DEPARTMENT OF GEOSCIENCES AND GEOGRAPHY A 1

Geospatial Environmental Data Modelling Applications Using Remote Sensing, GIS and Spatial Statistics

MIKA SILJANDER

UNIVERSITY OF HELSINKI
FACULTY OF SCIENCE

Geospatial environmental data modelling applications using remote sensing, GIS and spatial statistics

Mika Siljander

Department of Geosciences and Geography
Faculty of Science
University of Helsinki
Finland

Academic dissertation

To be presented with the permission of the Faculty of Science of the University of Helsinki, for public criticism in the Auditorium E204 of the Physicum building (Gustaf Hällströmin katu 2) On March 12th 2010 at 12 o'clock.

Helsinki 2010

Supervisor:

Dr. Petri Pellikka
Professor
Department of Geosciences and Geography
University of Helsinki
Finland

Pre-examiners:

Dr. P.K. Joshi
Associate Professor
Department of Natural Resources, TERI University
New Delhi
India

Dr. Tarmo Virtanen
University Lecturer
Department of Biological and Environmental Sciences
University of Helsinki
Finland

Opponent:

Dr. Risto Kalliola
Professor
Department of Geography
University of Turku
Finland

Publisher:

Department of Geosciences and Geography
Faculty of Science
PO Box 64, FI-000140 University of Helsinki
Finland

ISSN–L 1798–7911
ISSN 1798–7911 (print)
ISBN 978-952-10-6109-7 (paperback)
ISBN 978-952-10-6110-3 (PDF)
http://ethesis.helsinki.fi

Helsinki University Print
Helsinki 2010

Abstract

This thesis presents novel modelling applications for environmental geospatial data using remote sensing, GIS and statistical modelling techniques. The studied themes can be classified into four main themes: (i) to develop advanced geospatial databases. Paper (I) demonstrates the creation of a geospatial database for the Glanville fritillary butterfly (*Melitaea cinxia*) in the Åland Islands, south-western Finland; (ii) to analyse species diversity and distribution using GIS techniques. Paper (II) presents a diversity and geographical distribution analysis for Scopulini moths at a world-wide scale; (iii) to study spatiotemporal forest cover change. Paper (III) presents a study of exotic and indigenous tree cover change detection in Taita Hills Kenya using airborne imagery and GIS analysis techniques; (iv) to explore predictive modelling techniques using geospatial data. In Paper (IV) human population occurrence and abundance in the Taita Hills highlands was predicted using the generalized additive modelling (GAM) technique. Paper (V) presents techniques to enhance fire prediction and burned area estimation at a regional scale in East Caprivi Namibia. Paper (VI) compares eight state-of-the-art predictive modelling methods to improve fire prediction, burned area estimation and fire risk mapping in East Caprivi Namibia.

The results in Paper (I) showed that geospatial data can be managed effectively using advanced relational database management systems. Metapopulation data for *Melitaea cinxia* butterfly was successfully combined with GPS-delimited habitat patch information and climatic data. Using the geospatial database, spatial analyses were successfully conducted at habitat patch level or at more coarse analysis scales. Moreover, this study showed it appears evident that at a large-scale spatially correlated weather conditions are one of the primary causes of spatially correlated changes in *Melitaea cinxia* population sizes.

