ACC DAR Adapting to Climate Change in Coastal Dar es Salaam

WORKING PAPER

Development of a Methodology for Land Cover Classification in Dar es Salaam using Landsat Imagery

Rome, 27 January 2012

Grant Contract Beneficiary:	***Sapienza University of Rome***
Contact Person:	***Silvia Macchi***
Partner in the Action:	***Ardhi University Dar es Salaam***
Associate in the Action:	***Dar City Council***
Project title:	***Adapting to Climate Change in Coastal Dar es Salaam***
Project acronym:	***ACC Dar***
Contract number:	***2010/254-773***
Project duration:	***01/02/2011 – 31/01/2014***

Author(s): ***Congedo, Luca (Sapienza University of Rome)***

Munafò, Michele (ISPRA - Italian National Institute for Environmental Protection and Research)

How to quote: *Congedo, L. and Munafò, M. (2012) Development of a Methodology for Land Cover Classification in Dar es Salaam using Landsat Imagery. [pdf] Rome: Sapienza University. Available at: <http://www.planning4adaptation.eu/>*

The project is co-funded by European Union

Table of contents

Figures

Tables

Acronyms and Abbreviations

ACC Dar – Adapting to Climate Change in Coastal Dar es Salaam
CC – Climate Change
CNES – Centre National d'Études Spatiales
DN – Digital Number
DOS – Dark Object Subtraction
ETM – Enhanced Thematic Mapper
ETM+ – Enhanced Thematic Mapper Plus
EVI – Enhanced Vegetation Index
GCP – Ground Control Points
GIS – Geographic Information System
HRG – High Resolution Geometric
HRVIR – High-Resolution Visible and Infrared
LC – Land Cover
LCC – Land Cover Change
LDCM – Landsat Data Continuity Mission
LMI – Landscape Metrics Indices
LPGS – Level 1 Product Generation System
LU – Land Use
ML – Maximum Likelihood
MODIS – Moderate Resolution Imaging Spectroradiometer
NASA – National Aeronautics and Space Administration
NDVI – Normalized Difference Vegetation Index
NIR – Near Infrared
SLC – Scan Line Corrector
SPOT – Satellite Pour l'Observation de la Terre
TM – Thematic Mapper
UDEM – Urban Development and Environment Management
USGS – United States Geological Survey
WRS – Worldwide Reference System

Glossary

Georeferencing – The process of image registration to a map coordinate system, in order to have every pixel addressable in terms of east and north, or latitude and longitude (Richards & Jia, 2006).

Land Cover – The "physical material at the surface of the earth. It is the material that we see and which directly interacts with electromagnetic radiation and causes the level of reflected energy that we observe as the tone or the digital number at a location in an aerial photograph or satellite image. Land covers include grass, asphalt, trees, bare ground, water, etc." (Fisher, et al., 2005, p. 89).

Land Cover Change – The detection of changes in Land Cover, usually analysing multitemporal data; in remote sensing, Land Cover Change will result in changes in reflectance values (Lu, et al., 2011).

Land Use – The "description of how people use the land. Urban and agricultural land uses are two of the most commonly recognised high-level classes of use. Institutional land, sports grounds, residential land, etc. are also all land uses" (Fisher, et al., 2005, p. 89).

Radiance – The "flux of energy (primarily irradiant or incident energy) per solid angle leaving a unit surface area in a given direction", "Radiance is what is measured at the sensor and is somewhat dependent on reflectance" (NASA, 2011, p. 47).

Reflectance – The "ratio of reflected versus total power energy" (NASA, 2011, p. 47).

Remote Sensing – The measurement of the energy emanating from the earth's surface, using a sensor mounted on an aircraft or spacecraft platform, in order to obtain an image of the landscape beneath the platform (Richards & Jia, 2006).

Urban Sprawl – The unplanned, low-density urban expansion, characterized by a mix of land uses on the urban fringe (EEA, 2006).

Vulnerability – The "degree to which a system is susceptible to, or unable to cope with, adverse effects of climate change, including climate variability and extremes" (IPCC, 2001, p. 21).

Acknowledgements

Eng. Luca Congedo and Eng. Michele Munafò were the main authors of this study.

Landsat imagery was accessible from the USGS archive.

Thanks go to the DCC for providing the set of GIS data and remote sensing imagery related to Dar Municipality, and thanks go to Eng. Liana Ricci who collected that data.

Thanks go to Eng. Dionis Rugai for his collaboration and photointerpretation of images of Dar.

Foreword

This working paper presents a methodology developed for the Land Cover (LC) classification of Dar es Salaam, in order to assess the contribution of Land Cover Changes (LCC) to urban vulnerability to Climate Change (CC) in coastal Dar es Salaam (Tanzania).

The Dar es Salaam city is experiencing fast population growth and consequently the built-up area is expanding especially in informal peri-urban settlements that are growing persistently at the fringe.

This study is part of the “Adapting to Climate Change in Coastal Dar es Salaam” (ACC Dar) project. The study consists of the development of methodologies for monitoring spatial changes through Remote Sensing and Geographic Information System (GIS) techniques, which are tailored to the needs and resources of Dar City Council’s planning services.

The main objectives of the developed methodology is to monitor changes in Dar’s peri-urban settlements and increase the knowledge of peri-urban dynamics, in order to explore the impacts of urbanization on natural resources; the rapid changes in LC and Land Use (LU) patterns can increase CC effects on livelihoods of those living in Dar, in terms of sensitivity and exposure.

The developed methodology aims to reduce vulnerability to CC increasing adaptive capacity, especially that of Dar’s municipalities, which need to monitor LCC at a very low cost, and adjust their plans to those changes in a flexible way. This should make the planning process more effective in reducing vulnerability, providing a flexible framework of services and meeting the needs of inhabitants.

Remote sensing images are useful for mapping and analysing LCC. In this study, Landsat imagery was chosen for the characteristics of its resolutions, which allow for LC mapping in a semiautomatic process and of several years. Another advantage of this satellite is the availability of Landsat archive, where images are provided for free by the United States Geological Survey (USGS), which makes the developed methodology very affordable.

A workflow was developed in order to generate the LC maps of the Dar es Salaam region and analyse spatial variations during the recent decades. The workflow steps are: preprocessing and processing of Landsat data, correcting for atmospheric effects with a Dark Object Subtraction (DOS) model, masking clouds and their shadows in a semiautomatic way, and performing a semi-automatic LC classification.

In this Activity, a similar methodology for LC mapping was developed using SPOT images (provided for free by the European Space Agency), which is described in the working paper “Development of a Methodology for Land Cover Classification in Dar es Salaam using SPOT Imagery”.

Moreover, a methodology for the analysis of spatial patterns was developed, using Landscape Metrics Indices (LMI). It is described in the working paper “Development of a Methodology for Assessing Land Cover Fragmentation”.

Executive Summary

This study is part of the 2.1 Activity of the “Adapting to Climate Change in Coastal Dar es Salaam” (ACC Dar) project, which has the following objectives: enhancing the capacity of Dar's municipalities in understanding CC issues specific to coastal areas; assessing the CC impacts on the livelihood of those dwellers, partially or totally depending on natural resources, increasing knowledge of autonomous adaptive capacity; and developing methodologies for integrating adaptation activities into

strategies and plans for Urban Development and Environment Management (UDEM) in coastal unplanned and underserviced settlements.

Dar es Salaam is located in the east of Tanzania, on the Indian Ocean coast. Dar's population is rapidly growing, and the built-up area of the city continuously expanded during the last 20 years, especially at the fringe, because of the continuous growth of informal peri-urban settlements.

Vulnerability to CC, as stated by IPCC (2001), is a function of: sensitivity; exposure to climatic hazards; and adaptive capacity.

LCC effects can increase vulnerability to CC on living environment for people who rely on natural resources for their livelihood (Paavola, 2008). Often, people's actions of adaptation to local environmental issues can have in turn severe consequences on ecosystem (Metzger, et al., 2006).

In East Africa, LCC derives from the interactions of various agents, where the driving forces are both anthropogenic (urbanization, migration, land tenure, etc.) and environmental (climate, rainfall variability, soil and groundwater degradation, etc.) (Olson, et al., 2004).

In Dar, one of the causes of the rapid unplanned settlements growth is the type of regulatory framework, with administrative procedures taking too long to make land available to the seekers (Kironde, 2006), thus causing rapid LCC.

The relationship between the response to CC, and local LU changes is very difficult to understand (Lioubimtseva, et al., 2005), because CC is affected by many variables, related to natural resources and socio-political situation (Lioubimtseva & Henebry, 2009).

Remote sensing and GIS are useful instruments for mapping and analysing LCC (Africover, 2002), in order to understand and monitor spatial changes.

This study consists of the development of a methodology for monitoring spatial changes through Remote Sensing and GIS techniques, which are tailored to the needs and resources of the Dar city Council's planning services.

This study, aims to reduce vulnerability by increasing adaptive capacity, particularly of the Dar's municipalities, which will be able to monitor LCC over the years, in a very affordable fashion; the developed methodology could be integrated, with little effort, in strategic and planning activities, for monitoring very rapid LCC.

The Dar's municipalities, monitoring LCC, could provide a flexible framework for LU planning, and moreover integrate participatory processes in it (Halla, 2005; Sales Jr., 2009).

The developed methodology used Landsat satellite images, because of their spatial and spectral resolutions, multitemporal images availability, and particularly the free cost of data; in fact Landsat images are provided for free by the USGS, through an online archive.

A workflow has been designed for the elaboration of Landsat data, which includes the following steps:

a) Image selection;
b) Preprocessing (atmospheric effects correction and image preparation to classification);
c) Processing (image classification).

The correction of atmospheric effects was performed with a Dark Object Subtraction (DOS) model; that model does not require any in field measurement, therefore it is very affordable.

Because of the high cloud cover, a mosaic step was included in the workflow for masking clouds and their shadows using a semiautomatic process.

The classification of remote sensing images is a method of features identification in the scene. The method labels the pixels in the image through a classification algorithm, which is based on the pixels spectral characteristics, allowing for the thematic map creation (Richards & Jia, 2006).

The Maximum Likelihood (ML) algorithm, one of the most used supervised classifier, was chosen for image classification; that algorithm is based upon the Gaussian threshold, stored in each class signature, for assigning a class to every image pixel (Huang, et al., 2009).

A Knowledge-Base classification was chosen for enhancing the classification process, particularly for vegetation, with the inclusion of the layers: ML classification; NDVI; EVI; and Dar boundary shapefile.

The use of Landsat images, which are provided for free, reduced costs to only purchase the commercial image processing and GIS software. The LC classification, semi-automatically performed, reduced time and cost of LC maps production, especially for large areas like Dar city.

LC classifications of Dar es Salaam were performed for the years: 2002, 2004, 2007, 2009, and 2011.

The LC classification results show growing trend of new urbanization over the analysed years; urban sprawl is affecting the city, particularly along the main roads, and also new urban areas are growing quite far from the city centre.

The LCC between 2002 and 2011 shows an increment of the “Continuous Urban” class from 4.98% of Dar area, to 8.76%; also the “Discontinuous Urban” class augmented from 4.80% in 2002 to 14.01% in 2011.

The LC maps developed with this methodology can be upgraded and integrated in GIS of Dar’s municipalities. Through spatial analysis functions, the planning services could assess the environmental priorities and plan the needed infrastructures for inhabitants.

The major problem encountered in the classification process is the difficult identification in the images, of pixels representing the LC classes, because of the rapid LCC over the years. Another problem is the lack of reference images for the past years, which are images with higher spatial resolution that allow for the identification of LC classes in Landsat images.

The LC classifications are based upon image mosaics. The cloud cover issue has entailed the selection of images acquired in different months of the year; therefore, the mosaic pixels do not necessarily belong to the same month, depending on the location. That different seasonality causes fluctuations in reflectance values, especially for vegetation surfaces, because of the changes of phenological state that occur during the year.

Dar is a social and economic attractive region for Tanzania, and therefore the developed methodology for LC monitoring could also be adopted and upgraded by other Tanzanian administrations; that integration could facilitate the creation of urban scenarios and coordination of all levels of government planning, towards CC adaptation (Levina, et al., 2007). Moreover, planning process could be more effective in reducing vulnerability, if short-term decisions are adapted to the climate variability and extreme events, and if long-term decisions consider uncertainty (Hallegatte, 2009).

Future work of this study is the assessment of classification accuracy and the validation of LC maps. The LC validation phase will be also fundamental for the assessment of the present methodology, and to refine its processing steps.

The spatial resolution of Landsat (i.e. 30m) is good mainly at regional scale, because a single pixel of the image could include a mixture of cover types, causing the mixed pixel issue (Richards & Jia, 2006). Another methodology was developed by this ACC Dar project activity, for assessing the urban sprawl with a higher detail level, using SPOT images, which are provided without cost for research projects by the European Space Agency (ESA); that methodology is described in the working paper “Development of a Methodology for Land Cover Classification in Dar es Salaam using SPOT Imagery”.

Although the Landsat 5 satellite is no longer operational and the Landsat 7 is affected by SLC-off gaps, the developed methodology can rely on a new satellite (the Landsat Data Continuity Mission), planned by the collaboration between the NASA and USGS, which will be launched in 2013 and should provide images for future LC monitoring.

1. Introduction, Scope, and Motivation

The Dar es Salaam Municipality is located in the east of Tanzania, on the coast of the Indian Ocean, between longitudes 39°0’ - 33°33’ East and latitudes 6°36’ - 7°0’ South.

Dar es Salaam has an area of 1 800km^2 and is characterized by the coastal plain in the central part of the city, the middle plateau to the north, the Pugu Hills to the west, and eight offshore islands (United_Republic_of_Tanzania, 2004); the shaded relief realized with the SRTM DEM (data available from the USGS) is shown in Figure 1.

The city was established in 1862 as a port and trading centre; in 1891 became the national capital, in 1949 acquired municipal status, and in 1961 achieved city status. In the 1970s Dar es Salaam lost its official status of capital city, which now is Dodoma, but Dar remains the centre for the permanent central government bureaucracy (UN-HABITAT, 2009). Dar has three Districts: Ilala, Temeke and Kinondoni.

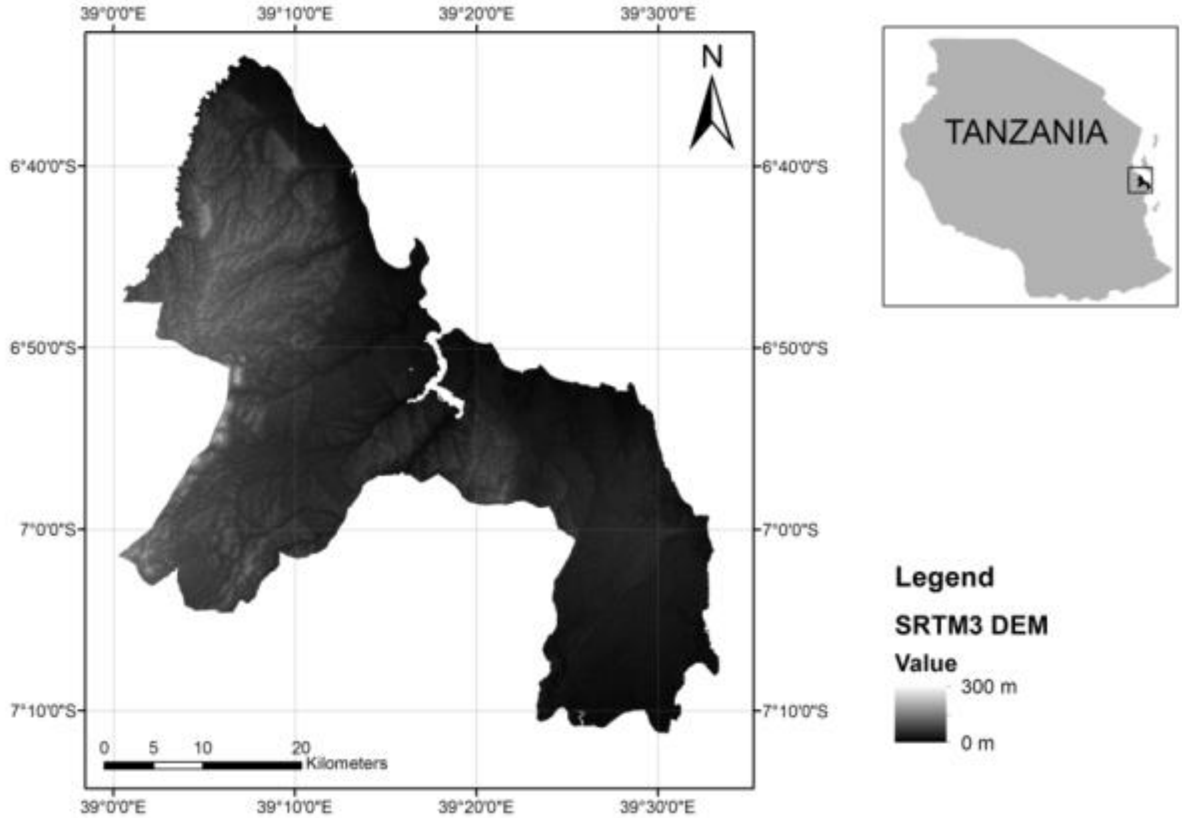

Figure 1: Dar es Salaam, shaded relief

1.1 Background

Dar es Salaam's population is growing very rapidly. In particular, during the last 20 years the built-up area of the city expanded, especially at the fringe, because of the continuous growth of informal peri-urban settlements.

