Potter, C.S., Prentice, K., Sellers, P.J., Townshend, J.R.G., Tucker, C.J., Ustin, S.L., & Vitousek, P.M. (1995). Mapping the land surface for global atmosphere-biosphere models: Toward continuous distributions of vegetation's functional properties *Journal of Geophysical Research-Atmospheres*, 100, 20867-20882
- Denslow, J.S. (1987). TROPICAL RAINFOREST GAPS AND TREE SPECIES DIVERSITY. Johnston, R. F. (Ed.). *Annual Review of Ecology and Systematics*, Vol. 18. ix+540p. Annual Reviews, Inc.: Palo Alto, California, USA. Illus, 431-452
- Di Vittorio, A.V., Negrón-Juárez, R.I., Higuchi, N., & Chambers, J.Q. (2014). Tropical forest carbon balance: effects of field- and satellite-based mortality regimes on the dynamics and the spatial structure of Central Amazon forest biomass. *Environmental Research Letters*, 9
- Dietze, M.C., & Clark, J.S. (2008). Changing the gap dynamics paradigm: Vegetative regeneration control on forest response to disturbance. *Ecological Monographs*, 78, 331-347
- Dozier, J. (1984). Snow Reflectance from LANDSAT-4 Thematic Mapper. *IEEE Transactions on Geoscience and Remote Sensing*, GE-22, 323-328
- Espirito-Santo, F.D.B., Gloor, M., Keller, M., Malhi, Y., Saatchi, S., Nelson, B., Junior, R.C.O., Pereira, C., Lloyd, J., Frohling, S., Palace, M., Shimabukuro, Y.E., Duarte, V., Mendoza, A.M., Lopez-Gonzalez, G., Baker, T.R., Feldpausch, T.R., Brienen, R.J.W., Asner, G.P., Boyd, D.S., & Phillips, O.L. (2014). Size and frequency of natural forest disturbances and the Amazon forest carbon balance. *Nature Communications*, 5, 3434-3434
- Fang, J.Y., Chen, A.P., Peng, C.H., Zhao, S.Q., & Ci, L. (2001). Changes in forest biomass carbon storage in China between 1949 and 1998. *Science*, 292, 2320-2322
- Fang, J.Y., Guo, Z.D., Piao, S.L., & Chen, A.P. (2007). Terrestrial vegetation carbon sinks in China, 1981-2000. *Science in China Series D-Earth Sciences*, 50, 1341-1350

FAO, & JRC (2012). Global forest land-use change 1990–2005. In. Rome: Food and Agriculture Organization of the United Nations and European Commission Joint Research Centre

Feng, M. (2015). Forest cover of North America in the 1970s mapped using Landsat MSS data. In, *AGU Fall Meeting Abstracts*

Feng, M., Huang, C., Channan, S., Vermote, E.F., Masek, J.G., & Townshend, J.R. (2012a). Quality assessment of Landsat surface reflectance products using MODIS data. *Computers & Geosciences*, 38, 9-22

Feng, M., Huang, C., Sexton, J.O., Channan, S., Narasimhan, R., & Townshend, J.R. (2012b). An approach for quickly labeling land cover types for multiple epochs at globally selected locations. In, *Geoscience and Remote Sensing Symposium (IGARSS), 2012 IEEE International* (pp. 6203-6206)

Feng, M., Sexton, J.O., Huang, C., Anand, A., Channan, S., Song, X.-P., Song, D.-X., Kim, D.-H., Noojipady, P., & Townshend, J.R. (2016). Earth science data records of global forest cover and change: Assessment of accuracy in 1990, 2000, and 2005 epochs. *Remote Sensing of Environment*, 184, 73-85

Fisher, J.I., Hurtt, G.C., Thomas, R.Q., & Chambers, J.Q. (2008). Clustered disturbances lead to bias in large-scale estimates based on forest sample plots. *Ecology Letters*, 11, 554-563

Foley, J.A., DeFries, R., Asner, G.P., Barford, C., Bonan, G., Carpenter, S.R., Chapin, F.S., Coe, M.T., Daily, G.C., Gibbs, H.K., Helkowski, J.H., Holloway, T., Howard, E.A., Kucharik, C.J., Monfreda, C., Patz, J.A., Prentice, I.C., Ramankutty, N., & Snyder, P.K. (2005). Global consequences of land use. *Science*, 309, 570-574

