The Remote Sensing Monitoring Program of Indonesia's National Carbon Accounting System: Methodology and Products

LEMBAGA PENERBANGAN DAN ANTARIKSA NASIONAL

Indonesian National Institute of Aeronautics and Space

Indonesia-Australia Forest Carbon Partnership

Australian Aid

BAPPENAS

LAPAN

The Remote Sensing Monitoring Program of Indonesia's National Carbon Accounting System: Methodology and Products

June 2014

Indonesia-Australia Forest Carbon Partnership | Australian Aid

BAPPENAS

LAPAN

CSIRO

ACKNOWLEDGEMENTS

The work summarised in this document has resulted from contributions from numerous people and institutions over many years. INCAS program was conceived in partnership between Indonesian and Australian government Agencies, with Kementerian Kehutanan (Ministry of Forestry/MoF) as the Indonesian Executing Agency. In 2009, MoF Director of Forest Resource Inventory and Monitoring in the Directorate General of Forest Planning, Dr. Ir. Hermawan Indrabudi M.Sc., invited Lembaga Penerbangan dan Antariksa Nasional (National Institute of Aeronautics and Space/LAPAN) to a key implementation meeting in recognition of LAPAN's remote sensing roles. LAPAN's Deputy Chairman for Remote Sensing Affairs, Ir. Nur Hidayat, Dipl.Ing, commited to provide technical staff and facilities to develop the remote sensing program. We are grateful for the sustained support of LAPAN management. The remote sensing program has also benefited from support and input from other agencies; Unit Kerja Presiden Bidang Pengawasan dan Pengendalian Pembangunan (Presidential Work Unit on Overseeing and Controlling Development/UKP4), Kementerian Perencanaan Pembangunan Nasional (Ministry of National Development Planning/Bappenas), MoF, Badan Informasi Geospasial (Geospatial Information Agency/BIG), and from experts based in provinces throughout Indonesia. INCAS has been managed and supported through Indonesia-Australia Forest Carbon Partnership (IAFCP) since 2009; the managers and staff of IAFCP played a crucial role in supporting the infrastructure and activities of the remote sensing program. CSIRO Australia has provided sustained technical support and training, and other international experts have provided valuable input. The authors and editors are grateful to all those who have contributed to the work and progress of the program.

This report should be referenced as follows:

LAPAN (2014). The Remote Sensing Monitoring Program of Indonesia's National Carbon Accounting System: Methodology and Products, Version 1. *LAPAN-IAFCP.* Jakarta.

Principal authors and reviewers of this document:

Dr. Orbita Roswintiarti and Ratih Dewanti (LAPAN) and Suzanne Furby and Jeremy Wallace (CSIRO).

Contacts:

Contacts: Arum Tjahjaningsih (arum.tjahyaningsih@lapan.go.id), Kustiyo (kustiyo.kustiyo@lapan.go.id).

Cover picture:

Forests cover and change map of Indonesia produced by LCCA

This list acknowledges the management and processing team who have contributed directly to the data processing, methods development and creation of the products described in this document:

1. **LAPAN**

Dr. Orbita Roswintiarti, Dr. Bambang Trisakti, Ratih Dewanti, Arum Tjahjaningsih, Kustiyo, D Heri Sulyantara, Tatik Kartika, Ita Carolita, Sri Harini Pramono, I Made Parsa, Dianovita, Mulia Inda Rahayu, Sukentyas Estuti Siwi, Marendra Eko Budiono, Siti Hawariyyah, Inggit Lolita Sari, Sigit Pranotowijoyo, Ediyanta Purba, Hedy Izmaya, Yusron, Novie Indriasari, Danang Surya Candra, Yudhi Prabowo, Alif Nurmareta, Fadilah Rahmawati W., Asma Ramli, Ellina Tria Novitasari, Salira Vidyan, Haryo Surya Ganesha, Yoyok Bambang Irawan, Purbo Alam Prakoso, Nedy Dwiyandi, Indra Stevanus, Paksy Premandika, Choirin Nisak, Fachrizal Ahmad Sumardjo, Fitri Anggorowati, Hardi Tristiono, M Ferdhiansyah Noor, Eddy Winarto, Supratikno, Ogi Gumelar, Babag Purbantoro, Heru Noviar, Mukhoriyah, Silvia Anwar, Soko Budoyo, Emiyati, Joko Santo Cahyono, Iskandar Effendy, Gagat Nugroho, Nelly Dyahwathi, Nursanti Gultom, Kuncoro T., Djahroni, Andri Susanto, Rossi Hamzah, Dwi Nurcahyo Ari Putro, Bayu Bajra and Sarip Hidayat.

2. **MoF**

Belinda Arunawati, Rinaldi Immanuddin, Wahyu Catur Adinugroho, Donny Wicaksono, Sudirman Sudradjat, Retnosari Yusnita, Muhammad Yazid, Anna Tosiani and Virni Budi Arifanti.

3. **BIG**

Sumaryono, Mulyanto Darmawan, Habib Subagio, Sri Hartini, Prita Brada Bumi and Jaka Suryanta.

4. **Local Expert (Authority of Forestry at Province and Regency level)**

Erizal, Hasvia, Rini Anggraeni, Melisa Elizabeth Pasalbessy, Fatur Fatkhurohman, Dheny Trie Wahyu Sampurno, Edy Mulyadi, Lyna Mardiana, Yuliarsah, Sutardi, Joseph Boseren, Sahabuddin, Rina Puspitasari, Era Rante, Irmadi Nahib, Rahmat Zukra, Philipus Maxi, Budi Rario, Yan Hendrie and C. John G. Andreys.

5. **IAFCP**

Anne Casson, Thomas Harvey, Septa Febrina Heksaputri and Nofaldi.

6. **CSIRO**

Suzanne Furby, Jeremy Wallace, Tony Traylen, Drew Devereux, Simon Collings and Nat Raisbeck-Brown.

TABLE OF CONTENTS

1. INTRODUCTION

The Land Cover Change Analysis program (LCCA) is the remote sensing monitoring component of Indonesia's National Carbon Accounting System (INCAS). The body of this report provides a summary of the methods and products of the LCCA. This section provides a brief background of INCAS and its LCCA.

Figure 1.1. Forest extent and change map 2000-2009 for Indonesia produced by the LCCA. Dark green indicates areas that were always forest from 2000 to 2009, red shows forest loss between 2000 and 2009 while yellow indicates forest gain in the same period. [source: INCAS poster, Workshop on Earth Observation Satellite Data to Support REDD+ Implementation in Indonesia, February 2014]

Planning for the INCAS program commenced in 2008 as a collaboration between the Indonesian and Australian Governments. The aim was to build a credible and sustainable system in Indonesia to account for greenhouse gas (GHG) and allow for robust emissions reporting from Indonesia's land sector, with full national coverage. It was undertaken in response to national and international policy drivers with a focus on forests and forest change. Indonesia's forests are globally significant in terms of carbon storage and other values. Estimates suggest that the land sector has been the largest contributor to the nation's GHG emissions. Accurate estimation of deforestation rates has been a focus for local, national and international interest.

Decreasing rates of forest loss, improving forest management and establishing reforestation programs all present opportunities for Indonesia to benefit from international initiatives such as REDD+ (Reducing Emissions from Deforestation and Forest Degradation and the role of conservation, sustainable management and enhancement of forest carbon stocks). REDD+ is intended to provide economic incentives for these actions. In 2009, President Yudhoyono pledged to reduce Indonesia's GHG emissions by up to 26% below business as usual levels in 2020, with consideration of increasing this to 41% with sufficient international support. Following this the Government of Indonesia (GoI) signed a Letter of Intent with the Government of Norway, entering into a partnership on REDD+. A national system for monitoring and reporting forest change is required for participation in such programs. National policy drivers and international reporting requirements on Monitoring, Reporting, and Verification (MRV) also require such a system.

Work on INCAS commenced in 2009 under the Indonesia-Australia Forest Carbon Partnership (IAFCP). The program consists of two major technical components; the remote sensing component, and the emissions estimation component. The remote sensing component, LCCA, provides spatially detailed monitoring for the whole country of changes in forest area over time using satellite remote sensing imagery. The emissions estimation component includes forest biomass measurement, forest disturbance mapping, and carbon stock assessment and emissions estimations to produce GHG accounts. In both components, the initial approach was to transfer and adapt knowledge and experience from Australia's national system (Caccetta et al 2013) to build operational systems and capacity in Indonesia. The Ministry of Forestry (MoF) is the lead GoI partner for the overall INCAS program and leader of the emissions estimation component. The LCCA remote sensing program is led by the Indonesian National Institute of Aeronautics and Space (LAPAN) in collaboration with MoF, the Indonesian Geospatial Information Agency (BIG), the IAFCP and others. Through IAFCP, international expertise has been provided to develop the LCCA program at LAPAN. CSIRO Australia has provided sustained technical support and training. The program has also had significant input and interaction with Professor Matthew Hansen of the University of Maryland and his group, who have conducted workshops and training with input from other international experts.

The monitoring system was designed in response to existing and anticipated international agreements and frameworks, including developments from the Kyoto protocol, IPCC guidelines and expectations for REDD+. The design requirements included (a) national coverage (b) sub-hectare spatial resolution (c) capacity to monitor historic changes over at least ten years, and to continue monitoring into the future. Landsat imagery, on account of its resolution and historic archive, was the only feasible data source to meet these requirements. Access to and processing of Landsat imagery were initial priorities for implementing the system.

The initial objective of the LCCA was to map the extent of forested land and the annual changes in the extent for the whole of Indonesia for the 10-year period from 2000-2009 to provide inputs for carbon accounting activities. For this purpose forest cover is defined as physical land cover irrespective of tenure; as a collection of trees with height greater than 5 metres and having greater than 30% canopy cover. Plantations of oil palm and coconut palm are considered as non-forest. All other land cover is considered non-forest. LCCA does not produce classifications of forest type from satellite imagery; forest type information in INCAS is provided from MoF during the the biomass and emissions estimation process.

Commencing in 2009, Landsat data were sourced, assembled and processed to meet this primary objective. The LCCA data and products have since been extended to include recent 'update' years and now cover the period 2000-2012, with a commitment to complete 2013. A consistent, systematic forest monitoring approach is being applied to the whole of Indonesia for this period. Historic data were sourced from archives in Thailand, Australia and the United States as well as LAPAN's own Landsat archive. Since the LCCA program began, LAPAN has greatly expanded and strengthened Indonesia's data reception and archiving capacities through relationships with international agencies including the United States Geological Survey (USGS). Data received and archived in Indonesia will be used for future updates of the LCCA. Landsat 8 imagery is already received by LAPAN and likely to be the main monitoring data source for coming years. Other data sources have been considered to continue and complement the program, and LCCA methods can be applied or adapted to other optical data. In addition to providing forest change products for carbon accounting, the image and mosaic products from LCCA will have wide application to land use management and local government spatial planning in Indonesia.

Since work on the LCCA remote sensing component commenced, there have been significant developments in the variety and availability of remote sensing data systems and in computational capacity. LCCA has incorporated a 'continuous improvement' approach to adapt and evolve the system while maintaining consistency for monitoring purposes. Internationally, the importance of forest monitoring has driven efforts to coordinate and improve access to satellite observational data for this purpose through the Global Earth Observation System of Systems (GEOSS) and the Global Forest Observations Initiative (GFOI). Indonesia has been a key contributor to these activities and its experience in developing the LCCA is highly relevant.

This document provides a detailed summary of LCCA data, methods and products. The protocols for quality assurance and archiving are also described. Detailed descriptions of individual programs with operator manual level information will be produced in a separate document; referred to here as the Operational Manual.

2. LCCA STATUS AND FUTURE DIRECTIONS

The INCAS-LCCA program is ongoing. The IAFCP has supported the development of the methodology as well as developing the capability, capacity and infrastructure at LAPAN to allow for the continuation of the LCCA as an operational program within LAPAN.

By the middle of 2014, annual forest extent and change products for Indonesia from 2000-2012 will have been produced. The effort to initiate the program and to process this historical data has been considerable, but from the middle of 2014 the program will move to a single-year 'annual update' mode where much less effort is required. There is a commitment by LAPAN to continue the processing to produce an update using 2013 data. Discussions around extending the time series back to 1990 are ongoing.

The technical capacity and data streams exist to continue the annual updates of LCCA into the future. Technical challenges of new data sources, such as Landsat 8 imagery, are being addressed.

Insitutional support is equally important for continuity of the program – it is important that a clear mandate for the LCCA exists and key to this will be an evident strong demand from stakeholders for the generation of credible land cover change products.

Resourcing levels to perform an annual update (the process of adding one year sequentially to the time series) can be estimated based on recent milestone progress in the LCCA. Currently, three new years (2010-2012) are being added across the whole country over a period of approximately seven months by an experienced team, many contributing in part-time roles. The total effort is equivalent to approximately 14 full time staff. This would indicate that an experienced team of around six staff should be able to complete an annual update within six months. In practice a larger team will be needed, as it will be vital to maintain skilled staff within the team, and to plan for training and succession of staff.

As well as routine update activities, an ongoing INCAS-LCCA should involve continuous improvement activities in a research component. This will include activities to evaluate new methods and to incorporate new data, and possibly to examine distribution of products in different forms. It will include research aimed at improving the accuracy of the products and at improving the efficiency of creating the products.

Activities for improving the accuracy of the products that are being discussed include:

- Adopting a new ortho-rectification base derived from SPOT 6 imagery currently being worked on by BIG.
- Using SPOT 6 and Pleiades high resolution image data (acquired through a collaboration with Airbus Defence and Space) to assess and improve the forest extent mapping.
- Identifying cloud gaps in the Landsat image mosaics in important areas of change and searching for alternative optical image sources to fill the gaps. SPOT 4 and SPOT 5 may do this for the more recent years.
- Using radar data sources to separate some land cover types that are more difficult to separate using optical data.

Activities for improving the efficiency of the creation of the products that are being discussed include:

- More automated cloud-masking, particularly taking advantage of the improved signal quality and new image bands available in Landsat 8 data.
- Porting the software for the mostly automated processing steps to run on new data servers with faster I/O capabilities and/or faster network connections.

To guide the continuous improvement process, feedback from the data users and data processing team should be sought to identify the issues that have the biggest effect on the efficiency or suitability of the products. Research tasks will be designed to develop and test new methods against the current results for both accuracy and efficiency. If new methods, or new data sources, prove to be better, training in the new processing will be provided to the processing teams and the operational methodology updated. This cycle of evaluating, testing and improving can be continued throughout the program life.

In parallel with improving the methodology, the infrastructure for data processing, data archiving and data delivery should be reviewed. Substantial improvements in technology have already been adopted in the LCCA through developments in LAPAN's Remote Sensing Technology and Data Centre. Multi-CPU blade server computational technology has become available for the most computationally intensive steps in the processing and the data archive is being transferred to new, faster servers in the Data Centre.

There is also a need to educate the user community about the current products – their strengths and limitations for particular purposes and seek feedback on how best to deliver information products for applications other than the original carbon accounting purpose.

The INCAS-LCCA is one of a number of programs being developed for forest monitoring purposes both within and outside Indonesia. A formal accuracy assessment process should be developed to compare different products, noting that they will most likely have different purposes and different policy drivers. For example, LAPAN are developing a rapid response, 'early warning' forest detection methodology. The temporal resolution is much finer and the spatial resolution is much coarser than the INCAS-LCCA. The accuracy requirement is also lower. All areas of possible forest clearing are detected and provided to local/regional agencies for on-ground verification. Only those areas that are confirmed by these sources are acted upon. Opportunities for coordination and cooperation between these two projects are being considered as well as with other research activities.

Finally, the LAPAN team now has the capability to consider developing new products to complement the current forest extent and annual change maps. Such developments must be undertaken in collaboration with the other stakeholders, and will typically involve other data in addition to remote sensing.

3. SAMPLE RESULTS

The primary objective of the LCCA is to produce annual forest extent and change products. Figure 1.1 in the first section shows the forest extent, forest loss and forest gain from 2000-2009. Figure 3.1 shows the 2000-2012 product in more detail for Kalimantan.

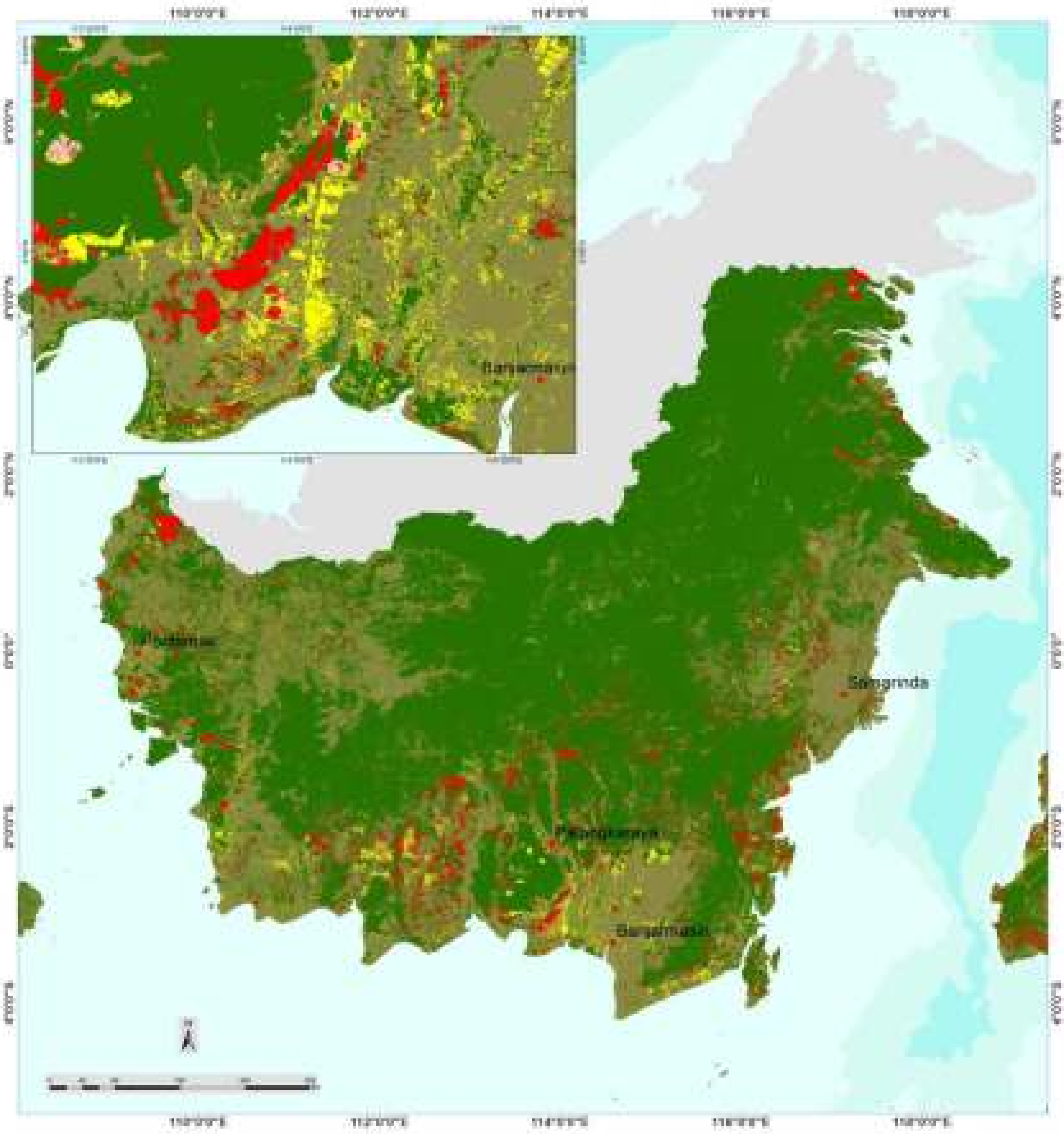


Figure 3.1. Forest extent and change map 2000-2012 for Kalimantan produced by the LCCA. Dark green indicates areas that were always forest from 2000 to 2012, red shows forest loss between 2000 and 2012 while yellow indicates forest gain in the same period. [source: INCAS poster, Workshop on Earth Observation Satellite Data to Support REDD+ Implementation in Indonesia, February 2014].

The change in forest extent is available on an annual basis, as shown in Figure 3.2. Statistical summaries of forest extent and change at national, regional and provincial levels can be produced from the digital products.

It is noted, in the context of carbon accounting, that not all change is relevant to reporting requirements. The identification of relevant land units for inclusion in a national carbon account is dependent on policy decisions from the national reporting agency.

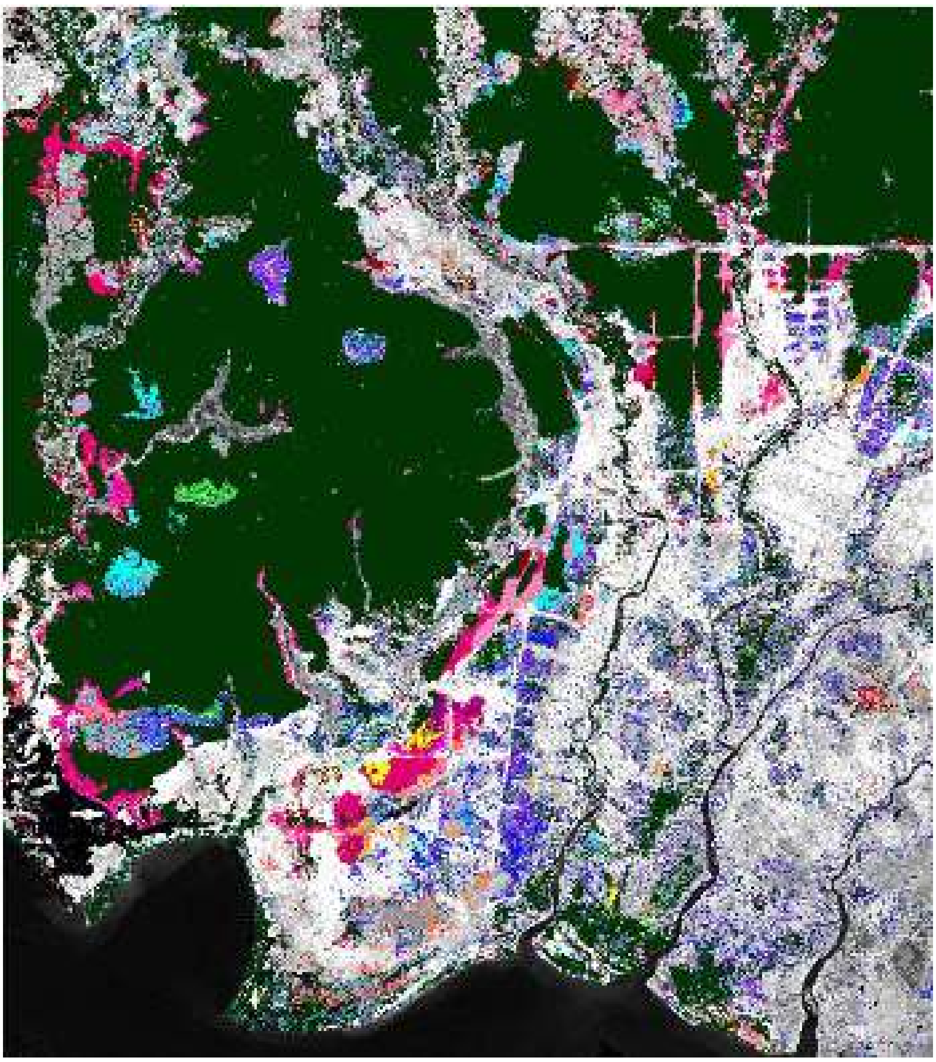


Figure 3.2. Forest extent and change map 2000-2012 for a region in Central Kalimantan produced by the LCCA. Dark green indicates areas that were always forest from 2000 to 2012, shades of red, orange, yellow and pink show forest loss between 2000 and 2012. Each shade corresponds to a different year (e.g 2000-2001, 2001-2002, ... , 2011-2012). Shades of green, blue and purple indicate reforestation in the same period. [source: INCAS poster, Workshop on Earth Observation Satellite Data to Support REDD+ Implementation in Indonesia, February 2014].

Some of the datasets assembled during the LCCA processing are useful products for a range of other applications. The individual geometrically and radiometrically corrected images are available along with corresponding cloud masks. Regional and national mosaics are also available. An example is shown in figure 3.3.

Figure 3.3. 2007 Landsat mosaic of Nusa Tenggara with bands 3,4,5 in BGR. The area shown is approximately 1500km by 500km. Black areas within the islands indicate areas of missing data due to cloud in 2007.

4. OVERVIEW OF OPERATING METHODOLOGY

There are a number of steps to produce the annual forest extent and change maps; the outputs from each step typically are required inputs for the subsequent step. The progression of processing steps is shown in the flowchart in Figure 4.1. Each of these steps is described in greater detail in sections below.

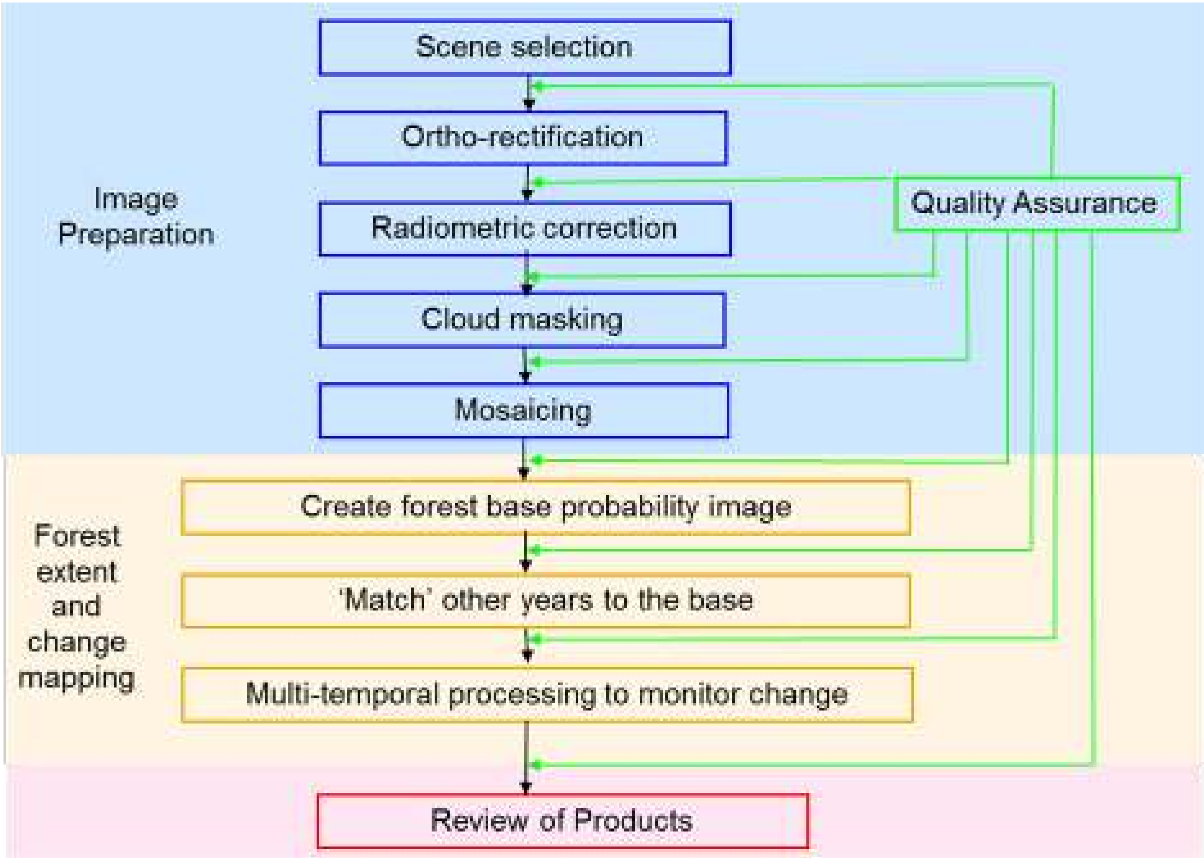


Figure 4.1. Flowchart of the steps in the INCAS-LCCA processing sequence. The steps in blue form the data preparation processing performed on individual images. The steps in orange form the forest extent and change mapping processing performed on the image mosaics. After every step a quality assurance process is performed to check the accuracy of that processing step. If accuracy requirements are not met, the data is reprocessed. The final step is a review of the products.

First the images to be used in the LCCA are selected. Not every image available in archives is suitable or needed for the processing. The selected images must be aligned geographically to each other and to other map data. Corrections to make the image values more consistent through time are then made. Contaminating data – such as cloud and shadow, haze, smoke and image noise - that obscure the ground cover are then masked from the images. The individual images are then mosaiced into larger units – mosaic tiles – to streamline the subsequent processing. Together these steps form the data preparation stage of the processing (Section 6 below). These preparations make the image data suitable for use for a range of applications.

There are three steps to making the annual forest extent and change products from the image mosaics (Section 7). Firstly ground-truth data – expert knowledge and high resolution images – are used to associate the image signals with forest and not forest cover to create a forest base for a single year in a very hands-on approach. Then a semi-automated matching process is used to 'match' the data for other years to the base. In the final step, knowledge of the temporal growth patterns in forest and non-forest cover types is used in a mathematical model to refine the single-date results to provide more reliable change detection.

The final step in the processing is to review the products, both to gain feedback on their accuracy and to understand their strengths and limitations for particular purposes. This review can suggest strategies for improving the products in the future.

After each step in the processing there is a quality assurance (QA) process to check that the method has been correctly applied and that the results meet required accuracy standards. If an image does not meet the standards for that step, the cause is investigated and the image reprocessed to correct the problem and checked again. The next step is not commenced until the current step is successfully completed. The quality assurance checks also ensure consistency between data processed by different team members and at different times during the project.

An image database identifying each selected image is created. This is used to track the progress of each image through the data preparation and quality assurance checks. Summary information from this database is used to report overall progress and identify bottlenecks or delays in the processing. The timely creation of the products relies on good management of the processing using such data.

A comprehensive data archive has been created in the LAPAN Data Centre for the LCCA. Systematic archives are created for output images and products for each processing step. The archive includes processing files which enable each step of the processing to be audited, and reproduced if necessary. The product archive includes a full record of each version of the processing; both the original 2000-2009 products and the improved and updated 2000-2012 products are archived so that any version of the products can be reproduced.

A processing team with experience in the use of satellite imagery has been trained to perform the processing steps and the quality assurance checks using the methodology described in this document, as per the detailed instructions in the 'Operational Manual'. Some team members specialise in particular steps and are able to train new team members. This ability to learn then teach the processing methodology is particularly important in long-running operational programs as some people will move on to other activities and be replaced by new team members.   This dedicated image analysis team is supported by people with local knowledge of the land cover in each region when the forest base is created. These local experts do not need experience with satellite imagery, although it is an advantage.

5. SOURCES OF DATA AND INFORMATION

5.1 Landsat imagery

As noted in Section 1, Landsat imagery was chosen as the only feasible data source to provide monitoring information for the implementation of LCCA. Landsat 5 (LS-5) and Landsat 7 (LS-7) were operational in the period. LS-5 is the preferred source for most of the period due to a technical problem with the scan line corrector (‘SLC-off’) which affected LS-7 from mid-2003. However in cases where cloud cover affects available LS-5 imagery, LS-7 imagery may be more useful. Both instruments have collected regular repeat coverage every 16 days over the period, but not all overpasses had been received and archived. Due to the frequency of cloud coverage over Indonesia, and other data quality problems it was desirable to have access to the most complete archive possible prior to selection of scenes for the LCCA (see Section 6.1 for description of scene selection).

The most complete archive of LS-5 imagery for western Indonesia for the period was held at Thailand’s GISTDA receiving station; Australia’s archive, held at Geoscience Australia (GA) covers far eastern Indonesia (Papua to eastern Nusa Tenggara) with LS-5 and LS-7 imagery. LAPAN’s receiving station at Parepare covers all of Indonesia, except for the very western tip of Sumatra, but only limited scenes had been archived. The main source of data for the central region was thus the USGS archive, which was far from complete for LS-5 as it consists of a sample of scenes selected for on-board storage and downloaded in the US. GA coordinated the image acquisition of Landsat imagery from these international data agencies in collaboration with the LCCA scene selection process described in Section 6.1. All selected data was delivered to Indonesia for processing within LCCA. Section 6.1 also provides detail on the numbers of scenes sourced from these various archives.

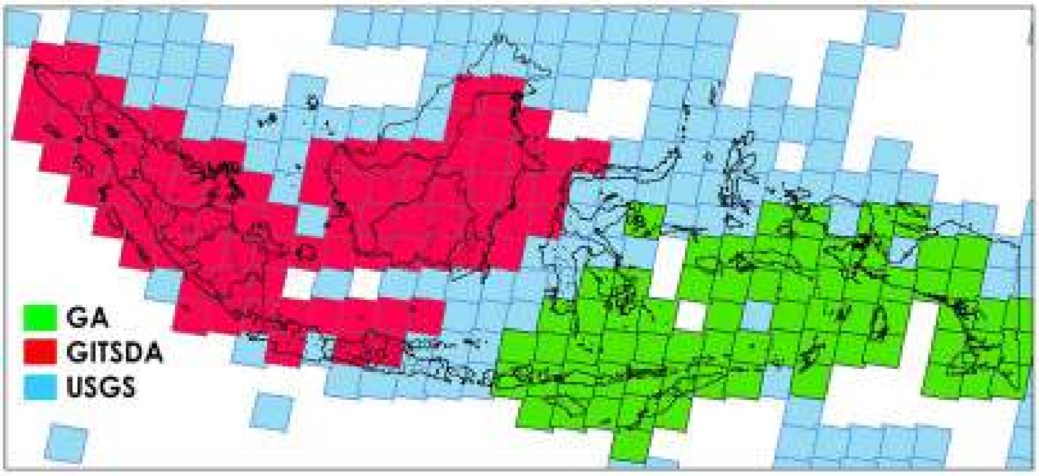


Figure 5.1. Indicative extent of spatial coverage of the GISTDA and GA Landsat archives. Lapan’s receiving station reception disk covers the full extent of Indonesia except for the northwest tip of Sumatra.

With respect to more recent and future data, it should be noted that the data reception, archiving and availability in Indonesia has advanced dramatically since the LCCA commenced. In part this progress is a result of the experience in implementing the LCCA. Landsat 8 (LS-8) is now operational and complete coverage of Indonesia including western Sumatra is received, archived and routinely ortho-rectified by LAPAN. This has arisen through collaboration with USGS which has assisted LAPAN in implementing an international-standard processing stream for the imagery. LAPAN has also negotiated reception of higher resolution SPOT-4, SPOT-5, SPOT-6 and Pleiades imagery in collaboration with Airbus Defence and Space.

5.2 High resolution imagery

Samples of high resolution satellite imagery were acquired for the LCCA. These were used for purposes of accurate interpretation of land cover in the forest mapping and forest review stages of the program. These tasks required an image resolution which could provide estimates of tree density, and indications of height from shadow. A resolution of two metres or better is required for these purposes. Air photography was not available, so commercial high resolution satellite imagery provided the only option.

Limited archives were available at the time from the Ikonos satellite (which commenced operation in 2006), GEO-eye, Quickbird and WorldView satellites. With additional support from IAFCP, samples of high resolution imagery were purchased over Kalimantan, Sumatra, Papua and Sulawesi for the forest base workshops conducted in 2011-2012. The criteria for image selection and the uses of the data are described in Section 6.2 below and in Section 7. All data were purchased with a multi-user licence which allowed distribution of copies to GoI agencies including LAPAN, MoF, BIG and the Presidential Working Unit for Supervision and Management of Development (UKP4). Further detail on the number and location of images is found in Section 6.2. Images were not purchased by the program for the remaining regions of Indonesia due to developments in data policy and data availability as described in Section 6.1.