As stated by Briggs and Mwamfupe (2000), the most significant developments of the city have taken place since 1967. Especially during the 1980s, urban expansion along the arterial roads has been faster because of people trying to reduce travel time to the city centre, which at that time was very long, because of public transportation issues. In 1990s there was better public transportation system and private transportation increased, thus the city also expanded away from the arterial roads, producing an irregular spatial pattern. Dar's population in 1988 was almost 1.36 million and in 2002 had become closer to 2.5 million (Kironde, 2006).

Kombe (2005) describes that migrants from upcountry, attracted by job opportunities in Dar, acquire land and build houses in poverty, bypassing formal urban land management, and create informal social networks, thus facilitating intra-urban and rural-urban migration. Those settlements lack of services like electricity, transportation networks (Olvera, et al., 2003), potable water (Kyessi, 2005), causing public health threats. Moreover the soil sealing (impermeabilization of soil surface) determined by urbanization can increase flooding risk (Swan, 2010).

According to Kironde (2006) one of the causes of that rapid growth of unplanned settlements is the type of regulatory framework, where administrative procedures take too long to make land available to seekers.

In peri-urban settlements the lack of planning is compensated by social networks, supported by cultural norms (Kombe, 2005). Nevertheless, unplanned LCC can generate environmental degradation, thus increase vulnerability to CC effects, which are particularly heavy for those inhabitants whose livelihood depends on natural resources (Paavola, 2008).

1.2 Goals and scope

The main goals of this study are in the context of the ACC Dar project objectives, which are: to enhance the capacity of Dar's municipalities in understanding CC issues, specific to coastal areas, and in assessing their impacts on the livelihood of those urban dwellers partially or totally depending on natural resources; to improve the knowledge on autonomous adaptive capacity, and developing methodologies for integrating adaptation activities into strategies and plans for Urban Development and Environment Management (UDEM) in coastal unplanned and underserviced settlements.

This study aims to improve the City Council’s planning services in understanding LC and LU patterns, developing methodologies for monitoring changes in peri-urban settlements.

The main goal of this study is to develop a methodology for LC monitoring, based on remote sensing and GIS techniques, which must be rapid and easy to update, in order to be consistent with the pace of growth of the city. The developed methodology must be suitable to needs and resources of Dar’s municipalities, and consequently must have very low cost.

1.3 Motivation

Continuous change in LC poses a challenge for urban planners and decision-makers, because also the lack of financial (Kironde, 2006). The main motivation of this study is to improve capacity building of Dar’s municipalities, providing a suitable methodology for LC monitoring, which should be tailored to equipment already available among municipalities, or that could be upgraded with little effort.

Dar’s municipalities will be able to adjust their plans according to LCC, and therefore to make planning processes more effective in providing services and reducing vulnerability to CC.

The final beneficiaries of this applied methodology will be Dar’s inhabitants, especially who live in unplanned areas and whose livelihood depends on natural resources. The lack of planned services potentially increases CC impacts; for instance, the CC related issue of flooding, could be aggravated in peri-urban settlements where pit-latrines are often used, because groundwater can be polluted when flooding occurs, causing epidemics (Paavola, 2003).

This methodology should improve the knowledge about LC, and help administrations to set a flexible planning framework that could increase adaptive capacity of Dar’s inhabitants and to provide needed services, while facilitating participatory processes (Halla, 2005; Sales Jr., 2009).

2. Approach and Methods

This study aims to develop a methodology for semi-automatic LC classification, using remote sensing imagery. The methodology allows for LCC monitoring over the years, providing useful information for assessing urban sprawl.

This study is part of the Activity 2.1 of the ACC Dar project. This activity has developed two similar methodologies for semi-automatic LC classification, using two different satellite sources (Landsat and SPOT), with different resolutions. Moreover, another methodology was developed for assessing LC classification accuracy. A methodology has been developed for assessing LC fragmentation and measuring landscape patterns. Those methodologies are described in as many working papers.

This paper presents the methodology for semi-automatic classification performed using Landsat images, for the purpose of monitoring Dar changes over the years (Figure 2).

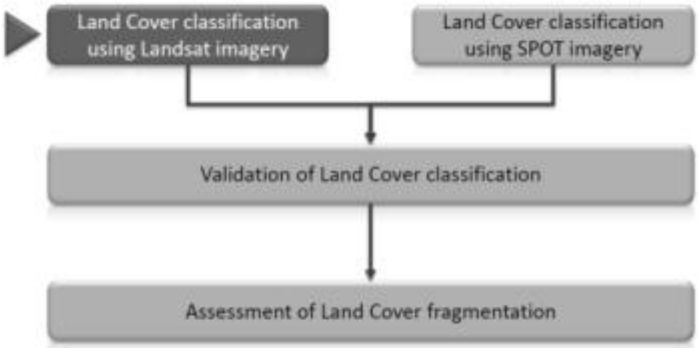

Figure 2: Developed methodologies of Activity 2.1; LC classification using Landsat imagery

2.1 Overall Approach

Vulnerability to CC, as stated by IPCC (2001), is a function of:

- the sensitivity, that is "*the degree to which a system will respond to a given change in climate, including beneficial and harmful effects*";
- the "*exposure of the system to climatic hazards*";
- the adaptive capacity, that is "*the degree to which adjustments in practices, processes, or structures can moderate or offset the potential for damage or take advantage of opportunities created by a given change in climate*".

This study aims to reduce vulnerability by increasing adaptive capacity, particularly of Dar's municipalities, while monitoring LCC in a very affordable fashion.

Remote sensing techniques are useful for LC classifications (Africover, 2002), allowing for the acquisition of multispectral images and their analyses using image processing software for semi-automatic classifications.

The classification of remote sensing images is a method of features identification in the study area; "*computer interpretation of remote sensing image data is referred to as quantitative analysis because of its ability to identify pixels based upon their numerical properties and owing to its ability for counting pixels for area estimates*" (Richards & Jia, 2006, p. 74).

Classification process, through a classification algorithm, labels the pixels in the image, basing on their spectral characteristics, allowing for the creation of thematic maps (Richards & Jia, 2006).

Landsat is a family of satellites launched by NASA; in this study Landsat 5 and 7 were considered, which are two satellites with similar sensors characteristics: every image has 7 multispectral bands with spatial resolution of 30m (Landsat 7 has an additional panchromatic band with 15m resolution), and image size at ground is 170km north-south by 183km east-west (NASA, 2011).

Landsat satellites were chosen because of their spatial and spectral resolution, multitemporal image availability, and particularly the free data acquisition cost, thus obtaining a very affordable methodology, suitable to the needs of Dar municipalities.

Landsat images are provided for free by the USGS (http://landsat.usgs.gov , accessed 27/01/2012), which made available over internet the archive of images acquired since 1984 over Dar.

In the developed methodology were used the reflected solar energy bands, excluding the thermal band, because in this range it is possible to identify materials by their spectral response, using supervised classification.

In order to maximize the spectral contrast between vegetated surfaces and impervious surfaces it should be preferable to classify images acquired during summer.

Landsat 7 images, which were acquired after 2003, are affected by a technical problem causing SLC-off gaps along the image, with stripes of null data; moreover USGS has stopped acquiring Landsat 5 from 2011/11/18 due to electronic problems (http://www.usgs.gov/newsroom/article.asp?ID=3040 , accessed 27/01/2012). More information about Landsat is provided in Appendix 1.

Remote sensing images acquired over Dar es Salaam are also affected by cloud cover that is present all over the year; therefore, cloud-masking and mosaic processes are needed in order to obtain cloud-free images of the whole area.

The need to mosaic images requires a high amount of data for each year, therefore the cost of commercial satellites images could be unaffordable; that is another reason to prefer Landsat data, because are free.

A workflow has been designed for the preprocessing of Landsat data, correcting for atmospheric effects with a Dark Object Subtraction (DOS) model, masking clouds and their shadows in a semiautomatic way, and processing data.

This Activity also developed a methodology for assessing Landscape patterns and urban fragmentation (described in another working paper), which should enhance the knowledge of local administrations about environmental issues related to soil sealing and urban sprawl.

2.2 Data Collection and Analysis Methodology

Landsat imagery is distributed by USGS at no charge, indeed "*there are no restrictions on Landsat data downloaded from USGS EROS, and it can be used or redistributed as desired. However, a statement of the data source when citing, copying, or reprinting USGS Landsat data or images is requested*" (from http://landsat.usgs.gov/Landsat_Search_and_Download.php , accessed 27/01/2012).

Dar es Salaam data can be downloaded from:

- ftp://ftp.glcf.umd.edu/glcf/Landsat/WRS2/p166/r065/ (accessed 27/01/2012), where limited datasets are available, referred to certain years, and no registration is needed;
- http://earthexplorer.usgs.gov/ (accessed 27/01/2012), where Landsat imagery is available from the archive, free for registered users.

Downloaded imagery is composed of a .tif file for each Landsat band, and an MTL.txt file which contains metadata information.

Images are already georeferenced in WGS 84 datum and UTM projection in a north up (map) orientation, and are of Level 1 of the Product Generation System (more information in Appendix 1 - Landsat Satellites, in Level 1 Product Generation System section).

It is possible to order for free the processing of images that are present in the USGS on-line archive, but not available for download; depending on the USGS queue for processing, images are generally processed in 1 to 3 days, and an e-mail confirm the process conclusion.

In this study the image classification process was based on the semi-automatic Maximum Likelihood (ML) algorithm, which allows for the identification of LC classes; the algorithm is based on training area collected over the image, which define the spectral signatures of classes.

3. Findings

The developed methodology requires the following steps:

a) Image selection;
b) Preprocessing:
   1. Georeferencing images in order to assign spatial coordinates to pixels;
   2. Creating masks of clouds and shadows, and applying those mask to the Landsat bands in order to exclude pixels belonging to clouds or shadows from LC classification;
   3. Converting the multispectral bands (1, 2, 3, 4, 5 and 7) from DN to reflectance, applying atmospheric correction;
   4. Mosaicking temporally different images, in order to obtain a cloud-free and gap-free image;
c) Processing:
   1. Classifying the image mosaic with Maximum Likelihood (ML) algorithm;
   2. Elaborating vegetation indices (NDVI and EVI), which are useful for classifying vegetation;
   3. Classifying the ML classification and the vegetation indices through a Knowledge-Base classification.

Figure 3 shows the developed methodology workflow.

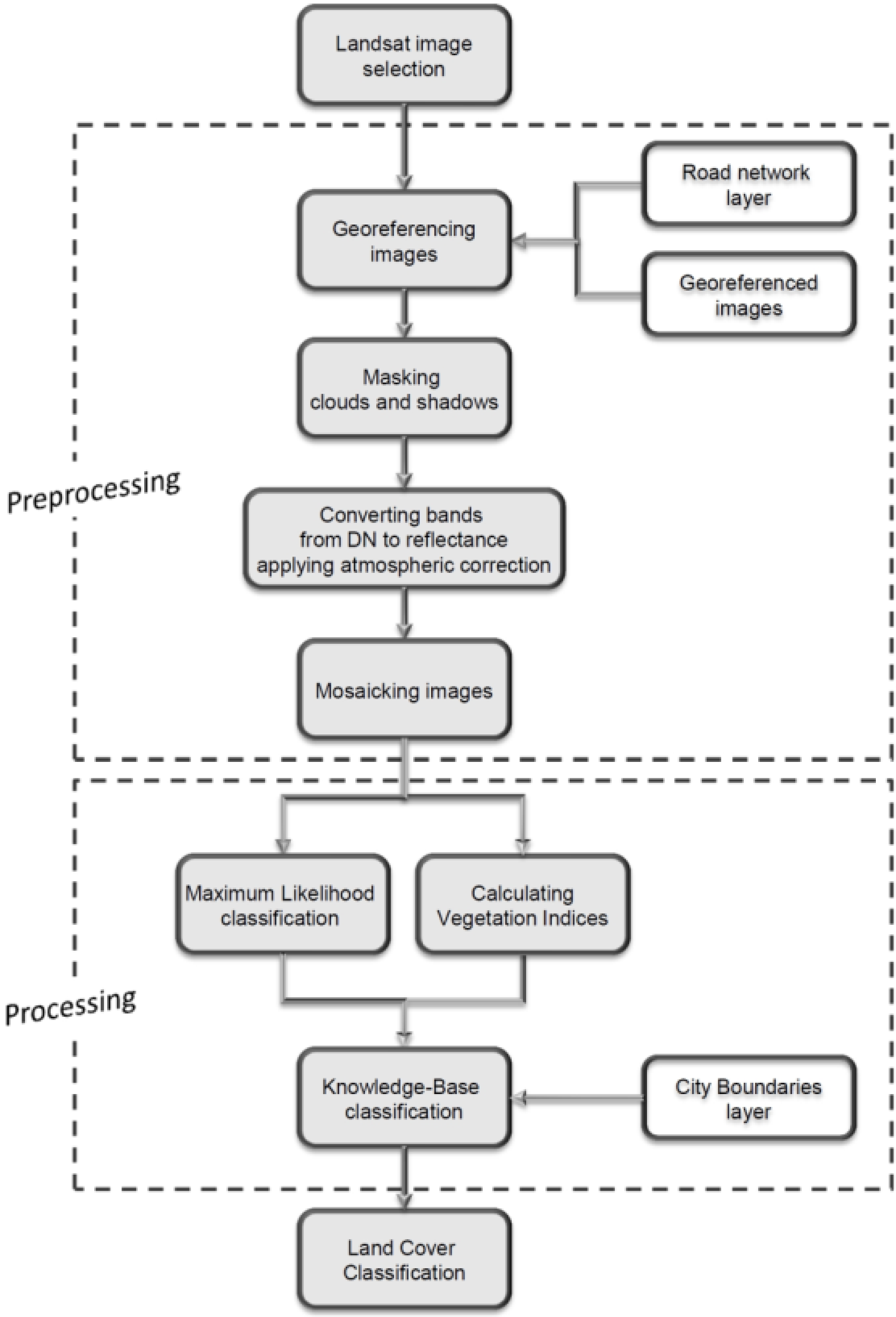

Figure 3: Methodology workflow

3.1 Image Selection

Worldwide Reference System "*is a global notation system for Landsat data. It enables a user to inquire about satellite imagery over any portion of the world by specifying a nominal scene center designated by PATH and ROW numbers*" (from http://landsat.gsfc.nasa.gov/about/wrs.html , accessed 27/01/2012).

Landsat 4, 5, and 7 images are referred to the WRS-2 (Worldwide Reference System – 2), which is an extension of WRS-1 used for Landsat 1-3; in that system, Dar es Salaam is located in:

- Path = 166;
- Row = 065.

Several Landsat images were downloaded from the USGS archive, and because of the high percentage of cloud cover, 23 images were selected to be processed in order to generate the mosaics.

The selected images are listed in Table 1, which shows the related percentage of useful area over Dar (area that is free of clouds and gaps); often the useful area is less than one half of the image.