Friedl, M.A., Sulla-Menashe, D., Tan, B., Schneider, A., Ramankutty, N., Sibley, A., & Huang, X.M. (2010). MODIS Collection 5 global land cover: Algorithm refinements and characterization of new datasets. *Remote Sensing of Environment*, 114, 168-182

Frolking, S., Palace, M.W., Clark, D.B., Chambers, J.Q., Shugart, H.H., & Hurtt, G.C. (2009). Forest disturbance and recovery: A general review in the context of spaceborne remote sensing of impacts on aboveground biomass and canopy structure. *Journal of Geophysical Research-Biogeosciences*, 114

Galiatsatos, N. (2009a). The shift from film to digital product : focus on CORONA imagery. *Photogrammetrie - Fernerkundung - Geoinformation.*, 251-260

Galiatsatos, N. (2009b). The Shift from Film to Digital Product: Focus on CORONA Imagery. *Photogrammetrie Fernerkundung Geoinformation*, 251-260

Galiatsatos, N., Donoghue, D.N.M., & Philip, G. (2008). High resolution elevation data derived from stereoscopic CORONA imagery with minimal ground control: An

approach using Ikonos and SRTM data. *Photogrammetric Engineering and Remote Sensing*, 74, 1093-1106

Gao, F., Masek, J.G., & Wolfe, R.E. (2009). Automated registration and orthorectification package for Landsat and Landsat-like data processing. *Journal of Applied Remote Sensing*, 3

Gao, J., & Liu, Y. (2012). Deforestation in Heilongjiang Province of China, 1896-2000: Severity, spatiotemporal patterns and causes. *Applied Geography*, 35, 345-352

GLCF, & GSFC (2014). Global Land Cover Facility (GLCF) and Goddard Space Flight Center (GSFC): GLCF Forest Cover Change 2000 2005. In. Global Land Cover Facility, University of Maryland, College Park

Gong, P., Wang, J., Yu, L., Zhao, Y., Zhao, Y., Liang, L., Niu, Z., Huang, X., Fu, H., Liu, S., Li, C., Li, X., Fu, W., Liu, C., Xu, Y., Wang, X., Cheng, Q., Hu, L., Yao, W., Zhang, H., Zhu, P., Zhao, Z., Zhang, H., Zheng, Y., Ji, L., Zhang, Y., Chen, H., Yan, A., Guo, J., Yu, L., Wang, L., Liu, X., Shi, T., Zhu, M., Chen, Y., Yang, G., Tang, P., Xu, B., Giri, C., Clinton, N., Zhu, Z., Chen, J., & Chen, J. (2013). Finer resolution observation and monitoring of global land cover: first mapping results with Landsat TM and ETM+ data. *International Journal of Remote Sensing*, 34, 2607-2654

Gutman, G., Byrnes, R., Masek, J., Covington, S., Justice, C., Franks, S., & Headley, R. (2008). Towards monitoring land-cover and land-use changes at a global scale: The Global Land Survey 2005. *Photogrammetric Engineering and Remote Sensing*, 74, 6-10

Gutman, G., Huang, C., Chander, G., Noojipady, P., & Masek, J.G. (2013). Assessment of the NASA-USGS Global Land Survey (GLS) datasets. *Remote Sensing of Environment*, 134, 249-265

Haddad, N.M., Brudvig, L.A., Clobert, J., Davies, K.F., Gonzalez, A., Holt, R.D., Lovejoy, T.E., Sexton, J.O., Austin, M.P., Collins, C.D., Cook, W.M., Damschen, E.I., Ewers, R.M., Foster, B.L., Jenkins, C.N., King, A.J., Laurance, W.F., Levey, D.J., Margules, C.R., Melbourne, B.A., Nicholls, A.O., Orrock, J.L., Song, D.-X., & Townshend, J.R. (2015). Habitat fragmentation and its lasting impact on Earth's ecosystems. *Science Advances*, 1

Hahn, G.J., & Shapiro, S.S. (1994). *Statistical Models in Engineering*. Wiley

Han, Y.K., Byun, Y.G., Choi, J.W., Han, D.Y., & Kim, Y.I. (2012). Automatic Registration of High-Resolution Images Using Local Properties of Features. *Photogrammetric Engineering and Remote Sensing*, 78, 211-221

Hansen, M.C., Defries, R.S., Townshend, J.R.G., & Sohlberg, R. (2000). Global land cover classification at 1km spatial resolution using a classification tree approach. *International Journal of Remote Sensing*, 21, 1331-1364