5.3 Expert knowledge and map data

Information from experts with knowledge of regional land cover and land use was a critical input to the forest base mapping (Section 7.1) and product review (Section 8) as successive regions were processed. Typically three to five ground experts formed part of the base mapping team for each region. The MoF assisted with provision of its staff and with suggestions for regionally based experts. Staff from BIG also assisted in this role for all regions. These ground experts were asked to bring any relevant existing GIS data, maps or ground site information to assist in the forest mapping processes. In combination with the high resolution imagery, and the available maps, this human expertise with regional knowledge provided essential guidance to the production and assessment of the forest extent base maps.

6. DETAILED METHODOLOGY – IMAGE PREPARATION

This section describes all stages of the selection, processing and QA of Landsat scenes which are conducted prior to the forest extent mapping stages. The products from the image preparation stages used in subsequent steps of LCCA are corrected annual mosaic images for all regions of Indonesia. Products produced within the image preparation include ortho-rectified and radiometrically corrected individual scenes, and processing files which record the details of all processing steps and which are sufficient to provide an audit record of the numerical processing, and to reproduce each step if required. All such files are in the LCCA archive at LAPAN.

6.1 Scene selection

To map annual forest extent and change we require one cloud-free view of the land cover each calendar year. A single clear image is ideal and sufficient, but due to cloud cover this is a rare occurrence in tropical regions like Indonesia. The aim of the scene selection is to provide annual coverages of Indonesia with maximum cloud-free land area using the minimum number of scenes. The selection criteria were designed to provide best inputs to the subsequent forest extent mapping process. The specifications allow for selection of several images within each path/row to achieve a composite mosaic for each year, but this number is limited in anticipation of the manual effort required in parts of the image preparation processing.

The land area of Indonesia is covered by approximately 220 Landsat scenes (Figure 6.1.1) from paths 100-131. Many of these scenes include large areas of water. The Landsat satellite revisits each orbit path every 16 days, so where the archive is complete, there are approximately twenty-two images per year for each path/row to choose from. The historic archives available to the LCCA typically contain less than this as described in Section 5.

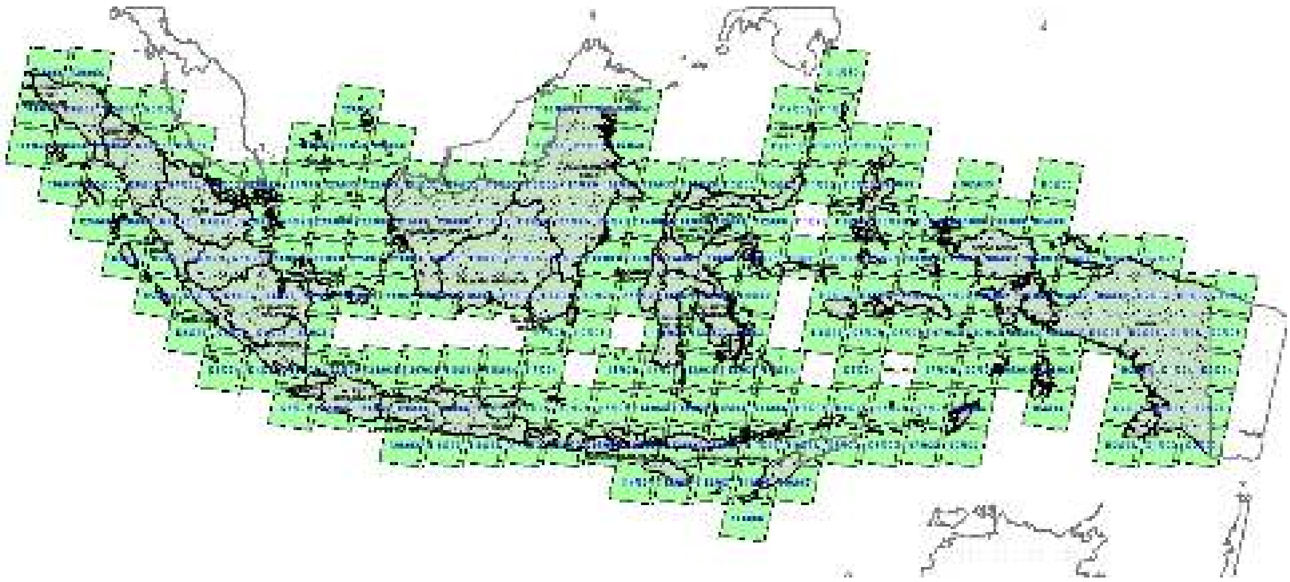


Figure 6.1.1. Landsat path/row coverage of Indonesia

Image selection is best done by visual assessment of high-quality browse images. The full set of browse images from the archives described in Section 5 was provided to the LCCA and is archived. Experience tells us that image cloud cover statistics typically provided with browse images are unreliable for scene selection. The summary statistics do not provide any information on the spatial distribution of the cloud in an image – and this is an important consideration. For example, an image with 50% cloud cover all in the north will be usable in the south and easy to 'cloud mask'. An image with patchy 50% cloud over the whole scene is likely to be contaminated by atmospheric effects and shadow and difficult to mask and use. More obviously, in images covering small islands, a small proportion of cloud over the land areas may make an image useless for LCCA purposes, while cloud areas over ocean are irrelevant.

To make the image selection, all image dates for the particular path/row and calendar year should be available as high-quality browse images. The best images are those that are completely free of any data problems, and which provide the best discrimination of forest cover. Besides cloud cover, other data quality issues include, but are not limited to, data errors (e.g. line drop-out), smoke and extensive flooding.

Sensor quality is also considered. Prior to August 1999, Landsat 5 is the only option. Until the Landsat 7 scan line corrector (SLC-off) problem, LS-5 and LS-7 both provide suitable imagery. LS-7 is the preferred choice where cloud and other data quality issues are acceptable. From July 2003, all LS-7 images have the SLC-off problem producing stripes of missing data. The preference is for LS-5 imagery, but where cloud problems are significant LS-7 images are considered. Between November 2011 and June 2013 only LS-7 SLC-off data is available. From July 2013, Landsat 8 is the preferred sensor.

In most parts of Indonesia for most years, cloud-free land area is the primary consideration for selection. Where there is a choice of cloud-free images, the time of year of the available images can also be considered. In many regions images from the dry season provide best discrimination between forest and non-forest land cover and are chosen where possible. In parts of Java and Nusa Tenggara the forest is deciduous, losing its leaves in the dry season, and images from a slightly wetter time are preferred.

Early in the project, the processing level of the source images was also considered. Path level images (L1G – nominal orientation) were preferred over automatically ortho-rectified images (L1T) as the accuracy and reliability of the L1T products did not meet the registration specifications. Over the life of the project, the automated processing systems have improved and processing level is less of an issue.

Generally, for each path/row and year up to 4 scenes are selected; more are considered with LS-7 SLC-off data and in areas where cloud-free imagery is known to be rare. A ‘primary’ image is selected with the biggest contiguous cloud-free areas. If this scene is completely cloud free over land areas, then no further images are required for that path/row and year. All subsequent images are chosen to complement the primary image, i.e. have clear data where the primary is cloudy, to build the composite with maximum cloud-free area. Where choice exists, temporal consistency down paths and close dates across rows are preferred. Haze requires attention – it is not always clearly visible in the browse imagery.

As an example of applying the scene selection criteria, Figure 6.1.2 shows the browse images for all available LS-5 scenes for path/row 100/065 in 2004. LS-5 is the preferred option as LS-7 is SLC-off. Clearly, cloud or data problems affect many images.−

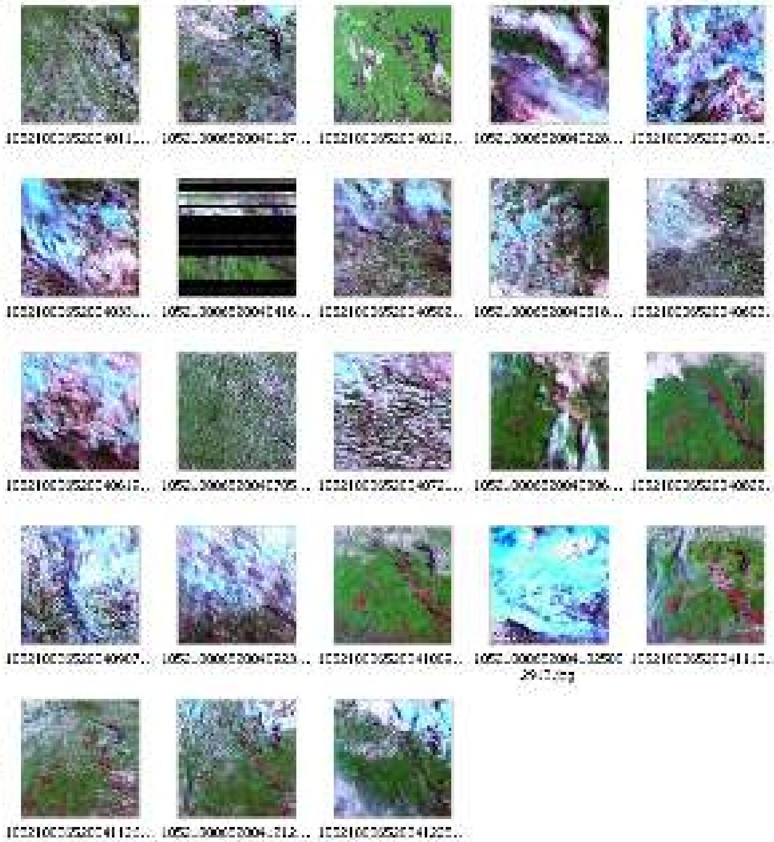


Figure 6.1.2. Browse images for all available LS-5 images in path/row 100/065 in 2004.

Figure 6.1.3 shows the three most likely candidate images for selection in more detail. The first candidate image is cloud-free in the north-east and approximately 70% cloud free in the south. The cloud in the north-west is in large patches rather than scattered which is easier to mask. Close inspection of the second candidate image shows that although the south appears cloud-free in the thumbnail, there is evidence of haze (bluish-purple smudges) in the larger view. The third candidate image is clear in the south and also in the central west where the first candidate image is cloudy. Together the first and third candidate images provide cloud free coverage for nearly the entire scene. The north-west has remaining areas of cloud. The review of the remaining browse images concentrates on this region. Unfortunately there are no candidates that fix this and so the scene selection stays at two for this path/row in 2004.

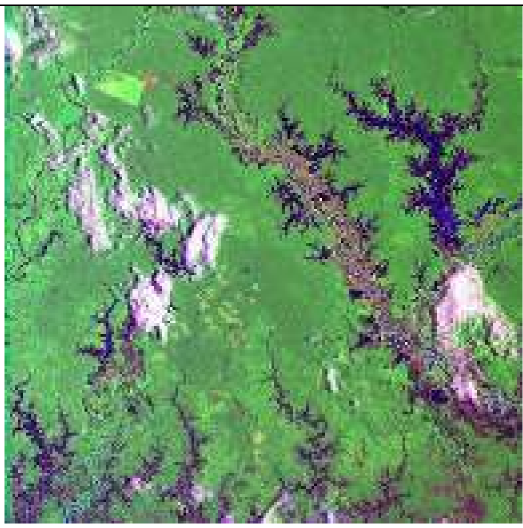
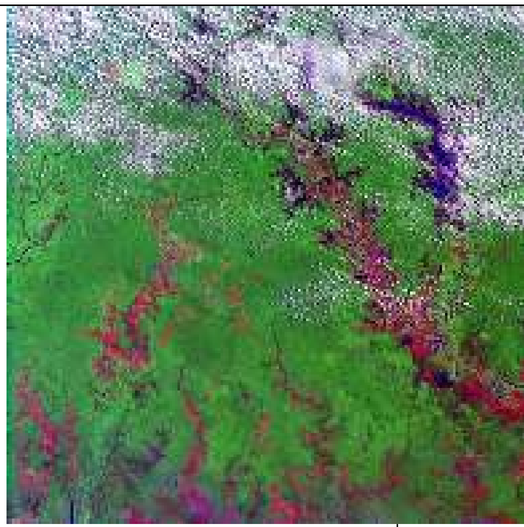
 **Candidate image 1** (12 Feb 2004) - top row centre in Figure 6.1.2	 **Candidate image 2** (9 Sep 2004) – 3rd row right in Figure 6.1.2
 **Candidate image 3** (22 Aug 2004) – 4th row centre in Figure 6.1.2	

Figure 6.1.3. The three most clear browse images for 100/65 in 2004.

The result of the scene selection process is a list of images to be used – for each path/row for each year. Large areas with no clear imagery can also be noted. For QA, the list of selections and the full set of browse images can be passed to another person for review.

Once the scene selections are final, the full versions of the selected images are obtained from the different source archives, each with their own data formats, and imported into a standard format for subsequent processing. The list of images is imported into the INCAS-LCCA database. This database records information about the image – location, date, and source – and monitors the progress of processing through the image preparation and QA steps. Figure 6.1.4 shows a snapshot of part of the database and Figure 6.1.5 shows a summary report derived from the database.

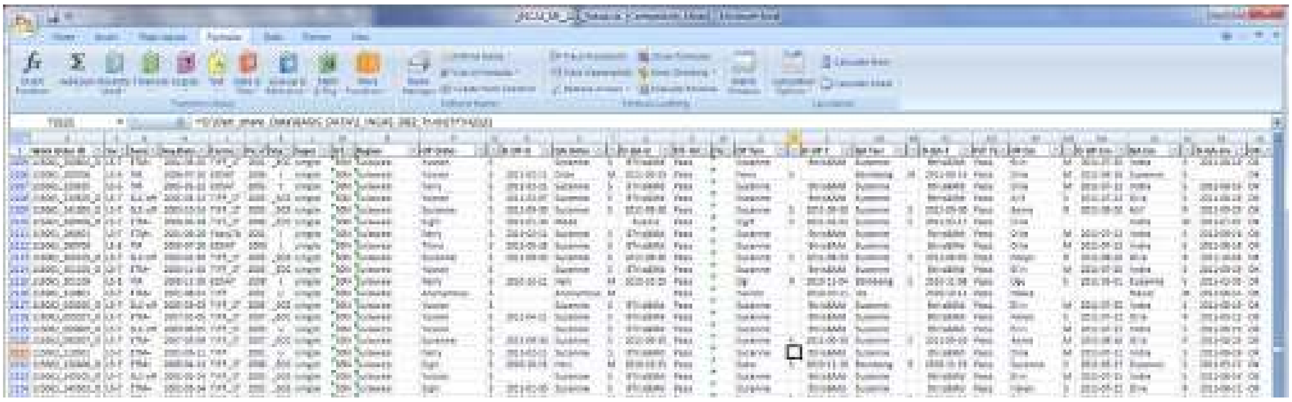


Figure 6.1.4: Snapshot of a page of the image database. As well as identifying all of the images selected, it tracks the process of the image through each processing and QA step.

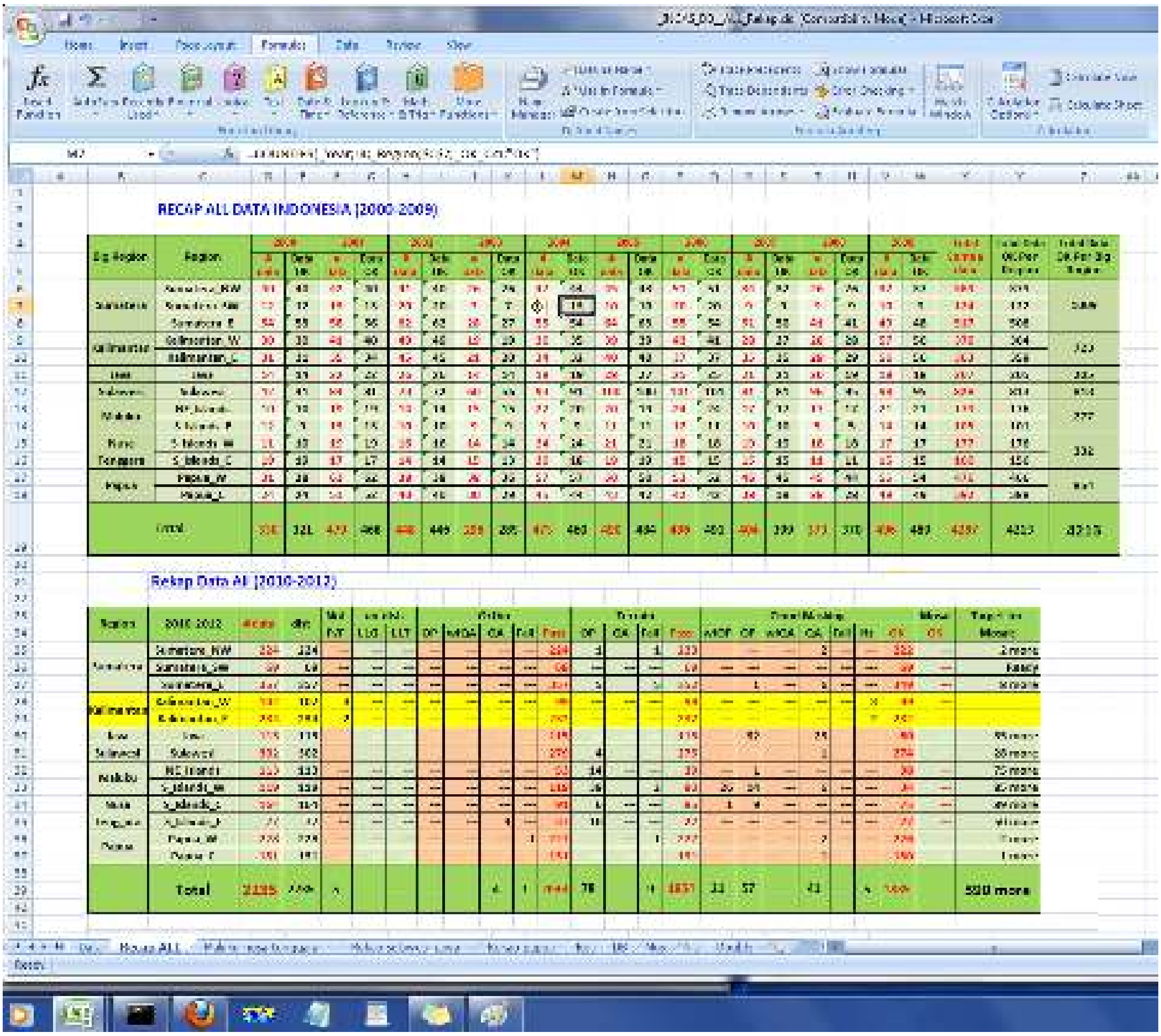


Figure 6.1.5. Sample summary reports derived from the database for tracking progress at a management level.

Summaries of the images used in the INCAS–LCCA can be extracted from this database. Figure 6.1.6 illustrates the number of images selected for each path/row across Indonesia in 2008. A small number of images over land areas indicates either that good clear images are available, or that few suitable images are available. The amount of clear land (after cloud masking) that results from these selections is shown in the mosaicing section (Section 6.7) of this document. Figure 6.1.7 provides a summary of the numbers of images by region.

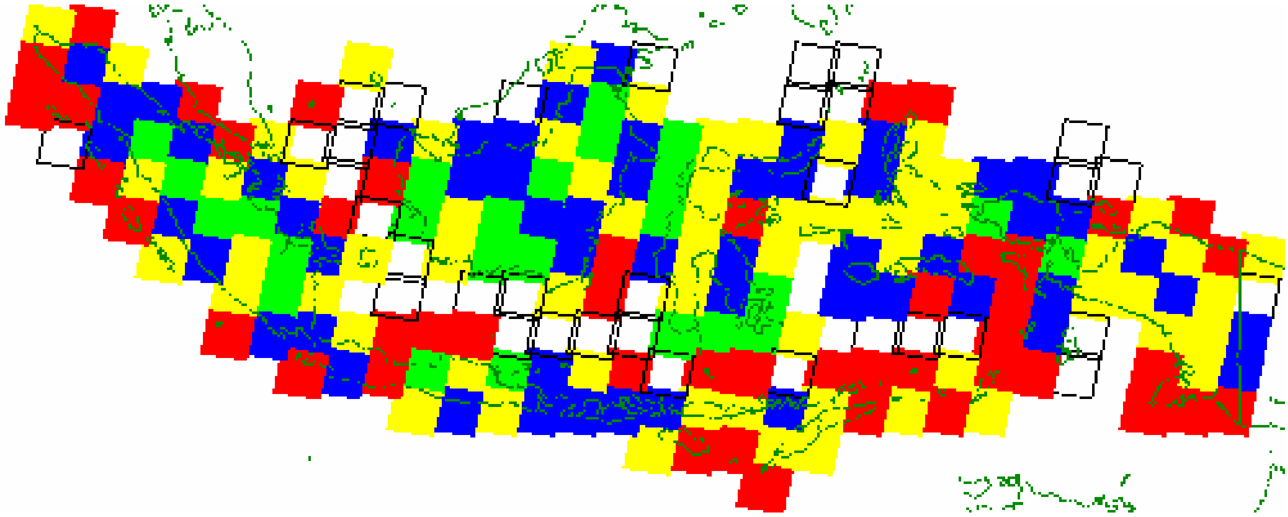


Figure 6.1.6. Number of images selected for 2008. White shows no suitable data available, red shows one selected image, yellow shows two selected images, blue shows three selected images and green shows four or more selected images.

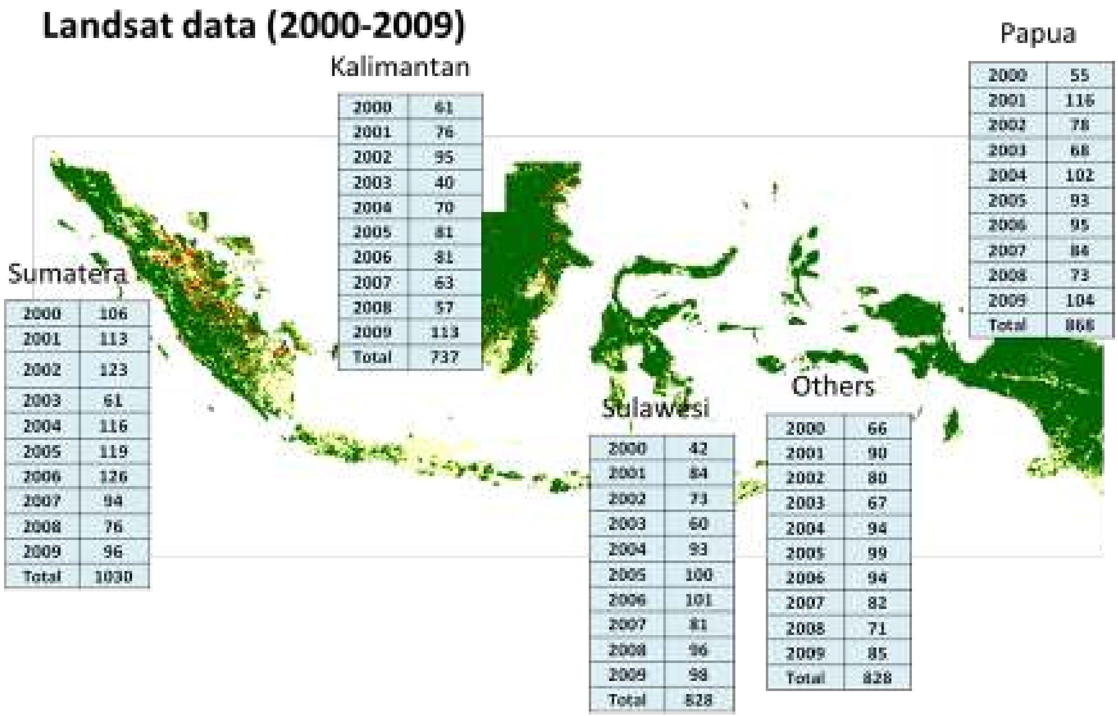


Figure 6.1.7. Number of Landsat images selected for each year from 2000-2009 across Indonesia

As at April 2014, there are 6553 Landsat images in the INCAS-LCCA archive for years 2000 to 2012. The Landsat image data was sourced from three outside archives (Thailand (GISTDA) – 893 scenes, Australia (GA) – 974 scenes, the United States of America (USGS) – 4436 scenes, as well as the Indonesian (LAPAN) archive – 250 scenes. For images acquired from December 2012, the LAPAN archive will be the primary data source.

6.2 High resolution imagery

Ground-truth information for training the forest extent mapping and for accuracy assessment is provided by local experts from each region of Indonesia. This is assisted by the interpretation of high resolution imagery.

The resolution, location and date of acquisition determine the usefulness of high resolution imagery for INCAS-LCCA. Visible and near infra-red imagery is suitable for determining land cover. Higher spectral resolution is of limited value for this purpose.

The spatial resolution must be fine enough to identify individual tree crowns and land use types. In fine resolution imagery the presence, or absence, of distinct shadows give an indication of vegetation height. A spatial resolution of 2m or finer is required. Suitable sources of imagery include aerial photography (variable), IKONOS (approx 4m multi-spectral, approx 1m panchromatic), Quickbird, WorldView 2, GeoEye1, Pleiades (approx 2m multi-spectral, 0.5m panchromatic) and Spot 6 (1.5m multi-spectral and panchromatic). Other sources such as SPOT-5 (10m multi-spectral, 5m regular panchromatic or 2.5m super mode panchromatic), ALOS (10m multi-spectral (AVNIR) and 2.5m panchromatic (PRISM)) and RapidEye (6.5m) are less suitable for identifying trees and tree height, but may be useful for broad land cover interpretation. Figure 6.2.1 shows some sample imagery.

For INCAS-LCCA forest extent mapping purposes, image dates closest in time to the date of the Landsat mosaic chosen for the 'forest base' are the most useful (Section 7.1). 2008 is generally the target year as it was the most recent for which Landsat mosaics were produced at the time the forest processing started. Choosing a recent year allows access to local memory of land cover and condition from site visits rather than having to rely exclusively on high resolution imagery for developing the classification.

Figure 6.2.1: Landsat and high resolution imagery for a sample area in central Kalimantan. Top left: 2009 Landsat TM image (25m) bands 3,4,5 in BGR. Top right: Pan-sharpened SPOT 5 image (2.5 and 10m). Bottom-left: Pan-sharpened SPOT 6 (1.5m). Bottom-right: Pan-sharpened Quickbird (0.6m and 2m). Individual trees can be seen in the Quickbird and SPOT 6 imagery which is suitable for LCCA purposes. Land cover can be inferred from the SPOT 5 imagery but more interpretation is required.

High resolution imagery is expensive to acquire and archives do not provide complete coverage. Sample images in key locations were selected in locations which were best for the purpose, considering images already held by GoI. At the beginning of the project little high resolution data was available. Images were selected and purchased by IAFCP for the INCAS-LCCA for Kalimantan, Sumatra, Papua and Sulawesi. For Java, imagery that had been purchased by the Ministry of Agriculture was available for use by LAPAN– the coverage is shown in Figure 6.2.2. Similar imagery held for Nusa Tenggara could not be used because of license restrictions. Extensive use was also made of the imagery available in Google Earth(TM) for Java and Nusa Tenggara. The resolution of the imagery in Google Earth(TM) varies significantly, but is high for much of Java and parts of Nusa Tenggara. Little high resolution imagery is available for Maluku, either commercially or publicly, and more emphasis was placed on seeking local expert input.

The high resolution images for Kalimantan, Sumatra, Papua and Sulawesi were selected to contain cloud-free samples of all important cover types. It is important that the images contain multiple cover types and/or a mixture of tree density so that boundaries between land cover classes can be interpreted. Extra images were purchased in areas of rapid change and uncertainty about cover type. High resolution image locations were selected by comparing a browse library of available images to the LCCA Landsat mosaics. Locations were selected with high local variability in the Landsat mosaics (multiple cover types) and relatively clear data in the high resolution imagery. As many land cover types and groupings were covered as possible within budget limitations. Figure 6.2.3 shows the high-resolution image locations for the four regions. In total, 271 high resolution images were purchased for the LCCA.

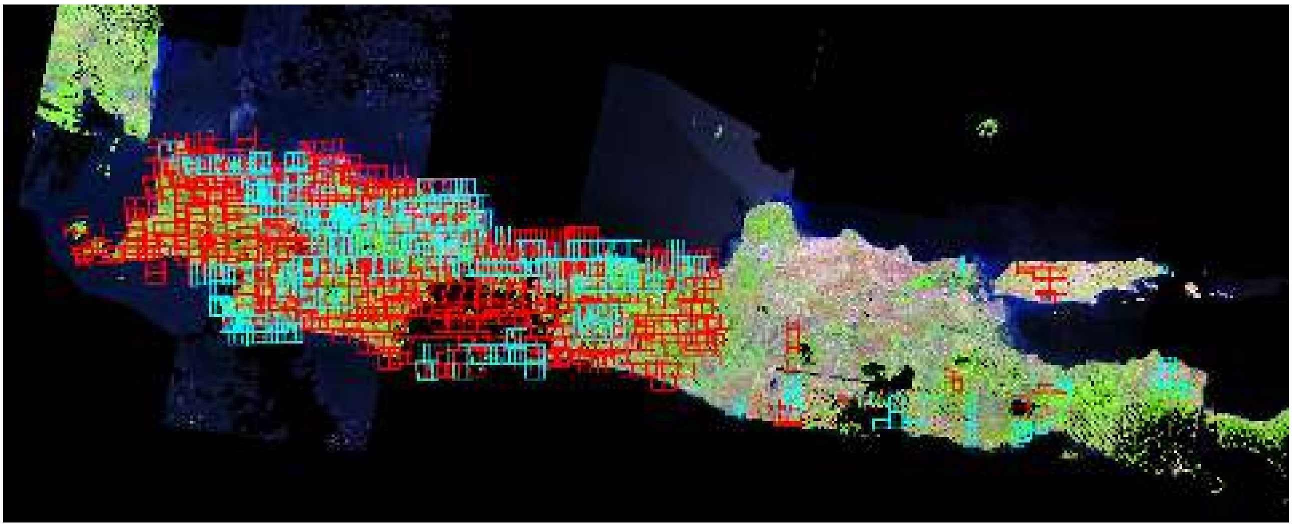


Figure 6.2.2. Locations of IKONOS (red) and GeoEye1 (light blue) imagery for Java available to LAPAN from the Ministry of Agriculture; background: 2008 Landsat TM mosaic (bands 3, 4, 5 in BGR).

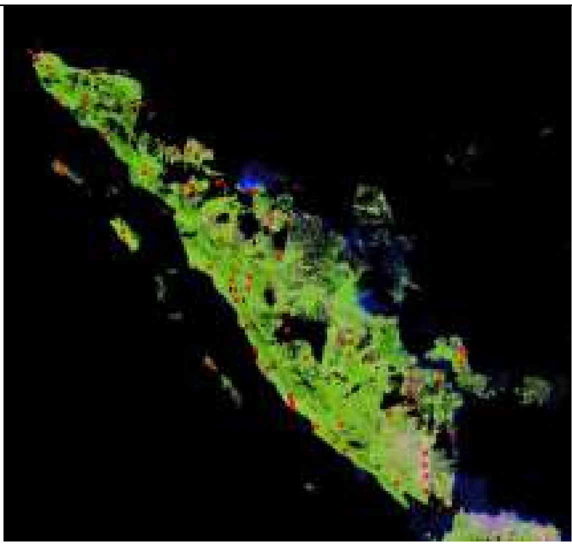
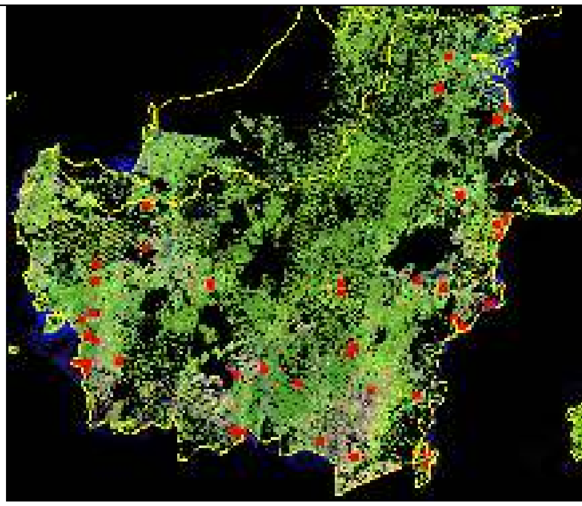
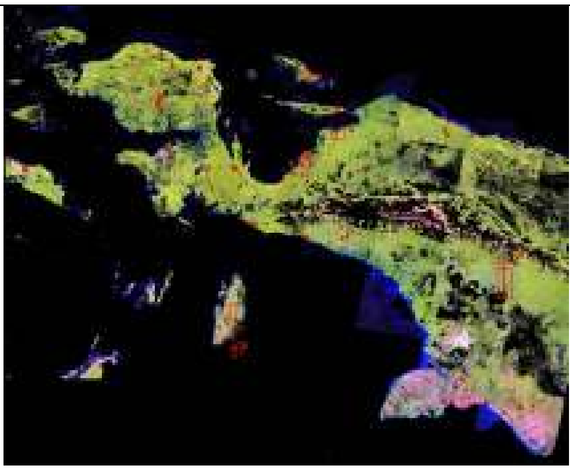
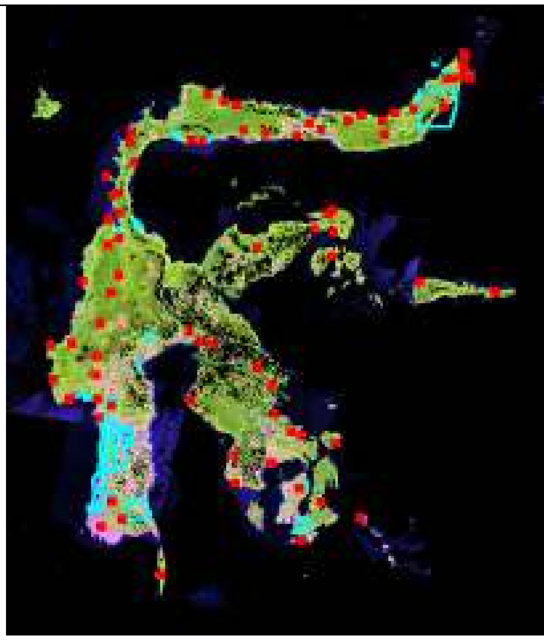
 **Sumatra:** 77 Quickbird, Ikonos, World View-2 (2006 to 2009).	 **Kalimantan:** 53 Quickbird, Ikonos, World View-2, and Geo Eye-1 (2001 to 2010, mostly 2007 to 2009).
 **Papua:** 62 Quickbird, Ikonos, World View-2, and GeoEye (2006 to 2011).	 **Sulawesi:** 79 Quickbird and World View-2 (2006 to 2011).

Figure 6.2.3. Locations and numbers of high resolution imagery (closed red squares) provided by IAFCP for four regions shown on 2008 Landsat TM mosaics (bands 3, 4, 5 in BGR). The open red and blue polygons show the locations of high resolution imagery already held by LAPAN.

6.3 Geometric base and correction

In recent years there have been rapid developments in programs to improve the availability and standards of rectified Landsat imagery. At the time of INCAS planning and commencement this was not the case; selection of a geometric reference base for Indonesia, and procedures for accurate geometric correction were recognised as of basic importance.