Table 1: List of Landsat images processed and related percentage of useful area

Satellite	Date of acquisition [YYYY-MM-DD]	Useful area [%]
Landsat 7	1999-11-19	44.42
Landsat 7	2000-05-29	29.8
Landsat 7	2000-06-30	60.78
Landsat 7	2000-10-20	44.91
Landsat 7	2001-02-09	38.03
Landsat 7	2001-04-30	40.15
Landsat 7	2001-11-24	48.79
Landsat 7	2002-03-16	29.84
Landsat 7	2002-04-17	23.96
Landsat 7	2002-07-22	25.21
Landsat 7	2002-08-23	29.59
Landsat 7	2002-11-11	29.24
Landsat 7	2002-12-13	33.47
Landsat 7	2003-11-14	64.12
Landsat 7	2004-07-11	58.28
Landsat 7	2006-02-23	51.62
Landsat 7	2007-06-18	66.04
Landsat 7	2008-02-29	65.55
Landsat 5	2009-07-01	97.4
Landsat 5	2010-05-17	68.77
Landsat 7	2011-01-20	55.74
Landsat 7	2011-02-05	56.8
Landsat 5	2011-07-07	80.8

3.2 Preprocessing

Before processing images using a LC classification algorithm, it is necessary to perform various preprocessing operations, in order to eliminate or reduce any possible source of error, like clouds and atmospheric effects, which modify the spectral values of pixels. In some case, it is needed to georeference images.

3.2.1 Georeferencing Images

Most of Landsat images are already georeferenced. Depending on the processing time, images from the archive could be processed with the Standard Terrain Correction (Level 1T, the most accurate), the Systematic Terrain Correction (Level 1GT), or the Systematic Correction (Level 1G) with lower precision.

Therefore, georeferencing is not always required because Landsat images are already georeferenced by USGS; however, images with high cloud cover could have low geometric accuracy, and therefore it is necessary to georeference them.

The typology of correction is described in the metadata (MTL.txt file) of each image; in case of L1GT or L1G correction, the error in pixel position (geometric accuracy) could also be of 250m (more information in the Appendix 1 - Landsat Satellites, in the Geometric Accuracy section).

Georeferencing is performed using one or more reference files, which are for example high resolution images, and identifying several Ground Control Points (GCPs), which are the inputs for the georeferencing transformation, on both the images.

In this study, georeferencing is performed using the road net shape file and other images as reference. The identification process of the GCPs can be difficult because of the cloud cover, which hides many useful points of the images. Moreover, the oldest images of Dar were acquired when several roads were still unpaved or not built, therefore their identification is difficult.

3.2.2 Cloud Cover and Clouds' Shadow Mask

Cloud cover is calculated for Landsat 7 by USGS with an Automated Cloud Cover Assessment (ACCA) algorithm; as described in NASA (2011), ACCA recognizes clouds by passing through the scene data twice, using twenty-six different filters, and it is based on the premise that clouds are colder than Earth surface features.

"*The algorithm works well in most cases, but occasionally, temperature inversions occur and invalid cloud cover assessments may result. While there are also intermittent problems with the ACCA detection of popcorn clouds and haze, the ACCA algorithm maintains a higher degree of overall accuracy than the previous methods employed for past Landsat processing systems*" (from http://landsat.usgs.gov/percentage_of_cloud_cover_calculated.php , accessed 27/01/2012).

For Landsat Data Continuity Mission (LDCM) a band for detecting cirrus cloud is planned (more information in the Appendix 1 – Landsat Satellites, in Landsat Data Continuity Mission).

Each USGS Landsat image is processed with the ACCA algorithm, in order to calculate the cloud cover percentage.

Other studies tested several algorithms on Landsat data, for example a modified version of a cloud masking algorithm originally developed for MODIS (Moderate Resolution Imaging Spectroradiometer) images, without using the thermal band (Oreopoulos, et al., 2011).

Martinuzzi et al. (2007) proposed a method based on band 1 (blue) and 6 (thermal):

- "*Brightness values for clouds were identified by visual analysis. For band 1, DN values between 120 and 255 include clouds as well as urban, barren, quarries, rocks, and sand. For band 6, DN values of 102 to 128 include both clouds and densely forested areas.*
- *A mask was created for band 1 and for band 6 by using the previous values. A three-pixel buffer was added to incorporate mixed pixels from cloud borders that could not be incorporated by brightness value analysis.*
- *The intersection of both masks results in a final clouds mask […]*".

Before processing images it is useful to create a cloud mask in order to eliminate this source of error for Landsat classifications. Landsat images with cloud cover < 10% are often acquired over Dar es

Salaam in January, February and July.

It is difficult to detect clouds automatically, because of their spectral characteristics, as they can be misclassified with ice, snow, rocks (NASA, 2011).

The Martinuzzi et al. (2007) method was applied in this study, since it is an efficient and relatively simple algorithm to create cloud mask for Landsat images.

A prior visual analysis is needed in order to identify clouds in the scene:

- For band 1 (blue) clouds have Digital Number (DN) ranging from a minimum value to 255 (saturation); the minimum value has to be detected in pixels where clouds are thinner, and a mask has to be created where:

DNmin ≤ “DN of pixel in band 1” ≤ 255 (1)

- For band 6 (thermal), clouds are generally colder than other surfaces; it is necessary to identify the maximum DN value where clouds have higher temperature, and a mask has to be created where:

1 ≤ “DN of pixel in band 1” ≤ DNmax (2)

- The intersection between the two masks created, and an additional buffer of 3 pixels, provides the final cloud mask, where pixel values of 0 refers to clouds and pixel values of 1 refers to non-clouds.
- The resulting cloud mask can be used to remove clouds from the image (bands * mask), allowing to fill cloud gaps where mask values are 0, and replacing pixels with those of another Landsat image (mosaic process, described in the following paragraph Image Mosaic).

For Landsat 7, masks of SLC-off gaps can be created easily when:

DN of pixel = 0 (3)

If clouds are present in the scene, their shadows can alter the radiance at surface; it’s preferable to remove those shadows, with masks.

One way to mask shadows is to perform a ML classification, drawing training areas over several types of shadow surfaces, in order to consider their spectral variability. Using that shadow classification it possible to create a mask, and assign the value of 0 to pixels where there are clouds’ shadows, and the value of 1 elsewhere.

3.2.3 Reflectance Conversion and Atmospheric Correction

Landsat 7 system records reflected solar energy for bands 1-5 and 7 and emitted energy for band 6.

The Spectral Radiance at the sensor's aperture (L_λ) is measured in [watts/(meter squared * ster * μm)] and is given by (NASA, 2011):

$$L_\lambda = G_{rescale} * Q_{CAL} + B_{rescale} \quad (4)$$

where:

- $G_{rescale}$ is the rescaled gain (the data product "gain" contained in the Level 1 product header or ancillary data record) in watts/(meter squared * ster * μm)/DN

$$G_{rescale} = (LMAX_\lambda - LMIN_\lambda)/(Q_{CALMAX} - Q_{CALMIN}) \quad (5)$$

- $B_{rescale}$ is the rescaled bias (the data product "offset" contained in the Level 1 product header or ancillary data record) in watts/(meter squared * ster * μm)

$$B_{rescale} = LMIN_\lambda - (LMAX_\lambda - LMIN_\lambda)/(Q_{CALMAX} - Q_{CALMIN}) * Q_{CALMIN} \quad (6)$$

Therefore, eq. 4 is also expressed as:

$$L_\lambda = ((LMAX_\lambda - LMIN_\lambda)/(Q_{CALMAX}-Q_{CALMIN})) * (Q_{CAL}-Q_{CALMIN}) + LMIN_\lambda \quad (7)$$

where:

- Q_{CAL} = the quantized calibrated pixel value in DN
- $LMIN_\lambda$ = the spectral radiance that is scaled to Q_{CALMIN} in watts/(meter squared * ster * µm)
- $LMAX_\lambda$ = the spectral radiance that is scaled to Q_{CALMAX} in watts/(meter squared * ster * µm)
- Q_{CALMIN} = the minimum quantized calibrated pixel value (corresponding to $LMIN_\lambda$) in DN, and for LPGS products is equal to 1
- Q_{CALMAX} = the maximum quantized calibrated pixel value (corresponding to $LMAX_\lambda$) in DN = 255

The Table 2 shows $LMIN_\lambda$ and $LMAX_\lambda$ values for every Landsat band.

Table 2: ETM+ Spectral Radiance Range [watts/(meter squared * ster * µm)] (NASA, 2011)

Band Number	Processed Before July 1, 2000				Processed After July 1, 2000			
	Low Gain		High Gain		Low Gain		High Gain	
	LMIN	LMAX	LMIN	LMAX	LMIN	LMAX	LMIN	LMAX
1	-6.2	297.5	-6.2	194.3	-6.2	293.7	-6.2	191.6
2	-6.0	303.4	-6.0	202.4	-6.4	300.9	-6.4	196.5
3	-4.5	235.5	-4.5	158.6	-5.0	234.4	-5.0	152.9
4	-4.5	235.0	-4.5	157.5	-5.1	241.1	-5.1	157.4
5	-1.0	47.70	-1.0	31.76	-1.0	47.57	-1.0	31.06
6	0.0	17.04	3.2	12.65	0.0	17.04	3.2	12.65
7	-0.35	16.60	-0.35	10.932	-0.35	16.54	-0.35	10.80
8	-5.0	244.00	-5.0	158.40	-4.7	243.1	-4.7	158.3

"*For relatively clear Landsat scenes, a reduction in between-scene variability can be achieved through a normalization for solar irradiance by converting spectral radiance, as calculated above, to planetary reflectance or albedo. This combined surface and atmospheric reflectance of the Earth is computed with the following formula*" (NASA, 2011, p. 119):

$$\rho_p = (\pi * L_\lambda * d^2)/(ESUN_\lambda * \cos\theta_s) \quad (8)$$

where:

- ρ_p = Unitless planetary reflectance
- L_λ = Spectral radiance at the sensor's aperture
- d = Earth-Sun distance in astronomical units from an Excel file
- $ESUN_\lambda$ = Mean solar exo-atmospheric irradiances
- θ_s = Solar zenith angle in degrees, which is the reciprocal of the sun elevation angle

The mean solar exo-atmospheric irradiances values for Landsat bands are reported in Table 3.

Table 3: ETM+ Solar Spectral Irradiances (NASA, 2011)
generated using the solar spectrum of Thuillier et al., (2004)

Band	Solar Spectral Irradiances [watts/(meter squared * µm)]
1	1997
2	1812
3	1533
4	1039
5	230.8
7	84.90
8	1362

The values of Earth-Sun distance are listed in Table 4, in astronomical units for every day of the year, as reported in a free spreadsheet file provided by NASA, available from the url: http://landsathandbook.gsfc.nasa.gov/excel_docs/d.xls (accessed 27/01/2012).

Table 4: Earth-Sun distance (d) in astronomical units for Day of the Year (DOY)

DOY	d	DOY	d	DOY	d	DOY	d	DOY	d	DOY	d	DOY
1	0.98331	**61**	0.99108	**121**	1.00756	**181**	1.01665	**241**	1.00992	**301**	0.99359	**361**
2	0.9833	**62**	0.99133	**122**	1.00781	**182**	1.01667	**242**	1.00969	**302**	0.99332	**362**
3	0.9833	**63**	0.99158	**123**	1.00806	**183**	1.01668	**243**	1.00946	**303**	0.99306	**363**
4	0.9833	**64**	0.99183	**124**	1.00831	**184**	1.0167	**244**	1.00922	**304**	0.99279	**364**
5	0.9833	**65**	0.99208	**125**	1.00856	**185**	1.0167	**245**	1.00898	**305**	0.99253	**365**
6	0.98332	**66**	0.99234	**126**	1.0088	**186**	1.0167	**246**	1.00874	**306**	0.99228	**366**
7	0.98333	**67**	0.9926	**127**	1.00904	**187**	1.0167	**247**	1.0085	**307**	0.99202	
8	0.98335	**68**	0.99286	**128**	1.00928	**188**	1.01669	**248**	1.00825	**308**	0.99177	
9	0.98338	**69**	0.99312	**129**	1.00952	**189**	1.01668	**249**	1.008	**309**	0.99152	
10	0.98341	**70**	0.99339	**130**	1.00975	**190**	1.01666	**250**	1.00775	**310**	0.99127	
11	0.98345	**71**	0.99365	**131**	1.00998	**191**	1.01664	**251**	1.0075	**311**	0.99102	
12	0.98349	**72**	0.99392	**132**	1.0102	**192**	1.01661	**252**	1.00724	**312**	0.99078	
13	0.98354	**73**	0.99419	**133**	1.01043	**193**	1.01658	**253**	1.00698	**313**	0.99054	
14	0.98359	**74**	0.99446	**134**	1.01065	**194**	1.01655	**254**	1.00672	**314**	0.9903	
15	0.98365	**75**	0.99474	**135**	1.01087	**195**	1.0165	**255**	1.00646	**315**	0.99007	
16	0.98371	**76**	0.99501	**136**	1.01108	**196**	1.01646	**256**	1.0062	**316**	0.98983	
17	0.98378	**77**	0.99529	**137**	1.01129	**197**	1.01641	**257**	1.00593	**317**	0.98961	
18	0.98385	**78**	0.99556	**138**	1.0115	**198**	1.01635	**258**	1.00566	**318**	0.98938	
19	0.98393	**79**	0.99584	**139**	1.0117	**199**	1.01629	**259**	1.00539	**319**	0.98916	
20	0.98401	**80**	0.99612	**140**	1.01191	**200**	1.01623	**260**	1.00512	**320**	0.98894	
21	0.9841	**81**	0.9964	**141**	1.0121	**201**	1.01616	**261**	1.00485	**321**	0.98872	
22	0.98419	**82**	0.99669	**142**	1.0123	**202**	1.01609	**262**	1.00457	**322**	0.98851	
23	0.98428	**83**	0.99697	**143**	1.01249	**203**	1.01601	**263**	1.0043	**323**	0.9883	
24	0.98439	**84**	0.99725	**144**	1.01267	**204**	1.01592	**264**	1.00402	**324**	0.98809	
25	0.98449	**85**	0.99754	**145**	1.01286	**205**	1.01584	**265**	1.00374	**325**	0.98789	
26	0.9846	**86**	0.99782	**146**	1.01304	**206**	1.01575	**266**	1.00346	**326**	0.98769	
27	0.98472	**87**	0.99811	**147**	1.01321	**207**	1.01565	**267**	1.00318	**327**	0.9875	
28	0.98484	**88**	0.9984	**148**	1.01338	**208**	1.01555	**268**	1.0029	**328**	0.98731	
29	0.98496	**89**	0.99868	**149**	1.01355	**209**	1.01544	**269**	1.00262	**329**	0.98712	
30	0.98509	**90**	0.99897	**150**	1.01371	**210**	1.01533	**270**	1.00234	**330**	0.98694	
31	0.98523	**91**	0.99926	**151**	1.01387	**211**	1.01522	**271**	1.00205	**331**	0.98676	
32	0.98536	**92**	0.99954	**152**	1.01403	**212**	1.0151	**272**	1.00177	**332**	0.98658	
33	0.98551	**93**	0.99983	**153**	1.01418	**213**	1.01497	**273**	1.00148	**333**	0.98641	
34	0.98565	**94**	1.00012	**154**	1.01433	**214**	1.01485	**274**	1.00119	**334**	0.98624	
35	0.9858	**95**	1.00041	**155**	1.01447	**215**	1.01471	**275**	1.00091	**335**	0.98608	
36	0.98596	**96**	1.00069	**156**	1.01461	**216**	1.01458	**276**	1.00062	**336**	0.98592	
37	0.98612	**97**	1.00098	**157**	1.01475	**217**	1.01444	**277**	1.00033	**337**	0.98577	
38	0.98628	**98**	1.00127	**158**	1.01488	**218**	1.01429	**278**	1.00005	**338**	0.98562	
39	0.98645	**99**	1.00155	**159**	1.015	**219**	1.01414	**279**	0.99976	**339**	0.98547	
40	0.98662	**100**	1.00184	**160**	1.01513	**220**	1.01399	**280**	0.99947	**340**	0.98533	
41	0.9868	**101**	1.00212	**161**	1.01524	**221**	1.01383	**281**	0.99918	**341**	0.98519	
42	0.98698	**102**	1.0024	**162**	1.01536	**222**	1.01367	**282**	0.9989	**342**	0.98506	
43	0.98717	**103**	1.00269	**163**	1.01547	**223**	1.01351	**283**	0.99861	**343**	0.98493	
44	0.98735	**104**	1.00297	**164**	1.01557	**224**	1.01334	**284**	0.99832	**344**	0.98481	
45	0.98755	**105**	1.00325	**165**	1.01567	**225**	1.01317	**285**	0.99804	**345**	0.98469	
46	0.98774	**106**	1.00353	**166**	1.01577	**226**	1.01299	**286**	0.99775	**346**	0.98457	
47	0.98794	**107**	1.00381	**167**	1.01586	**227**	1.01281	**287**	0.99747	**347**	0.98446	
48	0.98814	**108**	1.00409	**168**	1.01595	**228**	1.01263	**288**	0.99718	**348**	0.98436	
49	0.98835	**109**	1.00437	**169**	1.01603	**229**	1.01244	**289**	0.9969	**349**	0.98426	
50	0.98856	**110**	1.00464	**170**	1.0161	**230**	1.01225	**290**	0.99662	**350**	0.98416	
51	0.98877	**111**	1.00492	**171**	1.01618	**231**	1.01205	**291**	0.99634	**351**	0.98407	
52	0.98899	**112**	1.00519	**172**	1.01625	**232**	1.01186	**292**	0.99605	**352**	0.98399	
53	0.98921	**113**	1.00546	**173**	1.01631	**233**	1.01165	**293**	0.99577	**353**	0.98391	
54	0.98944	**114**	1.00573	**174**	1.01637	**234**	1.01145	**294**	0.9955	**354**	0.98383	
55	0.98966	**115**	1.006	**175**	1.01642	**235**	1.01124	**295**	0.99522	**355**	0.98376	
56	0.98989	**116**	1.00626	**176**	1.01647	**236**	1.01103	**296**	0.99494	**356**	0.9837	
57	0.99012	**117**	1.00653	**177**	1.01652	**237**	1.01081	**297**	0.99467	**357**	0.98363	
58	0.99036	**118**	1.00679	**178**	1.01656	**238**	1.0106	**298**	0.9944	**358**	0.98358	
59	0.9906	**119**	1.00705	**179**	1.01659	**239**	1.01037	**299**	0.99412	**359**	0.98353	
60	0.99084	**120**	1.00731	**180**	1.01662	**240**	1.01015	**300**	0.99385	**360**	0.98348	