Hansen, M.C., Potapov, P.V., Moore, R., Hancher, M., Turubanova, S.A., Tyukavina, A., Thau, D., Stehman, S.V., Goetz, S.J., Loveland, T.R., Kommareddy, A., Egorov, A., Chini, L., Justice, C.O., & Townshend, J.R. (2013). High-resolution global maps of 21st-century forest cover change. *Science*, *342*, 850-853

Hansen, M.C., Stehman, S.V., & Potapov, P.V. (2010). Quantification of global gross forest cover loss. *Proceedings of the National Academy of Sciences of the United States of America*, *107*, 8650-8655

Hansen, M.C., Stehman, S.V., Potapov, P.V., Loveland, T.R., Townshend, J.R.G., DeFries, R.S., Pittman, K.W., Arunarwati, B., Stolle, F., Steininger, M.K., Carroll, M., & DiMiceli, C. (2008). Humid tropical forest clearing from 2000 to 2005 quantified by using multitemporal and multiresolution remotely sensed data. *Proceedings of the National Academy of Sciences of the United States of America*, *105*, 9439-9444

Haralick, R.M., Shanmuga.K, & Dinstein, I. (1973). Textural features for image classification. *Ieee Transactions on Systems Man and Cybernetics*, *SMC3*, 610-621

Hayes, J.P. (2013). *A Change in Worlds on the Sino-Tibetan Borderlands: Politics, Economies, and Environments in Northern Sichuan*. Lexington Books

He, T., Liang, S.L., Wang, D., Cao, Y., Gao, F., & Yu, Y. (2016). Deriving a land surface albedo product from Landsat MSS, TM, ETM+, and OLI data based on the unified direct estimation approach. *Remote Sensing of Environment*, revised

Holden, C.E., & Woodcock, C.E. (2016). An analysis of Landsat 7 and Landsat 8 underflight data and the implications for time series investigations. *Remote Sensing of Environment*, *185*, 16-36

Huang, C.-D., Zhang, J., Yang, W.-Q., Tang, X., & Zhang, G.-Q. (2009a). Spatial difference characteristics of forest vegetation carbon stock in Sichuan Province. *Acta Ecologica Sinica*, *29*, 5115-5121

Huang, C.-D., Zhang, J., Yang, W.-Q., Tang, X., & Zhao, A.-J. (2008a). Dynamics on forest carbon stock in Sichuan Province and Chongqing City. *Acta Ecologica Sinica*, 966-975

Huang, C., Davis, L.S., & Townshend, J.R.G. (2002). An assessment of support vector machines for land cover classification. *International Journal of Remote Sensing*, *23*, 725-749

Huang, C., Goward, S.N., Masek, J.G., Thomas, N., Zhu, Z., & Vogelmann, J.E. (2010a). An automated approach for reconstructing recent forest disturbance history using dense Landsat time series stacks. *Remote Sensing of Environment*, *114*, 183-198

Huang, C., Thomas, N., Goward, S.N., Masek, J.G., Zhu, Z., Townshend, J.R.G., & Vogelmann, J.E. (2010b). Automated masking of cloud and cloud shadow for forest

change analysis using Landsat images. *International Journal of Remote Sensing*, 31, 5449-5464

Huang, C.Q., Kim, S., Song, K., Townshend, J.R.G., Davis, P., Altstatt, A., Rodas, O., Yanosky, A., Clay, R., Tucker, C.J., & Musinsky, J. (2009b). Assessment of Paraguay's forest cover change using Landsat observations. *Global and Planetary Change*, 67, 1-12

Huang, C.Q., Song, K., Kim, S., Townshend, J.R.G., Davis, P., Masek, J.G., & Goward, S.N. (2008b). Use of a dark object concept and support vector machines to automate forest cover change analysis. *Remote Sensing of Environment*, 112, 970-985

Hudak, A.T., & Wessman, C.A. (1998). Textural analysis of historical aerial photography to characterize woody plant encroachment in South African savanna. *Remote Sensing of Environment*, 66, 317-330

Hurt, G.C., Fisk, J., Thomas, R.Q., Dubayah, R., Moorcroft, P.R., & Shugart, H.H. (2010). Linking models and data on vegetation structure. *Journal of Geophysical Research-Biogeosciences*, 115