Accurate scene registration within and between years is essential for a monitoring program so that misregistration errors are not confused with land cover change (Townshend et al. 1992). The effects of misregistration in the time series of imagery are illustrated in Figure 6.3.1. Forest versus not forest maps from two years are compared to identify change. The two maps are well registered in the picture on left and one map has been shifted one pixel to the right and one pixel down in the picture on the right. The small and narrow areas of change indicated on the left image are likely to be correct, while more false change (due to misregistration) is shown on the right. Although the net area of land cover change (total clearing subtract total regrowth) would be the same, the maps and statistics of change from the misregistered image pair would be incorrect, and unsuitable for carbon accounting.

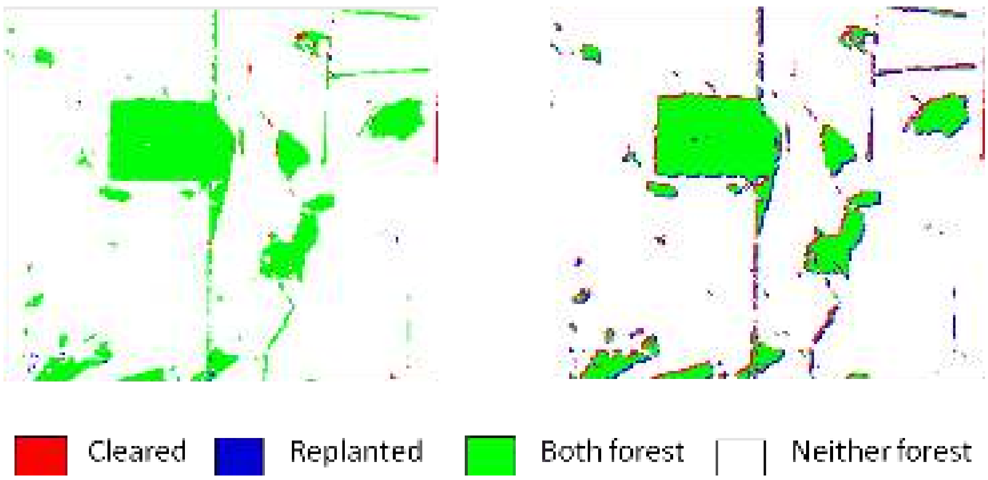


Figure 6.3.1. A comparison of forest and change maps from two time periods. The two images which produced the left map were well registered; the pair on the right have been produced from the same images where one has been shifted to simulate misregistration effects on the change maps.

Automated, or at least semi-automated, generation of ground control points (GCPs) by comparing a new image base to an agreed base image is the most efficient way of achieving good image to image registration. Ideally the base imagery would have high locational accuracy so that the image data can be compared to spatially referenced data from other sources.

As the equator passes through Indonesia a decision needed to be made within the LCCA about the projection for the individual images. For convenience and consistency all images within the LCCA are referenced to NUTM coordinates (WGS84 datum). In NUTM zones, the equator has a northing coordinate of zero. Areas south of the equator within the the LCCA thus have negative coordinates. If conversion to SUTM is required for any reason, only a simple constant shift (or offset) in northing is required. All images are resampled to 25m pixel size for simple alignment with metric map scales and area calculations. The final products are converted to geodetic projection to allow mosaicing across the whole country. A pixel size of 0.00025 degrees is selected.

6.3.1 Ortho-rectification base and digital elevation model

There was no GoI ortho-rectification image base for the whole of Indonesia when this project started in 2009. Therefore the Global Land Survey 2000 (GLS2000) data (USGS 2009, http://glcf.umd.edu/data/gls/) was selected for this purpose. The GLS2000 is formed from Landsat 7 ETM+ images and is a reprocessed version of the Global Land Cover Facility (GLCF) 2000 GeoCover(TM) collection (Tucker et al 2004) using Shuttle Radar Topography Mission (SRTM) digital topography and improved geodetic control (Gutman et al. 2008). The data is provided in GeoTIFF format with a UTM projection using the WGS-84 datum at 30m pixel resolution.

Individual images comprising GLS2000 were downloaded from the USGS by GA in May and June 2009 and mosaiced together within UTM zones. The associated metadata identifies the images as:

PRODUCT_NAME = "GLS-2000 Ver1.0"

PRODUCT_ELEVATION_DATA = "GLS-DEM Ver1.0"

The 2000 GeoCover(TM) product has been assessed as having a locational accuracy RMS error of 50m. Subsequent studies suggest that some images may exhibit errors considerably greater than 40-50m, particularly over mountainous areas and where the original DEM was of poor quality. The GLS2000 product was designed to address these limitations.

Some isolated image-to-image misregistration was discovered in adjoining images in parts of Indonesia during the INCAS–LCCA processing. In such regions, a new image for one or both path/rows was carefully processed and the well-registered images used to correct (overwrite) the GLS2000 product locally. Some parts of the GLS2000 base are quite cloudy. A mosaic was formed from cloud-masked images of Papua that have been processed as part of this project and used to replace cloudy data in the base for the LCCA.

At the present time the issue of a more accurate Indonesian ortho-rectification base is being pursued as part of other GoI projects. Once a new base is established, a transformation from the current LCCA base to the new one can be established for each region and the essential data and products transformed.

The digital elevation model (DEM) used in the ortho-rectification processing is the SRTM-DEM version 3 at 90m pixel resolution, downloaded in May 2009. It is also used in the terrain-illumination correct step.

6.3.2 Correlation matching and Master ground control points

To register a new image to the base, correlation matching is used to locate predefined physical features from the ortho-rectified base image in the new image. The coordinates of the center of these features are then used as the GCPs to formulate the transformation between the new image and the base.

A library of these predefined feature locations was created for every path/row for Indonesia. In the INCAS-LCCA these locations are called 'Master GCPs'. They are also referred to as 'GCP chips' in some applications. The latter terminology usually refers to a small image window or subset (say 64 lines by 64 pixels) that is stored in a database. In the LCCA, only the feature location is stored and the data from the image window is read from the base image as required during processing. This allows flexibility in the feature window size.

Traditionally, GCPs were points, such as a road intersection or a sharp corner of some region, which could be uniquely indentified in two images or in an image and a map. GCP features for correlation matching are larger clear physical shapes that can be identified in two images. If the images shapes are on top of each other, the correlation between the image windows is high. If the shapes are slightly shifted in any direction, the correlation between the image windows is lower. The higher the contrast between the shape and the background, the more quickly the correlation will decrease (or the more subtle a shift that can be detected). The shape must also be distinct in both east-west and north-south directions so that the change in correlation will be sensitive to misregistration in any direction. The shape must also be big enough compared to the window size to dominate in the correlation calculations. In the LCCA, Master GCP features are identified as the centre point of a ten line by ten pixel window. Figure 6.3.2 shows some examples of Master GCP features.

In correlation matching to locate the feature in a new image, the location to search for each feature is restricted to a twenty line by twenty pixel window around coordinates predicted by a simple linear relationship fitted to a small number of manually located starting GCPs. This is efficient computationally but also minimizes false matches to features with a similar shape elsewhere within the image.

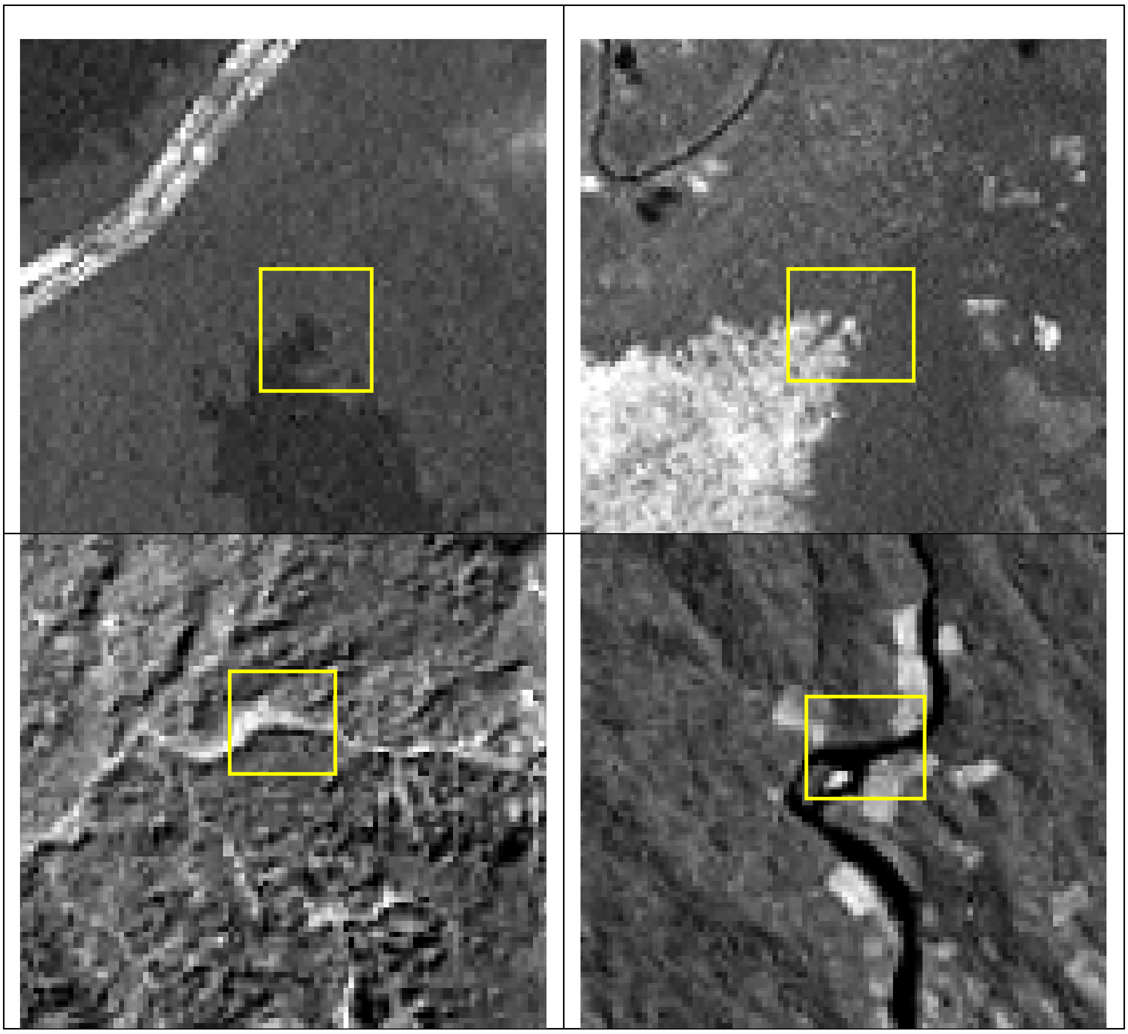


Figure 6.3.2. Some examples of Master GCP features from Kalimantan that are used in the correlation matching to generate GCPs. The whole region inside the yellow box forms the feature window. Band 7 of the ortho-rectification base image is shown in this display, but the correlation is performed over several image bands.

The Master GCPs used in the LCCA were collected manually by visual inspection of the mosaiced ortho-rectification base images. As well as being distinct shapes with high contrast to the background, Master GCPs also need to be features that are stable through time so that they can be used with any image dates. For example, natural clearings in a forest created by rocky outcrops are better than the edge of forest nearest to expanding agricultural lands. A distinct curve or peninsular in a rocky coastline is better than part of the waterline on a wide beach with a big tidal range. A bend in a river in the mountains that is tightly constrained by terrain is better than a bend in the same river in the low flat lands where flooding over successive wet seasons can change the main channel.

GCP chips were also obtained from USGS at the same time as the GLS2000 data was downloaded and evaluated for use in the LCCA. Inspection of these chips showed that many were fire fronts. They had very high spectral contrast within the chip image date, but no temporal stability and so were unsuitable for the LCCA. As a result, the LCCA adopted a manual selection process as described above. As locating Master GCP selection needs to be done once only, adopting a manual approach did not impose a high overhead.

A Master GCP file for an image is a thus text file which contains a list of geographic coordinates (easting and northing) in the UTM zone applicable for that image. The file name records the path/row and the UTM zone to which it applies. Where images overlap UTM zone boundaries, Master GCP files for both zones are stored.

Two sets of Master GCPs were selected. One – the Master Registration GCPs – serve to generate ground control points for fitting the ortho-rectification model, as described in the next section. The other – the Master Check GCPs – serve as sets of independent check points that are used during the quality assurance stage to assess how well the processed image is registered to the base. Typically about 100 points for each image were selected for Master Registration GCPs, and about 60 for Master Check GCP's. These numbers (if all are matched) are more than required for registration, but conditions in Indonesia (especially cloud) required such numbers for efficient rectification processing (below). The Master Registration GCP files and Master Check GCP files for all images are stored in the LCCA archive at LAPAN.

6.3.3 Ortho-rectification of individual images

Prior to subsequent processing in the LCCA, all Landsat images must be ortho-rectified and registered to the geometric base. The aim of this process is to produce individually rectified images, which will eventually result in a consistently-registered time series of imagery for all of Indonesia. As noted in Section 6.1, historic images were provided from different archives and with different levels of processing. The preferred processing level for historic imagery was path-oriented (nominal orientation) to enable full ortho-rectification correction within LCCA. Path-oriented images are referred to as 'raw' images in this section. In some cases, selected images were only available as Level 1T (L1T) ortho-corrected. This section describes the ortho-rectification process for path images. L1T images at the time were generally not registered to the base to LCCA standards and reprocessing was often required. Processing for these L1T images is also described below.

The ortho-rectification processing applies correlation matching to find in the overpass image the pixel locations which correspond to the base coordinate locations in the Master GCP file. The correlation calculations are performed using image bands 3, 4 5, and 7. Details of operator inputs to this process are described below. The result is a new file of GCP points for the specific image which contains only well-matched GCP locations in base map and raw image coordinates. Heights of the selected GCPs are then found from the DEM and attached.

These selected points are used in a single program ('geo-correct') which fits a sequence of three models in the actual ortho-rectification process of the raw image. The details are found in Wu (2008). This approach achieves accurate ortho-rectification without the requirement for sensor and orbital ephemeris information, only an approximate satellite height. The approach has been compared to established implementations using satellite orbital modelling and results agree very closely; for LCCA purposes they are identical. The resampling kernel used is 16-point sinc with Kaiser windowing.

The inputs to ortho-rectification are a raw overpass image, the base mosaic for the appropriate NUTM zone, the corresponding DEM, and the Master Registration GCP file for the overpass image. The major outputs are the ortho-rectified image, and the new final rectification GCP file with heights which was used. Other files including the interim GCP files and files providing diagnostics on the model fitting are also produced and forwarded for QA. Within the LCCA, software has been developed by LAPAN to simplify the operator processing and the generation of diagnostics for QA.

The steps for an operator in ortho-rectification of a raw image are summarized below. Full details of programs and operator instructions are found in the Operational Manual.

1. Display the raw image(s) and select 6 to 10 ground control points to link the raw image to the rectification base image. These GCPs may be any features, which can be seen clearly in both images and should be spread as widely as possible across the image. These points are used to fit a linear model in the correlation matching to locate the initial search window for matching each Master GCP. During this step any image quality issues – missing data, dropout, excessive striping - are noted.

2. Run automated correlation calculations to create an initial GCP file for the image. Image bands 3, 4, 5 and 7 are used in the matching calculations. GCPs matched with a low correlation (below 0.85) or with large residuals from a linear model fit (> 3 to 5 depending on the number of GCPs) are rejected. The output from this suite of programs is a matched set of control points in GCP file format and visual and quantitative diagnostics of the spatial distribution of GCP location and the spatial distribution of errors from a simple linear model fit. An example of a visual diagnostic produced is shown in Figure 6.3.3. Red symbols in the right picture indicate error greater than two pixels and indicate a need for review, especially if they occur in regions with few 'blue' GCPs. Both pictures show an adequate spatial distribution of GCPs.

3. Operator review to add/adjust GCP's to ensure an adequate spatial distribution and sufficient visual verification of all GCPs with large errors from a simple linear model fit. The result is a reduced set of GCPs for the image. At least 20 – 30 GCPs are expected.

4. Attach heights to the selected GCP's from the DEM (via programs). The result is the final set of GCPs for the image which is used in the ortho-rectification.

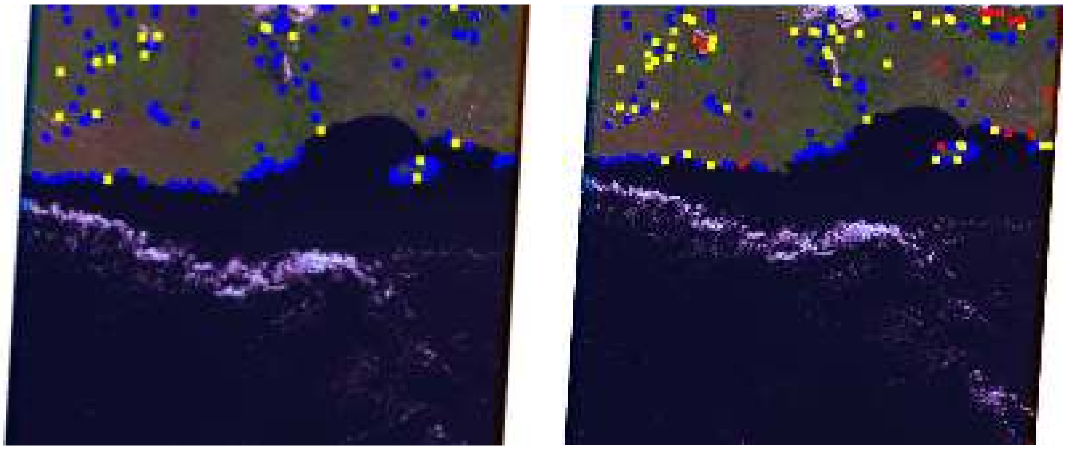


Figure 6.3.3. Examples of diagnostics produced for a scene for operator review of GCPs at step 3 above. Blue indicates GCPs with errors less than one pixel from a linear model fit; yellow indicates GCPs which have errors between one and two pixels; red indicates error greater than two pixels. The results in the left figure represent good quality of GCPs, whereas the right represents of poor GCPs which should be revised.

5. Run geo_correct to ortho-rectify the image. This is set up by running a preliminary program which sets up parameters and acts as a record of the input and output file.

6. Save all required files for QA. Specifications and names of required files are found in the Operational Manual.

Images which were supplied as L1T have already been through an ortho-correction process. After importing they are resampled to 25m pixel size and the QA process applied as described below for the images ortho-rectified by the LCCA team. Approximately 80% of L1T images were well registered to the base, although this percentage was lower for images with only small islands or extensive cloud with only limited clear land area. If an L1T image was found to be poorly matched to the base, a set of GCPs was derived for the image manually. A linear model only was fitted to these GCPs to re-rectify and resample the image. QA was then performed on the new result. This was found to produce satisfactory results in nearly all cases. In recent years, the consistency of automated Landsat processing by USGS and other agencies has improved following feedback from the LCCA program to USGS. LAPAN has implemented USGS processing of Landsat images currently received in Indonesia, and it is anticipated that images from this processing will be also submitted to the same QA process.

Ortho-rectification QA is carried out by a nominated group of independent staff. The final result is critically dependent on the operator's checking and decisions for the accepted GCPs during the processing. The requirement for acceptable registration to the base is that the overall misregistration be less than one pixel with no systematic patterns in the shifts between the new image and the base. For example, if the whole image is shifted half a pixel to the right of the base, then the registration is not acceptable even though the shift is less than one pixel. If the shifts are in random directions, overall residuals of half a pixel are acceptable.

The registration is assessed both visually and quantitatively. The correlation calculations are repeated between the ortho-rectified image and the base image at Master Check GCP locations. These check GCP locations are independent of the Master Registration GCP locations. The size of the correlation value and the size of the residuals from a simple linear fit are used to exclude poorly matched features from the statistical summaries. The RMS error and pictures of the spatial distribution of the well matched Check GCP locations, size of shifts and direction of shifts are produced – similar to those produced during the generation of the original GCPs for the model fitting. If these summaries show an acceptable registration, only a quick visual inspection is performed. If the summaries show possible problems a much more detailed visual inspection is performed to diagnose the problem and suggest possible solutions.

The final check of the rectified image is visual: one band of the rectified image (usually band 5 or 7) is overlaid on the corresponding band of the base in red and green, while standard image displays of both are viewed simultaneously. The QA operator will zoom to different parts of the image and check for any evidence of misregistration in unchanged features; the standard image displays provide guidance on cloud and changes in land cover which may be ignored. An example of the two-image red-green visual display is shown in Figure 6.3.4. Here dark and yellow colours indicate dark and bright areas which are 'unchanged' in brightness; red and green tones indicate change in brightness at that location. The displacement of roads and river edge in the left image is clear evidence of misregistration, while the right image appears well registered on these clear linear features. There are noticeable red areas in the river in the right image which would be carefully examined at QA. These are likely to result from changes in river level or turbidity rather than misregistration, in which case QA would pass this portion of the image. Note also the red and green coloured areas which result from land cover change. Where an image fails QA, it is returned to the ortho stage with comments and pictures such as shown in this figure, to assist in refinement of the selected GCPs.

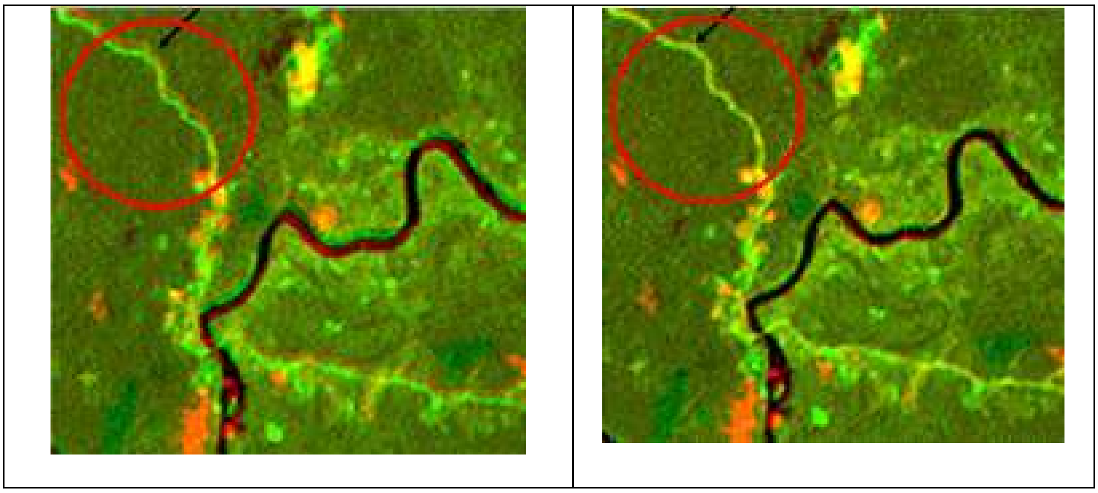


Figure 6.3.4. Detail of ortho-rectification QA visual checking, using a display of a rectified image band and base image band in red and green. Dark and yellow colours indicate dark and bright areas which are 'unchanged'; red and green tones indicate change in brightness at that location, The displacement of roads and river edge in the left figure is evidence of misregistration, while the right image does not show such displacement. The left ortho image would be failed at QA. These images are of Papua.

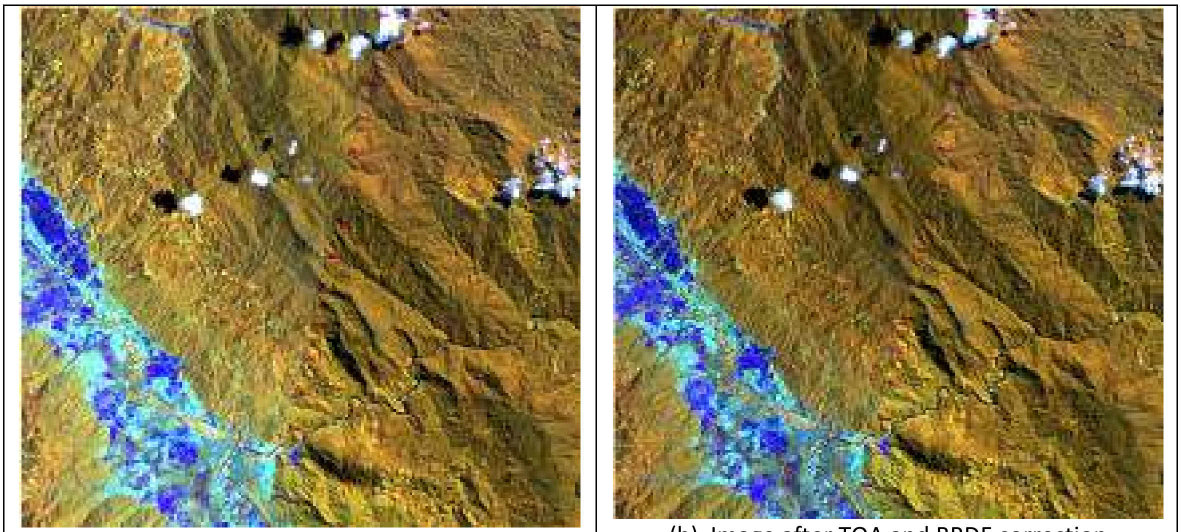
Archiving of files is the final stage of the ortho process; the files to be archived are described in the Operational Manual.

6.4 Radiometric correction - BRDF

The digital numbers recorded in Landsat images are affected by on-board processing parameters, the distance to the radiation source (the sun), angular effects due to variations in solar incidence angles and viewing angles. The variations due to angular effects are known as the bidirectional reflectance distribution function (BRDF) of the sensed surface, and are dependent on the land cover and wavelength. BRDF manifests as slight variation across an image and can be seen most obviously as 'edge effects' in uncorrected mosaics (Figure 6.4.1). In order to apply subsequent numerical classification processing in LCCA it is desirable that digital values are consistent over space and time.

In the LCCA program, correction procedures are applied to each image following ortho-rectification, using information extracted from the image metadata files; solar position and angles are calculated from the date and time of overpass. The required inputs are the ortho-rectfied Landsat image, and the image metadata file which is supplied in a range of formats depending on the original image source. The outputs are the corrected image and an ancillary file recording the correction parameters extracted from the metadata files in a common format.

The initial correction (implemented through the program sun_correct.exe) performs two steps. It corrects to scaled top-of-atmosphere (TOA) values using processing coefficients recorded in the metatdata and earth-sun distance calculated from the overpass date and time (Vermote et. al. 1994). The program then applies a BRDF correction to all bands of the image. A common two-kernel empirical BRDF function is applied to all images (Danaher et. al. 2001). The kernel functions and coefficients are the same as that used in Australia's NCAS processing (Furby 2002) which was optimised to correct for BRDF over forest land cover. The results were evaluated in the early stages of the LCCA on images from different parts of Indonesia and found to produce an acceptable correction. An example of two overlapping images before and after this correction is presented in Figure 6.4.1; the effect of the correction in reducing the edge between images is clearly seen. The result is a corrected image (referred to in the LCCA as the 'BRDF-corrected image') which is then submitted for QA.



(a) image before TOA and BRDF correction

(b) Image after TOA and BRDF correction

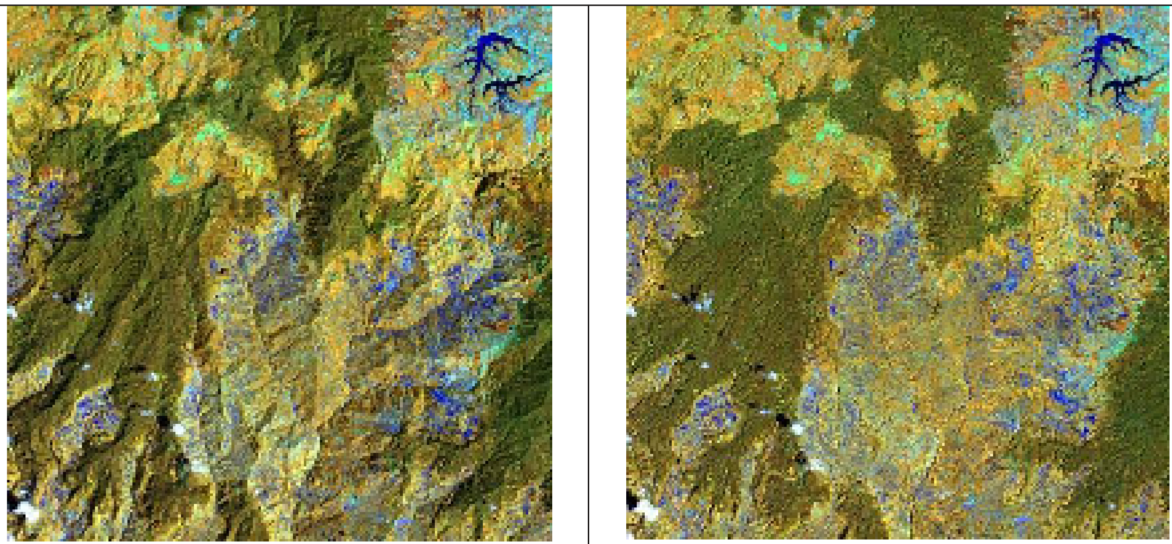
Figure 6.4.1. Detail of two overlapping images from south-west Java (bands 4,5,3 in RGB) before and after correction.

The QA process for BRDF-correction involves checking of the filenames and contents and visual inspection of the output images. The BRDF process is essentially fully automated and QA failures are few. The correction is applied immediately after the ortho-rectification step and it is the BRDF-corrected image that is viewed during the ortho-rectification QA.

6.5 Radiometric correction - terrain

Indonesia is a mountainous country and terrain illumination effects can be seen in most remote sensing images used in the LCCA. Within a single image, variations in slope and aspect will result in variations in incident energy, and thus affect the reflected energy and the recorded digital number in the Landsat image. This is seen in visual images of mountainous areas (e.g. Figure 6.4.1 above) - slopes facing away from the sun appear darker and the slopes facing towards the sun appear brighter, even when they have a common land cover. Sun angles vary with date, so that between images from different dates (e.g. adjacent images in mosaic) these terrain illumination effects will vary in direction, location and magnitude. It is essential to correct for these terrain effects before attempting to apply a numerical land cover classification to imagery from mountainous areas.

Terrain illumination correction is implemented in the LCCA and applied to individual images after the BRDF correction described above. The inputs required are the 'BRDF-corrected' image, the sun angle parameters in the common file format produced during the BRDF-correction and a digital elevation model (DEM). The DEM used in the LCCA is the SRTM-DEM at 90m resolution. This was the best available DEM with national coverage when the project commenced. The SRTM is generally co-registered well with the LCCA geometric base, but does have some large areas of low quality data (spatial artefacts). The principal output is the 'terrain-corrected' image. The correction parameters and a 'line-of-sight' image recording the angle below which the input solar illumination is blocked by terrain are also archived.



(a) Image before terrain illumination process

(b) Image after terrain illumination process

Figure 6.5.1: Sample image area from West Java before and after the terrain illumination correction. Bands 4, 5, 3 shown in RGB.

Line-of-sight (LOS) calculations are run in a separate step to identify areas of true shadow. In areas with very steep slopes, there may be areas of true shadow which do not receive direct illumination. Data from these pixels cannot be corrected and are essentially 'missing'. Their location can be determined from the DEM and solar position; they are identified in the 'line-of-sight' (LOS) calculation step and later replaced with null values.

The mathematical method to perform terrain corrections is described in Wu et al (2004); it implements a modification of the C-correction of Teillet et al (1982). If the angles (slope and aspect) of the surface for each pixel are known, and correction coefficients for the land cover type are known or can be estimated, correction can be applied. Slope and aspect are derived from the DEM. Estimation of correction coefficients for each band is a crucial step; these are ideally estimated for each image for the land cover type of greatest interest - forest cover in the LCCA. If images are cloud free, this coefficient estimation step can be automated using only the BRDF-corrected image, the DEM and a raster mask of forest extent image.

In Indonesia, cloud coverage affects almost every image, so operator intervention is required. Two approaches are available for coefficient estimation. In both cases the results are submitted for QA.

1. The original and preferred approach; coefficients are estimated for each scene from an operator-selected sample of pixels. The operator must select by digitising sample areas from the image; the sample pixels should include forest cover, cloud free and include areas of varying terrain (angles). A program to estimate coefficients is then run, coefficients are examined, the correction is applied and the results examined.

2. Use of default 'library' correction coefficients. This is applied when the standard approach fails for any reason (typically when cloud or haze coverage makes selection of sufficient samples for estimation impossible). A default set of coefficients is applied. Images are allocated to one of five classes according to their amount of terrain variation. An 'average' set of coefficients for each class has been created from parameters successfully estimated (passed QA) in the LCCA from different parts of Indonesia. These coefficient sets and the path/rows in each terrain class are recorded in the Operational Manual.

For an operator, the basic steps in performing terrain correction of an image using the standard approach are summarised below and in Figure 6.5.2. If library coefficients are used they will be input at step 6 below.

1) Ensure the BRDF-corrected images, the DEM image and a forest mask are available.
2) Calculate a line-of-sight image in preparation for terrain-illumination correction.
3) Create a vector file to identify clear pixels in region of terrain effects as training samples for the estimation of the coefficients for the correction.
4) Merge the training vector and forest masks to identify the candidate pixels to use in the coefficient estimation.
5) Estimate the parameters for the terrain illumination correction.
6) Apply terrain illumination correction to the image.
7) Prepare the data for QA
8) Archive the final image and ancillary data (as passes the QA assessment).

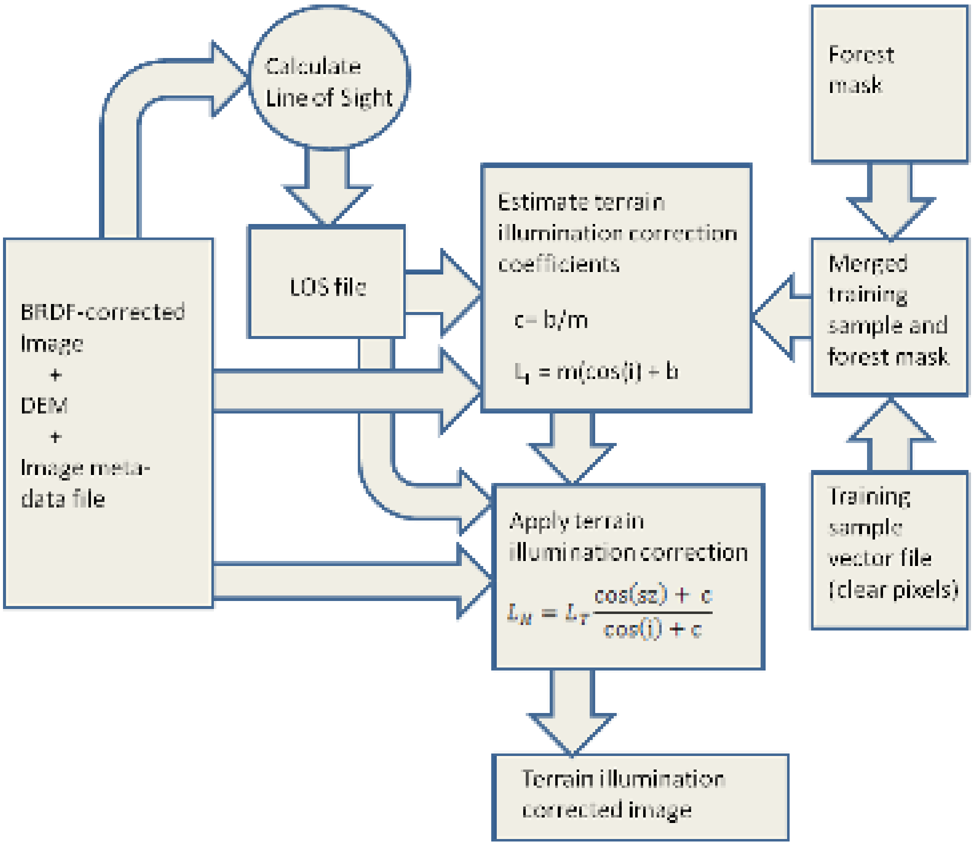


Figure 6.5.2. Flowchart of the terrain-illumination correction process

A small number of 'flat' regions in Indonesia have been identified where terrain variation is so small that correction is not required. The BRDF-corrected image is simply renamed. The path/row numbers of these images are listed in the Operational Manual.