Atmospheric correction is required for Landsat images before classification and change detection (Song, et al., 2001). As described by Zhang et al. (2010), land surface reflectance (ρ) can be estimated by the following equation:

$$\rho = [\pi * (L_\lambda - L_p) * d^2] / (T_v * F_d) \tag{9}$$

where:

- L_λ is the at-satellite radiance
- L_p is the path radiance
- d is the Earth–Sun distance in astronomical units
- T_v is the atmospheric transmittance in the viewing direction
- F_d is the irradiance received at the surface

The irradiance received at the surface is expressed by:

$$F_d = E_b + E_{down} \tag{10}$$

where:

- E_{down} is the downwelling diffuse irradiance
- E_b is the beam irradiance

The beam irradiance is defined as:

$$E_b = ESUN_\lambda * \cos\theta_z * T_z \tag{11}$$

where:

- $ESUN_\lambda$ is the mean solar exo-atmospheric irradiances
- θ_s is the solar zenith angle
- T_z is the atmospheric transmittance in the illumination direction

As originally described by Moran et al. (1992), the reflectance equation to convert from at-satellite radiances to surface reflectance, by correcting for both solar and atmospheric effects, is:

$$\rho = [\pi * (L_\lambda - L_p) * d^2] / [T_v * ((ESUN_\lambda * \cos\theta_s * T_z) + E_{down})] \tag{12}$$

The Dark Object Subtraction (DOS) atmospheric correction is an image-based technique, thus no in-situ measurements are needed during image acquisition.

Chavez (1996) explains that "*the basic assumption is that within the image some pixels are in complete shadow and their radiances received at the satellite are due to atmospheric scattering (path radiance). This assumption is combined with the fact that very few targets on the Earth's surface are absolute black, so an assumed one-percent minimum reflectance is better than zero percent*".

Assuming the existence of dark objects (surface reflectance ≈ 0), the minimum DN value is subtracted from all the pixels, removing atmospheric effects on the whole image.

The path radiance, as described by Sobrino et al. (2004), is calculated as:

$$L_p = L_{min} - L_{DO1\%} \tag{13}$$

where:

- L_{min} = "*radiance that corresponds to a digital count value for which the sum of all the pixels with digital counts lower or equal to this value is equal to the 0,01% of all the pixels from the image considered*" (Sobrino, et al., 2004, p. 437), therefore the radiance obtained substituting that digital count value (DNmin) in eq. 7;
- $L_{DO1\%}$ = radiance of Dark Object, assumed to have a reflectance value of 0,01

L_{min} and $L_{DO1\%}$ are expressed by the following equations:

$$L_{min} = G_{rescale} * DNmin + B_{rescale} \tag{14}$$

$L_{DO1\%} = 0,01* [(ESUN_\lambda * \cos\theta_s * T_z) + E_{down}] * T_v / (\pi * d^2)$ (15)

The path radiance is obtained substituting eq.14 and eq.15 in eq.13 resulting in the following equation:
$L_p = G_{rescale} * DNmin + B_{rescale} - 0,01* [(ESUN_\lambda * \cos\theta_s * T_z) + E_{down}] * T_v / (\pi * d^2)$ (16)

There are many ways to calculate the variables T_v, T_z and E_{down}; Song et al. (2001) compared various correction methods like DOS1, DOS2, DOS3 and DOS4, and concluded that the best correction was provided by DOS3, but very similar result was performed by DOS1.

Because of DOS3 is a more complex correction method than DOS1, as it requires an atmospheric radiative transfer model in order to calculate the variables T_v, T_z and E_{down} (Zhang, et al., 2010), in this study the DOS1 model (Chavez, 1996) is used.

DOS1 model assumes no atmospheric transmittance loss, and corrects for the spectral band solar irradiance and solar zenith angle, resulting in:

- $T_v = 1$
- $T_z = 1$
- $E_{down} = 0$

Substituting those values of T_v, T_z and E_{down} in eq.16, the path radiance is given by:
$L_p = G_{rescale} * DNmin + B_{rescale} - 0,01* ESUN_\lambda * \cos\theta_s / (\pi * d^2)$ (17)

Therefore, substituting also in eq. 12 those values of T_v, T_z and E_{down}, the land surface reflectance for Landsat images is:
$\rho = [\pi * (L_\lambda - L_p) * d^2] / (ESUN_\lambda * \cos\theta_s)$ (18)

where L_λ is defined by eq.7, L_p is defined by eq.17, d is calculated from Table 4, $ESUN_\lambda$ is found from Table 3, and $\cos\theta_s$ is the cosine of the solar zenith angle θ_s, which is reported in the image metafile.

Reflectance values should range from 0 to 1, while values above or below this range should be corrected using thresholds (below 0 = 0 and over 1 = 1).

3.2.4 Image Mosaic

Cloud cover over Dar es Salaam often makes classification process difficult, and it is necessary to mosaic several images, in order to obtain cloud free scenes. Mosaic is also needed for Landsat 7 images acquired after May 2003 (because of the SLC-off problem, see Appendix 1), where portions of image are null.

In order to create cloud free images of the whole study area, it is necessary to mosaic two or more images together, where cloud gaps of an image are replaced by the pixels of other images; therefore, perfect geometric registration between images is required.

For radiometric compatibility, it is important that mosaic is performed between images of the same season, in fact the phenological state of vegetation varies considerably during the year; therefore, all the images should be acquired in less than one month, or at least be acquired exactly in the same month of different years.

In order to understand the variation of vegetation state over the year, in Table 5 are listed the NDVI mean values (calculated excluding the ocean surface) of 6 images acquired in the year 2002; the NDVI mean value of April has the higher value, while July has the lower value.

Table 5: NDVI mean for 2002 images

Month	NDVI mean
March	0.3548
April	0.4089
July	0.2526

August	0.3310
November	0.2628
December	0.3132

If images are acquired from different months of the year, a radiometric normalization is needed because of different vegetation phenology and atmospheric effects (Helmer & Ruefenacht, 2007), in order to adapt the histogram of each image in the mosaic. Suddenly, image availability often did not allow for the mosaic of images of the same month, and therefore higher amount of spectral variability characterizes the mosaics.

Following (Figure 4) an example of image mosaic of two Landsat 5 images, in order to fill cloud mask gaps; although the mosaic process, some pixels remain null because the filling image has clouds in the same location of the first image. Therefore, mosaic process needs often more than 2 images, to eliminate all gaps.

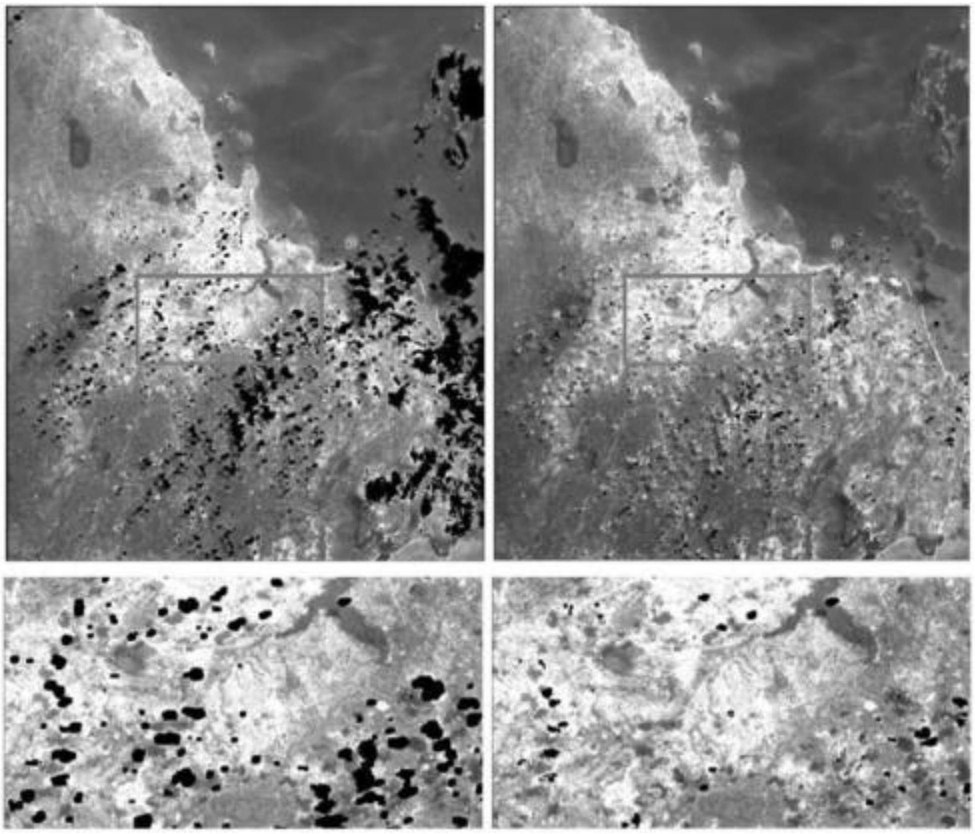

Figure 4: Mosaic example of Landsat 5 images (on the left the cloud masked image, on the right the image mosaic)

The last preprocessing step is to create the set of bands (an image file containing all the Landsat bands), ready for the classification process.

3.3 Classification Process

Data processing is the phase of image classification, based on the individuation of training areas.

3.3.1 Classification Algorithm

Classification algorithms are divided in supervised and unsupervised categories. The former requires training areas to be input by representing, in the image, the LC classes that are already known. The latter are based on *a posteriori* recognition of the classes, having no foreknowledge of their existence or names (Richards & Jia, 2006).

The ML algorithm is one of the most used supervised classifier, which uses the Gaussian threshold stored in each class signature to assign every pixel a class (Huang, et al., 2009).

ML classification assumes that the probability distributions for the classes are of the form of multivariate normal models (Richards & Jia, 2006). The discriminant function, as described by Richards and Jia (2006), is:

$$g_i(x) = \ln p(\omega_i) - ½ \ln |\Sigma_i| - ½ (x - m_i)^t \Sigma_i^{-1} (x - m_i) \qquad (19)$$

where:

- ω_i = class (where i = 1, . . . M and M is the total number of classes)
- x = pixel vector in n-dimension where n is the number of bands
- $p(\omega_i)$ = probability that the correct class is ω_i for a pixel at position x (if equal prior probabilities is assumed it can be omitted)
- $|\Sigma_i|$ = determinant of the covariance matrix of the data in class ω_i
- Σ_i^{-1} = inverse of the covariance matrix
- m_i = mean vector

therefore:

$$x \in \omega_i \text{ if } g_i(x) > g_j(x) \text{ for all } j \neq I \qquad (20)$$

ML has been employed in many LC change studies (Reis, 2008) and is implemented in several software programs. It is necessary to collect the training areas that define classes' statistics, in order to perform the LC classification.

For training areas collection, it is useful to view colour composite images, produced by the combination of three individual monochrome images, which highlight certain surfaces, and help photointerpretation; each band is assigned to a given colour: Red, Green and Blue (RGB) (NASA, 2011).

Some of the possible colour composites, in RGB order, are (adapted from http://grass.osgeo.org/gdp/html_grass64/i.oif.html , accessed 27/01/2012):

- 321: image in natural colour;
- 432: composite very sensitive to green vegetation, which is depicted in red in the image, and coniferous are darker red than deciduous forests;
- 742: composite very useful for forestry, for identifying recent harvest areas and road construction;
- 543: green vegetation is depicted in green and the shortwave band shows vegetation stress;
- 743: similar to 543, but burned areas are better recognizable;
- 754: this colour composite highlights soil texture classes (clay, loam, sandy).

Classification accuracy is assessed checking the coherence between the thematic map and reference data (i.e. ground truth), for a selected, preferably randomly, sample of pixels (i.e. test pixels).

Urban landscapes are a composite combination of buildings, roads, grass, trees, soil, water, and so on (Lu, et al., 2011).

The heterogeneous nature of landscape due to the variety of materials or surfaces makes it complex to define a spectrally distinct "built" class, with increasing difficulty when small, isolated patches of urban cover exist within a vegetated landscape, as is the case of periurban development (Shrestha & Conway, 2011).

Spectral similarities of endmembers make the classification process problematic, as bare soil and unpaved roads can be very similar to impervious surfaces, depending on the soil type (Van_de_Voorde, et al., 2008). Moreover, white soil and white roofs can be spectrally similar.

3.3.2 Spectral Vegetation Indices

Vegetation Indices are standardized methods, based on band ratios, which highlight vegetation dynamics (Song, et al., 2001).

One of the most used indexes is the Normalized Difference Vegetation Index (NDVI) that is a combination of the reflectance of Red and Near Infrared (NIR) wavelengths (Walthall, et al., 2004), and is defined as:

$$NDVI = (\rho_{NIR} - \rho_{RED}) / (\rho_{NIR} + \rho_{RED}) \qquad (21)$$

As described by Huang et al., (2009) "*NDVI range from −1.0 to 1.0. Higher index values are associated with higher levels of healthy vegetation cover, while index values near zero can be due to clouds and snow reflecting less green vegetation*".

Another vegetation index is the Enhanced Vegetation Index (EVI), described by Soudani et al. (2006) as:

$$EVI = G \, [(\rho_{NIR} - \rho_{RED}) / (\rho_{NIR} + C_1 * \rho_{RED} - C_2 * \rho_{BLUE} + L)] \qquad (22)$$

where *"G, C1, C2, and L are coefficients to correct for aerosol scattering, absorption, and background brightness (set at 2.5, 6, 7.5, and 1, respectively)*" (Soudani, et al., 2006, p. 166).

3.3.3 Knowledge-Base Classification

Image classification and pattern recognition can be performed with Knowledge-Base systems, which identify classes "*through the explicit representation of prior knowledge about their spectral, morphological or topological characteristics. Such knowledge, acquired from a human specialist, can reduce significantly the demand for training patterns*" (Costa, et al., 2010).

In particular, the inputs of the Knowledge-Base classification were the ML classification, NDVI, EVI, and Dar boundary shapefile.

NDVI and EVI were used for enhancing the classification process, particularly for vegetation and water identification.