Hurt, G.C., Thomas, R.Q., Fisk, J.P., Dubayah, R.O., & Sheldon, S.L. (2016). The Impact of Fine-Scale Disturbances on the Predictability of Vegetation Dynamics and Carbon Flux. *Plos One*, 11, e0152883

Isaev, A.S., Korovin, G.N., Bartalev, S.A., Ershov, D.V., Janetos, A., Kasischke, E.S., Shugart, H.H., French, N.H.F., Orlick, B.E., & Murphy, T.L. (2002). Using remote sensing to assess Russian forest fire carbon emissions. *Climatic Change*, 55, 235-249

Janssens, I.A., Freibauer, A., Ciais, P., Smith, P., Nabuurs, G.J., Folberth, G., Schlamadinger, B., Hutjes, R.W.A., Ceulemans, R., Schulze, E.D., Valentini, R., & Dolman, A.J. (2003). Europe's terrestrial biosphere absorbs 7 to 12% of European anthropogenic CO₂ emissions. *Science*, 300, 1538-1542

Jorgensen, P.V. (2000). Determination of cloud coverage over Denmark using Landsat MSS/TM and NOAA-AVHRR. *International Journal of Remote Sensing*, 21, 3363-3368

Kadmon, R., & Harari-Kremer, R. (1999). Studying long-term vegetation dynamics using digital processing of historical aerial photographs. *Remote Sensing of Environment*, 68, 164-176

Kennedy, R.E., Yang, Z., & Cohen, W.B. (2010). Detecting trends in forest disturbance and recovery using yearly Landsat time series: 1. LandTrendr - Temporal segmentation algorithms. *Remote Sensing of Environment*, 114, 2897-2910

Kennedy, R.E., Yang, Z.Q., Braaten, J., Copass, C., Antonova, N., Jordan, C., & Nelson, P. (2015). Attribution of disturbance change agent from Landsat time-series

in support of habitat monitoring in the Puget Sound region, USA. *Remote Sensing of Environment*, 166, 271-285

Kim, D.-H., Sexton, J.O., Noojipady, P., Huang, C., Anand, A., Channan, S., Feng, M., & Townshend, J.R. (2014). Global, Landsat-based forest-cover change from 1990 to 2000. *Remote Sensing of Environment*, 155, 178-193

Kim, D.H., Sexton, J.O., & Townshend, J.R. (2015). Accelerated deforestation in the humid tropics from the 1990s to the 2000s. *Geophysical Research Letters*, 42, 3495-3501

Kim, M., Warner, T.A., Madden, M., & Atkinson, D.S. (2011). Multi-scale GEOBIA with very high spatial resolution digital aerial imagery: scale, texture and image objects. *International Journal of Remote Sensing*, 32, 2825-2850

Kupfer, B., Netanyahu, N.S., & Shimshoni, I. (2015). An Efficient SIFT-Based Mode-Seeking Algorithm for Sub-Pixel Registration of Remotely Sensed Images. *Ieee Geoscience and Remote Sensing Letters*, 12, 379-383

Kushwaha, S.P.S., Kuntz, S., & Oesten, G. (1994). Applications of image texture in forest classification. *International Journal of Remote Sensing*, 15, 2273-2284

Lechner, A.M., Stein, A., Jones, S.D., & Ferwerda, J.G. (2009). Remote sensing of small and linear features: Quantifying the effects of patch size and length, grid position and detectability on land cover mapping. *Remote Sensing of Environment*, 113, 2194-2204

Li, Q.L., Wang, G.Y., Liu, J.G., & Chen, S.B. (2009). Robust Scale-Invariant Feature Matching for Remote Sensing Image Registration. *Ieee Geoscience and Remote Sensing Letters*, 6, 287-291

Liu, D.C. (2001). Tenure and management of non-state forests in China since 1950 - A historical review. *Environmental History*, 6

Liu, J., Li, S., Ouyang, Z., Tam, C., & Chen, X. (2008). Ecological and socioeconomic effects of China's policies for ecosystem services. *Proceedings of the National Academy of Sciences of the United States of America*, 105, 9477-9482

Liu, J.G., Linderman, M., Ouyang, Z.Y., An, L., Yang, J., & Zhang, H.M. (2001). Ecological degradation in protected areas: The case of Wolong Nature Reserve for giant pandas. *Science*, 292, 98-101