The QA process for terrain illumination correction is performed by an experienced independent team member. The operator submits output files, and any operator-defined vector files. These include outputs from the LOS processing.

The primary QA process is visual assessment of side-by-side images before and after correction, assessing the LOS areas and results in areas of different terrain for all bands. The correction coefficients are also checked against expected ranges. Effects of poor coefficient estimation are usually obvious and corrected by the operators prior to submission. QA will fail images if visually obvious terrain-illumination effects remain that do not correspond to artefacts in the DEM. It is noted here that DEM limitations in resolution, quality and in some cases registration, result is some local artefacts which cannot be corrected. Following QA, all files are archived. The terrain-corrected image is then used in subsequent processing steps.

6.6 Cloud masking

Cloud cover is a major problem for applications using optical imagery in Indonesia and other tropical regions. It was a major concern in the LCCA from the initial planning stages. Dense cloud cover obscures the land and makes the affected pixels unusable for land cover mapping. Thin cloud, cloud shadow and haze affect the digital reflectance from the land cover and make quantitative processing impossible. The frequency of cloud in many parts of Indonesia is such that cloud-free Landsat images are extremely rare, and even images with useably large cloud free areas are few (and in some cases nonexistent) within a calendar year. As discussed in Section 6.1, a response to this problem in the LCCA was the purchase of multiple scenes for a path/row within a year, in order to create a composite mosaic with the maximum cloud-free land area. To produce such mosaics, all cloud-, haze- and shadow-affected pixels must be removed. Pixels affected by smoke haze and data corrupted for other reasons must also be removed. Methods were required in the LCCA to identify and remove such pixels from Landsat scenes. In the Australian system, where cloud cover problems are much less, manual digitizing of cloud had been the chosen solution. It was recognised that an automated or semi-automated system was required for the LCCA.

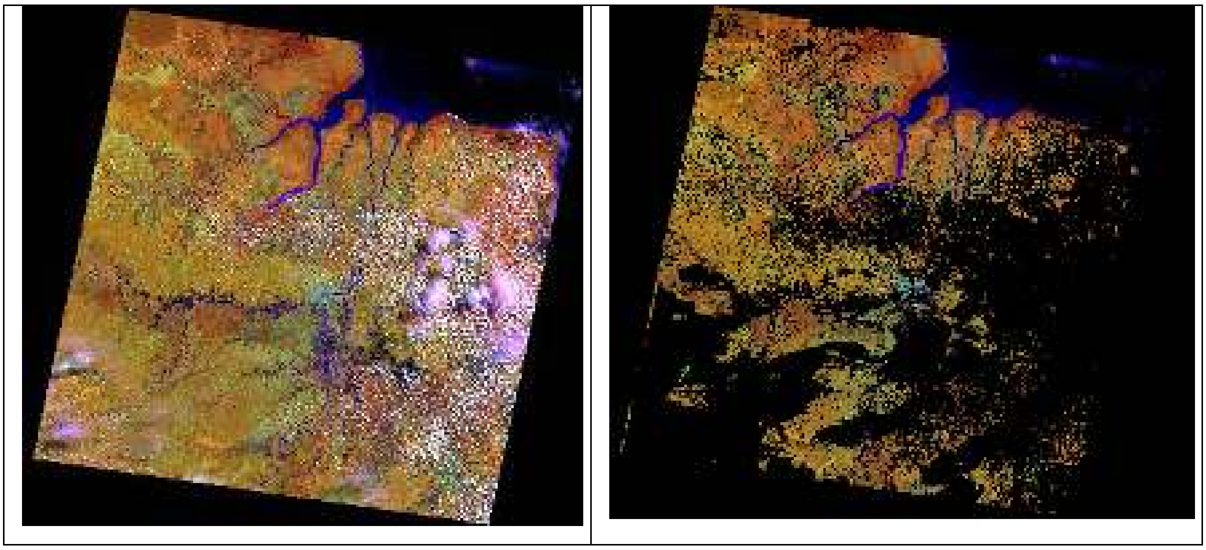


Figure 6.6.1. Example of a terrain corrected image and the cloud masked version of the same image. The image shown is Landsat 7 path/row 124/62 from 10 May 2001, bands 453 in RGB.

At the time of commencement of the LCCA, operational methods suited to the LCCA were not available. Cloud masking has been an area of research and development within the LCCA. The approach taken was to examine international research and best practice, to test and adapt these approaches to develop an effective method. The development of the current method has been conducted by LAPAN, with input from CSIRO and Professor Matthew Hansen (University of Maryland). The current approach is a combination of automated steps and manual refinement (digitising). The method has been developed and refined within the program in a process of continuous improvement to maximise accuracy and efficiency. This section provides a brief background on the methods development, and an overview of the current method.

In 2009-2010, effective operational cloud masking had been developed for AVHRR and MODIS processing streams, but cloud masking of Landsat was an active research area. Hansen et al (2008) identified cloud in Landsat using a tree classification method, with statistical calculation incorporating MODIS data (simultaneous with Landsat). Another method was suggested by Wang et al (1999), where cloud is identified using wavelet methods (Landsat bands 1 and 5 are used). Martinuzzi et al (2007) identify cloud on Landsat images by using Band 1, 4 and 6 and also use additional parameters (sun angle information and topography) to define shadow direction from identified cloud. Irish et al (2006) use a method to identify cloud which involves 5 bands of a Landsat image.

In terms of operating international systems at that time, Professor Matthew Hansen (then at South Dakota State University (SDSU)) was developing an automated tree-based cloud masking step within a Landsat time series processing stream, and was a collaborative partner in the INCAS-LCCA. His method was semi-supervised, involving the generation of trees from training data where each branch of the tree was a split on a spectral index and threshold. Eventually by these successive index-threshold splits, each pixel was classified as cloud-affected or not. The result was a cloud mask for the image. At SDSU, the index-threshold rules determined from training data were fixed and applied to calibrated imagery within a processing stream. In December 2010, through the INCAS collaboration, Hansen provided a training course at LAPAN in this tree-based approach, and how to train and implement it in commercial software package. In practice, the method proved difficult to implement in the LCCA; results (the masks from decision trees) from different operators proved highly variable. When trees derived from one set of images were applied to other images the results were often unsatisfactory. In the SDSU system, a common tree is derived from training on very numerous Landsat images. SDSU assisted further by applying their own software in the US to a sample of 25 LCCA terrain-corrected images. The results worked well in many but not all situations of cloud, shadow and land cover in Indonesia, and it was difficult to apply improvements directly to the results. This experience provided valuable learnings to the LCCA; in particular identifying the thermal band as important for detection of thin high cloud and also providing experience in applying indices and thresholds to the problem.

In response to these problems, LAPAN commenced development of a method which combines rules from spectral indices and angular information to produce a cloud mask. The development commenced with testing of indices and thresholds to identify cloud, haze and shadow from clear land cover using statistical methods. A combination of index-threshold rules were identified, thresholds selected and applied to images, and the results evaluated. This process identified cloud/shadow and land cover combinations where cloud masking was difficult, and those which were effectively masked by different indices. Land cover types which were erroneously masked (commission errors) were also identified. These findings were used to iterate the analysis process and to identify new, potentially more effective indices.

In general, it was found that bright clouds were effectively masked by brightness 'albedo' indices but that thresholds which identified all such cloud typically resulted in errors of commision by masking areas of bright land cover (such as roads or bare dry land). Cloud shadow areas are 'dark', but variable according to cloud density and land cover; masking such areas usually resulted in commission errors (over-masking) of 'dark' land covers such as dark forest or irrigated rice fields. Clouds and their shadows are of course linked and the location of one can be predicted from the other if the solar position and cloud height are known. This information was used to remove many errors of commission by using solar angle information (extracted from the satellite metadata), and estimates for the range of pixel displacement between a cloud and its shadow (arising from cloud height) derived from numerous images in Indonesia. The albedo values for smaller or thinner clouds or near the edges of thick cloud tend to overlap those for some ground covers. 'Possible' cloud is identified by a second, lower albedo threshold. Such areas are added to the mask if they are 'close' to the bright clouds or 'paired' with a shadow feature. The logical rules to implement this refinement may be described as follows:

1. Bright pixels are classified as 'cloud' using an albedo index-threshold rule; thresholds are chosen ('certain' and 'possible') to classify all visible cloud. Bare land is excluded from these pixels using another index-threshold rule. A threshold applied to the thermal image band is chosen to classify thin cloud as 'certain' cloud.

2. Dark pixels are classified as 'shadow' with water excluded using index rules, similarly.

3. Angular refinement. The regions found in (1) and (2) are 'grown' by a buffer within the program to include possible thin cloud/shadow and mixed pixels and to allow for varying cloud heights. The relationship between a cloud pixel and its shadow is calculated from sun angles as a displacement in number of pixels in x and y directions.

   (a) From pixels classified as 'possible cloud', the position of corresponding shadow is predicted, if these pixels are not classified as 'shadow' in (2), the cloud classification is assumed to be an error and removed.

(b) From pixels classified as 'shadow', the position of corresponding cloud is predicted, if these pixels are not classified as cloud in (1), the shadow classification is assumed to be an error and removed.

[Allowance for missing matches near image borders is made in this step; it is also noted that the missing data from LS-7 SLC-off images affects this part of the algorithm as matching pixels for cloud or shadow may be missing.]

4. Filling; small gaps within and between cloud and shadow regions are filled. Fixed thresholds define 'small'.

5. Remaining classified pixels are 'grown' by a buffer to ensure mixed pixels are included. The default buffer 'grow distance' is 3 pixels.

The rules are implemented in a single software program. In the initial development and testing of the approach, operators were required to set thresholds manually for each image, using interpretation on screen to classify cloud and shadow; the angular refinement correction was then run. With experience, default thresholds were identified, used as starting values and found often to be effective. Now three sets of default thresholds for the indices have been recorded. A program is run which automatically produces three versions of the automated mask. The operator selects the best for subsequent steps – or if none are satisfactory, is required to reset thresholds manually.

Following this processing, the resulting cloud and shadow masks are examined visually and a manual digitising step is typically carried out to remove any remaining unmasked bad data (most often haze); or to identify any large areas of commission. The results of the digitising and original masks are combined and then used to relabel pixels in a masked copy of the image using supplied programs.

The processing steps are described and illustrated below.

Step 1: Run a batch script that uses the cloud masking program (LPNcloudshadow_v8). Input data files are (1) the terrain corrected Landsat image (2) a text file generated at earlier steps which contains solar position information (3) the default thresholds sets. The outputs are separate cloud and shadow mask files; these are single band raster classification files. Three sets are initially produced using the default parameters; only the best is kept.

The index formulas used are as follows:

(a) To identify bright cloud, two indices are combined, a simple brightness and a second index which helps to retain bright bare ground areas

$$B1+B2+B3)/3 > T_{a1}, T_{a2} \text{ and } (2*B1-B2-B3+2*B4-2*B7) > T_b$$

Where B1...B7 are the digital numbers for the Landsat bands, T_{a1}, T_{a2} and T_b are certain and possible thresholds for the albedo and bare land indices respectively

(b) To identify high cold cirrus cloud, which may be thin and omitted by (a) above, a simple threshold (T_c) is applied to the Landsat thermal band

$$B6 < T_c$$

(c) To identify shadow (dark) two indices are used in combination. The second index helps to separate water from shadow

$$(B4+B5)/2 < T_s) \text{ and } (B2+B3-B5) < T_w$$

T_s and T_w are referred to as the shadow threshold and the 'water' threshold respectively

As noted above, sets of default thresholds have been defined for this step. They can be seen below with angular information from a particular image.

Table 6.6.1. Default threshold coefficient sets used in running the cloud masking program

_def1	_def2	_def3
Batas Awan Albedo bawah : 20 Batas Awan Albedo atas : 30 Batas Awan Lahan Terbuka : 130 Batas Awan Thermal : 125 Batas Bayangan Awan : 45 Batas Air : 25 AZIMUT : 100.7864 Posisi Bayangan dx-pixel : 98 Posisi Bayangan dy-pixel : 18 Tinggi awan Maximum-pixel : 40 Grow Distance Awan : 3 Grow Distance Bayangan : 3 Border SLCOFF : 1	Batas Awan Albedo bawah : 25 Batas Awan Albedo atas : 35 Batas Awan Lahan Terbuka : 130 Batas Awan Thermal : 125 Batas Bayangan Awan : 45 Batas Air : 25 AZIMUT : 100.7864 Posisi Bayangan dx-pixel : 98 Posisi Bayangan dy-pixel : 18 Tinggi awan Maximum-pixel : 40 Grow Distance Awan : 3 Grow Distance Bayangan : 3 Border SLCOFF : 1	Batas Awan Albedo bawah : 25 Batas Awan Albedo atas : 35 Batas Awan Lahan Terbuka : 130 Batas Awan Thermal : 120 Batas Bayangan Awan : 40 Batas Air : 25 AZIMUT : 100.7864 Posisi Bayangan dx-pixel : 98 Posisi Bayangan dy-pixel : 18 Tinggi awan Maximum-pixel : 40 Grow Distance Awan : 3 Grow Distance Bayangan : 3 Border SLCOFF : 1

The 'best' thresholds are not necessarily the ones that produce the most accurate masks. They are the ones that produce a result that can be corrected with the least amount of manual digitising effort. For example, small scattered clouds are very time consuming to digitise so it is better to set the thresholds to mask all such cloud and perhaps 'over-mask' roads or recently cleared land where these can be corrected with a few large polygons. Similarly, a large area of thin haze may be omitted from the automatic mask to prevent over masking of agricultural lands because it can be digitised with a single polygon.

Examples of thresholding setting for the components of the cloud and shadow masking indices are shown in figures below.

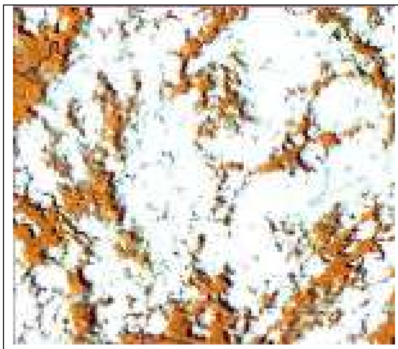


Image L5_116059_090606	Brightness index mask (red) with a threshold (T_{a1}) of 45

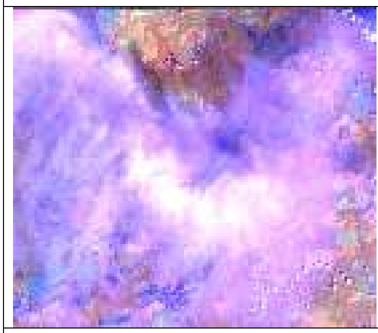



Image L5_116059_090606	Brightness index mask (red) with a threshold (T_{a1}) of 40

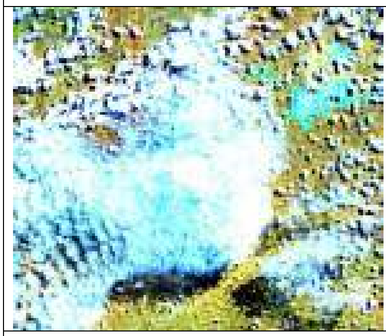



Image L5_116059_090606	Brightness index mask (red) with a threshold (T_{a1}) of 30

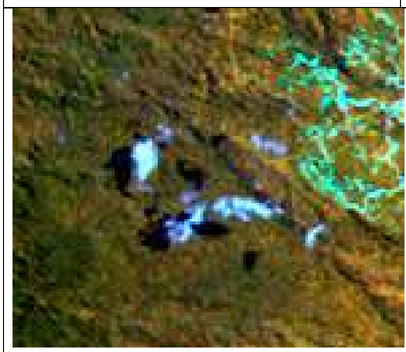
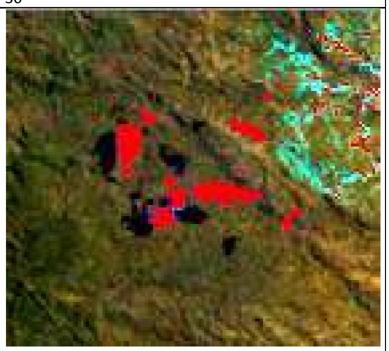

	

Image L5_122059_090807 original	Brightness index mask (red) with a threshold (T_{a1}) of 25

Figure 6.6.2. Paired input images (left) and masked brightness index (right) in red produced using the indicated brightness threshold. Bands 453 in RGB.

In Figure 6.6.2, it can be seen that haze in the second image may not be completely masked, while in the upper right of the two lower images some ‘over-masking’ of bright land as cloud (errors of commission) are apparent. In the bottom image this occurs in a cloud-free area and can be quickly corrected with manual digitizing at the next stage.

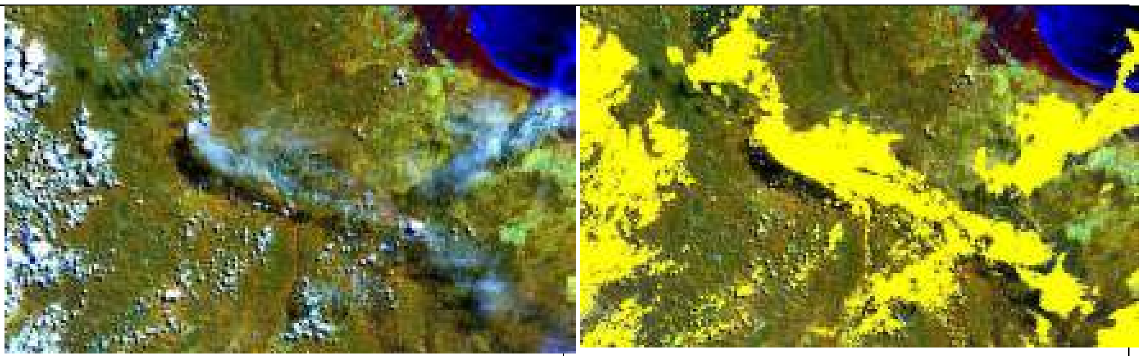
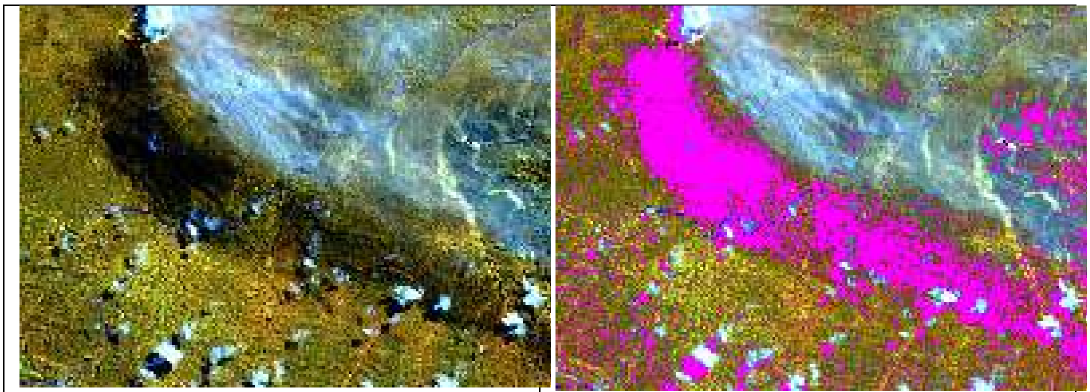


Image L5_116059_090606	Thermal mask (yellow) with a threshold (T_c) of 115

Figure 6.6.3. Paired input images (left) and masked thermal index (right) in yellow produced using the thermal image only. Bands 453 in RGB



L5_116059_090606	Shadow mask (pink) with a threshold (T_s) of 55

Figure 6.6.4 Paired input images (left) and masked shadow index (right) in pink produced using the indicated brightness threshold. Bands 453 in RGB

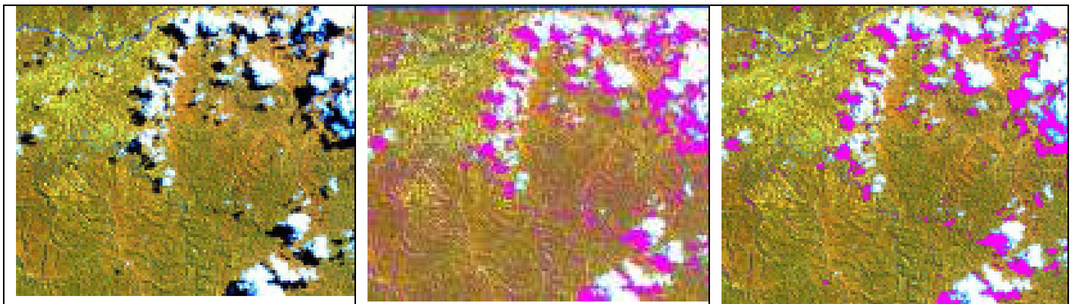


Image L5_116059_090606	Shadow mask (pink) with a threshold (T_s) of 55	Shadow mask (pink) with a threshold (T_s) of 45

Figure 6.6.5. Further example of an image and its masked shadow index showing the effect of varying the threshold. The mask is covering dark areas associated with terrain which have remained after terrain correction; adjusting the threshold removes most of this error while still masking the true shadow. Bands 453 in RGB.

Step 2. Operator must review the results by visual inspection. A good set of threshold-derived masks will cover all visible cloud and shadow with minimal commission error, or with errors which can be easily digitised and corrected in the subsequent steps. Errors such as shown in Figure 6.6.5 (centre) will indicate that revision of thresholds is required, and the operator will be required to perform manual threshold adjustment.

Step 3. Digitising to edit the cloud and shadow masks. The operator must check for errors of omission and commission. Vector outlines are digitised manually to add areas to the mask and delete errors of commission; the vectors are known as the 'add' and 'delete' vectors. All omission errors (bad data which remains after masking) are important and must be removed. It is desirable to remove commission errors as they represent masked good data which otherwise will not be used. Typical errors of omission will be thin haze and shadow; often these affect large contiguous areas and can be digitised quickly. Causes of commission errors are various, relating to land cover as well as cloud and shadow signatures, and may occur as many small areas. Where commission errors occur in cloud-free regions (Figure 6.6.2), they can be corrected easily. Where commission errors occur as small patches within patches of true cloud and shadow, the digitising effort may be considered too great. An example is shown below.

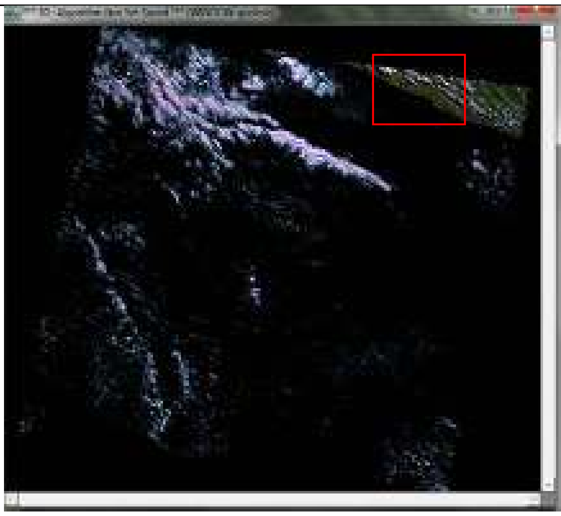
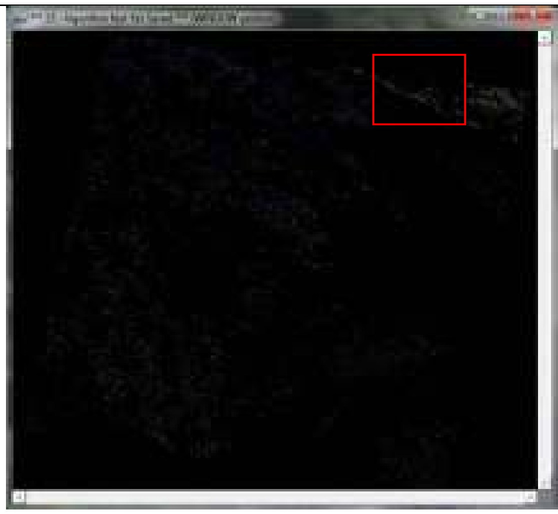
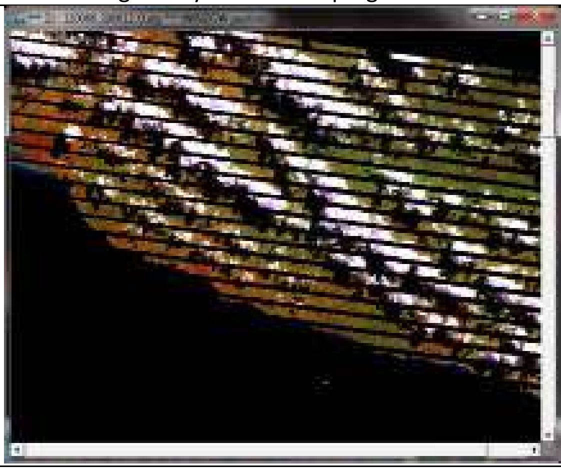
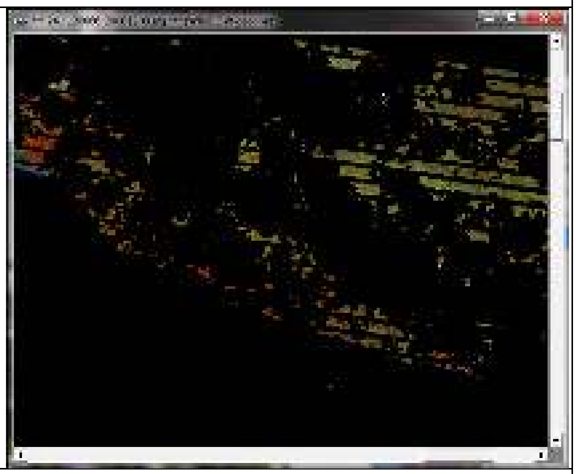
	
Original Image L7_120066_290612. This is an LS-7 SLC-off image. Only land is at top right.	The masked image image before editing
	
Detail of above: area indicated by the red rectangle above	Detail of above; same area as at left

Figure 6.6.6. Example. Input image and outputs from the automated cloud masking program prior to editing. Bands 453 in RGB.

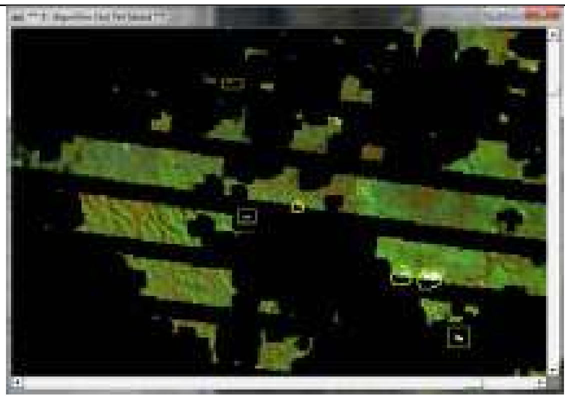
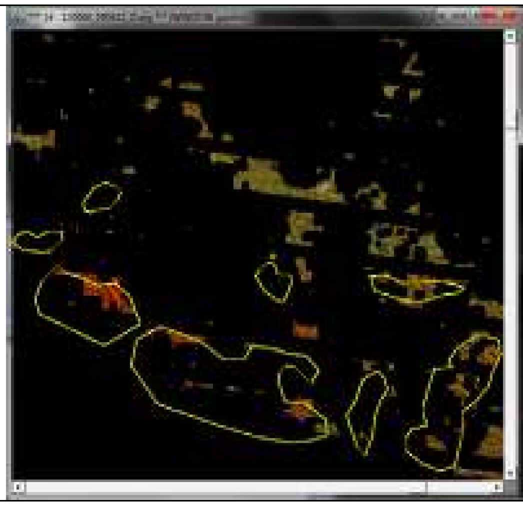
	
Further detail of original Image L7_120066_290612	Cloudmasked image for same area as at left. Yellow lines indicate digitised cloud areas to be added to the cloud mask
	
Detail of original image showing digitised commission errors which are deleted from the mask.	Same area and same vectors as at left on the maked image from thresholded masking program

Figure 6.6.7. Example of manual digitising correction of cloud masks. Input image and outputs from the automated cloud masking program together with manual edits (vectors). Bands 453 in RGB.

Step 4. Combining masks and creation of the final masked image for QA is done by running a program; the inputs are the terrain-corrected image, selected mask raster files and the two 'add' and 'delete' vector files. The output is the masked image, and new 'final mask' files produced from modifying the input masks with the vector deletions and additions. Creating final raster masks allows them to be applied to other versions of the image, for example to mask the BRDF-corrected image.

QA for cloud masking is conducted by an independent experienced operator who essentially begins at step 2 above, examing the image, the vectors, and the cloud masked image. Experienced operators can readily detect whether any problems are likely to be fixed by revising the thresholds for cloud or shadow. In such cases, the image will fail and the operator should be advised of the suggested changes to be tried. If the threshold classification appears adequate, and some errors remain, then vector editing is required. Any unmasked cloud or shadow data requires correction; and 'overmasking' will be assessed and, if significant and correctible with reasonable effort, the vector file must be revised. Revisions to the vectors are usually performed by the QA operator.

6.7 Mosaicing

The mosaic process produces a composite of all LCCA images within a fixed spatial extent for each year of LCCA coverage. These mosaic units form the basis for subsequent processing and products. Mosaicing can start when cloud masking of all LCCA images within the mosaic area for a year is completed. Within the mosaic process, multiple cloud-masked images within each path/row for that year are composited. The result is a single image, the 'mosaic image' for that year, covering the specified extents with minimum missing data. Only the six multi-spectral image bands are in the mosaic. The thermal image band is not retained after cloud masking is complete.

The mosaic extents are defined in NUTM projection. They have been modified from standard UTM map zones in order to include complete islands or island groups of Indonesia as far as possible. They are referred to as INCAS-LCCA mosaic tiles, and named for the standard map zone with greatest area within the mosaic. There are 13 standard LCCA mosaic tiles covering Indonesia (Figure 6.7.1).

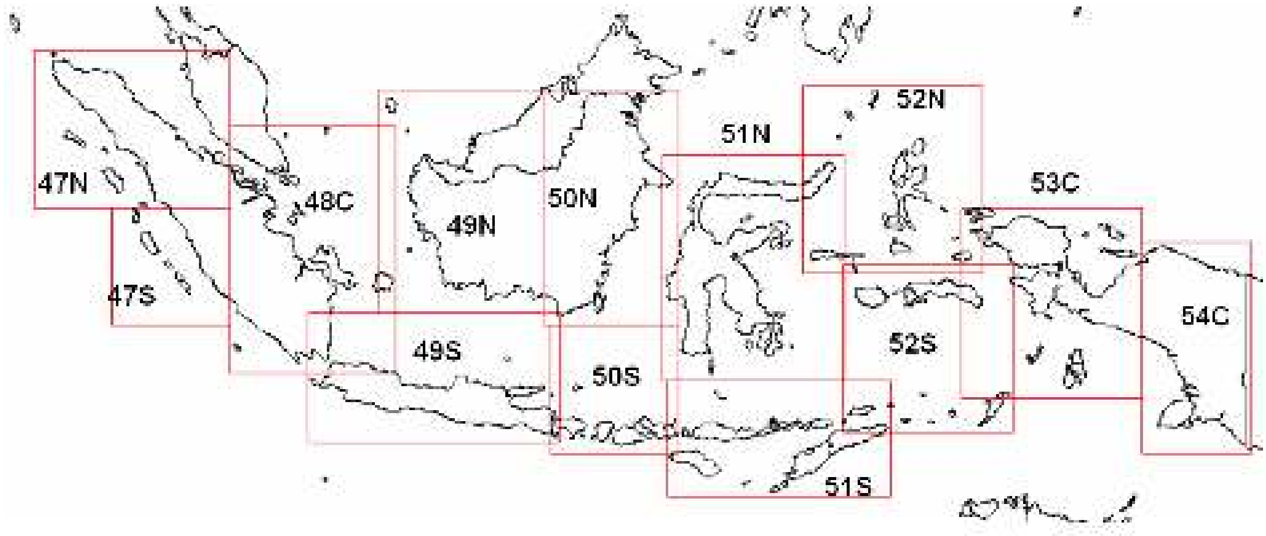


Figure 6.7.1: LCCA mosaic tiles for Indonesia

As with the individual images, all mosaics within the LCCA are referenced to NUTM coordinates (WGS84 datum). If conversion to SUTM is required for any reason, only a simple constant shift (or offset) in northing is required. Table 6.7.1 lists the extents of the standard LCCA mosaics in NUTM coordinates. These extents include an overlap with adjacent mosaic tiles. The file size for each mosaic tile as a six-band Landsat mosaic is very large. For convenience in storage and processing, each mosaic tile is divided further into standard 'quadrants' with a maximum file size of 2Gb (named e.g. 48C_nw for the northwest quadrant within mosaic tile 48C). The names and extents of these quadrants are found in the Operational Manual.

Table 6.7.1 . The extent of LCCA Mosaic Tiles; coordinates are all defined in WGS84, NUTM projection in the zone number indicated by mosaic name

Description	Map Sheet ID	Top Left	Bottom Right
Sumatra NW	47N	0, 750 000	950 000, -100 000
Sumatra SW	47S	300 000, 50 000	950 000, -600 000
Sumatra E	48C	100 000, 415 000	930 000, -750 000
Kalimantan W	49N	50 000, 550 000	930 000, -500 000
Java	49S	-210 000, -530 000	960 000, -1 000 000
Kalimantan E	50N	90 000, 550 000	800 000, -560 000
Nusa Tenggara W	50S	150 000, -450000	800 000, -1 100 000
Sulawesi	51N	-40 000, 280 000	875 000, -790 000
Nusa Tenggara E	51S	-10 000, -675 000	1 000 000, -1 250 000
Maluku N	52N	-100 000, 621 000	780 000, -350 000
Maluku S	52S	75 000, -250 000	920 000, -1 000 000
Papua W	53C	-100 000, 50 000	760 000, -840 000
Papua C	54C	-30 000, -110 000	580 000, -1 100 000

The creation of the mosaics proceeds as follows.

First, all of the cloud masked images within the mosaic extents from the chosen year are assembled. The majority will have been rectified to the required projection for that zone. Edge images and images in overlap areas which require reprojection are identified, and are reprojected using cubic convolution resampling. The operator will display all images from the same path/row and perform a visual assessment, with particular attention to any visual edge effects within an image and comparison of the location of clear areas. If any areas of data within a scene are judged to be poor quality for any reason, a simple vector is drawn and a program is run to mask these areas. The most common cause for this is subtle haze effects which are not detected in the masking step that are observed when observing data from the same area on two or more dates. Areas of scattered clear data where another image has complete clear data may also be manually excluded from the mosaic by this process. Low resolution 'snapshot pictures' are created from each image to allow faster comparison of images and the creation simple simulated mosaics.

The next step in mosaicing is to determine a priority order for mosaicing of images over the whole tile. Within path overlaps and where multiple images are available within a path/row, choices are made about the data from image is retained in the mosaic. The purpose of this process is to produce the best possible mosaic for subsequent numerical processing in the LCCA. The output from this step is a 'mosaic order' list which specifies which images are to be mosaiced on top of others. Data pixels from higher scenes will be used, but any missing data pixels in these will be filled from lower order scenes in the list where such data exists. The result is a composite mosaic for the year with minimal missing data – remaining areas of missing data will be those for which no good data exists in that year after the image selection and processing steps described above.