A total of 6 classes were identified in the scene, although the focus for LC was on urban patterns:

- "Continuous Urban", a very dense urbanization class, identified by ML classification;
- "Discontinuous Urban", a low density urbanization class, characterized by a mixed pixel of urban and vegetation or soil, identified by ML classification;
- "Full Vegetation", a vegetation class with:
  $NDVI \geq NDVImax$
  where NDVImax is a threshold value identified in each image, ranging between 0.65 and 0.75;
- "Most Vegetation", a vegetation class, identified by ML classification or with:
  $NDVImin \geq NDVI > NDVImax$
  where NDVImin is a threshold value identified in each image, ranging between 0.55 and 0.65;
- "Soil", identified by ML classification;
- "Water", identified by ML classification or with:
  $EVI < EVImax$
  where EVImax is a value ranging from 0 to 0.05.

Dar boundaries were used to limit LC classifications to the administrative area.

3.4 Results

Following, the results of LC classifications are described.

3.4.1 Land Cover Classifications

The developed methodology is a very affordable and semi-automatic LC classification, which allowed for the monitoring of LCC in Dar es Salaam during the last years; in particular, 5 classifications were performed for years: 2002, 2004, 2007, 2009, and 2011.

The LC classification results are listed in Table 6 and the maps are shown in Figure 5 (offshore islands were not considered in this statistics).

Table 6: Land Cover classification results

Class	2002 [ha]	2004 [ha]	2007 [ha]	2009 [ha]	2011 [ha]	2002 [%]	2004 [%]	2007 [%]	2009 [%]	2011 [%]
Continuous Urban	8415	10025	10447	12370	14808	4.98	5.93	6.18	7.32	8.76
Discontinuous Urban	8098	9134	12509	17318	23678	4.80	5.41	7.40	10.25	14.01
Soil	102079	95732	76011	57385	66791	60.46	56.66	44.98	33.96	39.52
Water	193	276	304	7	199	0.11	0.16	0.18	0.00	0.12
Full Vegetation	14887	13172	14905	26751	18195	8.82	7.80	8.82	15.83	10.77
Most Vegetation	35164	40631	54798	55144	45313	20.83	24.05	32.43	32.63	26.82

Urban sprawl is increasing at very fast pace in the last years, and also “Continuous Urban” class is increasing.

Between 2004 and 2007 there was just a little increment in the “Continuous” class; that could be because: new households were most in peri-urban area (“Discontinuous” class); or misclassification errors underestimated the “Continuous” class, but in this case the accuracy assessment will explain the error causes.

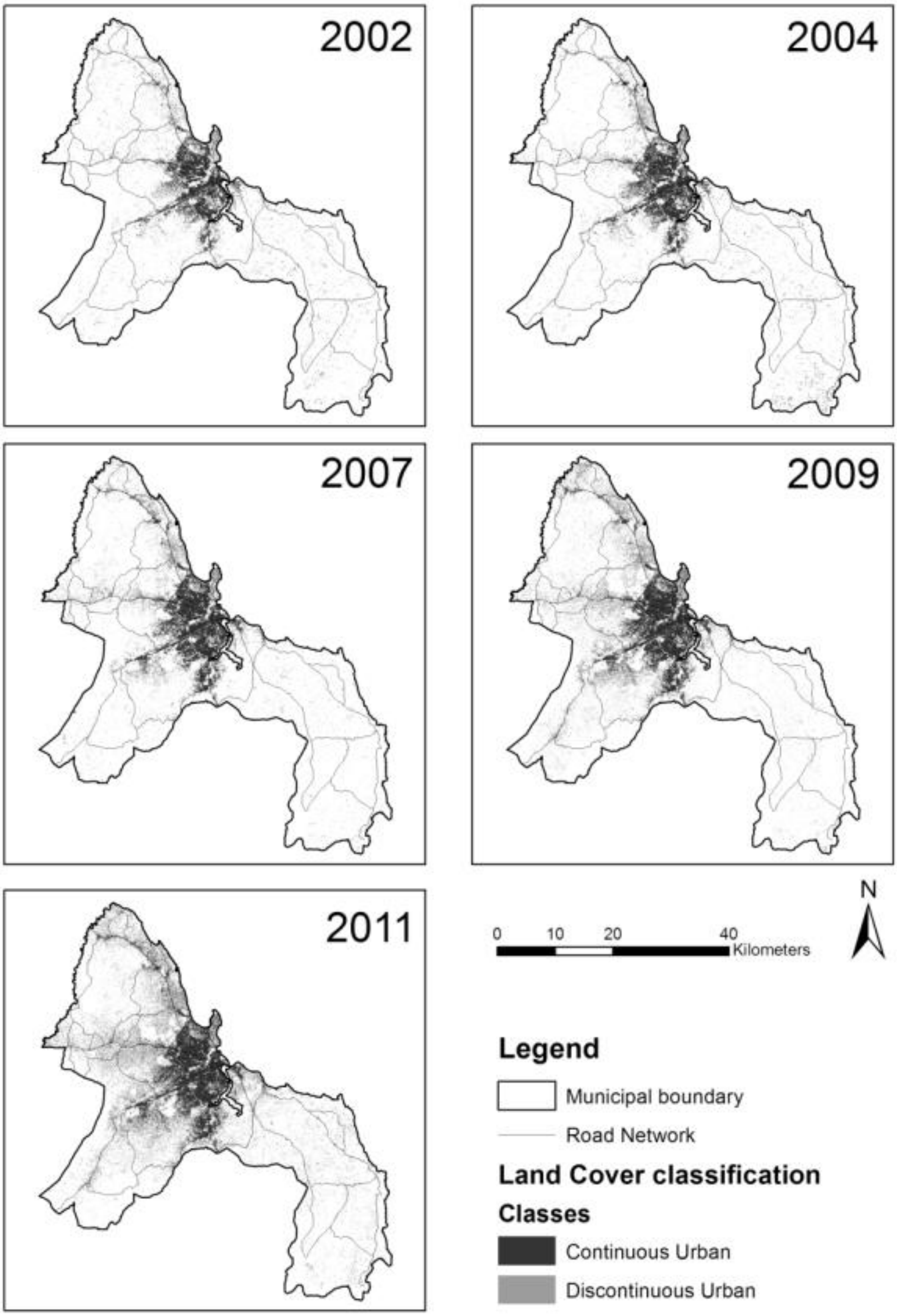

Figure 5: Land Cover Classifications of Dar es Salaam

4. Conclusions and Recommendations

Following, the conclusions derived from this study and the recommendations about the future steps of the project implementation.

4.1 Conclusions

This study developed a methodology for LC classification of Dar es Salaam using remote sensing imagery, in the context of the ACC Dar project objectives: to enhance the capacity of Dar's municipalities in understanding CC issues specific to coastal areas, and to assess their impacts on the livelihood of those urban dwellers partially or totally depending on natural resources; to increase the knowledge on autonomous adaptive capacity; to develop methodologies for integrating adaptation activities into strategies and plans, for UDEM, in coastal unplanned and underserviced settlements.

One of the main goals of this methodology is to be suitable to needs and resources of Dar's municipalities, because it is very affordable; the choice to use Landsat images, which are provided for free, had reduced the costs to the purchase of the commercial software for processing images.

LC classification is performed in a semi-automatic way, in order to reduce the time and cost of LC maps production, especially for large areas like Dar es Salaam; the methodology could be integrated in strategic and planning activities of Dar's municipalities with little effort. Furthermore, the ACC Dar project has planned a strong collaboration between the two partners, Sapienza University and Ardhi University, with training activities to be held in Dar, about Remote Sensing and the developed methodologies.

In particular in East Africa, LCC derives from the interactions of various agents, where the driving forces are both anthropogenic (urbanization, migration, land tenure, etc.) and environmental (climate, rainfall variability, soil and groundwater degradation, etc.) (Olson, et al., 2004); it is very difficult to understand the relationship between CC and local LU changes thereof (Lioubimtseva, et al., 2005) because CC is also affected by many variables, related to natural resources and socio-political situations (Lioubimtseva, et al., 2005).

The developed methodology aims to reduce vulnerability to CC by increasing adaptive capacity of Dar es Salaam's municipalities, which should be able to monitor LCC in a very affordable fashion, and assess those rapid changes.

Municipalities could adjust their plans in a flexible framework, providing the needed infrastructure and services to Dar's inhabitants, while taking into account a socially regularised land management, and addressing environmental problems (Kombe & Kreibich, 2000).

According to the LC classifications we can see a growing trend of new urbanization over the analysed years; urban sprawl is affecting the city, particularly along the main roads, but new urban areas are also growing far from the city centre.

The LCC from 2002 to 2011 shows an increment of the "Continuous Urban" class from 4.98% of Dar area, to 8.76%; also the "Discontinuous Urban" class augmented from 4.80% in 2002 to 14.01% in 2011.

Those urbanization trends confirm that Dar's inhabitants, and also migrants from upcountry, acquire land and build houses in poverty, bypassing formal urban land management (Kombe, 2005), adapting to local environment issues, for example with social organization in local informal institutions (Rodima-Taylor, 2012). Those unplanned settlements could have severe impacts on the ecosystem (Metzger, et al., 2006).

One of the causes of the rapid growth of unplanned settlements is the type of regulatory framework, with administrative procedures taking too long to make land available to the seekers (Kironde, 2006); therefore, it is important that Dar's municipalities have the instruments to constantly monitor LCC and to plan adapting to CC.

The LC maps developed with this methodology could be upgraded and integrated in GIS of Dar's municipalities; through spatial analysis functions the planning services could assess the environmental priorities and plan the needed infrastructures for inhabitants.

Remote sensing techniques are very useful for assessing landscape patterns without in situ measurements, but the atmosphere can also be a source of error, and can limit the applications; a

DOS model was used. in order to reduce atmospheric effects on the images, and the mosaic process tried to solve the issue of cloud cover.

The major problem encountered during the classification process was the difficulty in the identification of image pixels representing the classes, because of the very fast LCC over the years. The lack of reference images (images with higher spatial resolution acquired during the past years) limited the identification of mixed pixels in Landsat images, especially in the peri-urban areas.

Landsat spatial resolution (i.e. 30m) poses a challenge in identifying urban sprawl, because in a single pixel (mixed pixel) of the image there can be mixture of cover types (Richards and Jia, 2006), creating a mixed spectral signature depending on the percentage and the kind of materials are at ground; therefore not the whole area classified as “Discontinuous Urban” is really covered by impervious surfaces, because part of that area is covered by soil or vegetation.

Classifications are based on image mosaics, made of several images acquired in different months, because of the cloud cover issue; therefore, not all the pixels of a mosaic represent the same month of the year (often with different seasonality), causing fluctuations in reflectance values especially on vegetation surfaces, because of the phenological state changes during the year.

Moreover, the mosaic process, because of the multi-seasonality of the images, causes also a higher spectral variability, making the “Definition of the Training Areas” a critical step for successful classifications.

Other minor issues of the methodology are:

- The spectral similarity between white soil and white impervious surfaces, often leading to misclassification errors;
- The difficulty in identifying GCPs, during the georeferencing process of cloudy images, because of the limited visibility of ground due to clouds.

This methodology is similar to another one, developed by this ACC Dar project activity for LC classification based on SPOT images, which is described in the working paper “Development of a Methodology for Land Cover Classification in Dar es Salaam using SPOT Imagery”.

Table 7 shows a comparison of the Landsat and SPOT characteristics of images acquired over Dar es Salaam (see Appendix 1 – Landsat Satellites and Appendix 3 – SPOT Satellite Characteristics).

Table 7: Comparison of the Landsat and SPOT characteristics of images acquired over Dar es Salaam

Characteristics	Landsat	SPOT
Image availability	Landsat 4, 5, 7 Archive (images acquired since 1984)	SPOT 4, 5 Archive (images acquired since 1998)
Cost	Free from USGS	Mostly free from ESA (some images have an extra cost of 400 Euro for data repatriation)
Image size	185km x 185km	60km x 60km
Image per classification	1	3
Spatial resolution	30m	10m
Spectral resolution	7 multispectral + panchromatic	4 multispectral + panchromatic
Cloud cover	High for most of images	High for most of images

SPOT images are delivered for free by the ESA (although some of them entail an extra cost), just as Landsat’s are provided for free by the USGS.

About 3 SPOT images, acquired at the same time, are required for each classification, because of image size and satellite orbit; therefore, Landsat process needs lower number of images per classification, which is an advantage considering the requirement of creating image mosaics because of cloud cover.

The higher spatial resolution of SPOT allows for finer LC classifications, which are convenient for local studies; the coarse spatial resolution of Landsat is more suitable for regional studies.

Another advantage of Landsat images is spectral resolution, which allow for better discrimination of materials. Moreover, the thermal infrared band (which is not available in SPOT images) simplifies the creation of cloud masks.

The methodology for assessing the accuracy of LC classifications is described in the working paper "Development of a Methodology for Land Cover Classification Validation". The assessment of LC fragmentation, performed using LMI, is described in the working paper "Development of a Methodology for Assessing Land Cover Fragmentation".

4.2 Recommendations

The developed methodology has the main objective of monitoring the LCC of Dar, with the explicit requirements to be affordable and have low effort; these requirements aim to allow LC classifications to be updated by local government.

Because of Dar has important socio-economic role for Tanzania, the developed methodology for LC monitoring could be adopted and upgraded also by other Tanzanian administrations, in order to create urban scenarios and coordinate all levels of governmental planning for adaptation to CC (Levina, et al., 2007). In fact, planning processes could be more effective in reducing vulnerability, if short-term decisions are adapted to variability and extreme events of climate, while long-term decisions consider uncertainty (Hallegatte, 2009).

Future work of this study is the assessment of classification accuracy and the validation of LC maps, in order to evaluate the present methodology, and to refine its processing steps.

In order to further reduce the cost of this methodology, processing software could be selected from a valid list of open-source (and freeware) software, particularly: GRASS GIS (http://grass.osgeo.org/ , accessed 27/01/2012) and Quantum GIS (http://www.qgis.org/ , accessed 27/01/2012) software for preprocessing and processing images; InterIMAGE (http://www.lvc.ele.puc-rio.br/projects/interimage/ , accessed 27/01/2012) for creating Knowledge-Base classifications.

The spatial resolution of Landsat (i.e. 30m) is very good at regional level. Another methodology has been developed for assessing urban sprawl, with higher detail level (i.e. 10m), using SPOT images that provided by the ESA.

Although Landsat 5 is no longer operational and Landsat 7 is affected by SLC-off gaps, this methodology can rely on a new satellite, which is the Landsat Data Continuity Mission (LDCM), developed by the collaboration between NASA and USGS, and that will be launched in 2013 (for more information see Appendix 1 – Landsat Satellites, in Landsat Data Continuity Mission section).

Furthermore, ESA is going to launch in 2013 two new Sentinel-2 satellites (the second of five missions called Sentinels), which will provide "*enhanced continuity of SPOT- and Landsat-type data*" (http://www.esa.int/esaLP/SEMM4T4KXMF_LPgmes_0.html , accessed 27/01/2012).

Obviously the methodology will need to be updated, and possibly improved, according to the new conditions and technologies available in the future.

One of the main constraints in remote sensing applications over Dar is the cloud cover (in order to create mosaic and eliminate gaps, 62 Landsat images were acquired from USGS); therefore any future development of this methodology should find the best and automatic way in order to overcome this issue.

The developed methodology has a focus on urban sprawl patterns, but remote sensing data could be used also for monitoring crops, providing direct help and instructions to Dar's inhabitants who rely on urban agriculture for their livelihoods.

References

Africover, 2002. *Africover-Eastern Africa module. Land cover mapping based on satellite remote sensing.* /: FAO.

Briggs, J. & Mwamfupe, D., 2000. Peri-urban Development in an Era of Structural Adjustment in Africa: The City of Dar es Salaam, Tanzania. *Urban Studies,* 37(4), pp. 797-809.

Chavez, P. S., 1996. Image-Based Atmospheric Corrections - Revisited and Improved. *Photogrammetric Engineering and Remote Sensing,* 62(9), pp. 1025-1036.

Costa, G. et al., 2010. *Knowledge-Based Interpretation of Remote Sensing Data with the Interimage System: Major Characteristics and Recent Developments.* /, The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, p. 6.

EEA, 2006. *Urban sprawl in Europe - The ignored challenge,* Copenhagen: EEA/OPOCE.

Fisher, P., Comber, A. J. & Wadsworth, R., 2005. Land Use and Land Cover: Contradiction or Complement. In: P. F. Fisher & D. J. Unwin, eds. *Representing GIS.* Chichester, England: John Wiley \& Sons, pp. 85-98.

Halla, F., 2005. Critical elements in sustaining participatory planning: Bagamoyo strategic urban development planning framework in Tanzania. *Habitat International,* 29(1), pp. 137-161.