Lloyd, J., Gloor, E.U., & Lewis, S.L. (2009). Are the dynamics of tropical forests dominated by large and rare disturbance events? *Ecology Letters*, 12, E19-E21

Lowe, D.G. (1999). Object recognition from local scale-invariant features. In, *Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on* (pp. 1150-1157 vol.1152)

- Lowe, D.G. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60, 91-110
- Lu, A.F., Tian, H.Q., Liu, M.L., Liu, J.Y., & Melillo, J.M. (2006). Spatial and temporal patterns of carbon emissions from forest fires in China from 1950 to 2000. *Journal of Geophysical Research-Atmospheres*, 111
- Lu, D., Batistella, M., Moran, E., & de Miranda, E.E. (2008). A comparative study of landsat TM and SPOT HRG images for vegetation classification in the Brazilian amazon. *Photogrammetric Engineering and Remote Sensing*, 74, 311-321
- Macleod, R.D., & Congalton, R.G. (1998). Quantitative comparison of change-detection algorithms for monitoring eelgrass from remotely sensed data. *Photogrammetric Engineering and Remote Sensing*, 64, 207-216
- Malamud, B.D., Millington, J.D.A., & Perry, G.L.W. (2005). Characterizing wildfire regimes in the United States. *Proceedings of the National Academy of Sciences of the United States of America*, 102, 4694-4699
- Masek, J.G., Goward, S.N., Kennedy, R.E., Cohen, W.B., Moisen, G.G., Schleeweis, K., & Huang, C. (2013). United States Forest Disturbance Trends Observed Using Landsat Time Series. *Ecosystems*, 16, 1087-1104
- Masek, J.G., Vermote, E.F., Saleous, N.E., Wolfe, R., Hall, F.G., Huemmrich, K.F., Gao, F., Kutler, J., & Lim, T.K. (2006). A Landsat surface reflectance dataset for North America, 1990-2000. *Ieee Geoscience and Remote Sensing Letters*, 3, 68-72
- McDonald, R.A. (1995). CORONA - Success for space reconnaissance, a look into the cold-war, and a revolution for intelligence. *Photogrammetric Engineering and Remote Sensing*, 61, 689-720
- Mikolajczyk, K., & Schmid, C. (2004). Scale & Affine Invariant Interest Point Detectors. *International Journal of Computer Vision*, 60, 63-86
- Milojevic, S. (2010). Power Law Distributions in Information Science: Making the Case for Logarithmic Binning. *Journal of the American Society for Information Science and Technology*, 61, 2417-2425
- Moorcroft, P.R., Hurtt, G.C., & Pacala, S.W. (2001). A method for scaling vegetation dynamics: The ecosystem demography model (ED). *Ecological Monographs*, 71, 557-585
- Nabuurs, G.J., Masera, O., Andrasko, K., Benitez-Ponce, P., Boer, R., Dutschke, M., Elsiddig, E., Ford-Robertson, J., Frumhoff, P., Karjalainen, T., Krankina, O., Kurz, W.A., Matsumoto, M., Oyhantcabal, W., Ravindranath, N.H., Sanz Sanchez, M.J., & Zhang, X. (2007). Forestry. In B. Metz, O.R. Davidson, P.R. Bosch, R. Dave, & L.A. Meyer (Eds.), *Climate Change 2007: Mitigation. Contribution of Working Group III to the Fourth Assessment Report of the Intergovernmental Panel on Climate*

Change. Cambridge, United Kingdom and New York, NY, USA.: Cambridge University Press

Nagel, B.Y. (1991). Socioeconomic Differentiation among Small Cultivators on Paraguay's Eastern Frontier. *Latin American Research Review*, 26, 103-132

Nickson, A. (1981). Brazilian Colonization of the Eastern Border Region of Paraguay. *Journal of Latin American Studies*, 13, 111-131

NPIC (1967). The KH-4A Camera System. In: National Photographic Interpretation Center

Olofsson, P., Foody, G.M., Herold, M., Stehman, S.V., Woodcock, C.E., & Wulder, M.A. (2014). Good practices for estimating area and assessing accuracy of land change. *Remote Sensing of Environment*, 148, 42-57

Olofsson, P., Foody, G.M., Stehman, S.V., & Woodcock, C.E. (2013). Making better use of accuracy data in land change studies: Estimating accuracy and area and quantifying uncertainty using stratified estimation. *Remote Sensing of Environment*, 129, 122-131