A set of rules to determine the overlay order has been defined. The considerations include:

a) haze-free, dry images are preferred over cloudy or green images or images with other issues;

b) images with large unmasked data areas are selected above images with small areas of data, but...

c) images with contiguous clear areas (perhaps only in a portion of a whole image) may be selected above images with scattered missing data throughout;

d) LS-7 SLC-on images are preferred over LS-5

e) LS-5 images are preferred to LS-7 SLC-off images

The result of these decisions is the ‘mosaic order list file’ - which is a list of input images (cloud masked scenes) in mosaic order for the mosaic tile. This overlay order is copied for overlapping mosaic tiles so that the overlay order is consistent across the whole country within each calendar year.

The final step is to create the mosaic itself, by running the mosaicing program. The inputs to the program are the masked images and the order list file. Two output files are produced: the six-band mosaic for the tile in four or six ‘quadrant’ pieces depending on file sizes; and a two-band ‘mosaic-date’ image of the same dimension which records for each pixel the date and path/row of the source image from which its data has come. Even after using all available images, typical mosaics will still contain some missing data due to cloud. The results of the mosaic process for Sumatra in 2008, the mosaic image and raster mosaic dates image, can be seen in Figure 6.7.2. Figure 6.7.3 shows the mosaic for Java for 2008.

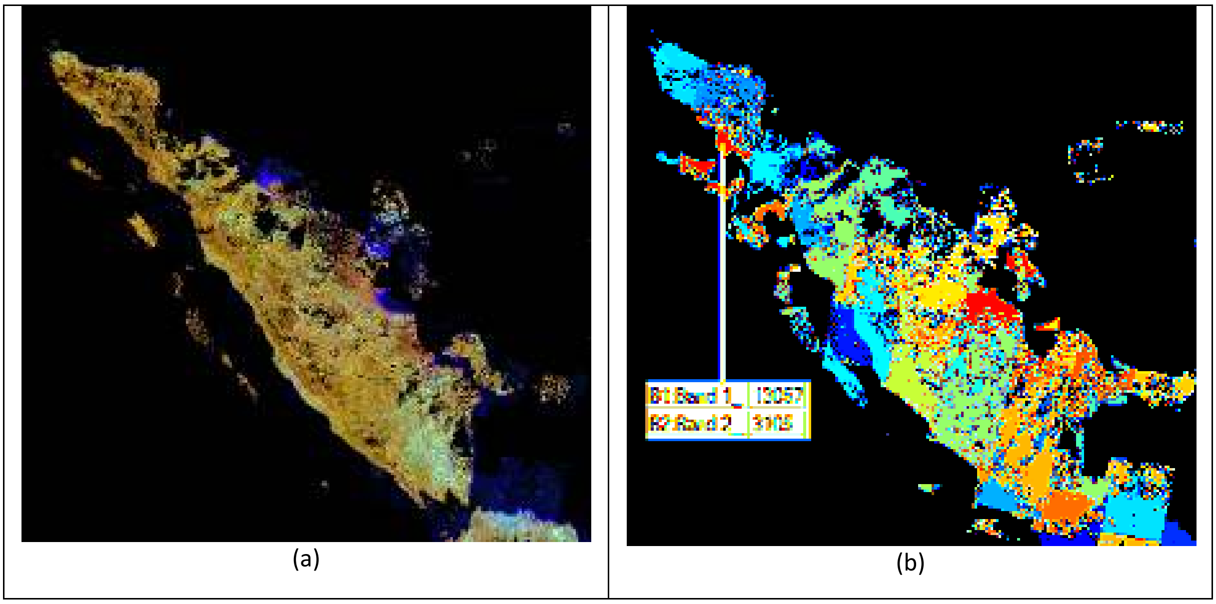


Figure 6.7.2. (a) Mosaic of Sumatra in 2008 (combined from 47N, 47C and 48C mosaic tiles); bands 453 in RGB. (b) Mosaic date information displayed in colour (the example pixel values indicate the source was from path/row image 130/57 acquired on 31 May 2008).

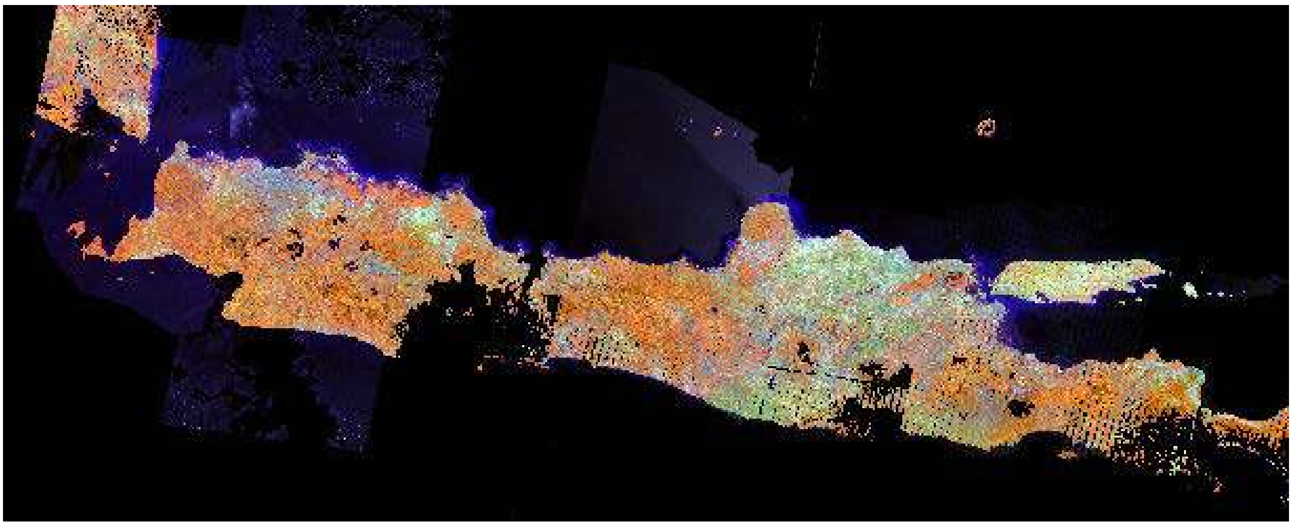


Figure 6.7.3. Display of mosaic Landsat Image for Java in 2008 (mosaic tile 49S); bands 453 in RGB, black areas over land indicate missing data for that year.

Mosaicing QA is performed by the QA team. The QA process involves first checking the completeness of the data file and the overlap order file of image dates. The next stage is to check visually the Landsat image mosaic by displaying the result of mosaic along with the date image mosaic. The final mosaic is compared to the simulated mosaic from the 'snapshot' pictures to ensure that no data has been omitted or incorrectly located. Adjoining quadrants are compared with each other and with adjoining mosaic tiles that are already complete to ensure the same overlay order in the overlaps. The QA team may reject and return to operator if any problems are detected. Provided all available scenes have been used, QA may suggest a change in order of mosaic, or removal (masking) of specific problem areas from an input image.

Once the mosaic has passed QA, it will be archived for use as an input for the next stages. Terrain-corrected mosaic images have already been produced for all of Indonesia for each year from 2000 to 2009. Mosaics for 2010 to 2012 will be complete in the middle of 2014. These are archived at LAPAN.

The frequency of cloud-free data can also be summarised. The missing data in mosaics over the 13 years can be used to estimate and illustrate the frequency of cloud problems within a year for data in various areas within the LCCA processing region. The data may also be summarised year by year to provide the proportion or area of cloud free data for each island group. In combination with summaries of the numbers of input images, these data might be used to predict the proportion of cloud-free imagery if full archives are available (as will be the case with LS-8) and if more frequent optical data were available.

Figures 6.7.4 and 6.7.5 show spatial summaries of the amount of clear land in the ten years from 2000 to 2009. Table 6.7.2 shows area summaries. Each pixel is coloured according to the number of years in which there was clear data suitable for forest mapping. Shades of green indicate pixels with data in all or most of the 10 years. Shades of blue indicate pixels with data in around half of the 10 years. Shades of yellow (3), orange (2), pink (1) and red (0) indicate pixels with data in few of the 10 years. Not surprisingly, the drier regions of Indonesia have clear data in more years. Some of the mountainous areas have the least data available. In parts of Papua, this may be the result of over-masking of snow and ice. The Riau and Jambi provinces in Sumatra also show low data coverage, most likely due to smoke haze. Such summaries can be used to prioritise searches for alternative data sources. For example, due to altitude and/or slope considerations broad scale land use change is unlikely in mountainous regions and so may be a low priority.

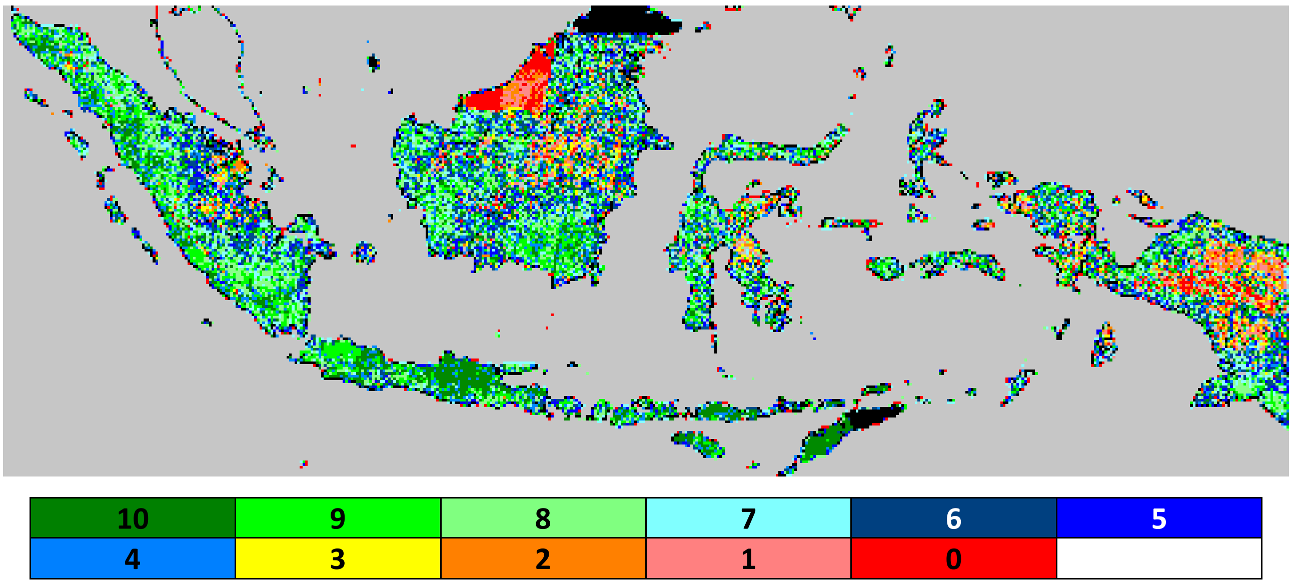


Figure 6.7.4. Display of the number of years from 2000 to 2009 with data suitable for forest mapping for the whole of Indonesia. As shown in the legend, shades of green indicate that there is clear data in most years; shades of blue indicate that there is clear data in around half the years; shades of yellow, orange and red indicate clear data in only a few years.

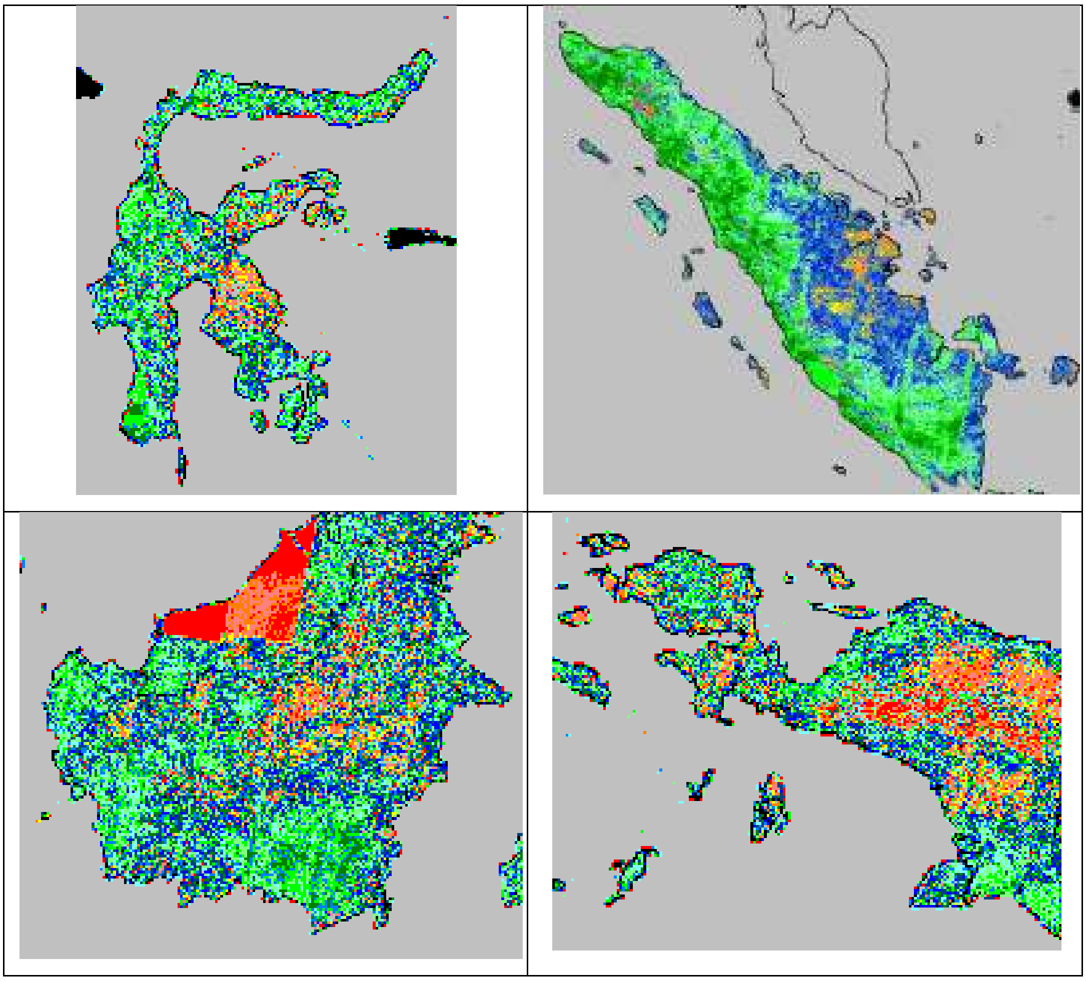


Figure 6.7.5. Display of the number of years from 2000 to 2009 with data suitable for forest mapping for sample regions in more detail. As shown in the legend in Figure 6.7.4, shades of green indicate that there is clear data in most years; shades of blue indicate that there is clear data in around half the years; shades of yellow, orange and red indicate clear data in only a few years.

Table 6.7.2 . Percentage land area with equal to or greater than 'years of clear data' data density by region

Years of Clear Data	Whole of Indonesia	Sumatra	Kalimantan	Java	Sulawesi	Nusa Tenggara	Maluku	Papua
1	98.44	99.81	99.34	99.99	99.22	99.99	99.36	94.55
2	95.85	99.17	97.30	99.95	96.44	99.97	97.62	87.81
3	91.33	97.32	92.54	99.72	91.60	99.867	94.20	78.46
4	84.16	93.05	83.72	99.08	84.71	99.51	88.09	66.65
5	74.30	85.87	70.71	97.54	75.69	98.70	78.42	53.28
6	62.24	75.67	54.81	94.33	64.13	97.11	65.28	39.39
7	48.80	62.16	38.40	89.24	49.97	94.11	49.60	26.01
8	35.54	46.82	23.93	81.87	34.32	89.44	33.05	14.75
9	23.21	30.30	12.59	70.07	19.10	81.32	18.14	6.51
10	10.94	12.03	4.28	42.90	6.15	63.848	6.82	1.50

The current LCCA mosaics produced from the terrain corrected and cloud-masked data are the inputs to forest mapping processes. Using the mosaic list file and data archived at LAPAN it is also possible to produce mosaic products using the BRDF-corrected imagery prior to terrain correction as inputs. These mosaics may be preferred for some purposes including visual interpretation.

7. DETAILED METHODOLOGY - FOREST EXTENT AND CHANGE MAPPING

This section describes the LCCA processing methods to produce the forest extent maps from all years of the mosaiced data. Three major stages are described. The first stage is the creation of a 'forest base' probability image for a chosen single year. In this stage, a 'first' forest probability map is created using supervised classification methods applied to the LCCA mosaics for that year. The process requires expert input to advise on land cover, and application of statistical procedures to derive the classified forest probability map. Operator judgment is used to evaluate and refine threshold parameters in the image classification process. The statistical procedures are essentially multivariate discriminate analyses; in particular canonical variate analysis (CVA) and related procedures. CVA is described in Campbell and Atchley 1981; and further described in applications using image data in Caccetta et al 2007, Furby et al 2010. In the second stage of processing, referred to as 'Matching', this base probability image is then the input to a semi-automated classification which produces forest probability images for the remaining years in the time series. In the third stage, these single-date forest probability images are then combined using a fully automated multi-temporal classification model to produce refined forest probability maps for each year. The mathematical framework for this process is known as a conditional probability network (CPN) (Lauritzen 1992, Lauritzen & Spiegelhalter 1988). Development and application of CPNs to land and forest monitoring is described in Kiiveri & Caccetta (1998) and Kiiveri et al (2001, 2003). Furby (2002) describes in detail the application of CVA and CPN within Australia's forest monitoring system.

Numerous methods exist for image classification. In the LCCA, a simple index and threshold image classification methodology is used to produce the single-date probability images. The approach is optimised locally for different vegetation and land use communities at the base mapping stage, through application of CVA and related routines. One advantage of this approach is that the methodology is much more easily understood and adjusted to local circumstances by the operators at the base mapping stage. A second advantage of this approach is that it can be applied through the matching process to other years in the automated matching process. This is both efficient and critically important in maintaining consistency over time. Temporal consistency is a paramount consideration in a monitoring system of this kind; within the LCCA, temporal consistency is further improved by the multi-temporal processing stage. It has been shown that, once combined with the multi-temporal processing model in the third stage, the index-threshold classifier used in the first and second stages provide results as accurate as applying far more complex initial classification methodologies (O'Connell and Caccetta 2009).

7.1 Forest base mapping

The base mapping uses a combination of data inputs and knowledge to guide and tune the local classifications (see Sections 5, 6.2). Experts with knowledge of ground cover in the region provided critical inputs through interpretation of high resolution images and Landsat mosaics, supported by maps and ground photographs where available. The base mapping process for each region was conducted as a dedicated workshop typically over three weeks, where image processing and statistical experts worked closely with the regional experts in an integrated team. The forest base workshop is a manually intensive process, but is carried out only once for each region to create the base. For adding other years to an existing time series, only the second and third stages, which are semi and fully automated, need be performed.

The first step in forest base mapping is to select the 'base year' for the process. In the LCCA, at least three years of mosaics were required to be completed prior to the base mapping; typically mosaics for 2008, 2006 and an earlier year (2000 or 2001) were available. Year 2008 was generally the desired choice as it was close in date to most of the high resolution images. Further, as the most recent available mosaic, it was most relevant to recent experience and memory of the ground experts. Missing data in mosaics was however an important consideration, as it is desirable that the base mapping be as complete as possible. Where cloud cover was significant, an alternative base year was chosen, sometimes only for sub-regions in the mosaic.

There are three main steps in creating a base forest probability image:

1. Dividing the region into stratifications zones. The forest base analysis and classification are carried out separately within each stratification zone.
2. Deriving indices to discriminate between forest and non-forest cover for each stratification zone. This process applies statistical analyses to training samples covering the range of forest and non-forest covers within the zone.
3. Setting thresholds for these indices that in combination identify certain forest cover, certain non-forest cover and the uncertain spectral region for each stratification zone for the base year. Different thresholds may be used for some image dates within the base year mosaic.

Each of these steps will be discussed in more detail below.

7.1.1 Stratification.

The land use and land cover (including forest and non-forest vegetation) varies greatly across Indonesia – from the drier tropical savannas in Nusa Tenggara to the wet peat forest of Sumatra and Kalimantan and the high altitude vegetation in Papua. A single image classifier cannot successfully separate all these different land cover types simultaneously. Instead, the approach adopted in the LCCA is to subdivide (or stratify) the whole area into smaller regions (zones) within which the mix of land covers and land uses are much simpler. The classifier can then be directed (by analysis) to separate the forest and non-forest cover types which occur within that region; e.g. dryland forest and dryland non-forest land uses, or coconut palms from mangrove and wetland forests, or rice crops from wetland forest.

Figure 7.1.1 shows the stratification zones for Kalimantan. The light orange 'background' colour shows the central mountainous dryland forest stratification zone. The darker orange zone at the bottom right was wetland forest in the 1990s which was cleared for agriculture as part of the larger 'transmigration' area on Kalimantan. Little natural forest remains in this zone. The red area next to it is predominantly the original wetland forest with only small patches of agriculture and settlements. The small purple zones next to the dark orange zone are regions of rubber plantations. Rubber is very different in the images to other forest plantations. The colours in the satellite image on the right of Figure 7.1.1 look quite different between these four zones and so it logical to treat them separately in the forest mapping process.

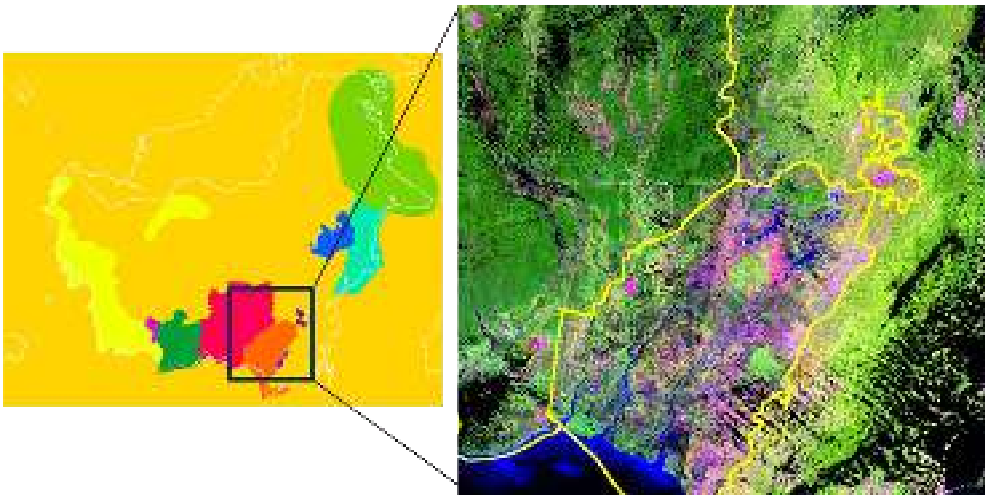


Figure 7.1.1. LCCA final stratifications zones for Kalimantan (left) and a zoomed-in display of the boundaries overlayed on the 2008 Landsat TM mosaic image (bands 5, 4, 3 in RGB) on the right.

Soil, climate, terrain and geology feature prominently in the factors that cause regional differences in land cover and land use. Access to datasets or expert knowledge of these can assist in the stratification process. Broad variations in vegetation spectral response and agricultural land use can be identified initially from inspection of the image mosaics. Regional experts with knowledge of land use and forest types provide critical input to the stratification of a region. Stratification zones are iteratively refined within a base workshop. Analysis of training sites from within the zones and review of initial forest mapping (described below) may suggest addition of further stratification zones or changes in boundaries to improve the accuracy of the forest maps. The final stratification zones, as vector and raster maps, are the archived outputs of this step.

7.1.2 Deriving indices for separating forest and non-forest cover

Indices - linear combinations of the Landsat image bands for separating forest and non-forest land covers - are optimised locally for each stratification zone. The index derivation process is driven by ground truth information – sample areas of forest and non-forest vegetation cover are selected from the image mosaic of the base year using ground truth information. Comprehensive samples of forest and non-forest cover are required within each stratification zone. Ground truth data includes high resolution satellite imagery (1-2m resolution) in which tree and non-tree cover is evident (Figure 7.1.2). Local experts who have a deep familiarity with the region and its land cover are critical. Based their knowledge, and viewing the high resolution data along with Landsat mosaics, interpretations of land cover are extrapolated spatially. These sample areas are indentified in the Landsat image mosaic as training site polygons and labeled. The analyses are directed towards discriminating between the particular training sites used. Provided that the sites are representative of the cover in the images, the indices and thresholds that separate the training sites will separate all forest and non-forest cover in that zone.

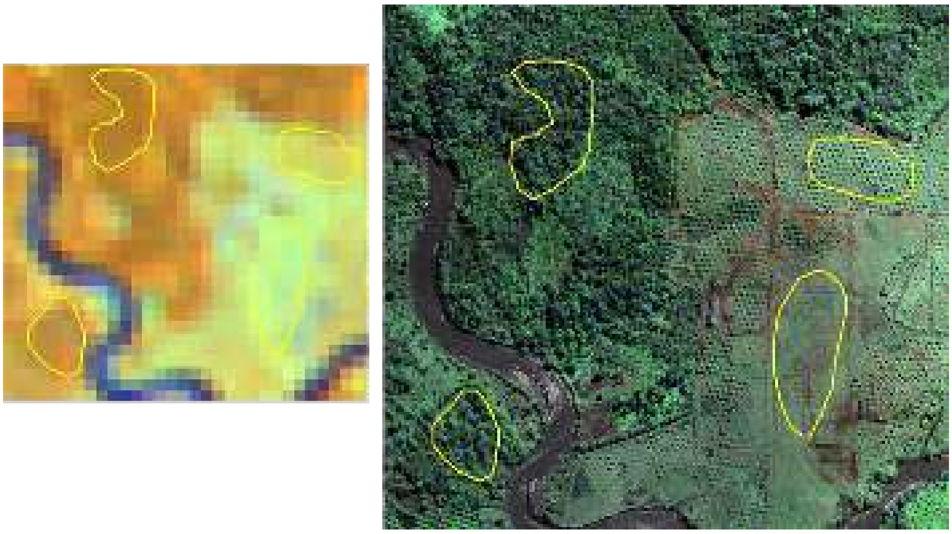
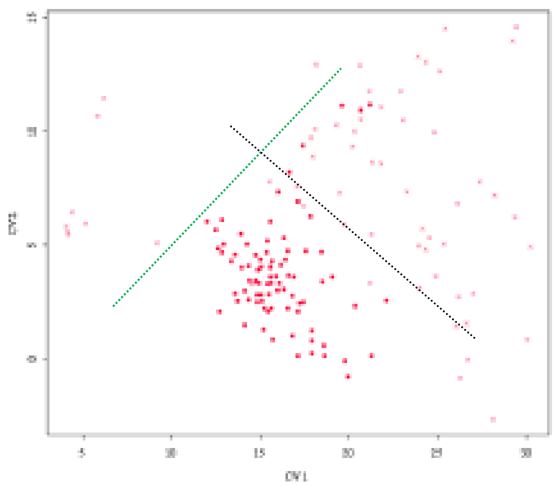
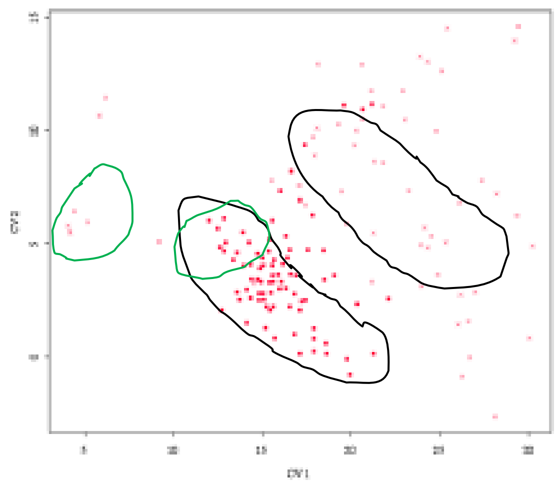


Figure 7.1.2. Landsat TM satellite image (left) with bands 4, 5, 3 in RGB and a high resolution image of the same area. The yellow outlines show forest and not forest training samples (sites) identified in each image.

As noted above, the image year selected for the forest base analyses is typically the one closest in date to most of the ground truth imagery and as close to the present as possible, to relate best to the memory of the local experts. One or two other image years may be included in these index derivation analyses so as to represent the seasonal variation in the time series. The advantages of including the extra images in the initial processing are that indices are found that work over a range of years, not just optimised for the conditions present in any particular year, or the need to account for variations in conditions between years is established very early in the process.

Statistical analyses are applied to identify the required spectral indices. Canonical variate analyses are performed using the training data to derive suitable indices for discriminating between forest and non-forest cover. A CVA finds the linear combinations of the image bands that maximise the differences between training sites relative to the variation within the sites. The canonical vectors give the directions of maximum site separability and the canonical roots give a measure of the amount of site separation in these directions. The canonical vectors, which are linear combinations of image bands, form the basis of the indices to be used. The usual canonical variate analysis will find the canonical vectors that best separate each of the training sites from every other training site. The canonical variate analysis can be modified to focus on the separation of particular groups of reference sites (forest and non-forest) rather than between all the sites. This is done using contrast vectors to focus on supplied groupings of the reference sites.





Index Coefficients	Proportion of Separation Provided Image A	Proportion of Separation Provided Image B
-0.035 0.34 -0.068 -0.078	100%	-
-4 34 -7 -8	99.92%	-
-1 8 -2 –2	99.25%	-
0 4 -1 –1	99.34%	74.22%
0 1 0 0	74.10%	97.58%
-0.068 0.26 0.013 0.003	-	100%
-7 26 1 0	-	99.96%
-1 3 0 0	75.28%	99.29%
-1 4 -1 -1	96.61%	65.59%
-1 4 1 1	38.15%	92.38%
-1 4 0 0	75.37%	99.74%
-1 5 0 0	75.27%	99.67%
0 3 -1 –1	92.29%	55.50%
0 5 -1 –1	99.25%	83.24%
0 6 -1 –1	97.34%	87.99%
0 7 -1 –1	95.17%	90.74%
0 8 -1 –1	**93.17%**	**92.45%**

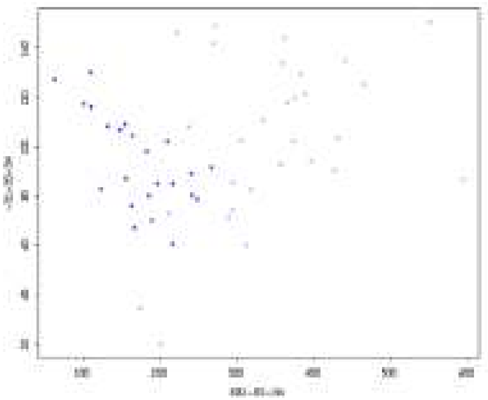


Figure 7.1.3. Top left: Sample canonical variate means plot. Site that are plotted close together are spectrally similar and sites are plotted further apart are more separable. The dashed lines separate most of the forest sites from most of the non-forest sites. Top Right: Contrast groups to derive indices for the lines of best separation. Bottom Left: Table showing index smoothing for robustness through time. Bottom Right: Resulting index plot.

The canonical vectors found by this process relate specifically to the particular training sites and image dates used in the analyses. The vectors are simplified, or smoothed, to make them more robust over the range of cover types and image dates. In this process, smoothed vectors, or indices, are sought that perform well over all of the selected representative image dates. Information statistics allow alternative indices to be compared. Figure 7.1.3 summarises outputs from the CVA and index-smoothing routines. This process can also be used to compare global vegetation indices (e.g. NDVI) as well as indices from similar stratification zones (Figure 7.1.4).

As well as producing a particular index or indices for separating forest and non-forest cover, the analyses also provide an objective measure of the number of indices required for adequate discrimination. The analyses have shown that for Landsat TM images the majority of the separation is obtained with a single index. However, a second masking index is always required to separate particular problematic cover types. Occasionally three indices are required.

It is noted here, and expanded below, that perfect separation is rarely provided by the indices for all pixels. The analysis indicates index values where classification is 'uncertain' (Figure 7.1.4). This is reflected in the classification as probabilities as discussed below.

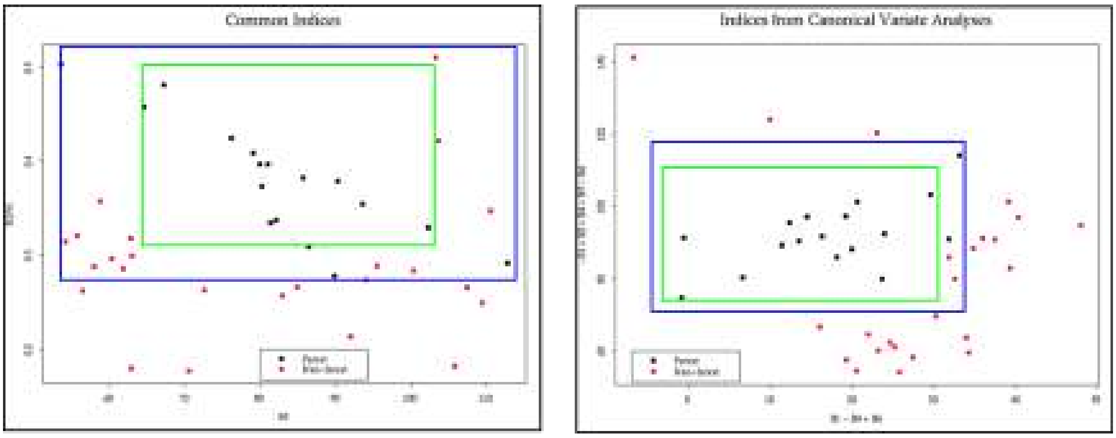


Figure 7.1.4. Index Plots from global indices (NDVI and Landsat TM band 5) and smoothed locally optimal indices. The green box in each is the certain forest spectral region – it contains only forest training sites. Outside the blue box is the certain non-forest spectral region – it contains only non-forest training sites. Between the two boxes is the uncertain spectral region – it contains both forest and non-forest training sites. The better the indices separate forest and non-forest cover, the smaller the uncertain region will be. Here the local indices perform better than the global indices.

7.1.3 Setting thresholds for separating forest and non-forest cover

Having found appropriate indices to separate forest and non-forest cover for a zone, it remains to set thresholds for each year to allow pixels to be assigned probabilities of forest (and non-forest) cover based on the values of the indices. These thresholds are set manually for the base year.

It will generally be the case that there is some spectral overlap between forest and non-forest training sites in each stratification zone, or at least a continuum of index values. There will be a range of index values that correspond to pixels that can confidently be labelled as forest, (or conversely, couldn't possibly be non-forest cover). There will also be a range of index values that correspond to pixels that can clearly be only non-forest cover. In between, there will also be a range of index values, where a pixel could be either forest or non-forest – a region of uncertainty. An example is shown in Figure 7.1.4. The green rectangle shows the certain forest region. The blue rectangle shows the combined certain forest and uncertain region. Outside the blue rectangle is the certain non-forest region.

As well as defining the edge of these three partitions of the image, the thresholds determine the probability assigned to each pixel. Pixels with index values in the certain forest spectral region are assigned a probability of forest cover of one (100%). Pixels with index values in the certain non-forest region are assigned a probability of forest cover of zero. Pixels with index values in the uncertain spectral region are assigned a probability of forest cover between one and zero based on the closeness to the certain forest thresholds. The thresholds need to be set ensure that the probabilities are high (or low) enough within the uncertain area. Initial thresholds are set by inspection of the index plots.