Hallegatte, S., 2009. Strategies to adapt to an uncertain climate change. *Global Environmental Change,* 19(2), pp. 240-247.

Helmer, E. H. & Ruefenacht, B., 2007. A comparison of radiometric normalization methods when filling cloud gaps in Landsat imagery. *Canadian Journal of Remote Sensing,* 33(4), pp. 325-340.

Huang, S.-L., Wang, S.-H. & Budd, W. W., 2009. Sprawl in Taipei's peri-urban zone: Responses to spatial planning and implications for adapting global environmental change. *Landscape and Urban Planning,* 90(1-2), pp. 20-32.

IPCC, 2001. *Climate Change 2001: Impacts, Adaptation, and Vulnerability: Contribution of Working Group II to the Third Assessment Report of the IPCC.* Cambridge: Cambridge University Press.

Kironde, J., 2006. The regulatory framework, unplanned development and urban poverty: Findings from Dar es Salaam, Tanzania. *Land Use Policy,* 23(4), pp. 460-472.

Kombe, W., 2005. Land use dynamics in peri-urban areas and their implications on the urban growth and form: the case of Dar es Salaam, Tanzania. *Habitat International,* 29(1), pp. 113-135.

Kombe, W. J. & Kreibich, V., 2000. Reconciling informal and formal land management: an agenda for improving tenure security and urban governance in poor countries. *Habitat International,* 24(2), pp. 231-240.

Kyessi, A., 2005. Community-based urban water management in fringe neighbourhoods: the case of Dar es Salaam, Tanzania. *Habitat International,* 29(1), pp. 1-25.

Levina, E., Jacob, J., Ramos, L. E. & Ortiz, I., 2007. Policy Frameworks for Adaptation to Climate Change in Coastal Zones: The Case of the Gulf of Mexico. *OECD Papers,* 7(1), p. 68.

Lioubimtseva, E., Cole, R., Adams, J. & Kapustin, G., 2005. Impacts of climate and land-cover changes in arid lands of Central Asia. *Journal of Arid Environments,* 62(2), pp. 285-308.

Lioubimtseva, E. & Henebry, G. M., 2009. Climate and environmental change in arid Central Asia: Impacts, vulnerability, and adaptations. *Journal of Arid Environments,* 73(11), pp. 963-977.

Lu, D., Moran, E. & Hetrick, S., 2011. Detection of impervious surface change with multitemporal Landsat images in an urban–rural frontier. *ISPRS Journal of Photogrammetry and Remote Sensing,* 66(3), pp. 298-306.

Martinuzzi, S., Gould, W. A. & Ramos, O. M., 2007. *Creating Cloud-Free Landsat ETM + Data Sets in Tropical Landscapes : Cloud and Cloud-Shadow Removal. General Technical Report PNWGTR521,* Rio Piedras, PR: U.S.: United States Department of Agriculture.

McGarigal, K. & Marks, M., 1995. *FRAGSTATS: spatial pattern analysis program for quantifying landscape structure. Gen. Tech. Rep. PNW-GTR-351..* Portland, OR: Pacific Northwest Research Station.

Mendoza, M. E. et al., 2011. Analysing land cover and land use change processes at watershed level: A multitemporal study in the Lake Cuitzeo Watershed, Mexico (1975–2003). *Applied Geography,* 31(1), pp. 237-250.

Metzger, M. et al., 2006. The vulnerability of ecosystem services to land use change. *Agriculture, Ecosystems \& Environment,* 114(1), pp. 69-85.

Moran, M., Jackson, R., Slater, P. & Teillet, P., 1992. Evaluation of simplified procedures for retrieval of land surface reflectance factors from satellite sensor output. *Remote Sensing of Environment,* 41(2-3), pp. 169-184.

NASA, 2011. *Landsat 7 Science Data Users Handbook.* Maryland: Landsat Project Science Office at NASA's Goddard Space Flight Center in Greenbelt.

Olson, J. M. et al., 2004. *The Spatial Patterns and Root Causes of Land Use Change in East Africa. I*: Land Use Change, Impacts and Dynamics (LUCID).

Olvera, L. D., Plat, D. & Pochet, P., 2003. Transportation conditions and access to services in a context of urban sprawl and deregulation. The case of Dar es Salaam. *Transport Policy,* 10(4), pp. 287-298.

Oreopoulos, L., Wilson, M. J. & Varnai, T., 2011. Implementation on Landsat Data of a Simple Cloud-Mask Algorithm Developed for MODIS Land Bands. *IEEE Geoscience and Remote Sensing Letters,* 8(4), pp. 597-601.

Paavola, J., 2003. *Vulnerability to Climate Change in Tanzania: Sources, Substance and Solutions. I*: A paper presented at the inaugural workshop of Southern Africa Vulnerability Initiative (SAVI) in Maputo, Mozambique June 19-21, 2003.

Paavola, J., 2008. Livelihoods, vulnerability and adaptation to climate change in Morogoro, Tanzania. *Environmental Science Policy,* 11(7), pp. 642-654.

Reis, S., 2008. Analyzing Land Use/Land Cover Changes Using Remote Sensing and GIS in Rize, North-East Turkey. *Sensors,* 8(10), pp. 6188-6202.

Richards, J. A. & Jia, X., 2006. *Remote Sensing Digital Image Analysis: An Introduction.* Berlin, Germany: Springer.

Rodima-Taylor, D., 2012. Social innovation and climate adaptation: Local collective action in diversifying Tanzania. *Applied Geography,* 33(0), pp. 128-134.

Sales_Jr., R., 2009. Vulnerability and adaptation of coastal communities to climate variability and sea-level rise: Their implications for integrated coastal management in Cavite City, Philippines. *Ocean Coastal Management,* 52(7), pp. 395-404.

Schwarz, N., 2010. Urban form revisited—Selecting indicators for characterising European cities. *Landscape and Urban Planning,* 96(1), pp. 29-47.

Shrestha, N. & Conway, T. M., 2011. Delineating an exurban development footprint using SPOT imagery and ancillary data. *Applied Geography,* 31(1), pp. 171-180.

Sobrino, J., Jiménez-Muñoz, J. C. & Paolini", L., 2004. Land surface temperature retrieval from LANDSAT TM 5. *Remote Sensing of Environment,* 90(4), pp. 434-440.

Song, C. et al., 2001. Classification and Change Detection Using Landsat TM Data When and How to Correct Atmospheric Effects?. *Remote Sensing of Environment,* 75(2), pp. 230-244.

Soudani, K. et al., 2006. Comparative analysis of IKONOS, SPOT, and ETM+ data for leaf area index estimation in temperate coniferous and deciduous forest stands. *Remote Sensing of Environment,* 102(1-2), pp. 161-175.

Swan, A., 2010. How increased urbanisation has induced flooding problems in the UK: A lesson for African cities?. *Physics and Chemistry of the Earth Parts ABC,* 35(13-14), pp. 643-647.

Syrbe, R., Bastian, O., Roder, M. & James, P., 2007. A framework for monitoring landscape functions: The Saxon Academy Landscape Monitoring Approach (SALMA), exemplified by soil investigations in the Kleine Spree floodplain (Saxony, Germany). *Landscape and Urban Planning,* 79(2), pp. 190-199.

Thuillier, G., 2004. Solar irradiance reference spectra for two solar active levels. *Advances in Space Research,* 34(2), pp. 256-261.

UN-HABITAT, 2009. *Tanzania: Dar es Salaam City Profile. I*: The United Nations Human Settlements Programme UN-HABITAT.

United_Republic_of_Tanzania, 2004. *Dar es Salaam City Profile. I*: Prepared by the City Council, with advice from Cities and Health Programme, WHO Centre for Development, Kobe, Japan.

Van_de_Voorde, T., Vlaeminck, J. & Canters, F., 2008. Comparing Different Approaches for Mapping Urban Vegetation Cover from Landsat ETM+ Data: A Case Study on Brussels. *Sensors,* 8(6), pp. 3880-3902.

Walthall, C. et al., 2004. A comparison of empirical and neural network approaches for estimating corn and soybean leaf area index from Landsat ETM+ imagery. *Remote Sensing of Environment,* 92(4), pp. 465-474.

Zhang, Z., He, G. & Wang, X., 2010. A practical DOS model-based atmospheric correction algorithm. *International Journal of Remote Sensing,* Volume 31, p. 2837–2852.

Appendix 1 – Landsat Satellites

This Appendix contains some additional information about Landsat satellites.

"*Landsat satellites have been providing multispectral images of the Earth continuously since the early 1970's. […]*

Landsat 7 is the latest satellite in this series. The first was launched in 1972 with two Earth-viewing imagers - a return beam vidicon and an 80-meter multispectral scanner (MSS).

Landsat 2 and 3, launched in 1975 and 1978 respectively, were configured similarly. In 1984, Landsat 4 was launched with the MSS and a new instrument called the Thematic Mapper (TM). Instrument upgrades included improved ground resolution (30 meters) and 3 new channels or bands. In addition to using an updated instrument, Landsat 4 made use of the multimission modular spacecraft (MMS), which replaced the Nimbus, based spacecraft design employed for Landsats 1-3. Landsat 5, a duplicate of 4, was launched in 1984 and even today after 26 years - 21 years beyond its 5-year design life - is still returning useful data.

Landsat 6, equipped with a 15-meter panchromatic band, was lost immediately after launch in 1993" (NASA, 2011, p. 3).

Landsat 7 was launched in 1999 and has a 705km, sun synchronous, earth mapping orbit with a 16-day repeat cycle, with an orbit inclination of 98.2 degrees (NASA, 2011).

Landsat 4 and 5 TM sensor

"*Landsat Thematic Mapper (TM) images consist of seven spectral bands with a spatial resolution of 30 meters for Bands 1 to 5 and 7. Spatial resolution for Band 6 (thermal infrared) is 120 meters, but is resampled to 30-meter pixels. Approximate scene size is 170 km north-south by 183 km east-west*" (from http://landsat.usgs.gov/band_designations_landsat_satellites.php , accessed 27/01/2012).

Table 8: Landsat 4-5 Thematic Mapper (TM) sensor (NASA, 2011)

*TM Band 6 was acquired at 120-meter resolution, but products processed before February 25, 2010 are resampled to 60-meter pixels. Products processed after February 25, 2010 are resampled to 30-meter pixels

	Wavelength [µm]	**Resolution** [m]
Band 1	0.45-0.52	30
Band 2	0.52-0.60	30
Band 3	0.63-0.69	30
Band 4	0.76-0.90	30
Band 5	1.55-1.75	30
Band 6	10.40-12.50	120* (30)
Band 7	2.08-2.35	30

Landsat 7 ETM+ sensor

Landsat 7's sensor named Enhanced Thematic Mapper Plus (ETM+) "*is a derivative of the Thematic Mapper (TM) engineered for Landsats 4 and 5, but is more closely related to the Enhanced Thematic Mapper (ETM) lost during the Landsat 6 failure*" (NASA, 2011, p. 19).

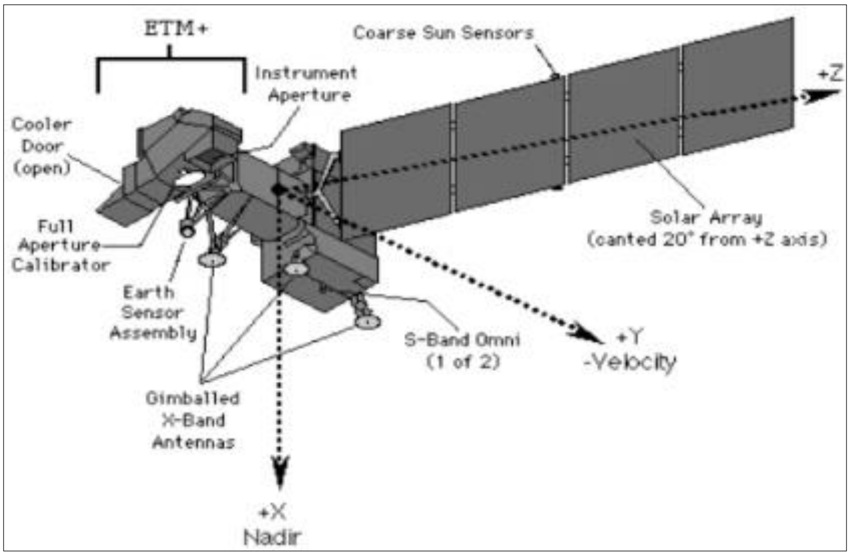

Figure 6: The Landsat 7 satellite as viewed from the sun side (NASA, 2011)

"*Landsat 7 images consist of eight spectral bands with a spatial resolution of 30 meters for Bands 1 to 7. The resolution for Band 8 (panchromatic) is 15 meters. All bands can collect one of two gain settings (high or low) for increased radiometric sensitivity and dynamic range, while Band 6 collects both high and low gain for all scenes. Approximate scene size is 170 km north-south by 183 km east-west*" (from http://landsat.usgs.gov/band_designations_landsat_satellites.php , accessed 27/01/2012).

Table 9: Landsat 7 ETM+ sensor (NASA, 2011)
* ETM+ Band 6 is acquired at 60-meter resolution. Products processed after February 25, 2010 are resampled to 30-meter pixels.

	Spectral Response	**Wavelength** [micrometers]	**Resolution** [meters]
Band 1	Blue-Green	0.45-0.52	30
Band 2	Green	0.52-0.60	30
Band 3	Red	0.63-0.69	30
Band 4	Near IR	0.77-0.90	30
Band 5	Mid-IR	1.55-1.75	30
Band 6	Thermal IR	10.40-12.50	60* (30)
Band 7	Mid-IR	2.09-2.35	30
Band 8	Pan	0.52-0.90	15

"*The bidirectional scan mirror assembly (SMA) sweeps the detector's line of sight in west-to-east and east-to-west directions across track, while the spacecraft's orbital path provides the north-south motion. A Ritchey-Chretien telescope focuses the energy onto a pair of motion compensation mirrors (i.e. scan line corrector) where it is redirected to the focal planes. The scan line corrector is required due to the compound effect of along-track orbital motion and crosstrack scanning which leads to significant overlap and underlap in ground coverage between successive scans*" (NASA, 2011, pp. 19-20).

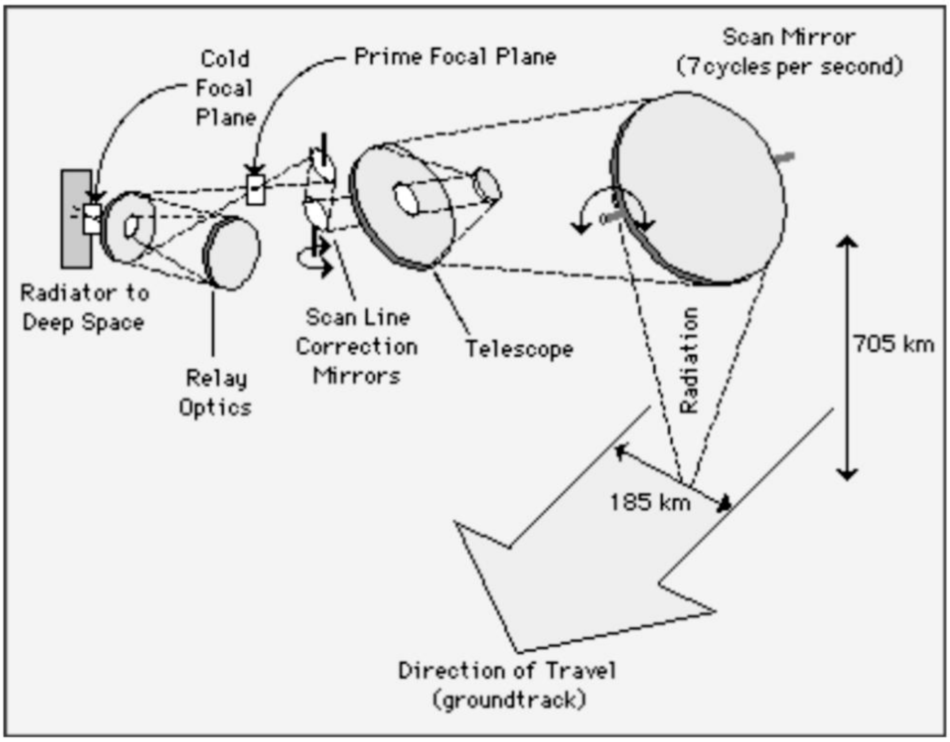

Figure 7: ETM+ Optical Path (NASA, 2011)

Landsat 7 SLC-off

"*On May 31, 2003, the Scan Line Corrector (SLC), which compensates for the forward motion of Landsat 7, failed. Subsequent efforts to recover the SLC were not successful, and the failure appears to be permanent. Without an operating SLC, the Enhanced Thematic Mapper Plus (ETM+) line of sight now traces a zig-zag pattern along the satellite ground track. As a result, imaged area is duplicated, with width that increases toward the scene edge*" (from http://landsat.usgs.gov/products_slcoffbackground.php , accessed 27/01/2012).