Olofsson, P., Stehman, S.V., Woodcock, C.E., Sulla-Menashe, D., Sibley, A.M., Newell, J.D., Friedl, M.A., & Herold, M. (2012). A global land-cover validation data set, part I: fundamental design principles. *International Journal of Remote Sensing*, 33, 5768-5788

Pal, M., & Mather, P.M. (2005). Support vector machines for classification in remote sensing. *International Journal of Remote Sensing*, 26, 1007-1011

Palace, M., Keller, M., Asner, G.P., Hagen, S., & Braswell, B. (2008). Amazon forest structure from IKONOS satellite data and the automated characterization of forest canopy properties. *Biotropica*, 40, 141-150

Pan, Y.D., Birdsey, R.A., Fang, J.Y., Houghton, R., Kauppi, P.E., Kurz, W.A., Phillips, O.L., Shvidenko, A., Lewis, S.L., Canadell, J.G., Ciais, P., Jackson, R.B., Pacala, S.W., McGuire, A.D., Piao, S.L., Rautiainen, A., Sitch, S., & Hayes, D. (2011). A Large and Persistent Carbon Sink in the World's Forests. *Science*, 333, 988-993

Piao, S., Fang, J., Ciais, P., Peylin, P., Huang, Y., Sitch, S., & Wang, T. (2009). The carbon balance of terrestrial ecosystems in China. *Nature*, 458, 1009-1018

Pimm, S.L., Jenkins, C.N., Abell, R., Brooks, T.M., Gittleman, J.L., Joppa, L.N., Raven, P.H., Roberts, C.M., & Sexton, J.O. (2014). The biodiversity of species and their rates of extinction, distribution, and protection. *Science*, 344, 987-+

Puetz, S., Groeneveld, J., Henle, K., Knogge, C., Martensen, A.C., Metz, M., Metzger, J.P., Ribeiro, M.C., de Paula, M.D., & Huth, A. (2014). Long-term carbon loss in fragmented Neotropical forests. *Nature Communications*, 5

Putz, F.E. (1983). TREEFALL PITS AND MOUNDS, BURIED SEEDS, AND THE IMPORTANCE OF SOIL DISTURBANCE TO PIONEER TREES ON BARRO-COLORADO ISLAND, PANAMA. *Ecology*, 64, 1069-1074

Richards, J.A., & Jia, X. (2006). Error Correction and Registration of Image Data. *Remote Sensing Digital Image Analysis* (pp. 27-65): Springer Berlin Heidelberg

Richards, P.D. (2011). Soy, Cotton, and the Final Atlantic Forest Frontier. *Professional Geographer*, 63, 343-363

Rigina, O. (2003). Detection of boreal forest decline with high-resolution panchromatic satellite imagery. *International Journal of Remote Sensing*, 24, 1895-1912

Roy, D.P. (2000). The impact of misregistration upon composited wide field of view satellite data and implications for change detection. *IEEE Transactions on Geoscience and Remote Sensing*, 38, 2017-2032

Sexton, J.O., Noojipady, P., Song, X.-P., Feng, M., Song, D.-X., Kim, D.-H., Anand, A., Huang, C., Channan, S., Pimm, S.L., & Townshend, J.R. (2016). Conservation policy and the measurement of forests. *Nature Clim. Change*, 6, 192-196

Sexton, J.O., Song, X.-P., Feng, M., Noojipady, P., Anand, A., Huang, C., Kim, D.-H., Collins, K.M., Channan, S., DiMiceli, C., & Townshend, J.R. (2013a). Global, 30-m resolution continuous fields of tree cover: Landsat-based rescaling of MODIS vegetation continuous fields with lidar-based estimates of error. *International Journal of Digital Earth*, 6, 427-448

Sexton, J.O., Song, X.-P., Huang, C., Channan, S., Baker, M.E., & Townshend, J.R. (2013b). Urban growth of the Washington, D.C.–Baltimore, MD metropolitan region from 1984 to 2010 by annual, Landsat-based estimates of impervious cover. *Remote Sensing of Environment*, 129, 42-53

SFA (2003). State Forestry Administration of China: Regulations for National Forest Inventory in China. In

SFA (2005). State Forestry Administration of China: the Sixth National Forest Resource Inventory Report (1999-2003). In. Beijing, China

SFA (2009). State Forestry Administration of China: the Seventh National Forest Resource Inventory Report (2004-2008). In. Beijing, China