Refinement of the initial estimates of the thresholds is conducted using image displays as in Figure 7.1.5. Operators assess whether there are cover types (image pixels) that are not adequately represented in the training data. This is an important process in which ground knowledge provides important input; operators may view, assess and discuss results of adjusting thresholds to minimise misclassification and set appropriate probabilities within the ‘uncertain’ zone. This assessment, by comparison of results from neighbouring zones, effectively optimizes use of ground truth over a wider area and may generate refinement of zone boundaries.

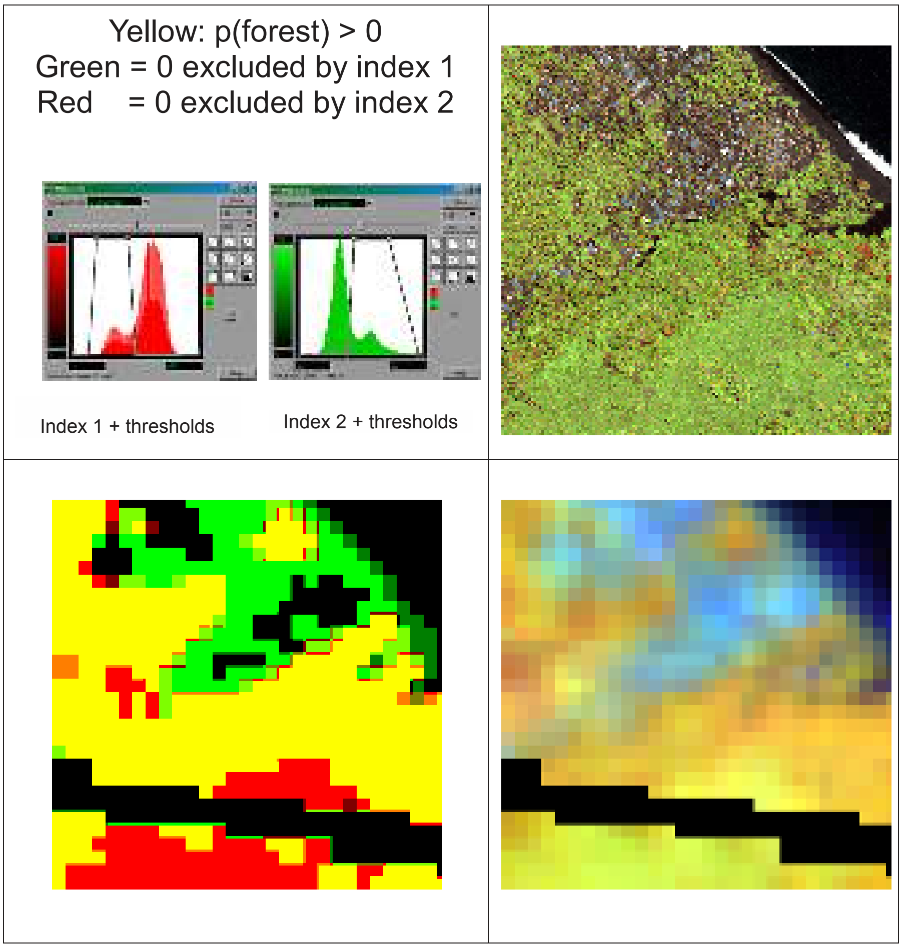


Figure 7.1.5. Illustration of threshold refinement from image displays which combine interpretation of ground truth with classification results (lower left). Thresholds are adjusted until best results are achieved.

Thresholds are set separately for each stratification zone, even if the same indices are applied to several stratification zones. It may also be necessary to locally adjust the thresholds for different image dates within a stratification zone, if the seasonal or atmospheric conditions are very different. Sub-zones are defined by intersecting the image date boundaries for the mosaic with the stratification zone boundaries.

When a satisfactory set of thresholds are obtained for a zone, it is saved and a forest probability image is created from this file and the mosaic image, and reviewed. The forest cover probability images for the individual zones are overlaid as part of final QA for the base. Zones and thresholds may be refined at this stage.

Within each zone, **QA is essentially carried out by the team during this stage as a continuous process** by viewing alternative results on screens, with reference to ground experts and to results from neighbouring zones. Examples of displays used in this process are shown in Figure 7.1.6. Beyond the areas of the LCCA high resolution coverage, Google Earth(TM) is used in this assessment.

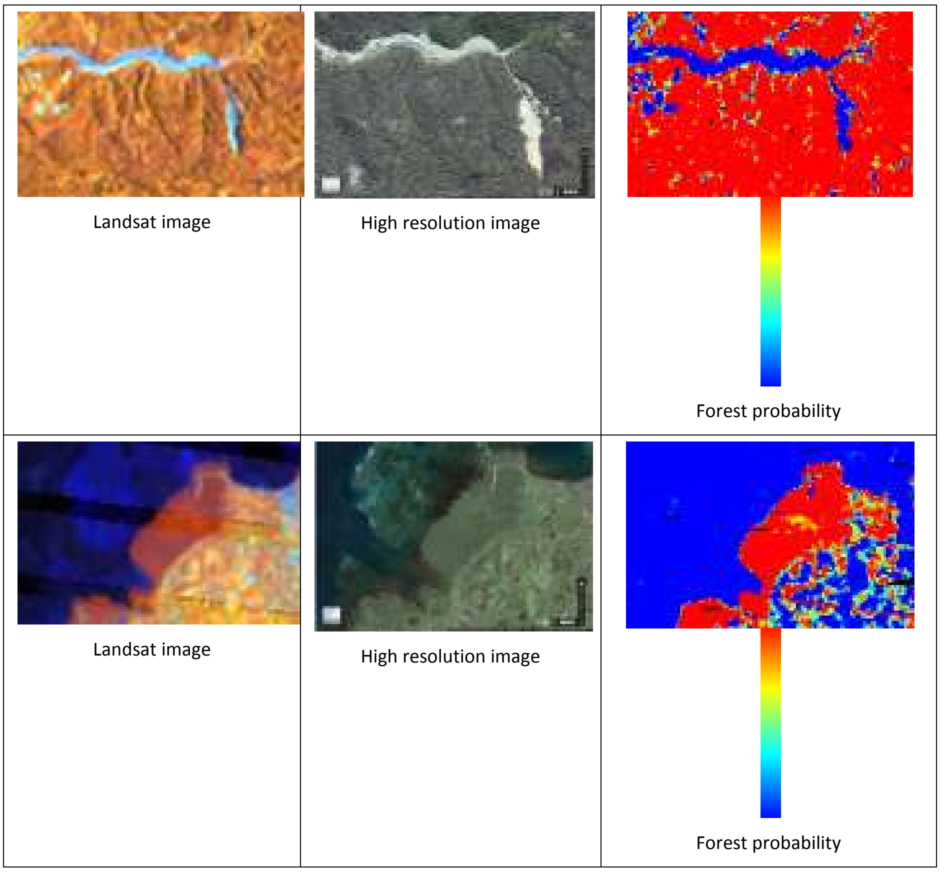


Figure 7.1.6. Two examples from Sulawesi for QA checking of classification results. The base forest probability image is compared to Landsat image and high resolution image. In these examples the high resolution imagery is from Google Earth(TM)

The final outputs from the base mapping are: a final set of stratification zones in raster and vector format; a set of index-thresholds files for each zone (and subzone); and the final map of probabilities for the region. QA is required to check these files, the mosaic control program and the mosaic results. Examples of results are shown below.

Spectral separation is impossible for some desired classes of forest (or non-forest) where the spectral signals overlap. In such cases other data or knowledge must be employed to correctly label such cover classes. An additional process in the base mapping process and QA, which continues in subsequent QA steps, is the identification of cover classes in different regions which result in errors in the forest probabilities. Examples of problems of this type are:

- Some kinds of non-forest plantation such as oil-palm - mature palms are identified as forest in probability image.
- Deciduous forest, e.g. cinnamon and teak forest, which may be incorrectly classified depending on the season of imagery.
- Rice crops in some growing phases sometimes have high probability and may be identified as forest.
- Fluctuating water levels in wetlands create varying signals, some of which may be confused with forest.

In general, the location of these areas is restricted, and correction approaches using other data may be employed – see Section 8.

7.1.4 Base mapping results: example

The results of the base mapping for Sulawesi, produced from base year 2008, are illustrated below. Figure 7.1.7 shows the mosaic and the stratification zones for the island. Figure 7.1.8 shows the base forest probability image. Figures 71.9, 7.1.10 and 7.1.11 show samples of the forest base map and high-resolution imagery. Displays like these are used in the QA of the forest base probability image.

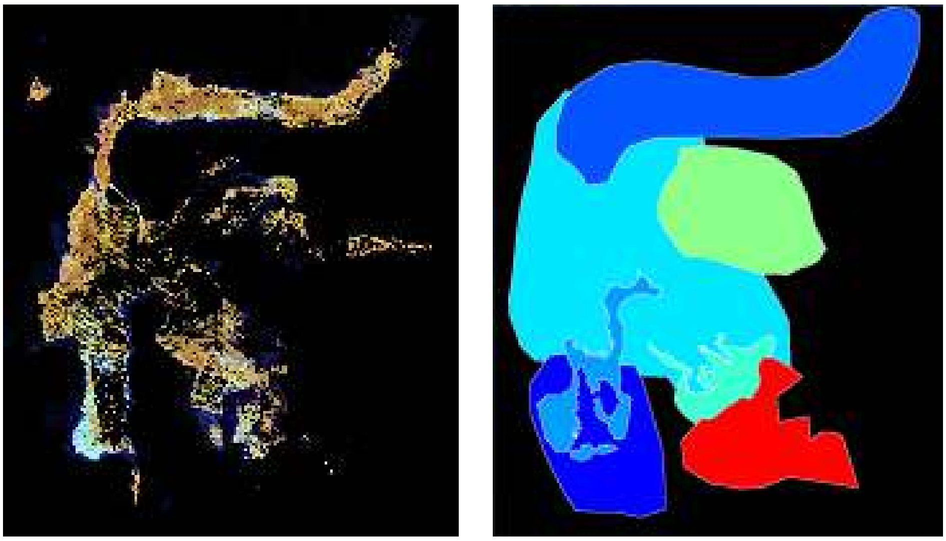


Figure 7.1.7. Sulawesi mosaic 2008 (left, bands 453 in RGB) and stratification zones (right)

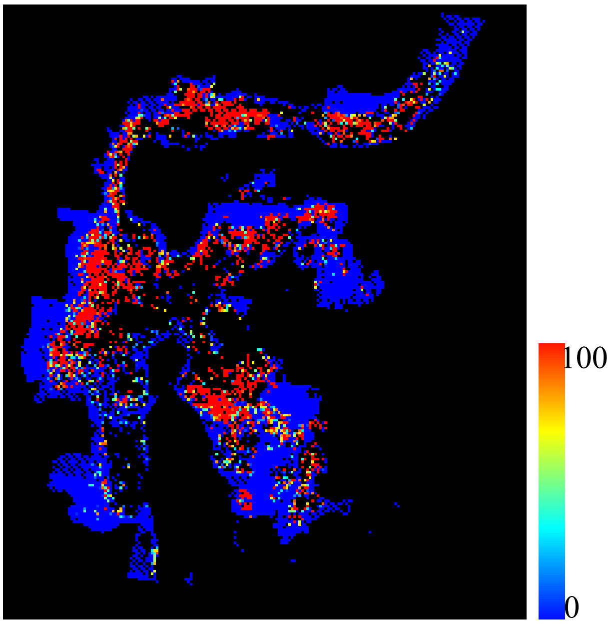


Figure 7.1.8. Base forest probability image for Sulawesi. It is formed mostly from the 2008 Landsat mosaic with zone 6 from the 2006 Landsat mosaic. The red colour indicates probability calculated as 100% in base mapping; the blue colour indicates probability 0 (non-forest). Other colours indicate the probability is between 0 and 100% as per the scale bar shown.

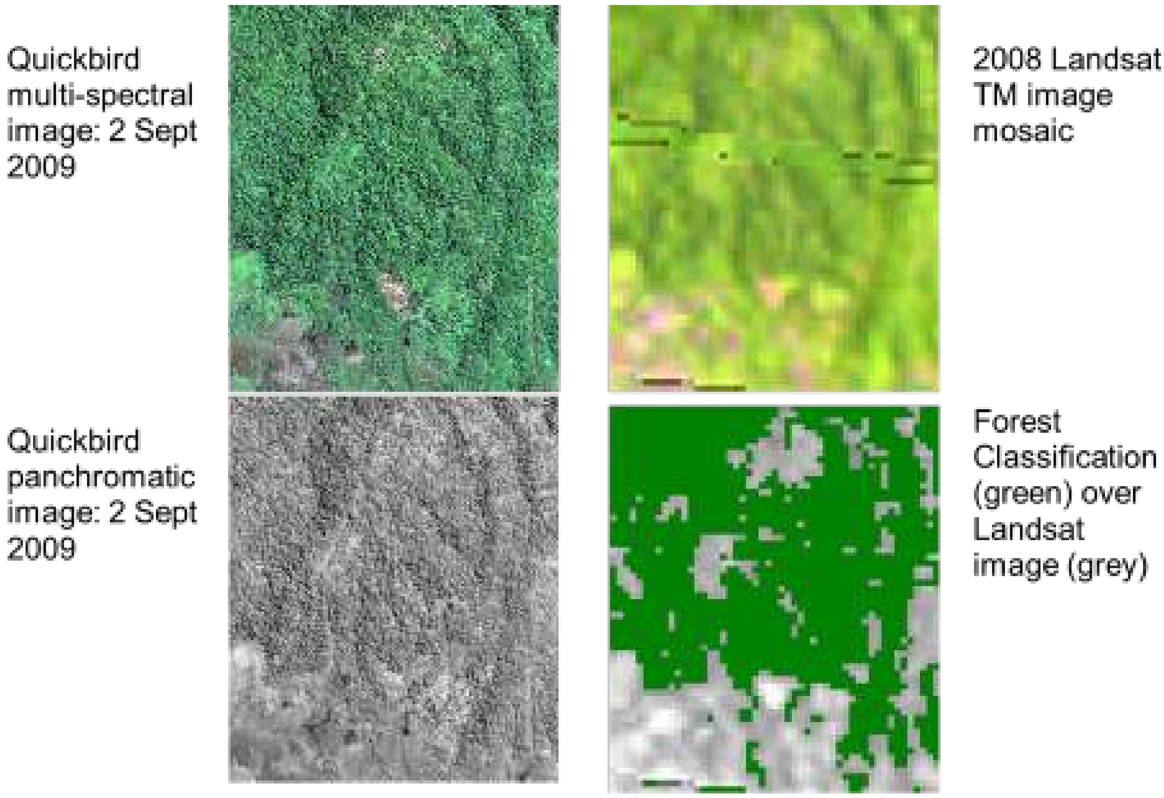


Figure 7.1.9. Example: Base forest extent (probability>50%) compared to high-resolution imagery. Detail from Sulawesi.

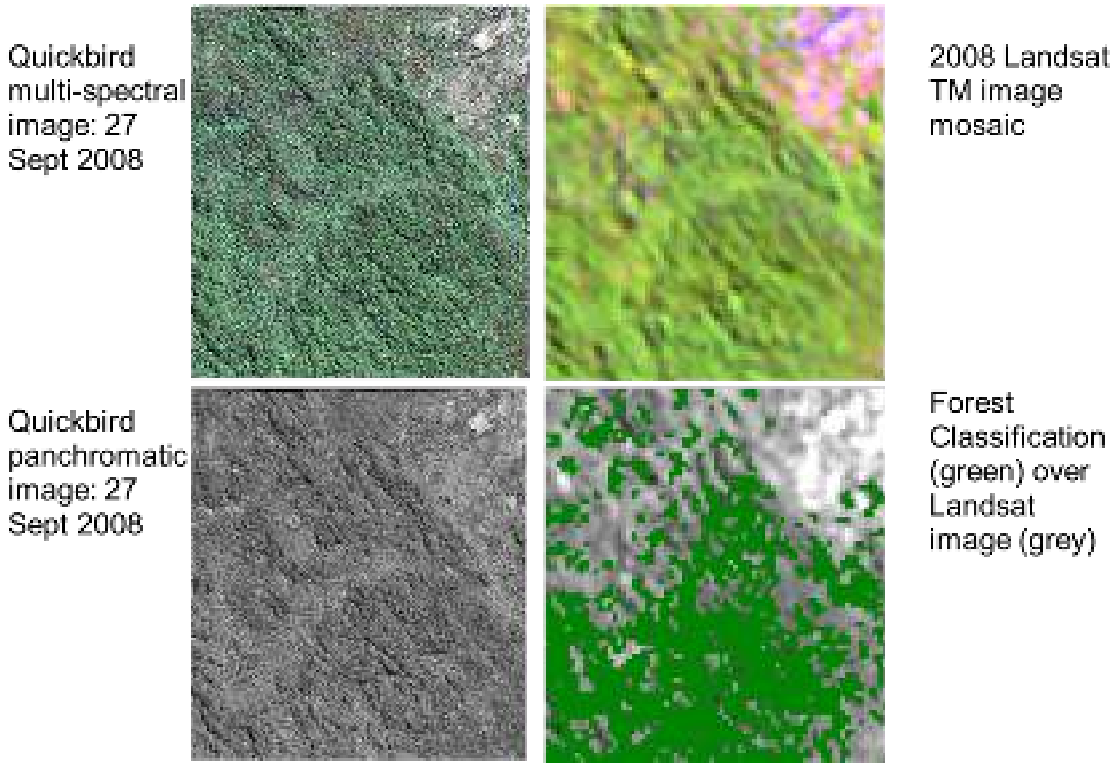


Figure 7.1.10. Example: Base forest extent (probability>50%) compared to high-resolution imagery. Detail from Sulawesi.

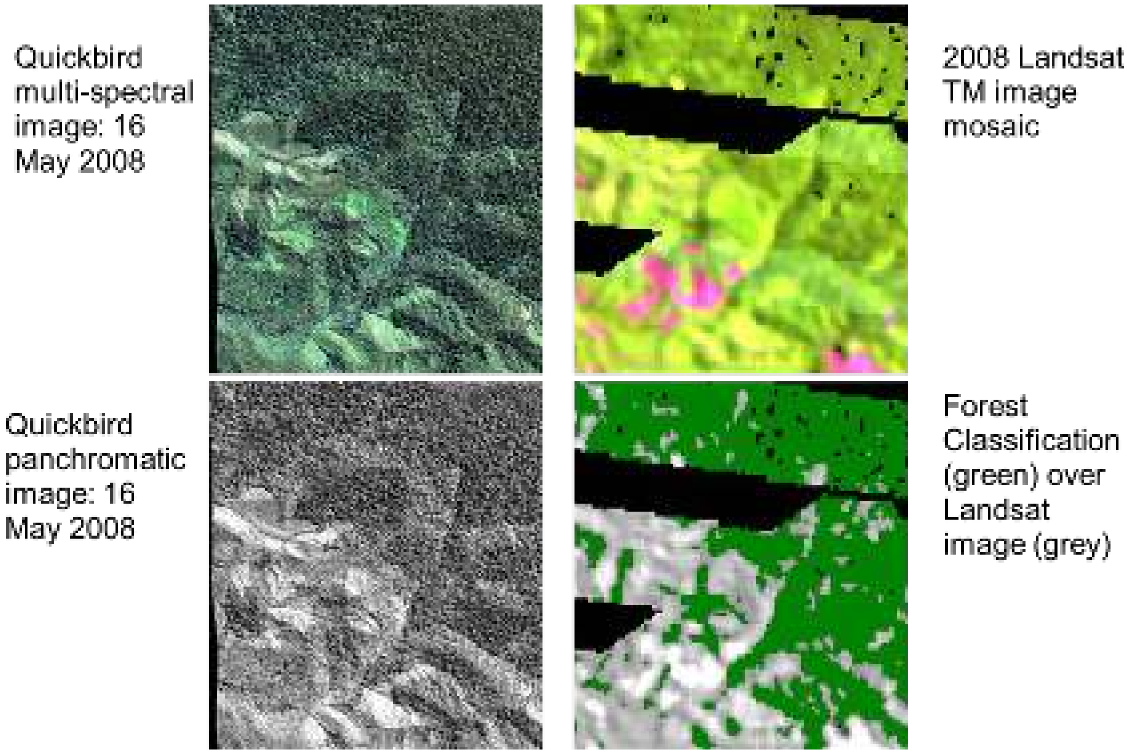


Figure 7.1.11. Example: Base forest extent (probability>50%) compared to high-resolution imagery. Detail from Sulawesi.

7.2 Matching: forest classification for other years

Following the base mapping, a semi-automated matching process is applied to the LCCA mosaics to derive forest cover probabilities for the other years. As described above, the indices used in the base mapping are developed by analysis; which typically considers more than one year. In the automated matching, it is assumed that the derived indices can be applied to their zone in all years – the matching process is run to set appropriate thresholds for each of the years, which can be applied to produce the classification for that year. The matching process is applied in the Australian forest monitoring system (Caccetta et al. 2007, 2013); it has advantages of efficiency, of removing operator judgement and maintaining temporal consistency. Occasionally the automation does not produce adequate results and some manual adjustment of thresholds is required.

This process requires as inputs the index-threshold files for each zone, as well as the Landsat mosaic for the new year. The forest map probabilities from an existing year are also required; initially this is the base map as described above. Training regions are selected by operators for each zone, and within these regions the matching algorithm adjusts thresholds to produce a best match of probabilities to the base using a criterion of minimum sum of absolute differences. These matched thresholds allow probabilities to be assigned to pixels to produce a forest probability map for the new date. Figure 7.2.1 illustrates this process.

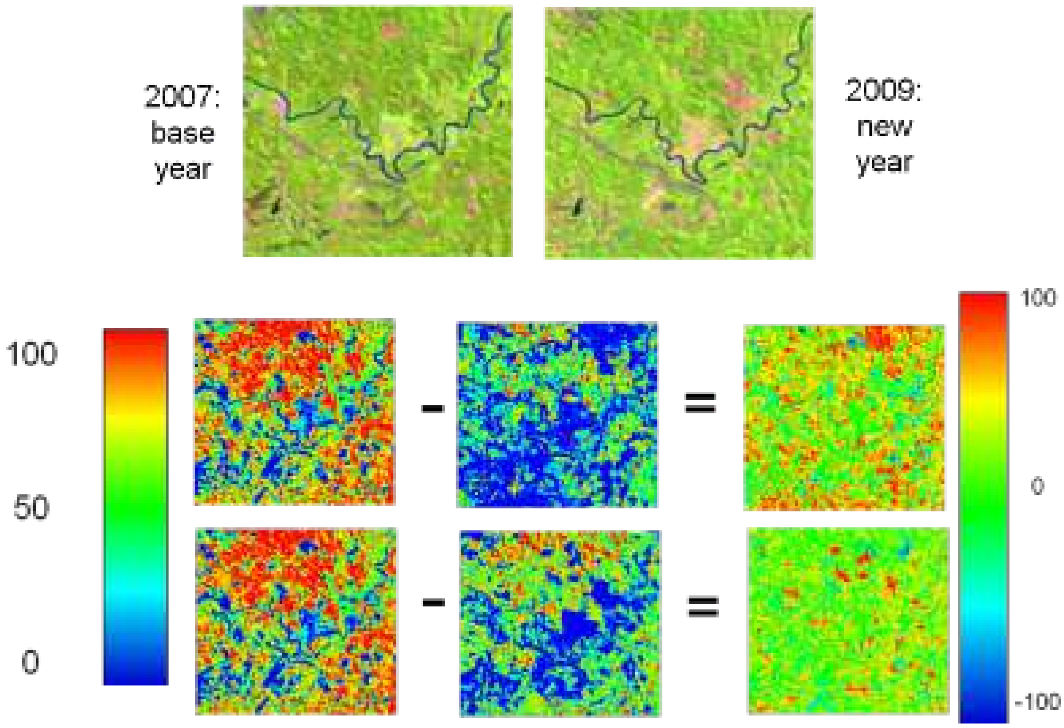


Figure 7.2.1. Illustration of the matching process. The top row shows the 2007 base year for a small region in Sulawesi and the same area in the 2009 image mosaic. A matching region is also shown by the yellow box. The two rows below show probability images on the left. The left most probability image is for the 2007 base and is the same in both rows. The middle probability image is for the 2009 image. At the top, the probability image is calculated by applying the thresholds for the 2007 image. Below is the result of adjusting the thresholds using the matching process. The image on the right in each row is the difference between the 2009 and 2007 probability image. Shades of green indicate little difference between the two probability images. The remaining red and blue areas in the bottom row correspond to areas of change observed between the two Landsat image mosaics.

The selection of an existing date of probabilities to which other years are matched is done by considerations such as:

- In regions where a lot of land use change is expected between the images in the time series, the ideal base image is one closest in time to the 'new' image data so that there is the least change between the images being compared. We may prefer to match the 2009 imagery to a 2008 base, then 2010 to 2009. This is the strategy that was used for Kalimantan and Sumatra.
- In areas where the spectral change is mostly seasonal change in non-forest land use, a fixed (or common) base is preferred for overall consistency. All years would be matched to the first forest base, 2008 say. This strategy was used for the remaining regions of Indonesia.

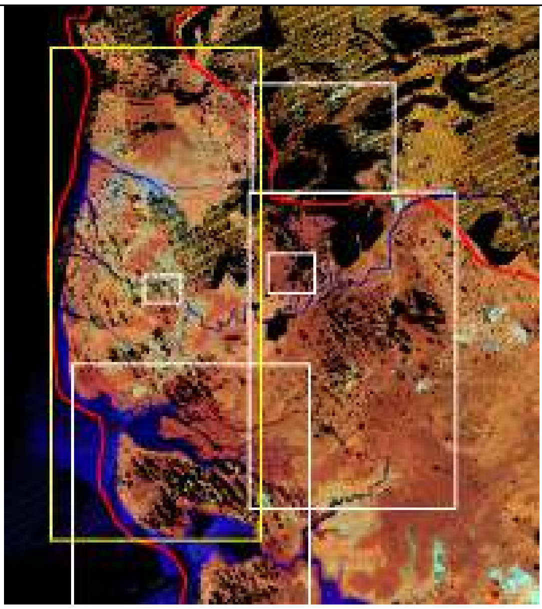
The QA team advises and may review the most suitable forest base probability image for each 'new' year.

Effectively, for the training 'matching region' and indices supplied, the matching program produces the best possible reproduction of the base probability image, according to the 'minimum difference' criterion. The program requires reasonable starting values of thresholds and these are provided from existing base thresholds files for the zone. Thresholds are iteratively adjusted until the best match (minimum sum of absolute differences) is found.

Matching regions are selected for each zone. They need to be created by operators, and may produce varying results. In practice, since the matching is automated, a number of regions for each zone are typically selected, and the best results chosen. Criteria for selecting matching regions to produce good results applicable to whole zone are as follows:

- they should be representative of the whole sub-zone,
- contain sufficient areas of forest and non-forest cover (in both the new and base image); and
- should not include a large proportion of areas of change in forest extent between the current and base years whenever possible.

An example of matching regions is shown in Figure 7.2.2.


(a)


(b)

Figure 7.2.2. Example of matching regions (rectangles) for zone 4 (red boundary) in Kalimantan overlaid on (a) base mosaic 2008 , and (b) mosaic from 2009. The yellow box is likely to be best because it include various types of land cover, others are dominated by a reduced set of land cover such as mangrove, primary forest or areas outside the zone. Bands 453 in RGB.

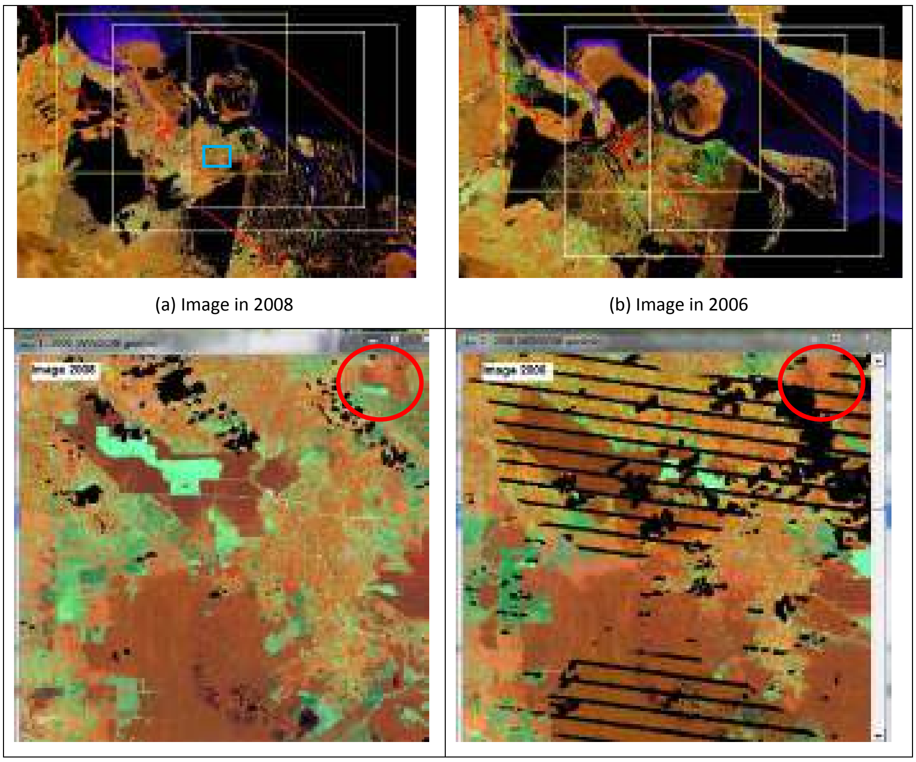
The outputs of the matching program are the index-threshold files for all input training regions, along with an automatically-generated corresponding single-band forest cover probability image for each. For a zone, alternative outputs are quickly compared to the base or reference image in image display. Areas of difference in probability are immediately highlighted and compared to the satellite image data to check for areas of omission or commission in extent and change. If a satisfactory output has been produced, it is usually quickly identified and retained. If not, the training region and/or starting estimates can be modified and the matching repeated to try and improve the forest cover probabilities calculated. As a last resort, the thresholds can be adjusted manually to see if particular errors can be reduced and if so, what the consequences for other parts of the sub-zone or other cover types might be. This manual adjustment process is the same as described in Section 7.1.3.

As with creating the forest base, it may be necessary to locally adjust the thresholds for different image dates within a stratification zone, if the seasonal or atmospheric conditions are very different. One or more matching regions are selected just within the particular image date to derive these local thresholds. Sub-zones are defined by intersecting the image date extents for the mosaic with the stratification zone boundaries.

To perform the matching process, the principal operator input is definition of matching regions and then assessment and selection of results for submission to QA. Various support programs are run to set up the matching and produce list files for archiving. The matching program is then run and results assessed as described above. The final outputs for QA include the vector file of all matching regions and the selected best index-threshold file for all zones (and any subzones).

QA for matching will check files and probabilities generated from each zone. Particular attention is paid during QA to making sure the results are consistent across all years processed. Once all zones and subzones thresholds files pass QA, a probability mosaicing programs is run to produce the forest probability image for that year for each quadrant of the mosaic tile.

Figures 7.2.3 to 7.2.6 show an example of forest probabilities generated by matching and selection by the operator.



(a) Image in 2008

(b) Image in 2006

Figure 7.2.3. Example of matching regions (yellow and white rectangles) for zone 12 (red boundary) for Riau province Sumatra overlaid on mosaics for 2008 (base) and 2006. More detail for the area in the blue box is shown in the bottom row. Bands 453 in RGB.

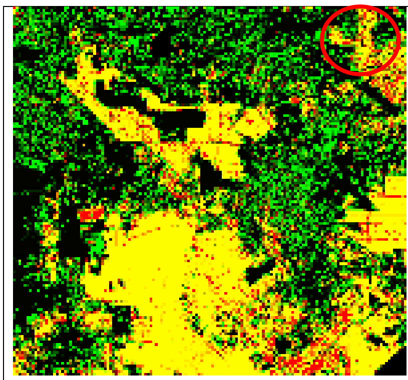
 Index display from 2008 (forest base) in red and green.	 Index display from 2006 in red and green from matching region 121.
	 Index display from 2006 in red and green from matching region 123.

Figure 7.2.4. Example of index displays for the 2008 forest base and two outputs for 2006 from the matching process. The display shown combines in red and green the results from two indices (see Figure 7.1.5). Yellow areas indicate high forest probabilities from both indices; black zero probability; areas of red and green are excluded by one index and would get zero probability of forest cover. The area in the red circle in the lower map (version123) has under-classified forest in this area compared with the base and the imagery. The top map (version121) is a better result.

Based on comparison illustrated above, the results from the version 121 matching region were selected in this case. The matching program saves the indices and thresholds that correspond to each result in a text file like that shown in Figure 7.2.5. Figure 7.2.6 shows the corresponding probability images calculated from the index displays in Figure 7.2.4 for the 2008 base and the 'best' 2006 result.

```
BASE         Z:\_CLASSIFICATION_\_Sum\4_Final_Single\2008\prob_images\ltm_2008_INCAS_nutm47n_se_prbifor25
IMAGE        Z:\_CLASSIFICATION_\_Sum\0_Sat_Images\2006\47N\ltm_2006_INCAS_nutm47n_ne_ter
INDICES 2
 0.00 0.00 -1.00 1.00 2.00 0.00
 0.00 1.00 1.00 -1.00 1.00 0.00
 91.9 111.8 147.7 186.2
 -31.1 -19.5 11.2 14.6
REGION       643244.0 277676.0 834171.0 146573.0
BOOST        1.00
RESIDUAL     5.2
```

Figure 7.2.5. Example of an index-threshold text file output from matching

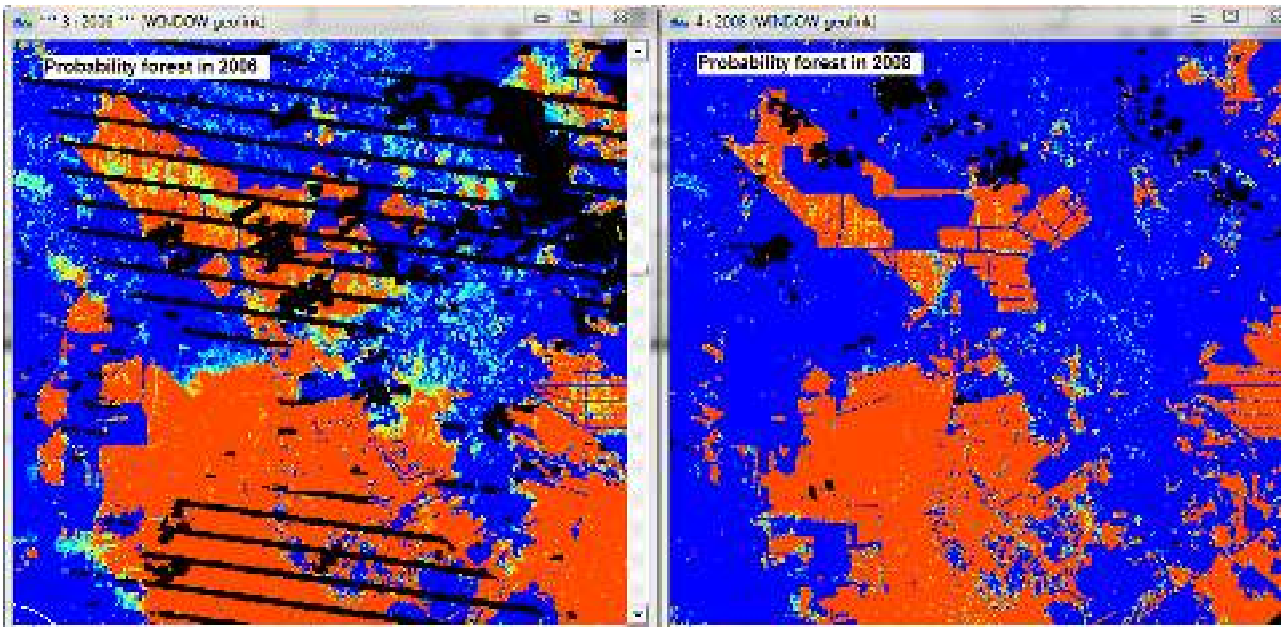


Figure 7.2.7. Example. Output forest probability map for 2006 produced by matching for the area above, compared to the 2008 base probabilities. Red indicates high probabilities (near 100%) while blue indicates probability zero. Missing data is shown in black.