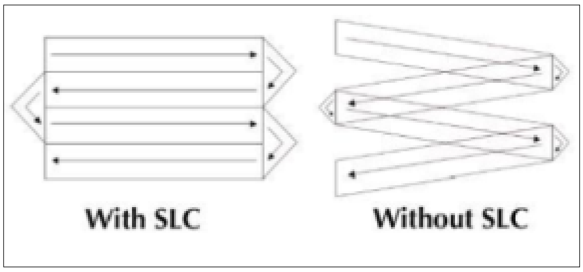

Figure 8: SLC Failure
(from http://landsat.usgs.gov/products_slcoffbackground.php , accessed 27/01/2012)

"*The Landsat 7 ETM+ is still capable of acquiring useful image data with the SLC turned off, particularly within the central part of any given scene. The Landsat 7 ETM+ therefore continues to acquire image data in the "SLC-off" mode. All Landsat 7 SLC-off data are of the same high radiometric and geometric quality as data collected prior to the SLC failure.*

The SLC-off effects are most pronounced along the edge of the scene and gradually diminish toward the center of the scene […]. The middle of the scene, approximately 22 kilometers wide on a Level 1 (L1G, L1Gt, L1T) product, contains very little duplication or data loss, and this region of each image is *very similar in quality to previous ("SLC-on") Landsat 7 image data. […]*

An estimated 22 percent of any given scene is lost because of the SLC failure. The maximum width of the data gaps along the edge of the image would be equivalent to one full scan line, or approximately 390 to 450 meters. The precise location of the missing scan lines will vary from scene to scene." (from http://landsat.usgs.gov/products_slcoffbackground.php , accessed 27/01/2012).

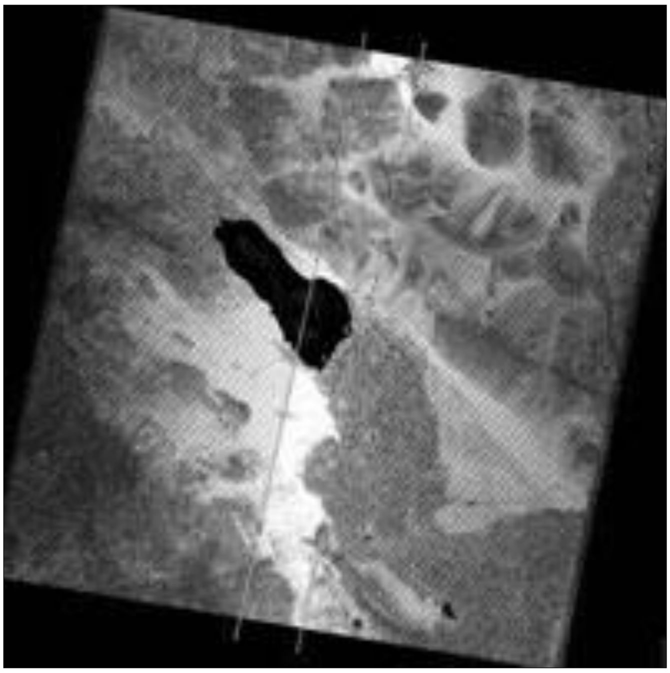

Figure 9: Complete Landsat 7 scene showing affected vs. unaffected area
(from http://landsat.usgs.gov/products_slcoffbackground.php , accessed 27/01/2012)

It is necessary to fill the SLC-off gaps in the image. Alternative processes are:

- Create a gap mask and mosaic two Landsat 7 SLC-off images, because generally the gaps are not in the same position;
- Create a gap mask and mosaic a Landsat 7 SLC-off with a Landsat 5 image.

The problem in mosaic process could be the different vegetation status because of the seasonal change or a different urban land cover; therefore, the images should be acquired very near in time.

Landsat Data Continuity Mission

"*The Landsat Data Continuity Mission (LDCM), a collaboration between NASA and the U.S. Geological Survey, will provide moderate-resolution (15 m–100 m, depending on spectral frequency) measurements of the Earth's terrestrial and polar regions in the visible, near-infrared, short wave infrared, and thermal infrared. […]*

The LDCM satellite payload consists of two science instruments—the Operational Land Imager (OLI) and the Thermal Infrared Sensor (TIRS). These two sensors will provide seasonal coverage of the global landmass at a spatial resolution of 30 meters (visible, NIR, SWIR); 100 meters (thermal); and 15 meters (panchromatic). The spectral coverage and radiometric performance (accuracy, dynamic range, and precision) are designed to detect and characterize multi-decadal land cover change in concert with historic Landsat data. Coordinated calibration efforts of USGS and NASA will again be part of the LDCM calibration strategy. The LDCM scene size will be 185-km-cross-track-by-180-km-along-track. The nominal spacecraft altitude will be 705 km. Cartographic accuracy of 12 m or better (including compensation for terrain effects) is required of LDCM data products. LDCM includes evolutionary advances in technology and performance. The OLI provides two new spectral bands, one *tailored especially for detecting cirrus clouds and the other for coastal zone observations, and the TIRS will collect data for two more narrow spectral bands in the thermal region formerly covered by one wide spectral band on Landsats 4–7. Additionally, LDCM is required to return 400 scenes per day to the USGS data archive (150 more than Landsat 7), increasing the probability of capturing cloud-free scenes for the global landmass. […]*

LDCM is now in the final phase of design and fabrication. Due to the high national importance of the mission and the need to maintain the continuity of the Landsat data record, NASA and USGS will strive to launch LDCM in January 2013." (from http://landsat.gsfc.nasa.gov/about/ldcm.html , accessed 27/01/2012).

Table 10: Landsat Data Continuity Mission sensor
(from http://landsat.usgs.gov/band_designations_landsat_satellites.php , accessed 27/01/2012)

Bands	**Wavelength** [micrometres]	**Resolution** [metres]
Band 1 – Coastal aerosol	0.433 – 0.453	30
Band 2 – Blue	0.450 – 0.515	30
Band 3 – Green	0.525 – 0.600	30
Band 4 – Red	0.630 – 0.680	30
Band 5 – Near Infrared (NIR)	0.845 – 0.885	30
Band 6 – SWIR 1	1.560 – 1.660	30
Band 7 – SWIR 2	2.100 – 2.300	30
Band 8 – Panchromatic	0.500 – 0.680	15
Band 9 – Panchromatic	1.360 – 1.390	30
Band 10 – Thermal Infrared (TIR) 1	10.3 – 11.3	120
Band 11 – Thermal Infrared (TIR) 2	11.5 – 12.5	120

Level 1 Product Generation System

Nearly all USGS Landsat standard data products are generated by Level 1 Product Generation System (LPGS).

"*Scenes processed after April 5, 2004: LPGS scale Level 1 products to a range of 1-254. DN values of 0 are reserved for scan gap and flag fill. DN values of 255 are reserved for saturation.*

Scenes processed prior to April 5, 2004: LPGS scaled Level 1 products to a range of 1-254 with values of 0 and 255 set aside to flag fill and high saturation, respectively" (from http://landsat.usgs.gov/products_IP_LPGSvsNLAPS.php , accessed 27/01/2012).

The "PROCESSING_SOFTWARE" line, in the metadata file, describes the software version of LPGS.

Geometric Accuracy

USGS delivers Landsat imagery already geometrically and radiometrically corrected by one of the following methods (from http://landsat.usgs.gov/products_productinformation.php , accessed 27/01/2012):

- "*Standard Terrain Correction (Level 1T) - provides systematic radiometric and geometric accuracy by incorporating ground control points while employing a Digital Elevation Model (DEM) for topographic accuracy. Geodetic accuracy of the product depends on the accuracy of the ground control points and the resolution of the DEM used:*
  - *Ground control points used for Level 1T correction come from the GLS2005 data set. DEM sources include SRTM, NED, CDED, DTED, and GTOPO 30*
- *Systematic Terrain Correction (Level 1Gt) - provides systematic, radiometric, and geometric*

accuracy, while employing a Digital Elevation Model (DEM) for topographic accuracy.

- *Systematic Correction (Level 1G) - provides systematic radiometric and geometric accuracy, which is derived from data collected by the sensor and spacecraft. Geometric accuracy of the systematically corrected product should be within 250 meters (1 sigma) for low-relief areas at sea level*".

The "PRODUCT_TYPE" line of the metadata file, provided with the downloaded images, describes the level of correction.

High and Low Gain

The ETM+ images are acquired in either a low or high gain state, and gain selection for a scene is controlled by the "Mission Operations Center" located at Goddard Space Flight Center in Greenbelt, Maryland.

Gain selection "*is performed by changing the reference voltage for the analog to digital convertor. This occurs in the preceding scene. The science goal in switching gain states is to maximize the instrument's 8 bit radiometric resolution without saturating the detectors. This requires matching the gain state for a given scene to the expected brightness conditions. For all bands, the low gain dynamic range is approximately 1.5 times the high gain dynamic range. It makes sense, therefore, to image in low gain mode when surface brightness is high and in high gain mode when surface brightness is lower*" (NASA, 2011, p. 53).

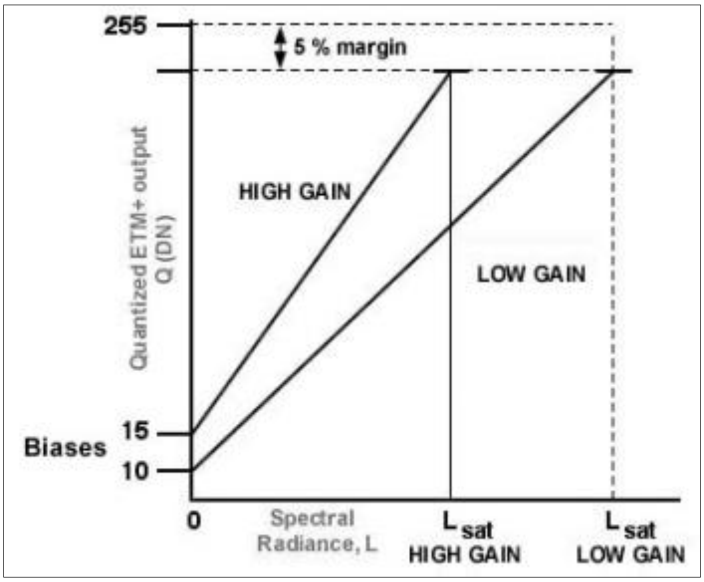

Figure 10: Design ETM+ Reflective Band High and Low gain Dynamic Ranges (NASA, 2011)

The NASA (2011) describes for Land (non-desert, non-ice) these gain settings:

- Bands 1-3 set to high gain;
- Band 4 set to high gain except where sun elevation is greater than 45°(set to low gain) - to avoid dense vegetation (reflectance > 0.66) saturation. (At this sun angle high gain in band 4 saturates at about a reflectance of 0.66, so switching to low gain keeps targets at this reflectance or below from saturating.);
- Bands 5, 7 set to high gain;
- Band 8 set to low gain.

Gain settings of downloaded imagery are located in the metadata file. The following is an example of gain settings for Landsat 7 image acquired over Dar es Salaam on 05/02/2011 (where H = high gain and L = low gain):

- BAND1_GAIN = "H";
- BAND2_GAIN = "H";
- BAND3_GAIN = "H";
- BAND4_GAIN = "L";
- BAND5_GAIN = "H";
- BAND6_GAIN1 = "L";
- BAND6_GAIN2 = "H";
- BAND7_GAIN = "H";
- BAND8_GAIN = "L".

Appendix 2 – Land Cover Classifications

This Appendix contains the figures representing the LC classifications of Dar for years: 2002, 2004, 2007, 2009, and 2011.