- SFA (2014). State Forestry Administration of China: Commercial logging-ban in natural forests: Sichuan. In: <http://www.forestry.gov.cn/main/72/content-677314.html> (access date: 2016.8.24)
- Shaban, M.A., & Dikshit, O. (2001). Improvement of classification in urban areas by the use of textural features: the case study of Lucknow city, Uttar Pradesh. *International Journal of Remote Sensing*, 22, 565-593
- Shimada, M., Itoh, T., Motooka, T., Watanabe, M., Shiraishi, T., Thapa, R., & Lucas, R. (2014). New global forest/non-forest maps from ALOS PALSAR data (2007-2010). *Remote Sensing of Environment*, 155, 13-31
- Shure, D.J., Phillips, D.L., & Bostick, P.E. (2006). Gap size and succession in cutover southern Appalachian forests: an 18 year study of vegetation dynamics. *Plant Ecology*, 185, 299-318
- Singh, A. (1989). DIGITAL CHANGE DETECTION TECHNIQUES USING REMOTELY-SENSED DATA. *International Journal of Remote Sensing*, 10, 989-1003
- Smil, V. (1984). *The bad earth. Environmental degradation in China*. Zed Press
- Sohn, H.G., Kim, G.H., & Yom, J.H. (2004). Mathematical modelling of historical reconnaissance corona KH-4B imagery. *Photogrammetric Record*, 19, 51-65
- Song, D.-X., Huang, C., Noojipady, P., Channan, S., & Townshend, J.R. (2014). Comparison of remote sensing based forest area and change estimation with national forestry inventory between 2000 and 2005 in China. In, *Geoscience and Remote Sensing Symposium (IGARSS), 2014 IEEE International* (pp. 4268-4271)
- Song, D.-X., Huang, C., Sexton, J.O., Channan, S., Feng, M., & Townshend, J.R. (2015). Use of Landsat and Corona data for mapping forest cover change from the mid-1960s to 2000s: Case studies from the Eastern United States and Central Brazil. *ISPRS Journal of Photogrammetry and Remote Sensing*, 103, 81-92
- Song, W., Weicheng, F., Binghong, W., & Jianjun, Z. (2001). Self-organized criticality of forest fire in China. *Ecological Modelling*, 145, 61-68
- Stehman, S.V. (1999). Basic probability sampling designs for thematic map accuracy assessment. *International Journal of Remote Sensing*, 20, 2423-2441
- Stehman, S.V. (2014). Estimating area and map accuracy for stratified random sampling when the strata are different from the map classes. *International Journal of Remote Sensing*, 35, 4923-4939
- Stehman, S.V., & Czaplewski, R.L. (1998). Design and analysis for thematic map accuracy assessment: Fundamental principles. *Remote Sensing of Environment*, 64, 331-344

Tappan, G.G., Hadj, A., Wood, E.C., & Lietzow, R.W. (2000). Use of Argon, Corona, and Landsat imagery to assess 30 years of land resource changes in west-central Senegal. *Photogrammetric Engineering and Remote Sensing*, 66, 727-735

Thematic Database for Human-Earth System. Forest Inventory Data. In: <http://www.data.ac.cn/index.asp> (access date: 2016.8.24)

Tomppo, E., Gschwantner, T., Lawrence, M., & McRoberts, R.E. (2010). National Forest Inventories pathways for common reporting. In. New York: Springer

Tomppo, E., Olsson, H., Stahl, G., Nilsson, M., Hagner, O., & Katila, M. (2008). Combining national forest inventory field plots and remote sensing data for forest databases. *Remote Sensing of Environment*, 112, 1982-1999

Townshend, J.R., Masek, J.G., Huang, C., Vermote, E.F., Gao, F., Channan, S., Sexton, J.O., Feng, M., Narasimhan, R., Kim, D., Song, K., Song, D., Song, X.-P., Noojipady, P., Tan, B., Hansen, M.C., Li, M., & Wolfe, R.E. (2012). Global characterization and monitoring of forest cover using Landsat data: opportunities and challenges. *International Journal of Digital Earth*, 1-25

Townshend, J.R.G., & Justice, C.O. (1988). SELECTING THE SPATIAL-RESOLUTION OF SATELLITE SENSORS REQUIRED FOR GLOBAL MONITORING OF LAND TRANSFORMATIONS. *International Journal of Remote Sensing*, 9, 187-236