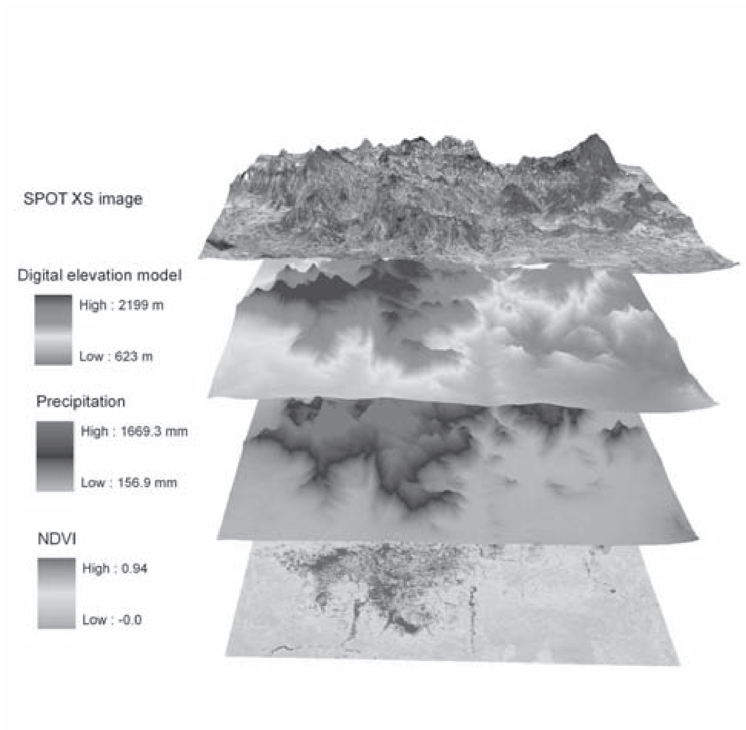
In Paper (II) spatiotemporal characteristics of Socupulini moths description, diversity and distribution were analysed at a world-wide scale and for the first time GIS techniques were used for Scopulini moth geographical distribution analysis. This study revealed that Scopulini moths have a cosmopolitan distribution. The majority of the species have been described from the low latitudes, sub-Saharan Africa being the hot spot of species diversity. However, the taxonomical effort has been uneven among biogeographical regions.

Paper III showed that forest cover change can be analysed in great detail using modern airborne imagery techniques and historical aerial photographs. However, when spatiotemporal forest cover change is studied care has to be taken in co-registration and image interpretation when historical black and white aerial photography is used.

In Paper (IV) human population distribution and abundance could be modelled with fairly good results using geospatial predictors and non-Gaussian predictive modelling techniques. Moreover, land cover layer is not necessary needed as a predictor because first and second-order image texture measurements derived from satellite imagery had more power to explain the variation in dwelling unit occurrence and abundance.

Paper V showed that generalized linear model (GLM) is a suitable technique for fire occurrence prediction and for burned area estimation. GLM based burned area estimations were found to be more superior than the existing MODIS burned area product (MCD45A1). However, spatial autocorrelation of fires has to be taken into account when using the GLM technique for fire occurrence prediction.

Paper VI showed that novel statistical predictive modelling techniques can be used to improve fire prediction, burned area estimation and fire risk mapping at a regional scale. However, some noticeable variation between different predictive modelling techniques for fire occurrence prediction and burned area estimation existed.



Geospatial map layers for Taita Hills Kenya

Publications

This thesis is a summary of the following articles, which are referred to in the text by their Roman numerals. Article IV contains colour figures, which have been printed here in greyscale.

I Nieminen, M., Siljander, M. & Hanski, I. (2004). Structure and dynamics of *Melitaea cinxia* metapopulations. In: Ehrlich, P. R. & Hanski, I. (eds.) *On the wings of checkerspots: a model system for population biology*. Oxford University Press, Oxford, pp. 63–91.

II Sihvonen, P. & Siljander, M. (2005). Species diversity and geographical distribution of Scopulini moths (Lepidoptera: Geometridae, Sterrhinae) on a world-wide scale. *Biodiversity and Conservation* 14, 703–721.

III Pellikka, P.K.E., Lötjönen, M., Siljander, M., & Lens, L. (2009). Airborne remote sensing of spatiotemporal change (1955-2004) in indigenous and exotic forest cover in the Taita Hills, Kenya. *International Journal of Applied Earth Observation and Geoinformation* 11, 221–232.