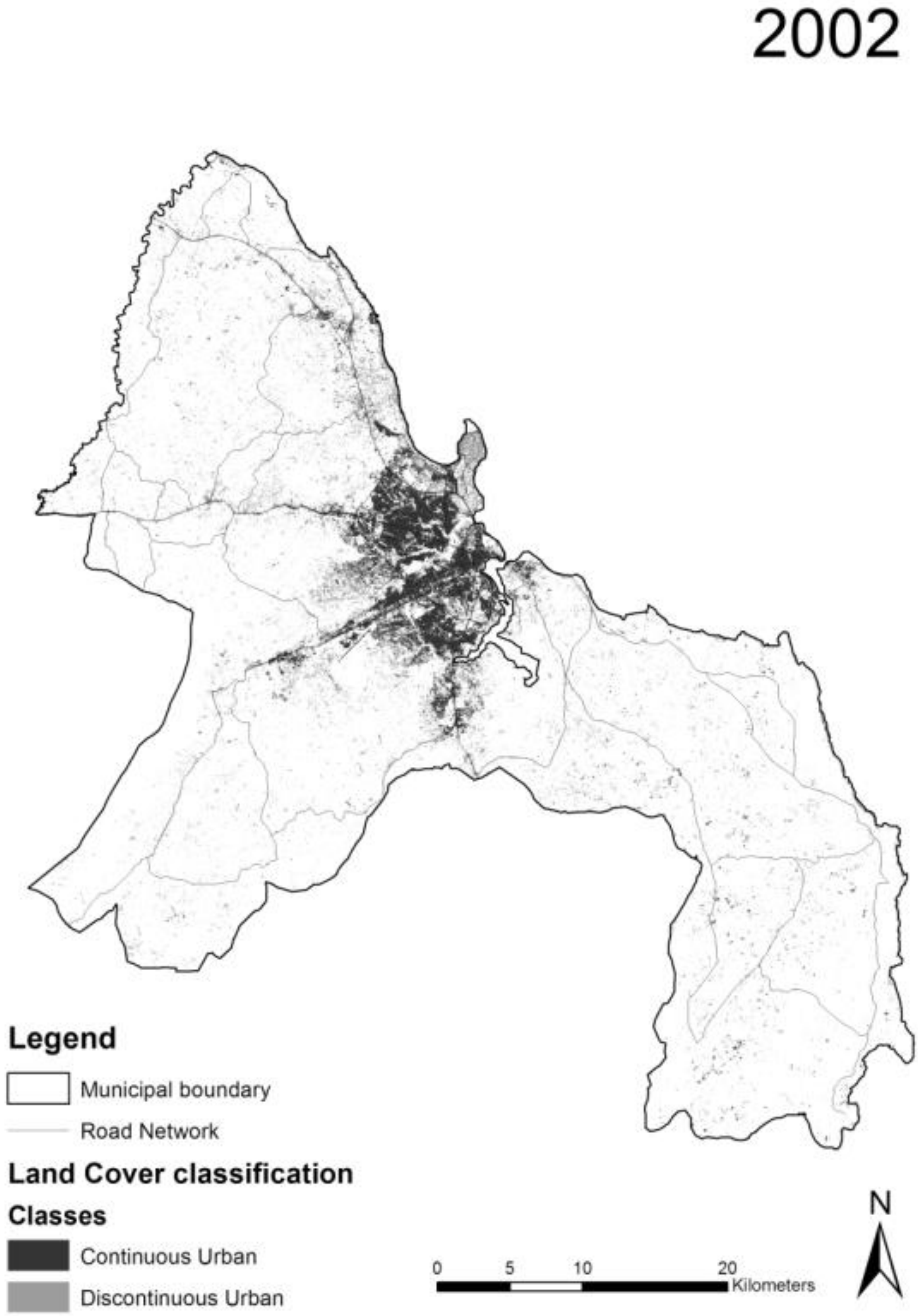

Figure 11: Land Cover classification for year 2002

2004

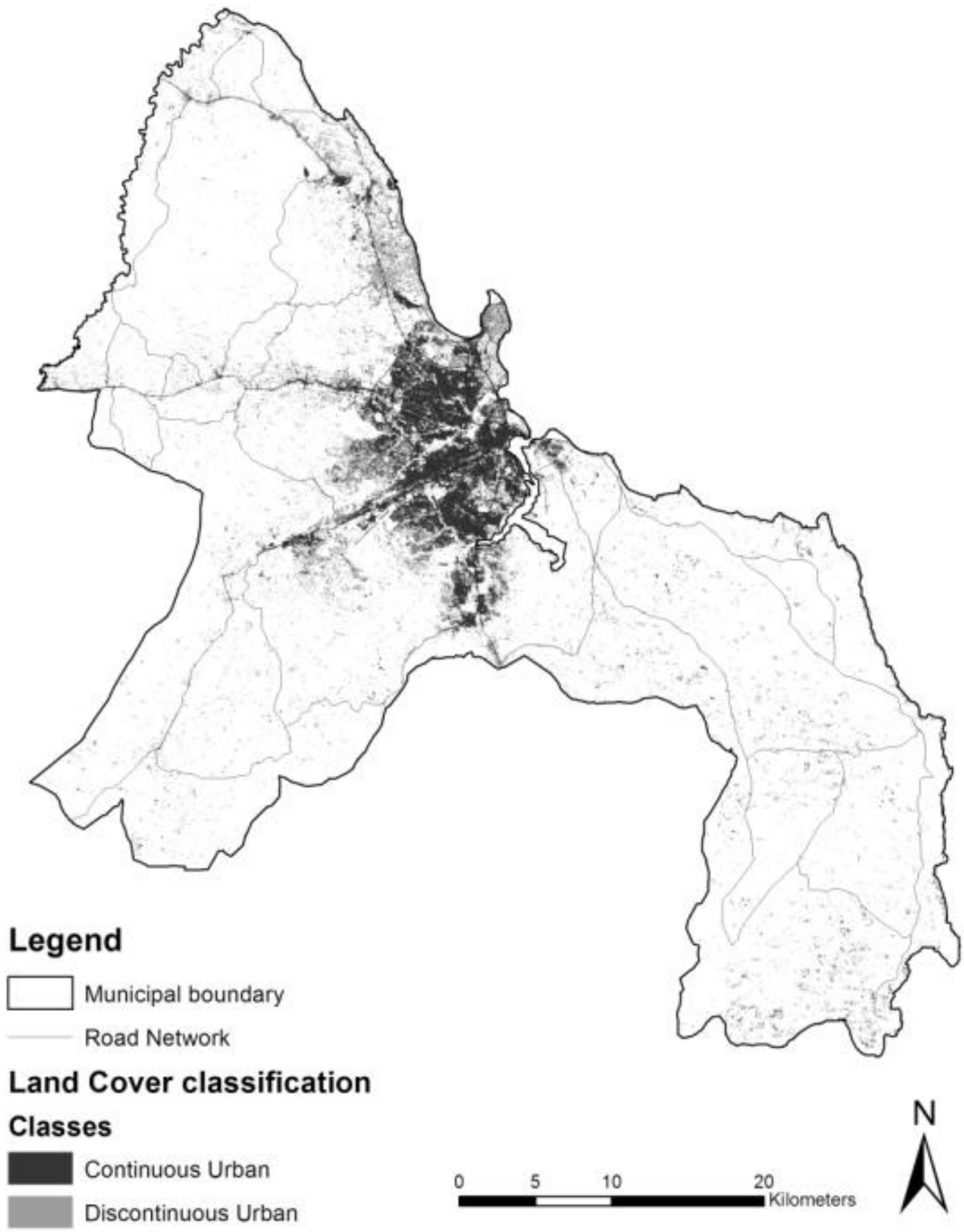

Figure 12: Land Cover classification for year 2004

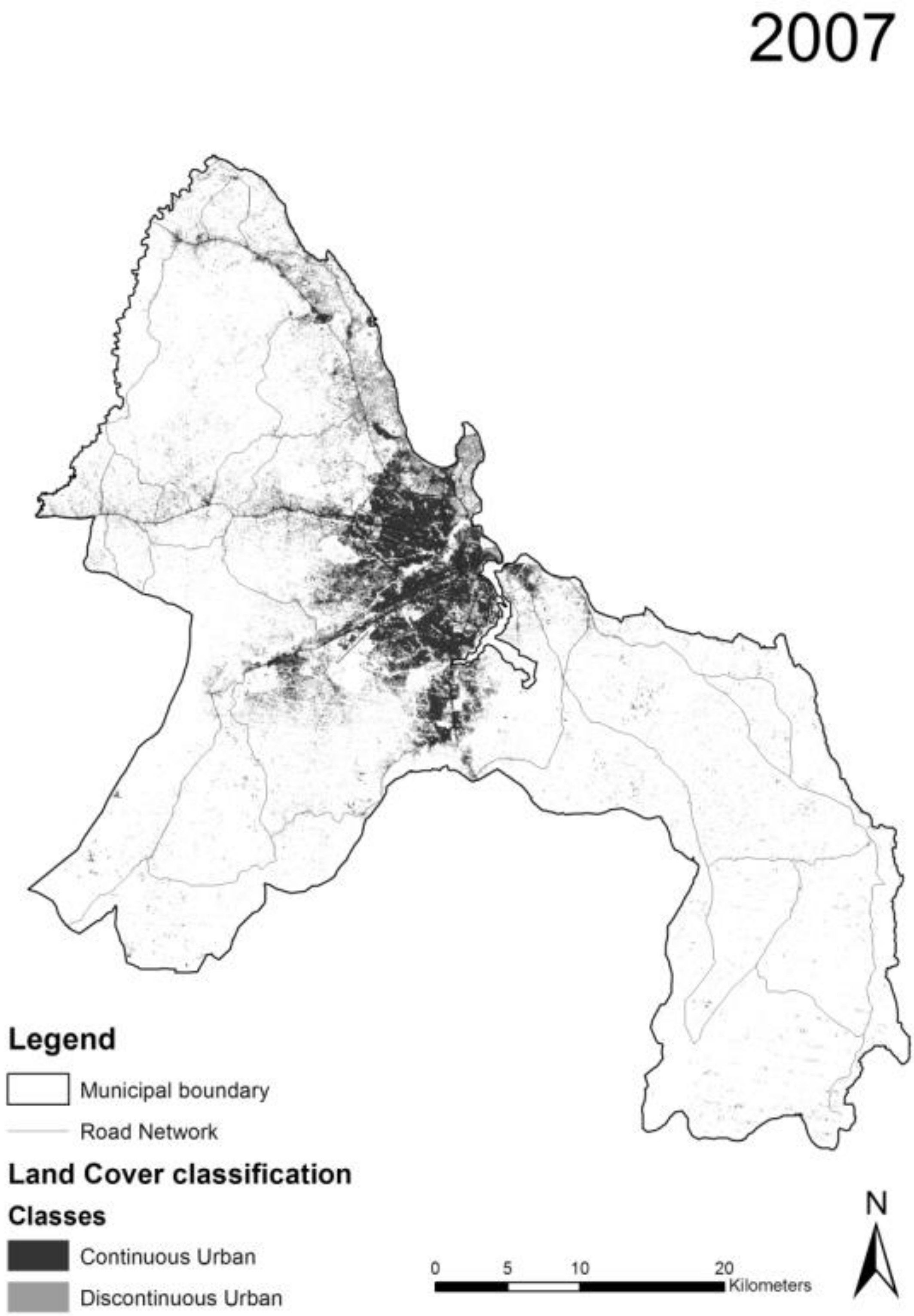

Figure 13: Land Cover classification for year 2007

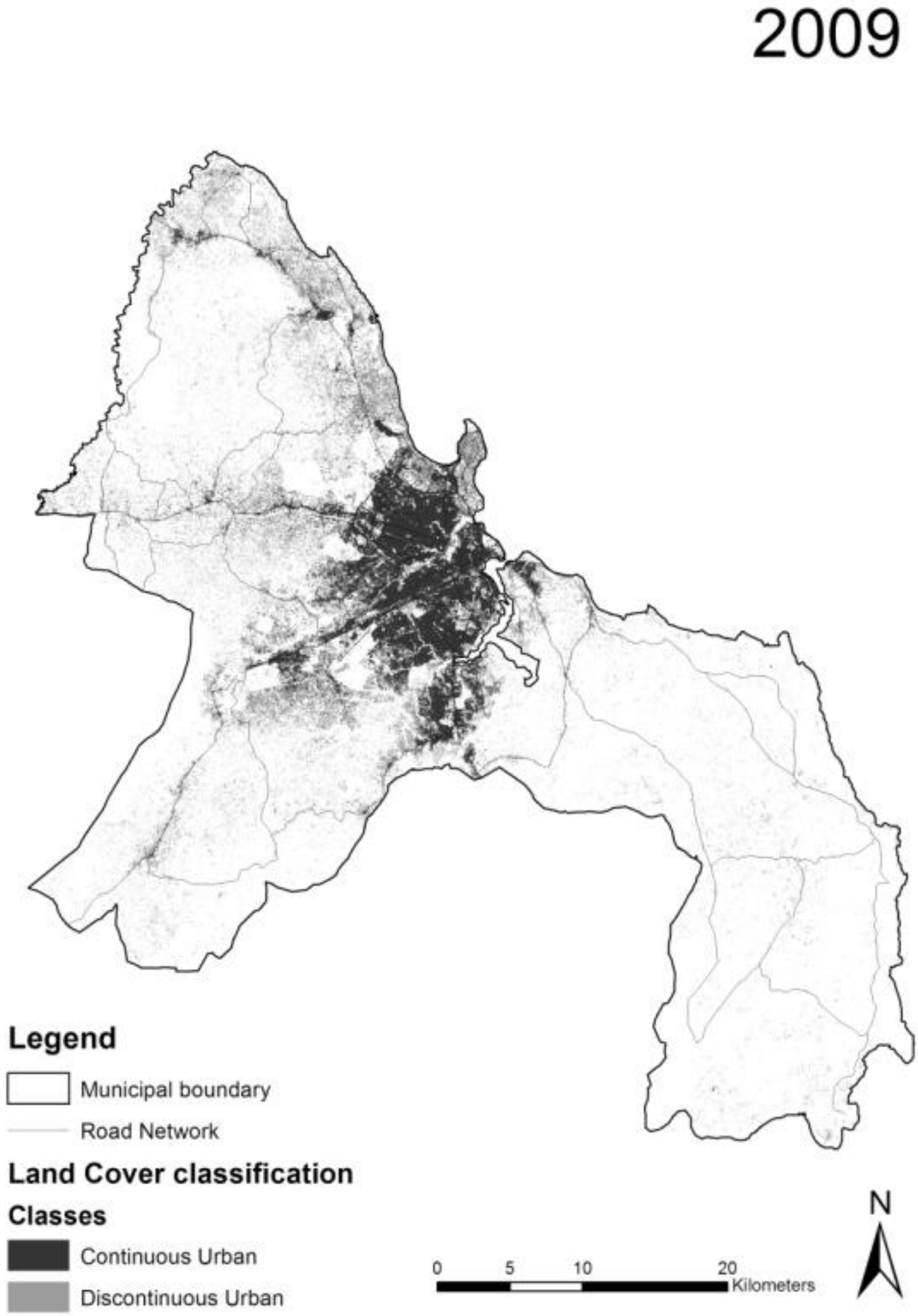

Figure 14: Land Cover classification for year 2009

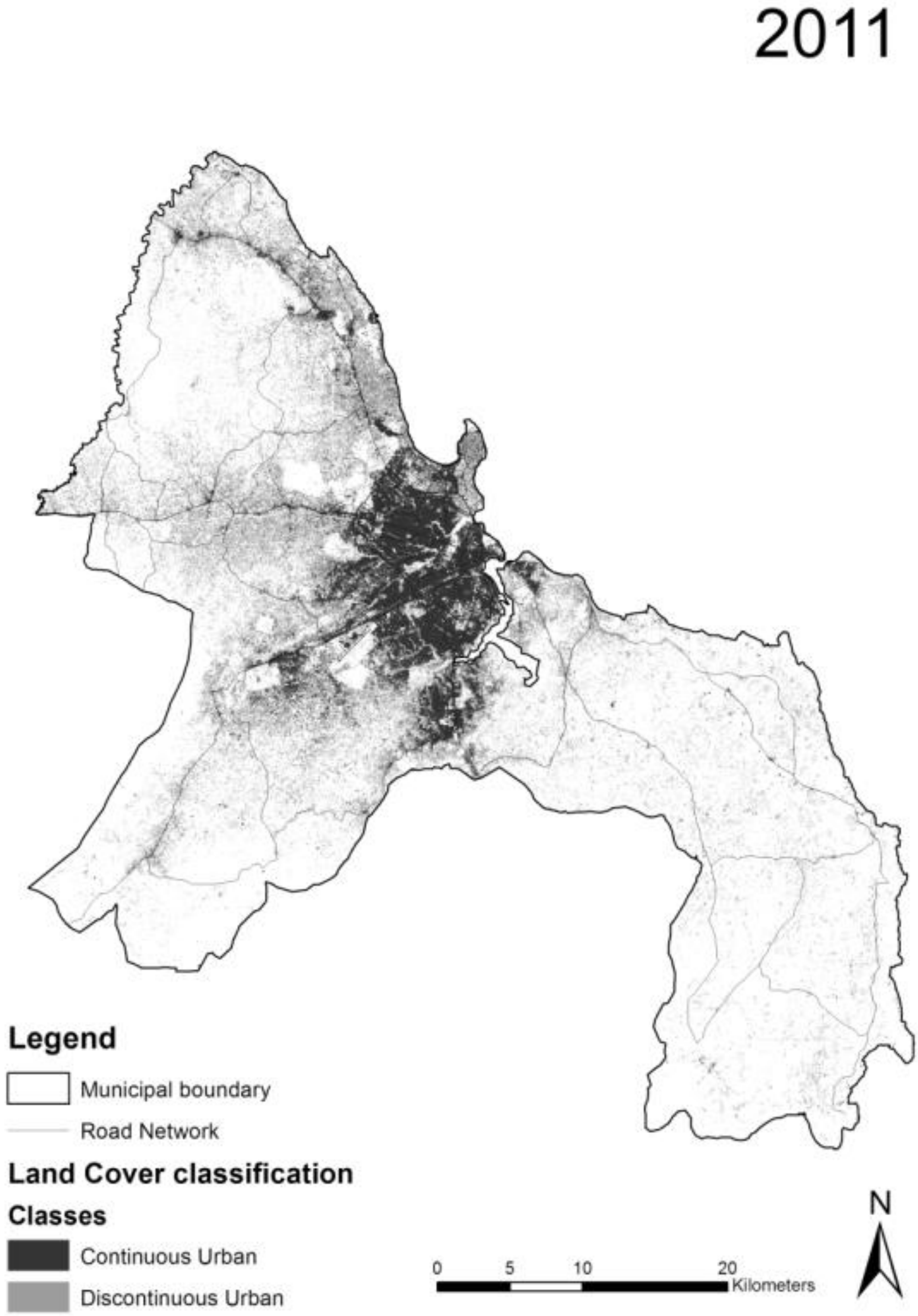

Figure 15: Land Cover classification for year 2011

Appendix 3 – SPOT Satellite Characteristics

SPOT is a family of satellites designed and developed by the French space agency Centre National d'Études Spatiales (CNES).
SPOT 4 was launched in 1998, and its sensor is called High-Resolution Visible and Infrared (HRVIR). Sensor characteristics thereof are displayed in Table 11.

Table 11: HRVIR sensor spectral bands and resolutions

Electromagnetic spectrum	**Pixel size [m]**	**Spectral bands [μm]**
Monospectral	10	0.61 – 0.68
B1 : Green	20	0.50 – 0.59
B2 : Red	20	0.61 – 0.68
B3 : Near-Infrared (NIR)	20	0.78 – 0.89
B4 : Short-Wave Infrared (SWIR)	20	1.58 – 1.75

SPOT 5 was launched in 2002. Its sensor is called High Resolution Geometric (HRG) and has the characteristics listed in Table 12.

Table 12: HRG sensor spectral bands and resolutions

Electromagnetic spectrum	**Pixel size [m]**	**Spectral bands [μm]**
Panchromatic	2.5 or 5	0.48 – 0.71
B1 : Green	10	0.50 – 0.59
B2 : Red	10	0.61 – 0.68
B3 : Near-Infrared (NIR)	10	0.78 – 0.89
B4 : Short-Wave Infrared (SWIR)	20	1.58 – 1.75