Townshend, J.R.G., Justice, C.O., Gurney, C., & McManus, J. (1992). The impact of misregistration on change detection. *IEEE Transactions on Geoscience and Remote Sensing*, 30, 1054-1060

Tucker, C.J., Grant, D.M., & Dykstra, J.D. (2004). NASA's global orthorectified Landsat data set. *Photogrammetric Engineering & Remote Sensing*, 70, 313-322

Turner, M.G., Baker, W.L., Peterson, C.J., & Peet, R.K. (1998). Factors influencing succession: Lessons from large, infrequent natural disturbances. *Ecosystems*, 1, 511-523

USGS (2004). Shuttle Radar Topography Mission, 1 Arc Second scene, SRTM_u03_n008e004, Unfilled Unfinished 2.0, February, 2000. In. College Park, Maryland, February 2000 Global Land Cover Facility, University of Maryland

USGS (2009). Fact Sheet of Declassified Images. In. U.S. Geological Survey, URL: <http://eros.usgs.gov/> (access date: 2013.5.2)

Vina, A., Bearer, S., Chen, X., He, G., Linderman, M., An, L., Zhang, H., Ouyang, Z., & Liu, J. (2007). Temporal changes in giant panda habitat connectivity across boundaries of Wolong Nature Reserve, China. *Ecological Applications*, 17, 1019-1030

- Vina, A., McConnell, W.J., Yang, H.B., Xu, Z.C., & Liu, J.G. (2016). Effects of conservation policy on China's forest recovery. *Science Advances*, 2
- Wang, L., Niu, Z., Wu, C., Xie, R., & Huang, H. (2012). A robust multisource image automatic registration system based on the SIFT descriptor. *International Journal of Remote Sensing*, 33
- Wang, S., van Kooten, G.C., & Wilson, B. (2004). Mosaic of reform: forest policy in post-1978 China. *Forest Policy and Economics*, 6
- Wang, X.K., Feng, Z.W., & Ouyang, Z.Y. (2001). The impact of human disturbance on vegetative carbon storage in forest ecosystems in China. *Forest Ecology and Management*, 148, 117-123
- White, M.A., Nemani, R.R., Thornton, P.E., & Running, S.W. (2002). Satellite evidence of phenological differences between urbanized and rural areas of the eastern United States deciduous broadleaf forest. *Ecosystems*, 5, 260-273
- Wu, J., & Li, H. (2006). Perspectives and methods of scaling. In J. Wu, K.B. Jones, H. Li, & O.L. Loucks (Eds.), *Scaling and Uncertainty Analysis in Ecology*: Springer Netherlands
- Xu, J.T., Yin, R.S., Li, Z., & Liu, C. (2006). China's ecological rehabilitation: Unprecedented efforts, dramatic impacts, and requisite policies. *Ecological Economics*, 57, 595-607
- Yin, R., & Newman, D.H. (1997). Impacts of rural reforms: the case of the Chinese forest sector. *Environment and Development Economics*, 2, 291-305
- Yu, L., Zhang, D., & Holden, E.-J. (2008). A fast and fully automatic registration approach based on point features for multi-source remote-sensing images. *Computers & Geosciences*, 34, 838-848
- Zeng, W., Tomppo, E., Healey, S.P., & Gadov, K.V. (2015). The national forest inventory in China: history - results - international context. *Forest Ecosystems*, 2, 23- Article No.: 23
- Zhang, P.C., Shao, G.F., Zhao, G., Le Master, D.C., Parker, G.R., Dunning, J.B., & Li, Q.L. (2000a). Ecology - China's forest policy for the 21st century. *Science*, 288
- Zhang, Y., & Song, C. (2006). Impacts of afforestation, deforestation, and reforestation on forest cover in China from 1949 to 2003. *Journal of Forestry*, 104, 383-387
- Zhang, Y.Q., Uusivuori, J., & Kuuluvainen, J. (2000b). Econometric analysis of the causes of forest land use changes in Hainan, China. *Canadian Journal of Forest Research-Revue Canadienne De Recherche Forestiere*, 30, 1913-1921

Zhao, F., Huang, C.Q., & Zhu, Z.L. (2015). Use of Vegetation Change Tracker and Support Vector Machine to Map Disturbance Types in Greater Yellowstone Ecosystems in a 1984-2010 Landsat Time Series. *Ieee Geoscience and Remote Sensing Letters*, 12, 1650-1654