7.3 Multi temporal classification

The individual year classifications of forest probabilities are combined and refined in this stage using a mathematical approach which combines the entire time series of probabilities, along with probabilistic weights relating to accuracy, temporal change and each pixel's neighbouring probabilities. The output is a second set of probability images, the final forest probabilities, which can be thresholded and differenced to provide extent and change maps. Models which describe probabilistic relationships between data of this kind are known as Bayesian networks. Conditional probability networks (CPNs) are used to perform the multi-temporal processing. The CPN provides a computational framework to combine uncertain data (relative likelihoods derived from satellite image data from individual years). CPNs allow for the assessment and propagation of uncertainty from multiple sources of data of varying quality or accuracy (Caccetta, 1997). The scheme for combining data was based upon techniques presented by Lauritzen and Spiegelhalter (1988) and Lauritzen (1992).

The motivation for this processing step in a monitoring context can be summarised as follows. Each classification map will contain some errors. For detecting changes between two or more maps, any errors in each map will result in erroneous changes when differencing of hard classification labels is used. Other information may be used to modify the probabilities. Knowledge of temporal growth patterns in forest and non-forest land cover types provides the opportunity to improve the results of the 'single-date' classifications using a multi-temporal processing scheme. The CPN processing aims to

- improve consistency of the forest mapping from year to year so that there is less error in the land cover change products due to errors in the individual years;
- infer the most likely land cover type or land cover change year in areas of missing data due to cloud cover (missing data will be filled with a best estimate using the time series); and
- resolve the uncertainty in the land cover at some dates by considering the information from the full time series.

A key concept in using a multi-temporal processing scheme is that trees take time to grow. A forest cannot grow and be harvested in only one year, or even two or three years. If the spectral signal is similar to forest in one year, but not the year before nor the year after, it is unlikely to be forest. If a region is truly forest it will be forest for many years. It may stay forest (always high forest probability) or be cleared for another land use (a sequence of high forest probabilities followed by a sudden change to a sequence of low or zero forest probabilities). A new forest plantation or a region of forest recovering from a severe fire will have a period of low or zero forest probabilities followed by a year or two of increasing forest probabilities followed by several years of high forest probability. Each of these circumstances has a clear temporal signature, as illustrated in Figure 7.3.1.

A signature that varies very rapidly between forest and non-forest cover or shows forest cover only for a single year is a very unlikely long-term trend for genuine forest cover. Temporal rules are used to weight against such areas being labelled as forest cover in any year. This strategy significantly reduces the amount of false change detected when comparing any two years. Similarly the temporal rules use the whole temporal sequence of probabilities to infer the cover type of areas with 'uncertain' probabilities in one or more years.

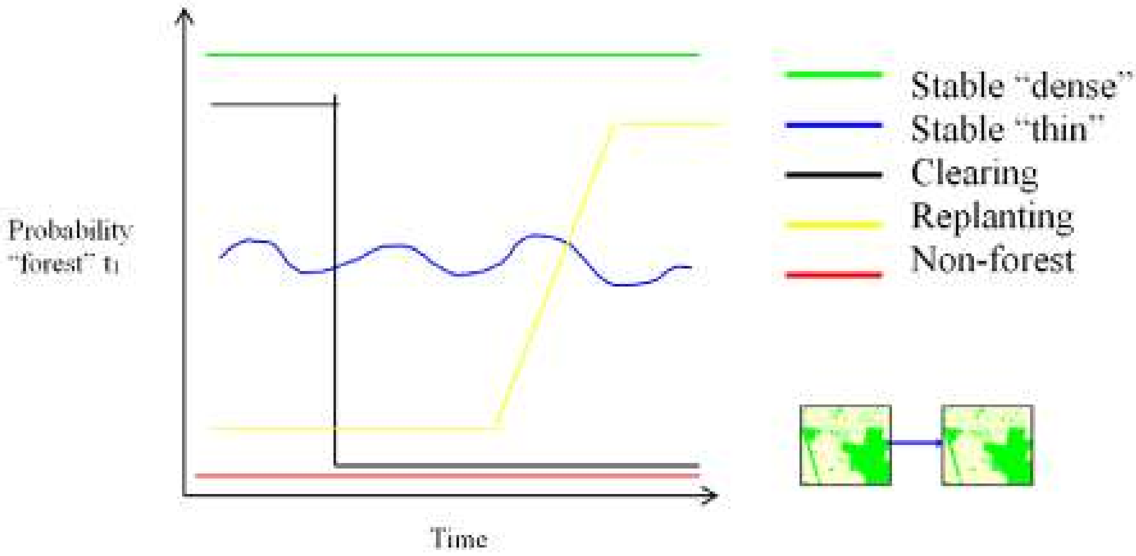


Figure 7.3.1. Schematic diagram of temporal patterns of probabilities over time for forest and non-forest cover, and for clearing and regrowth of forest. The graphic bottom right shows where this temporal pattern information is used in the multi-temporal classification model (see Figure 7.3.2).

The multi-temporal processing also uses the probabilities for the whole time series to infer the most likely land cover or land cover change date for years when there is no data such as in cloud gaps.

The CPN model used in the LCCA can be represented by a graph (Figure 7.3.2), where the nodes of the graph represent variables, and the edges (arrows in Figure 7.3.2) of the graph represent (conditional) independence assumptions between the variables. Note that in this example, as we do in practice, the structure of the network is assumed. Figure 7.3.2 is further discussed below.

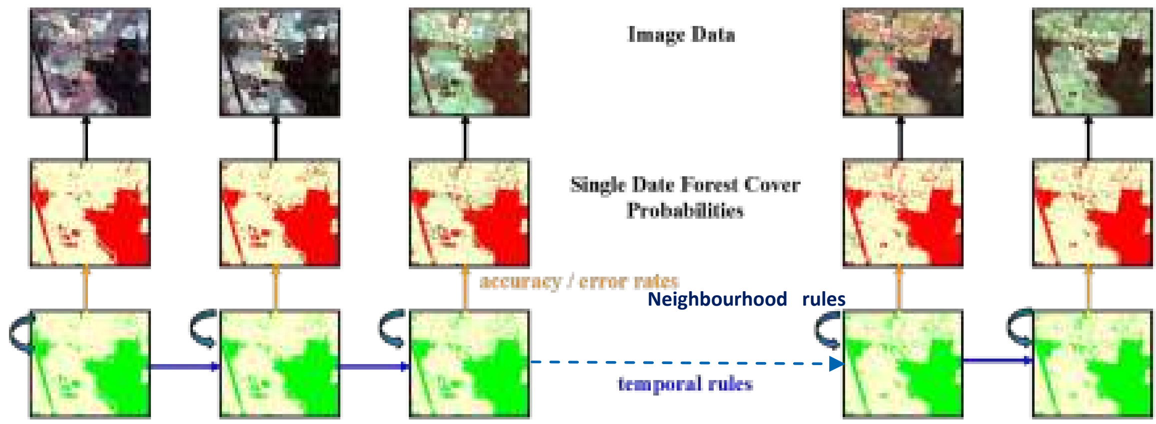


Figure 7.3.2. A pictorial representation of a conditional probability network as a graph as applied to the forest extent and change mapping analyses. The annual sequence is represented in columns from left to right.

The bottom row of Figure 7.3.2 represents the true maps of forest cover at each year; these are the desired outputs but cannot be observed directly. Values for the true maps can be inferred from the other variables. The top row is the satellite imagery. From the satellite imagery we obtain the middle row - the forest probabilities produced from the individual year mosaics (estimates of the true forest cover maps) – through the forest base and between year matching processing.

The graph edges or 'arrows' represent relationships between the variables. Observing a particular value for a variable provides some information about all the other variables to which it is connected. For example, in Figure 7.3.2, a vertical orange arrow ('edge') connecting a single-year classification with the true map of the same year represents the estimated accuracies for cover types in that 'single-date' classification (see below). The horizontal blue arrows represent knowledge about the temporal progressions in land cover, as per Figure 7.3.1. In the case of forest cover mapping, the strength of the relationships depends on the time interval between the image dates. For example, less change would be expected between images a few months apart than between images several years apart.

The rules, or the relationships between the variables, are expressed in terms of conditional probability tables. These tables (or parameters) need to be specified. There are three types of rules, or tables, in the CPN model being used. Error rates tables link the estimated forest cover to the true forest cover at each date (vertical arrows in Figure 7.3.2). Temporal rules link the true forest cover maps through time (horizontal arrows in Figure 7.3.2). Neighbourhood rules link a pixel to its neighbouring pixels (curved arrows in Figure 7.3.2).

The temporal rules describe the relationship between the successive true forest cover maps; and represent probabilities that (e.g.) a forest pixel may be converted to non-forest within one year. In the current LCCA processing, these values are fixed. The probability of land cover change and the probability of land cover not changing are 0.06 and 0.94 respectively for both forest and non-forest. These parameters have been set by experimentation to minimise the mapping of false change without omitting too much true change.

The error rates tables describe the accuracy of the forest cover maps derived from the indices and thresholds. Formally they are the probability of labelling a pixel as forest given that the true ground cover is forest and the probability of labelling a pixel as non-forest given that the true ground cover is non-forest. Both of these probabilities have been set to 0.88 for Indonesia. Due to the way the thresholds are set to avoid omission and commission errors, this constant high value is appropriate. However this value may be modified as discussed below.

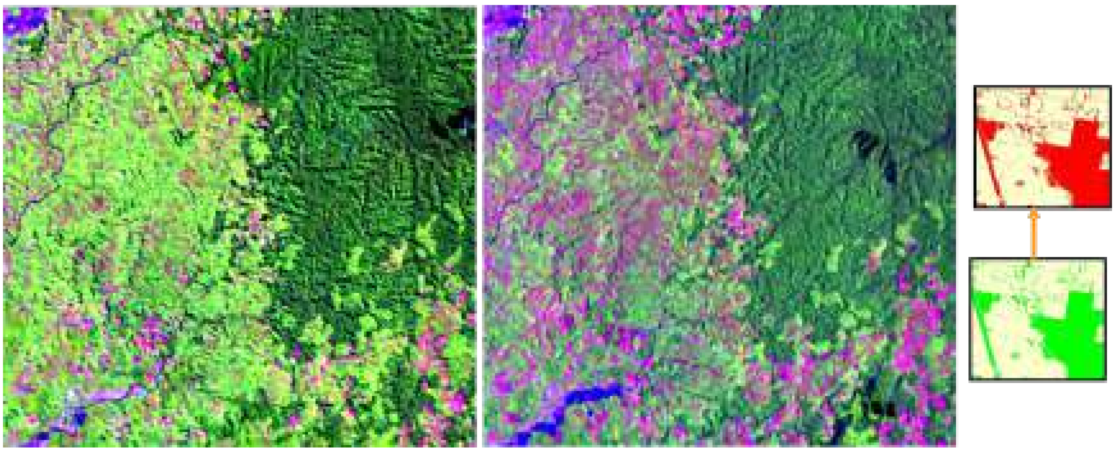


Figure 7.3.3. An image pair from the wet (left) and dry (right) seasons. Landsat TM bands 5,4 and 3 are shown in RGB. The graphic at the far right shows where the single-date classification accuracy is used in the multi-temporal classification model (Figure 7.3.2).

If an image from a non-optimal time of year is used, as illustrated in Figure 7.3.3, then the discrimination between forest and non-forest cover may be affected. The forest appears in shades of darker green in both the images and occurs mostly in the top right of the area displayed. The image to the left is from the wet season. There is a mixture of shades of green in the forest as well as a range of shades of green throughout the adjoining agricultural lands. The image to the right is from the dry season. There are fewer shades of green in the forest and much less green in the agricultural lands. There is much greater separation between the forest signal and the non-forest signal in the dry season image and there will be much less uncertainty and error in the classification. To represent this in the multi-temporal classification model the error rates for the wet season image can be reduced to 0.83 or 0.75 for one or both cover types, if required. This modification is only applied for years where the analyses showed reduced discrimination between the forest and non-forest sites for most stratification zones. A value of 0.83 is tested first, then 0.75 or 0.65.

Spatial probability rules (neighbourhood rules) are also implemented within the CPN. Small areas of one or two pixels classified differently from the majority of neighbours (e.g. as non-forest surrounded by forest) are likely to be mixed pixels or incorrect and to result in false changes over time. Neighbourhood rules (weights based on neighbouring pixels, Kiiveri and Caccetta, 1998) modify the output probabilities of such pixels. The weightings used are the same as implemented in the Australian forest monitoring system.

Figure 7.3.4 shows the effect of the CPN on the single-date probabilities, and on filling cloud gaps. Data from the five years from 2001 to 2005 are displayed. The image data shows an area in Central Kalimantan which contains wetland forest. The south-east (bottom right) has been cleared for agriculture in the 1990s. Small areas are still being cleared in the remaining natural forest. The second row in the figure shows the single-date forest probabilities for this region for each year. Red is certain forest and dark blue is certain non-forest. The other colours show uncertain regions. Yellows and oranges indicate high probability of forest cover, light blues low probability of forest cover and green is about 50%. Black shows areas with no data – cloud gaps. The third row in the figure shows the refined probabilities from the CPN processing, using the same colour display as the single-date probabilities. The uncertain areas have been resolved (only red and dark blue colours are seen) and the most likely probabilities now fill the cloud gaps.

The fourth row in the figure shows the forest extent and change products created from the CPN-refined probabilities.

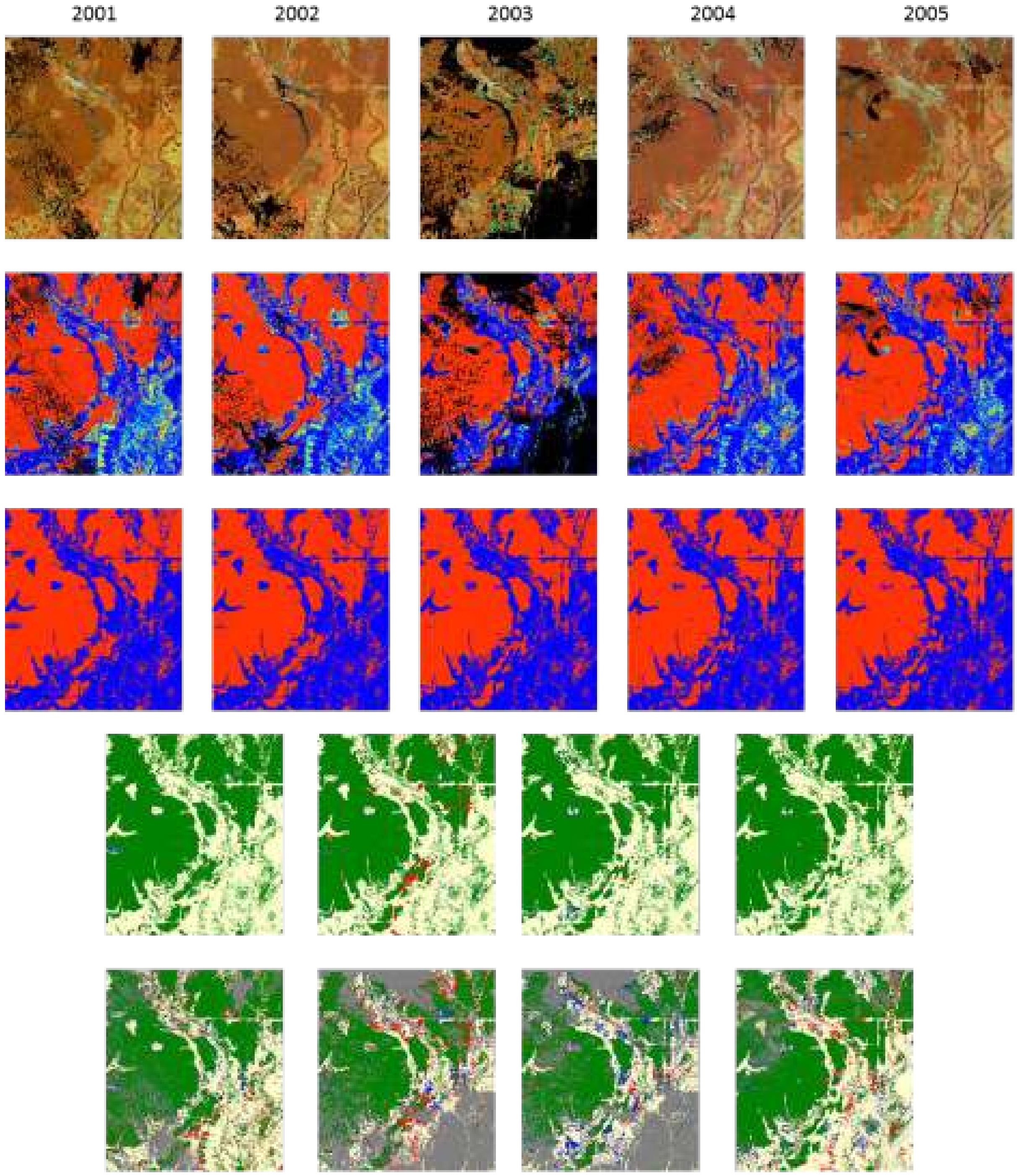


Figure 7.3.5. A display of satellite images (top row – bands 453 in RGB), single-date forest cover probabilities (second row – red is certain forest, through orange, yellow, green and light blue to dark blue as certain non-forest; black is no data due to cloud), refined forest cover probabilities from the CPN (third row – same colours as second row), forest extent and change products derived from the CPN probabilities (fourth row – green is forest in both years, red is forest loss, blue is forest gain and the light background is non-forest in both years) and forest extent and change products derived from the single-date forest probabilities (bottom row – same colours as fourth row with grey showing areas of missing data where change could not be calculated). The region shown is from the Central Kalimantan province.

Probabilities above 50% are labelled as forest and the remainder non-forest within each year; the class labels are then compared to show forest loss and gain. Dark green areas are forest in both years compared, with forest loss shown in red and forest gain shown in blue. The light background colour is non-forest in both years. Distinct areas of forest loss and gain are apparent in each annual interval, most particularly forest loss between 2002 and 2003. For comparison, the bottom row in the figure shows the forest extent and change displays calculated directly from the single-date input probabilities using the same threshold for forest and non-forest. Areas where change could not be calculated because one or other of the years has missing data are shown in grey. More change is shown in these 'single-date' change comparisons and many of these areas not also identified as change from the CPN probabilities flicker between forest and non-forest cover in each year. The missing data and 'flicker change' problem in comparison with the CPN outputs clearly illustrate the value of the CPN processing for monitoring.

This processing is very computationally intensive (more than a day on a regular desktop computer) and so is performed on a high-performance blade server cluster computer system where multiple processors perform the computations in parallel. This reduces the compute time to less than one hour. Spatially the processing units are the quadrants for each mosaic tile to fit within memory limitations on the cluster.

CPN processing commences when the matching process is complete for all time periods for the mosaic. In update mode, the new dates are added and the entire time series is reprocessed. The outputs from the multi-temporal processing are the temporal set of forest probability images, one for each year, where the probabilities of forest cover have been refined by the rules to provide temporally consistent estimates.

The products for forest extent and change maps which are derived from these outputs are described in Section 7.4. The feedback from reviews of the 2000-2009 products is discussed in Section 8.

7.4 Description of Products

As noted above, the outputs from the multi-temporal processing are raster probability images, one for each year, where the probabilities of forest cover have been refined by the rules to provide temporally consistent cover estimates.

The final products required are maps (or masks) that are indicators of forest extent at each year, and forest loss (clearing) and forest gain (revegetation) between successive years. The time series of probabilities are processed together to produce these products as binary raster maps; coded as (1) or (0). These can be displayed as conventional maps (below). In the conversion process, thresholds are applied to each year's probabilities to produce the extent map and differencing of these results to produce the annual change products. It is noted here that the CPN processing fills almost all areas where data were missing in the mosaics, but that in this conversion, any pixels for which there was no input data in any year are recoded to 99 in the conversion. The conversion process is fully automated.

The products are produced in both the local NUTM projection and geodetic projection. The tiles in the geodetic projection are mosaiced into whole island regions. The resolution of all LCCA products in NUTM is 25m, in geodetic it is 0.000025 degrees.

The key products created are:

- Forest extent maps for each year (2000, 2001, 2002, ...)
- Forest loss maps for each annual period (2000-2001, 2001-2002, ...)
- Forest gain maps for each annual period (2000-2001, 2001-2002, ...)
- Year of first forest loss
- Year of first forest gain

The year of forest first forest loss and gain are stored as integers counted from the beginning of the time series. For example, a change in the 2000-2001 interval is stored as 1 and 2001-2002 is stored as 2.

Statistical summaries at national, regional and local scales can be calculated directly from these products in GIS systems. Statistical and map summaries may be produced by a range of users for various purposes. Statistical summaries are not standard LCCA products, but examples are provided here for illustration.

An example of each product is presented and discussed below.

Annual Forest Extent. Figure 7.4.1 shows the 2009 forest extent product for Indonesia, and for Kalimantan as a regional example. There is a forest extent product for each year in the time series.

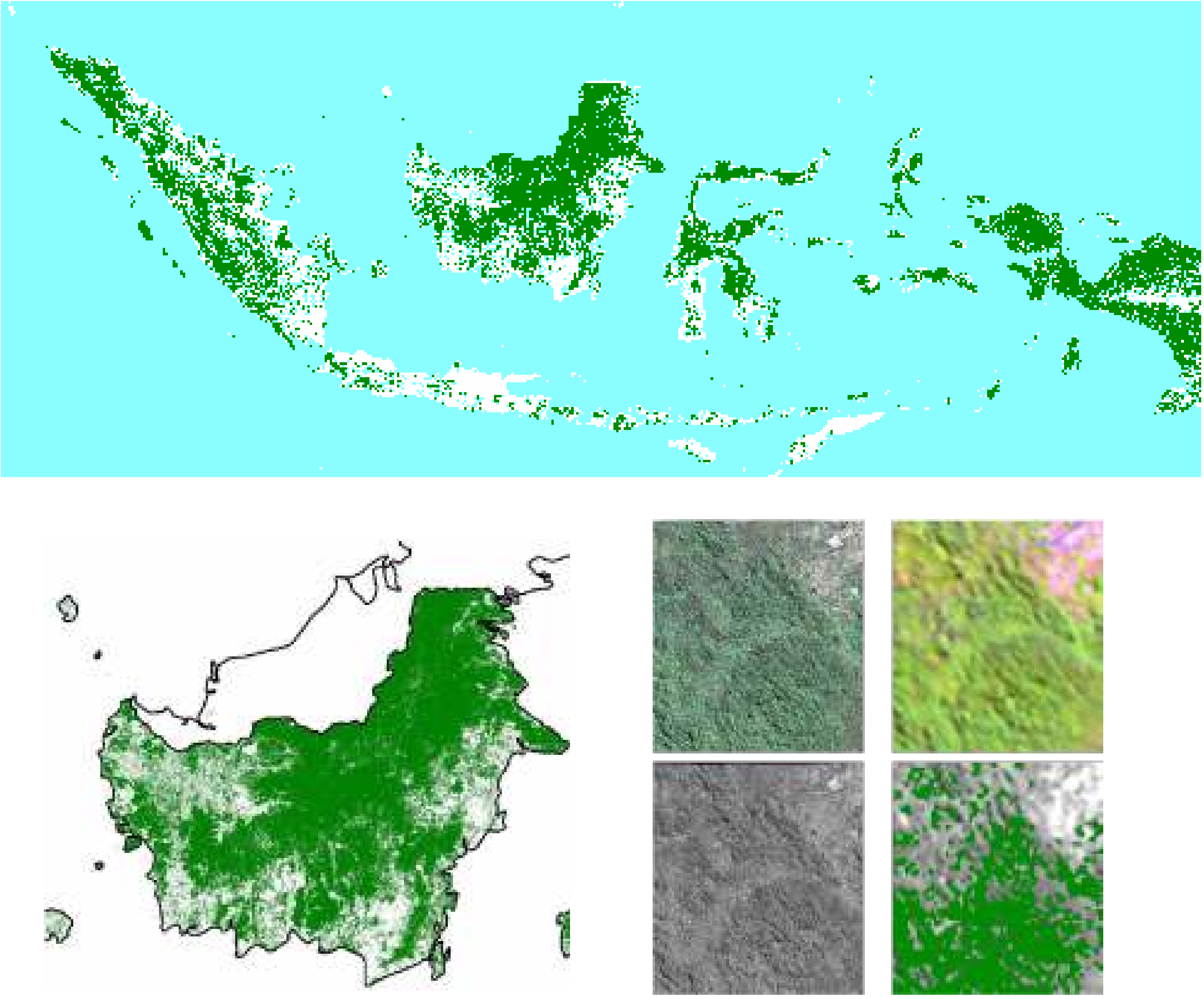


Figure 7.4.1. Illustration of forest extent (2009) at national, regional and local scale. The local scale includes comparison with Landsat and high resolution imagery.

Annual forest loss and gain. Forest loss (clearing or harvesting) and forest gain (regrowth or replanting) products are produced for each annual interval in the time series, e.g. 2000-2001, 2001 – 2002. This can be displayed as total forest change over a number of years as in Figure 7.4.2, and Figure 7.4.3, or year by year change maps as in Figure 7.4.4.

Figure 7.4.2. Total forest loss and gain 2000-2009 at national scale from the LCCA. Dark green indicates areas that were always forest from 2000 to 2009, red shows forest loss between 2000 and 2009 while yellow indicates forest gain in the same period. [source: INCAS poster, Workshop on Earth Observation Satellite Data to Support REDD+ Implementation in Indonesia, February 2014]

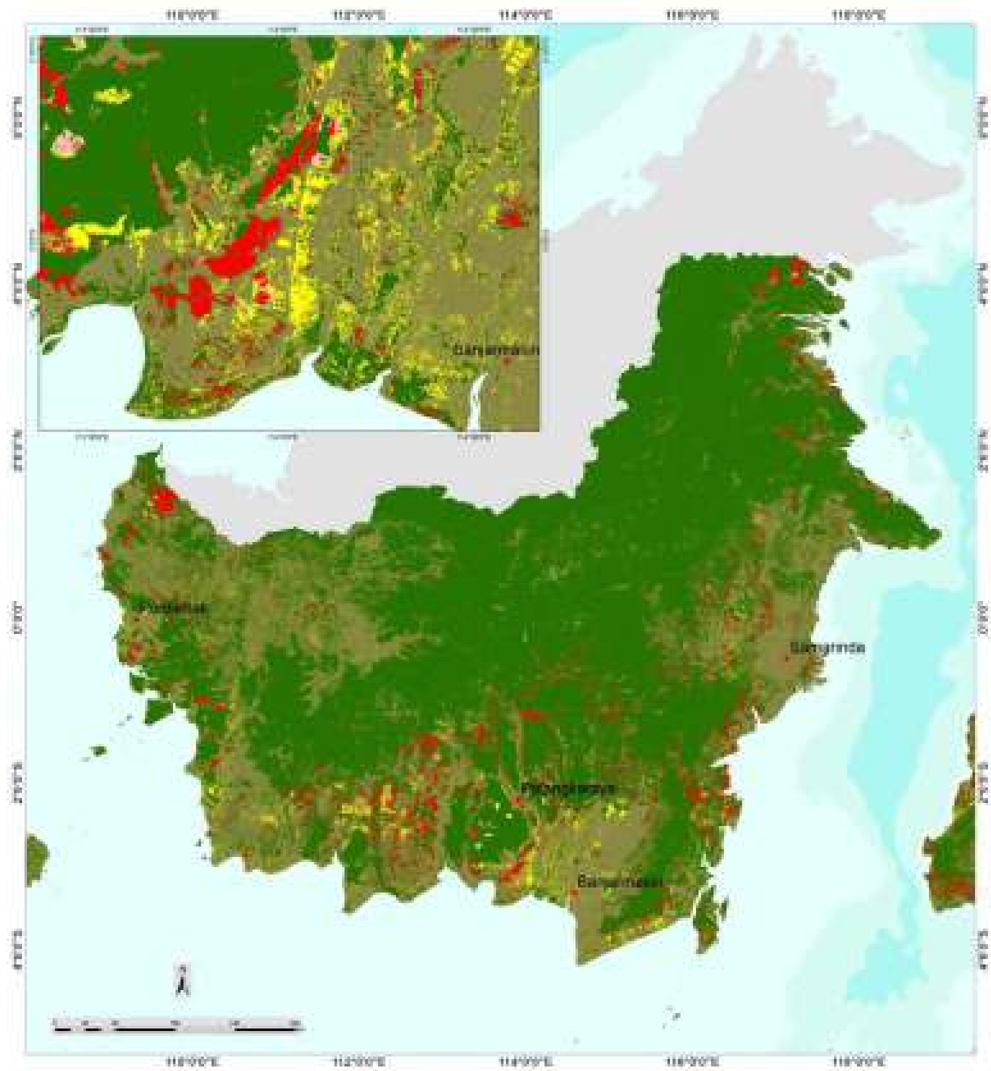


Figure 7.4.3 Forest loss and gain for 2000-2009 at regional scale. Dark green indicates areas that were always forest from 2000 to 2009, red shows forest loss between 2000 and 2009 while yellow indicates forest gain in the same period. [source: INCAS poster, Workshop on Earth Observation Satellite Data to Support REDD+ Implementation in Indonesia, February 2014]

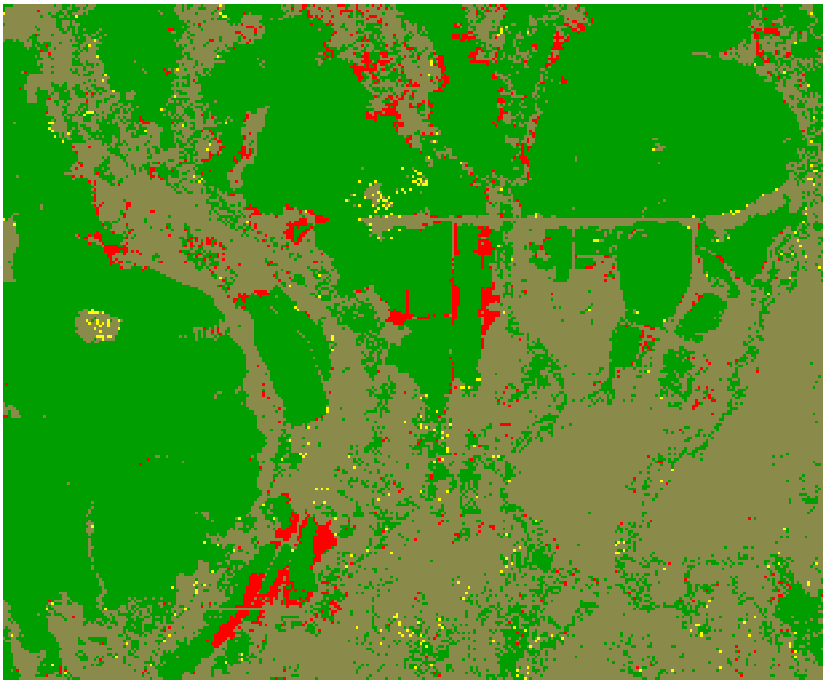


Figure 7.4.4 Forest loss and gain for 2002-2003 at local scale. The region shown is in the Central Kalimantan province. As in the figures above, green is forest in 2002, red is forest loss from 2002 to 2003 and yellow is forest gain from 2002 to 2003. The background colour is non-forest.

Year of first forest loss and gain. This product identifies the year of first change in forest cover within the time series. It was developed specifically for the carbon accounting stage of the INCAS program. In the forest gain product, it provides a planting date for new plantations or a regrowth date for disturbed forest. In the forest loss product it provides a year of disturbance. Deciding whether the date is the first disturbance of natural forest or the harvest date of already disturbed forest or a plantation requires some outside knowledge of the state of the forest at the beginning of the time series.

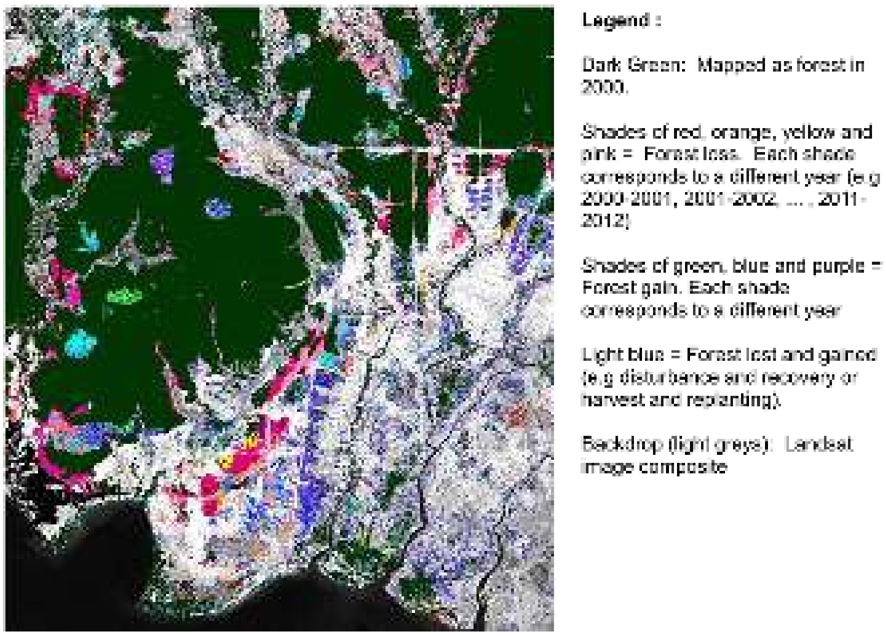


Figure 7.4.5. Year of first forest loss and gain at regional scale coloured by year of change as indicated in the legend above.

Statistical summaries. The area of forested land or forest extent change can be calculated from the LCCA products and presented in tabular and graph formats. Once again, these summaries can be made at national, regional and local levels as shown in Table 7.4.1 and Figure 7.4.6. These examples are provided for illustration purposes only and are not final results.

Table 7.4.1: Sample Forest Change Summary for Sulawesi

Time Period	Forest Loss (hectares)	Forest Gain (hectares)
2000-2001	46002.977	75737.929
2001-2002	111206.196	104178.523
2002-2003	155917.199	102938.227
2003-2004	145241.838	111251.447
2004-2005	119986.428	114067.985
2005-2006	195822.295	60392.266
2006-2007	183959.252	57955.607
2007-2008	231188.88	40375.692
2008-2009	238340.948	64201.398

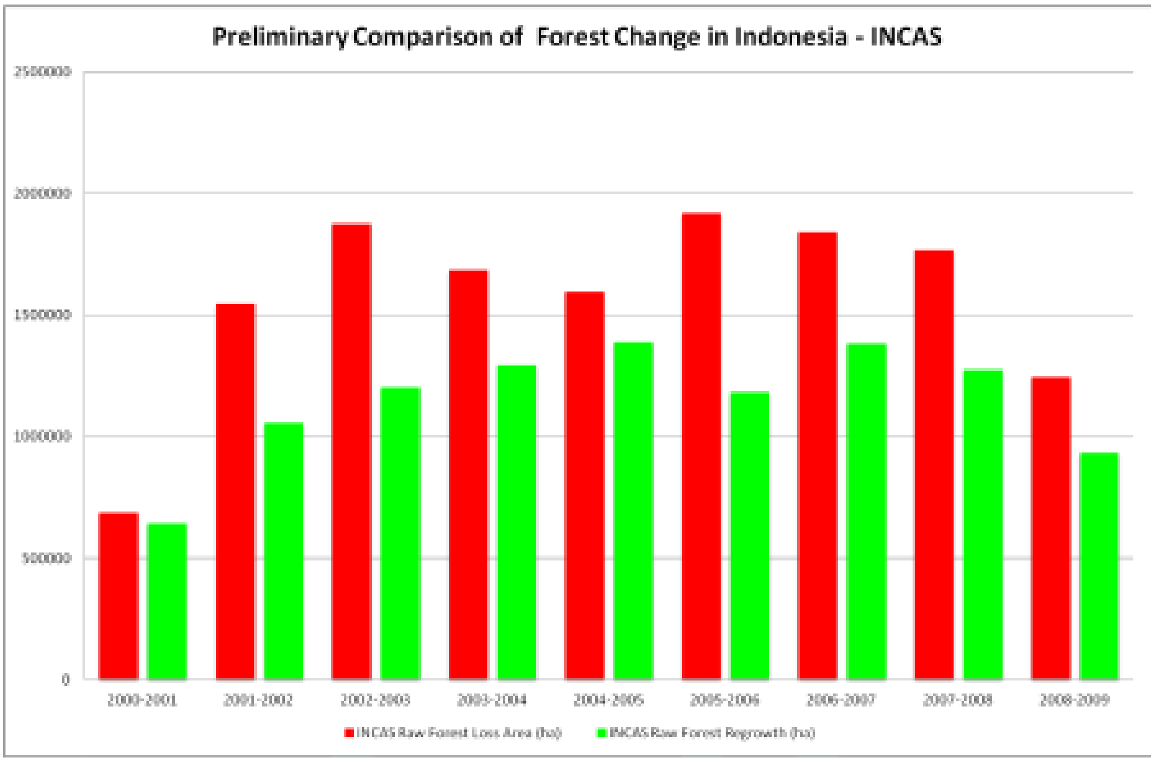


Figure 7.4.6. Sample graph of total areas of national forest loss and gain by year, 2000-2009.

8. REVIEW OF THE PRODUCTS

A key element of any mapping program is the accuracy of the maps produced. A formal accuracy assessment methodology – for both forest extent and change – is yet to be agreed. Understanding of accuracy, uncertainty and error is critical for use of maps in different policy contexts. Understanding of error also may feed back into a continuous improvement process. In the LCCA, there has been formal interaction with local experts to get feedback on the 2000-2009 version of the products. We are acting on the issues identified in this feedback to improve the accuracy of the 2000-2012 version of the products. The review process, which combines ground knowledge and the image processing experience which generated the products, fulfils two major roles within continuous improvement. They are:

1. it may identify errors which can be corrected by revising steps in the current process (e.g. stratification or thresholds for a zone), or it may identify errors or uncertainties which suggest need for investigation of new methods.

2. it may identify and locate particular ground cover types where spectral data cannot adequately classify the desired cover types, and errors in the extent of change results. Location of these cover types (e.g. by drawing vectors) provides a basis to remove particular errors of this kind, and, where required for particular purposes, to seek other information to identify and reduce such errors as far as possible.

This section of the documentation describes in some detail the review process to date. It provides summaries and examples of the major issues identified and how this information has been used to improve the new products.

8.1 Expert review

A product review workshop was held for each island region, namely Kalimantan, Sumatra, Papua, Sulawesi, Java, Nusa Tenggara and North and South Maluku. There were three main aims for these workshops:

(i) to get feedback on the accuracy of the forest extent and change products, identifying issues that we could fix and issues that must be accepted as limitations of the methodology;

(ii) for the first prospective users to get an understanding of the strengths and limitations of the products; and

(iii) to start to understand what ancillary data is required to use these products for the purpose of carbon accounting.

Carbon accounting rules typically require consideration of why or how the land cover changes – is it human-induced land use change (perhaps from natural forest to rice cropping) or land cover change caused by nature (perhaps a tsunami washes away mangrove forest). Our satellite mapping 'sees' what has changed and when, but has less information on how or why the change occurred. This 'attribution' of the land cover change is considered more completely as part of the carbon accounting part of the INCAS program, but some of the ideas were introduced at these workshops to discover if there were datasets available that would aid in the attribution.

The main outcomes from the forest review workshops are annotated forest change maps and an action list for the processing team of issues that should be considered when the products are next updated. During the workshop, the annotation occurs on paper maps printed from the digital products as can be seen in Figure 8.1.1. These regions are later digitised for easy overlay on the electronic products. Issues for action are identified in lists like that shown in Table 8.1.1.

Figure 8.1.1. Annotated forest change map (detail) from the Sumatra Forest Review workshop and some of the team who created it.

Table 8.1.1: Review actions for Sulawesi

Stratification Zone	Comments	Action
1	Why so much regrowth in 2001?	Review probabilities around these dates
	Blue regrowth is never forest	
2	Too much forest and regrowth through agricultural areas - worst in northeast	Review forest probabilities, consider splitting into more zones
	2001 regrowth in south west is always forest	
	Stripes of clearing in 2009	
3	As good as it gets for wetlands	
4	Too much clearing in 2009 in many places	Review 2009 and temporal consistency of probabilities. Consider new zone in south-west.
	Patches of mixed agriculture called regrowth	
	Too much forest in south-west	
	Small wetland in north is false change	
	Mature oil palm is incorrect but cannot be fixed.	
5	Minor false clearing in 2009	Review if caused by bright / hazy images
6	Too much forest and regrowth in islands to south-east.	Reset base using new ground truth on prints and recalculate. Consider new zone for south-east if needed
	Similar on mainland and islands to the north but lesser extent	Review 2009.
	2009 clearing issues on edge with zone 4??	
11	Clearing (2009) and regrowth (2001) in stripes is mostly likely wrong	Review probabilities near these dates
	Lots of 'blue' regrowth that is never forest	

Forest review workshops are a collaborative effort between expert groups. Participants include:

- People from LAPAN who are part of the processing teams who make the products. They are the experts in how the products were made, and what cover types could and couldn't be easily separated in the available image data.
- People from the MoF representing the teams who will be potential users of the data for both carbon accounting and other purposes.
- People from national and local agencies who know the land cover and the land use in the regions being reviewed.

Some of the people from the MoF and from the regions were part of the team providing local input to the creation of the forest base and some were people new to the project at this review stage.

Maps of initial forest extent and the year of first forest loss or forest gain (see Figure 8.1.2) were printed in large format to cover the whole region being reviewed at a scale where local change events and landmarks could be recognised by the team.

The experts come together around the printed maps. The ground experts consider where the maps are right and where the maps are wrong:

- Is the forest extent broadly correct? Are there areas of forest that are omitted? Are there regions mapped as forest that are not forest?
- Is the change extent and timing broadly correct? Are there areas mapped as change that were always forest or never forest? Are the areas of known change included in the maps?
- Is there ancillary data that can be accessed by the processing team to support the expert opinions? Can this data be used to help with attributing the cause of the change?

The data analysis experts consider the areas where the maps are found to be inaccurate. Together with the ground experts, they consider if the errors can be fixed. Are they caused by limitations in the separability of some cover types in the available satellite imagery? Is it because of how the forest (or non-forest cover) grows, e.g. deciduous forest, or how the forest is managed, e.g. teak and cinnamon? Is it because of some error or misunderstanding in the data processing and the methodology can be improved? The action lists (Table 8.1.1) are the summaries of these discussions.

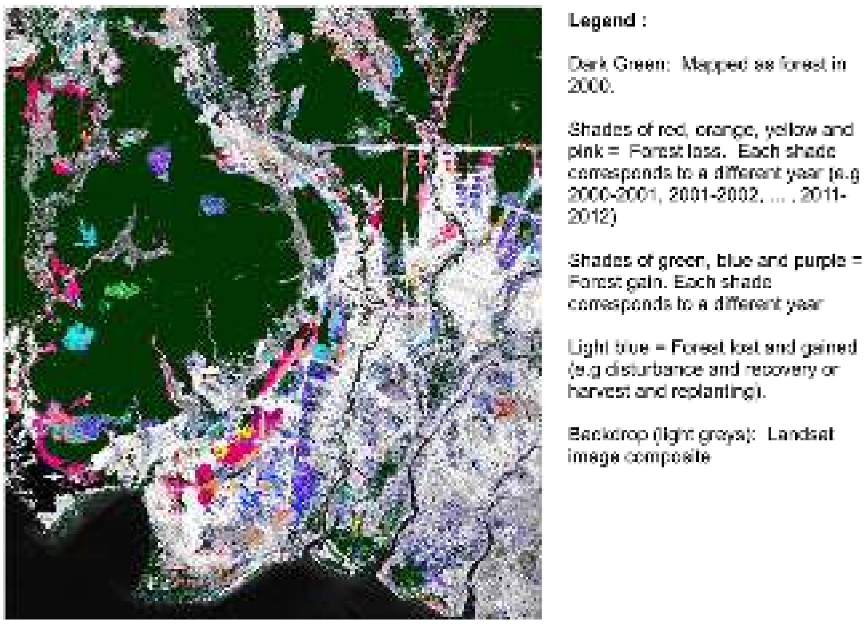


Figure 8.1.2: Year of first forest loss and gain at regional scale coloured by year of change as indicated in the legend above.

The conversations are recorded by annotating the printed maps (Figure 8.1.3). This includes an outline of the region being discussed and a very short description of the issue (a true cover type, date of change or confirmation of the change and the new cover type). Any ancillary datasets relating to the issue are collected. A digitising person is assigned to each working group who will later convert the drawings to vectors. Often the printed maps are scanned as an additional record of the process.

Although they do not provide a formal accuracy assessment, these forest review workshops provide invaluable feedback on errors in the products, both in forest extent, land use change and the timing of the change.

Figure 8.1.3. Annotation of forest extent and change maps during the forest review workshops.

8.2. Issues and actions identified

The main issues recognised during the forest review workshops are discussed below. They relate to problematic cover types (palms, wetlands, shrubs, deciduous forest, some forest management practices) where operational solutions are suggested, and to 'processing learning errors' which highlighted where current operator processes could be improved.

8.2.1 Cover type - palms

Young palms are spectrally separable from forest cover types in the Landsat image data, but the spectral signal of mature palms typically becomes more like the spectral signal of some of the true forest. This is essentially a limitation of the current products, but various strategies are considered to restrict the consequences of the error.

In Kalimantan and Sumatra the most common types of palms are oil palms (sawit). The clearing of forest prior to the planting of new oil palm is accurately detected and the land cover of the new oil palm plantation is mapped as non-forest for the first three to five years. After that, the plantation signal becomes more like forest and plantations appear as forest regrowth in the LCCA products. The simplest strategy here is to obtain or produce an oil palm extent map from another source. The oil palm map can be used to mask the regrowth in the LCCA product and the LCCA product can return a date of planting for new oil palm plantations. An alternative is to investigate palm mapping using other data sources (e.g. radar data) and integrate such data products into the LCCA processing stream.

Coconut palms (kelapa) are common in Sumatra and other parts of Indonesia. There are far fewer new plantations than for oil palm so most of the existing palm fall into the 'mature' category. They are also frequently grown together with tree crops in 'mixed' agricultural areas. Stratification zones are used to separate the major coconut palm cropping areas from the surrounding forest. Large commercial plantings predominantly occur in low lying coastal areas with boundaries that can be relatively easily defined. These regions may include coastal mangroves; however mangrove forests and coconut palms are spectrally separable. In the smaller, 'mixed plantings' regions the forest extent is either under- or over-estimated depending on whether there are more tree crops / natural forest or more palms.

Sagu and nipah palms are found naturally mixed into the forest in Papua. Spectrally they are very similar to forest, particularly when it is wet, and their co-occurrence with trees tests 'canopy density' limits in the definition of forest. They are part of the long-term natural cover and so temporal patterns cannot be used to separate the palms from the forest. Typically the regions are labelled forest which avoids inconsistencies causing false change.

8.2.2 Cover type -wetlands

In the wet season when wetlands are full of water, the spectral signal is very different to forest. In the dry, when the surface is bare of water, the spectral signal is also very different to forest. In between these two times, the cover is often green vegetation (grasses or reeds) growing in shallow surface water. This particular mixture of green vegetation and dark water signals appears very similar to the spectral signature of wetland forest types. What makes the forest different to wetlands is that the spectral signal from the wetland forest is relatively constant from year to year and from season to season but the signal of the other wetland vegetation varies from season to season and year to year. However, if the best available images are during the transition from wet to dry for several years in sequence, a sequence of 'forest-like' signals build and wetlands can be misclassified as forest. As this signal doesn't persist for as long as true forest, the products tend to show multiple regrowth and clearing events over many wetland areas.

Similar spectral confusion can arise in rice growing areas. When the paddy is flooded or completely dry, the spectral signal is very different to forest. When there is a mixture of water and green shoots, the spectral signal becomes more like forest. When the water recedes and there is just green crop, the signal is different to forest again. Unlike wetlands, major rice growing areas are geographically well separated from forest and stratification zone boundaries can usually be used to provide separation from forested areas. Wetlands tend to be much more ‘mixed’ into natural forest environments and stratification is not an efficient option. However, knowledge of these wetland locations as a spatial database can be built up using imagery and other data. Sourcing or developing a separate wetlands map (perhaps considering only wet season imagery and terrain) and using it to mask the false change is currently the recommended strategy.

8.2.3 Cover type - shrublands

The main difference between trees and shrubs is height; hence shrublands are excluded from forests by the height component of the forest definition. The Landsat imagery provides no information on vegetation height. Ground truth information (the shadows in high resolution imagery give an indication of feature height) and local knowledge are used to identify the major regions of shrubs with little (below 30%) or no tree cover. Stratification zone boundaries are again used to separate such regions from forest (more than 30% trees even if some shrubs are also present). Where there is a natural separation of trees and shrubs or historical land clearing has created an artificial separation, this strategy is effective. Where there is a continuum of cover from trees to shrubs, a choice is made to map the area as either forest or not forest depending on which minimises the error in the overall forest extent.

8.2.4 Cover type - deciduous forest

Deciduous forests lose their leaves during some part of the year. In Indonesia such species usually lose their leaves during the dry season. This occurs with teak (jati) forests, predominantly in the eastern half of Java and other species in eastern Java, Bali and parts of Nusa Tenggara and Maluku. If imagery is captured during the dry season, there is little or no canopy and the majority of the signal is from the ground below the tree. The signal is very unlike forest. In the wet season, there is full canopy cover and a strong forest signal. As several image dates are required to build cloud-free composite images within each year, there is limited ability to select seasonally-consistent image dates. Local knowledge has been used to create stratifications zones that separate deciduous forest from other forest types. Within these zones we try to use the understorey signal to map deciduous forest as forest even in the dry season. Even with these strategies, the amount of false change mapped is higher in deciduous forests than in non-deciduous forests.

8.2.5 Cover type and management - forest management practices

There are two main examples of forest management practices affecting the accuracy of the forest extent and change mapping. The first example is teak (jati) forest management. The quality of the wood depends on the amount of sunlight it receives. Trees are planted further apart to minimise the shadow of one tree falling on another tree, decreasing canopy density in a plantation. The ground beneath and between the trees is used for other agricultural crops, further confusing the signal received. Combined with the deciduous nature of the teak forests in some regions, these forests are difficult to map consistently. There are errors in both forest extent and change in such forests. Stratification zone boundaries are used to separate these forests from other cover, but the recommended solution is to seek or develop an independent teak forest extent map. As a very high value timber resource, information on its extent is likely to be available. Depending on quality and currency, such a map may be used to mask false change rather than adjusting the forest extent.

The other example of forest management practices causing more error is in cinnamon forests. Coppicing - cutting young tree stems down to near ground level to encourage many new shoots - is used to maximise the area of new young bark that contains the cinnamon spice. The cutting in cinnamon forests can occur every two to three years. Images acquired after cutting show bare stumps and little or no canopy. Large plantation areas have little or no temporal consistency in their forest / not forest signal. Again, stratification zones are used to separate the major plantations in Sumatra and attempts made to treat the ‘cut’ signal as forest. Natural cinnamon has a less regular cutting pattern and there is usually enough canopy and surrounding trees to allow more reliable mapping.

8.2.6 Learning errors – processing and skills improvement

It should be noted that the first national products (2000-2009) were created by teams trying to both master new processing methodologies and adapt those methodologies from an Australian / international application to Indonesian specific conditions while being required to meet tight processing deadlines.

Processing errors were made initially that the now-experienced teams would detect and correct before product creation, but the learning experience requires finding the consequences of those mistakes to develop the judgement and understanding of how to look for the errors and either prevent them or discover and fix them in a timely fashion.

Many of the identified errors all fall into the ‘easily fixable’ category. Strategies, in the form of processing software, operator review and QA, have been developed to prevent many of these errors. In the update processing (for years 2010-2012) these strategies have already been implemented and the errors themselves are being corrected at the same time as the new data is being processed. Improvements in efficiency and product consistency have resulted.

8.3 Product improvement

There are two aspects to a 'continuous improvement' program. One aspect is to take feedback from product review and any formal accuracy assessments and review the data processing to correct as many of the errors as possible. How this is being done in the LCCA program is discussed here. The second aspect of 'continuous improvement' is to review the methodology and the efficiency of its implementation to make the processing as efficient and accurate as possible. The plans for this are discussed as part of Section 2 'Status and Future Directions'.

The relevant outcomes from the Forest Review workshops are a list of problems found in the products for a particular region and the spatial location of the errors (Figure 8.1.1 and Table 8.1.1 in the preceding section). From these lists we can prioritise actions to address the problems. The factors considered in deciding what action to take and when include:

- the spatial extent of the problem;
- the availability of a solution;
- the time required to implement the solution; and
- the sensitivity of the subsequent purpose (or purposes) to the problem.

Fixing a problem that covers a large area spatially will provide a greater improvement to the products overall than working first on problems that are very localised. However, the size of the error is not the only consideration. Often local problems have quite simple solutions that are quick to implement. For example, several areas of mangrove forest along the south-east coast of Sumatra were mapped as non-forest in the 2000-2009 (version 1) products. The extent of the area omitted from the forest extent is quite small compared to the area of false change in cinnamon forests in south western parts of Sumatra (Figure 8.3.1).

Figure 8.3.2 shows the original stratification zone map for Sumatra. There are stratification zones for mangroves in the north-east (zone 28) and mangroves versus coconut palms along the eastern part of the north coast (zone 2), but no such stratification zone in the south-east. The solution to the omitted mangrove forests is simply to create a new mangroves zone (now called zone 29) and copy the indices from the existing mangrove zone. Experience in other parts of Indonesia shows that mangrove forest can be mapped very accurately when suitable stratification zones are used.

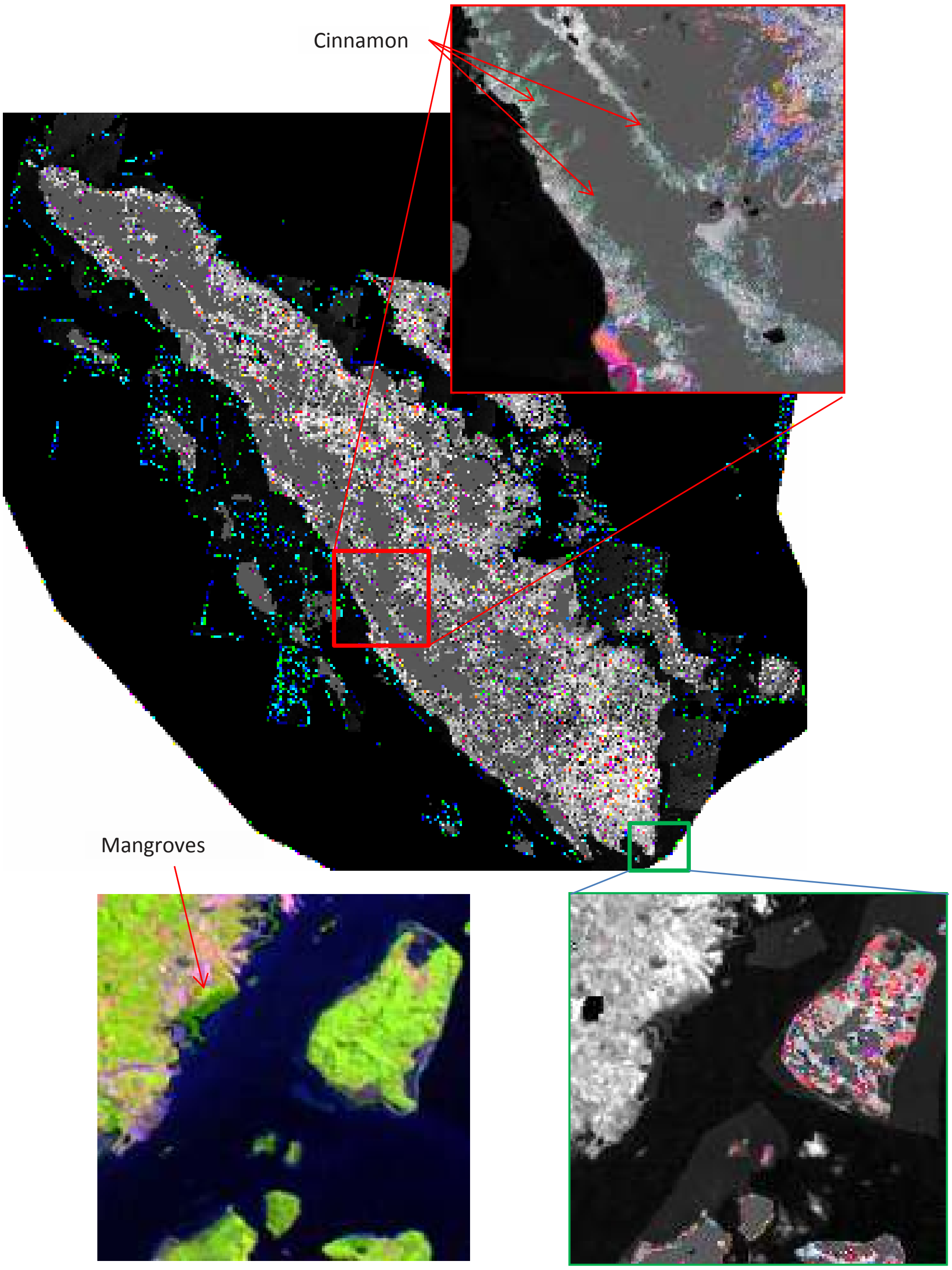


Figure 8.3.1: 2000-2009 year of first change map with insets showing the cinnamon forest false change and omitted mangrove forest. The change map shows forest extent in dark grey, forest loss in shades of red, pink and orange and forest gain in shades of green and blue. The Landsat satellite image bottom left is from 2009 with bands 543 in RGB.

A stratification zone (zone 11, Figure 8.3.2) was created in Sumatra for the larger commercial cinnamon plantations, but most of the areas identified as false change in cinnamon forests are in neighbouring zone 1. Zone 1 includes a number of dryland forest types and adjacent agricultural regions. The advice from the experts is that the cinnamon in zone 1 is mixed into the edges of the forest, between the more mountainous forest and the agricultural regions. Extending or adding stratification zones is not an option. Instead, a full review and adjustment of the time series of forest probabilities, and potentially derivation of new indices better suited to mapping cinnamon, is required. This is a much more time consuming task, and with the complications of the management of coppicing in the management of cinnamon forest, the results are likely to be only 'better', not 'perfect'.

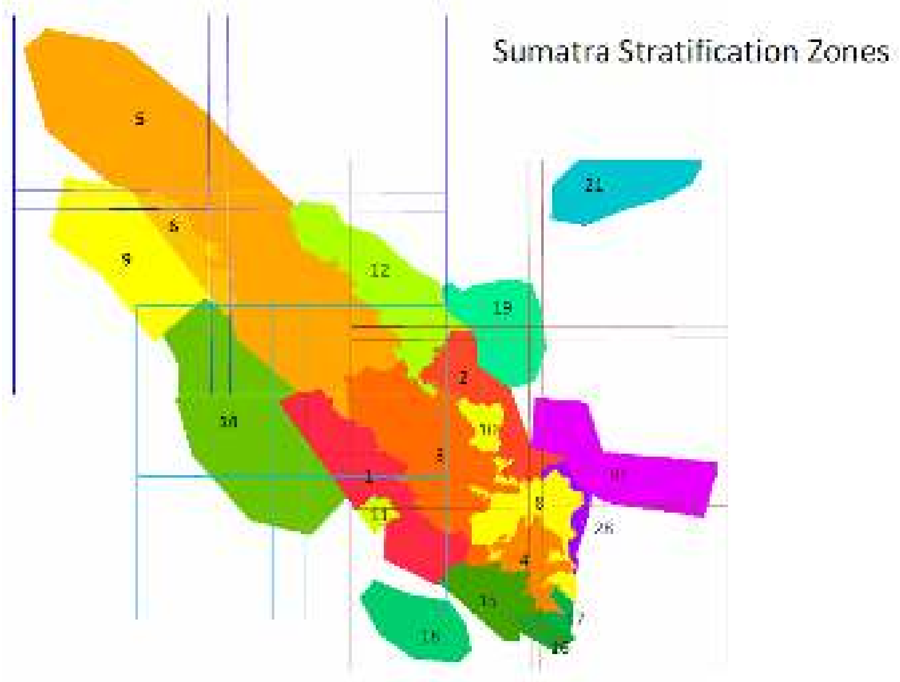


Figure 8.3.2. Stratification zone map from creating the Sumatra 2000-2009 (version 1) products.

Both of these problems were addressed and improved in creating the 2000-2012 (version 2) products for Sumatra. Figure 8.3.3 shows the new products.

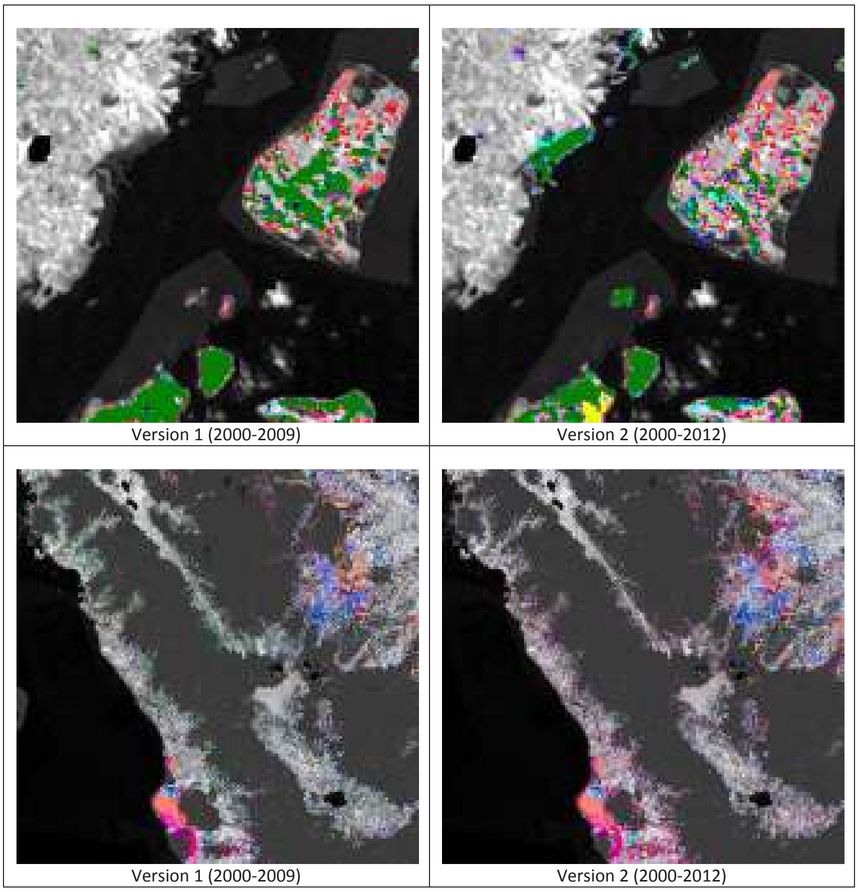


Figure 8.3.3. Original and improved year of first change maps for the inset areas in Figure 8.3.1. The change map shows forest extent in dark green (top) or dark grey (bottom), forest loss in shades of red, pink and orange and forest gain in shades of green and blue. In version 2, the mangrove forests are no longer omitted and the cinnamon forests appear as stable forest instead of forest regrowth.

The most efficient time to make corrections to the processing and products is when an additional year is being added to the time series. Corrections to the version 1 products were made for each island group as part of the update when data for 2010, 2011 and 2012 were added to the time series. The expert reviews can be revisited, comparing the original and new products. The lists of issues with the products are updated to recognise the corrections, and perhaps new issues that are now apparent once other issues are resolved are added. Policy considerations of different users may provide clear priorities for efforts to implement corrections, either within the image processing stream or by using other data. For carbon accounting, a means to identify oil palm from forest (as discussed above) may be a high priority, while for coastal habitat management, accurate maps of mangroves may be a priority.

This intention is an ongoing program of reviewing the products, both informally and formally when possible, and acting on the feedback from those reviews to improve the products over time.

9. REFERENCES

Caccetta, P. A., Furby, S. L., O'Connell, J., Wallace, J. F. and Wu, X. (2007). Continental Monitoring: 34 Years of Land Cover Change Using Landsat Imagery. Paper presented at the 32nd International Symposium on Remote Sensing of Environment, June 25-29, 2007, San Jose. Costa Rica.

Caccetta P.A. 1997, *Remote sensing, geographic information systems (GIS) and Bayesian knowledge-based methods for monitoring land condition*. Ph.D. thesis. School of Computing, Curtin University of Technology, Perth, WA, Australia.

Caccetta, P.A. Furby, S.L, Wallace, J.F., Wu, X, Richards, G., and Waterworth, R. 2013. Long-Term Monitoring of Australian Land Cover Change Using Landsat Data: Development, Implementation, and Operation. In: Achard F. & Hansen M.C (eds). Global Forest Monitoring from earth Observation, chapter 13. Taylor & Francis. New York.

Campbell, N.A., and Atchley W.R. 1981. The geometry of canonical variate analysis. Syst. Zool. 30(3): 268-280.

Danaher, T., Wu, X. and Campbell, N. A. (2001). Bidirectional Reflectance Distribution Function Approaches to Radiometric Calibration of Landsat TM Imagery. In: Proceedings of the International Geoscience and Remote Sensing Society Symposium, Sydney, Australia, July 2001.

Furby, S.L. 2002. Land Cover Change: Specification for remote sensing analysis. National Carbon Accounting System Technical Report 9. Australian Greenhouse Office, Canberra.

Furby, S.L, Caccetta, P.A. and Wallace, J.F. 2010. Salinity Monitoring in Western Australia using Remotely Sensed and Other Spatial Data. *J Environ*. Quality, 39, pp16-25.

Gutman, G., Byrnes, R., Masek, J., Covington, S., Justice, C., Franks, S., and Headley, R.. 2008. Towards Monitoring Land-cover and Land-use Changes at a Global Scale: The Global Land Survey 2005. Photogrammetric Engineering & Remote Sensing, January 2008.

Hansen M.C, Roy D.P., Lindquist E., Adusei B., Justice C.O., and Altstatt, A. 2008. A Method For Integrating MODIS and Landsat Data for Systematic Monitoring of Forest Cover Change in The Congo Basin. Remote Sensing of Environment, 112, 2495-2513.

Irish, R.R., Barker, J.L., Goward S.N.,and Arvidson, T. 2006. Characterization of the Landsat -7 ETM+ Automatwd Cloud-Cover Assessment (ACCA) Algorithm. *Photogrammetric Engineering and Remote Sensing*, 72, no 10, pp1179-1188.

Kiiveri, H.T., and Caccetta, P.A. 1998. Image fusion with conditional probability networks for monitoring salinisation of farmland. *Digial Signal Processing*. 8(4):225–230.

Kiiveri, H.T., Caccetta, P.A., and Evans F.H. 2001. Use of conditional probability networks for environmental monitoring. *International Journal of Remote Sensing*. 22(7):1173–1190.

Kiiveri, H.T., Caccetta, P.A., N.A. Campbell, Evans F.H., Furby S.L., and Wallace, J.F. 2003. Environmental monitoring using a time series of satellite images and other spatial data sets. p. 49–62. In: D.D. Denison et al (eds) *Nonlinear estimation and classification by lecture notes in statistics*. Springer Verlag, New York.

Lauritzen, S.L. 1992. Propagation of probabilities, means and variances in mixed graphical association models. J*ournal of the American Statistical Association*. 87(420):1098-1108.

Lauritzen, S.L., and D.J. Spiegelhalter. 1988. Local computations with probabilities on graphical structures and their application to expert systems. *Journal of the Royal Statistical Society*. Series B 50(2):157-224.

Martinuzzi S, Gould, W.A., and Ramos González O.M. 2007. Creating Cloud-Free Landsat ETM+ Data Sets in Tropical Landscapes: Cloud and Cloud-Shadow Removal, USFS IITF-GTR32.

O'Connell, J. and Caccetta, P. A. (2009), Testing of Alternate Classification Procedures within an Operational, Satellite based, Forest Monitoring System. In: '*Innovations in Remote Sensing and Photogrammetry*', *Lecture Notes in Geoinformation and Cartography*, 121-132, Springer Verlag Berlin Heidelberg, ISBN 978-3-540-88265-7.

Teillet, P.M., Guindon, B., Goodenough, D.G., 1982, On the Slope-Aspect Correction of Multispectral Scanner Data, *Canadian Journal of Remote Sensing*, 8(2), 84-106.

Townshend, J. R. G., C. O. Justice, W. Li, C. Gurney, and J. McManus, 1992. The impact of misregistration on the detection of changes in landcover. *IEEE Transaction on Geoscience and Remote Sensing,* 30:1054–1060.

Tucker, Compton, J., Grant, Denelle M., and Dykstra, Jon D. (2004). NASA's Global Orthorectified Landsat Data Set, *Photogrammetric engineering and Remote Sensing*, Vol 70, No. 3, March 2004, pp.313-322.

USGS 2009, Global Land Survey, 2000, Landsat ETM+, USGS, Sioux Falls, South Dakota.

Vermote, E., D. Tanré, J. L. Deuzé, M. Herman, and J.J. Morcrette, *Second Simulation of the Satellite Signal in the Solar Spectrum (6S), 6S User Guide Version 0*, April 1994.

Wang, B, Ono, A, Muramatsu, K, and Fujiwara, N. Automated Detection and Removal of Clouds and Their Shadow from Landsat TM Images. *IEICE Trans. Inf. & Systems*, 28, E82-D, pp453-460.

Wu, X., Furby, S. and Wallace, J.F. 2004. An approach for terrain illumination correction. In: Proceedings of the 12th Australasian Photogrammetry and Remote Sensing Conference. Fremantle Western Australia.

Wu, X. (2008). A Simplified Ortho-Rectification Approach for Satellite Imagery. In: Proceedings of the XXI Congress of the International Society for Photogrammetry and Remote Sensing Vol.IIIa, pp.1-6, 3-11 July, 2008.Beijing, China.

BAPPENAS

LAPAN