IV Siljander, M., Clark, B. & Pellikka, P.K.E. A predictive modelling technique for human population distribution and abundance estimation using remote sensing and geospatial data in a rural mountainous area in Kenya. *International Journal of Remote Sensing* (in press).

V Siljander, M. (2009). Predictive fire occurrence modelling to improve burned area estimation at a regional scale: A case study in East Caprivi, Namibia. *International Journal of Applied Earth Observation and Geoinformation* 11, 380–393.

VI Siljander, M. The performance of eight state-of-the-art predictive modelling techniques for improving fire occurrence prediction, burned area estimation and fire risk mapping: A case study of East Caprivi, Namibia (submitted manuscript).

THE AUTHOR'S CONTRIBUTION

In Paper (I) Siljander jointly designed and independently created the *Melitaea Cinxia* geospatial database and performed the necessary database operations and created the maps and spatial illustrations. Siljander collected, processed and partly analyzed the climatological data and wrote jointly the GIS-database and climate sections of the book chapter with M.N. and I.H. In Paper (II) Siljander designed and created the Scopulini moths geospatial database and made the species distribution maps and participated in spatial analysis. P.S. collected the species data and was the main author for the article. In Paper (III) Siljander jointly designed and created the geospatial database with P.K.E.P. and M.L. and contributed in statistics- and change detection analysis. In Paper (IV) Siljander designed and created the Taita Hills geospatial database except B.C. was responsible for the SPOT land cover model. Siljander analyzed the data and was the main author for the article. P.K.E.P. contributed mainly in introduction, study area and conclusion sections. In Papers (V and VI) Siljander designed and created the East Caprivi geospatial fire database, analyzed the data and was the sole author for the manuscripts. The original publications have been reprinted with kind permission from the publishers, Oxford University Press (I), Springer (II) and Elsevier (III, V).

Abbreviations

AIC	Akaike's Information Criterion
AML	ARC/INFO macro language
ANN	Artificial neural networks
AUC	Area under the curve
AVHRR	Advanced Very High Resolution Radiometer
ATSR	Along Track Scanning Radiometer
CASE	Computer aided software engineering
CTA	Classification tree analysis
DEM	Digital elevation model
ESRI	Environmental Systems Research Institute
ETM+	The Landsat Enhanced Thematic Mapper Plus
FAO	Food and Agriculture Organization of the United Nations
FINNIDA	Department for International Development Cooperation
FRA	Global Forest Resources Assessment
GAM	Generalized additive model
GCP	Ground control point
GPS	Global Positioning System
GBM	Generalized boosting model
GIS	Geographical Information Systems
GLM	Generalized linear model
GRASP	Generalized Regression Analysis and Spatial Prediction
IDW	Inverse distance weighted
MARS	Multiple adaptive regression splines
MDA	Mixture discriminant analysis
MIR	Middle infrared
MODIS	Moderate Resolution Imaging Spectroradiometer
MS	Multispectral
NDVI	Normalized Difference Vegetation Index
NIR	Near infrared
NOAA	National Oceanic and Atmospheric Administration
ODBC	Open Database Connectivity
PAN	Panchromatic
RDBMS	Relational database management system
RF	Random forest
ROC	Receiver operator curve
RS	Remote sensing
SA	Selective Availability
SQL	Structured query language
SPOT	Satellite Pour l'Observation de la Terre
TIN	Triangulated Irregular Network
TIR	Thermal infrared
TWI	Topographic wetness index
UNEP	United Nations Environment Program
UML	Unified Modelling Language
WIST	Warehouse Inventory Search Tool

Contents

Preface

“The Long and Winding Road” would be best to describe the process creating this thesis. “Long” because it has taken some years to finally finish it and “Winding”, mainly because it has taken me to so many places both thematically and from a geographical point of view. This thesis contains various study themes in the fields of geoinformatics and it covers studies from three different study areas, one is located in Finland and two in Africa, in addition one study is made at a world–wide analysis scale. For unknown strange reason, my destiny has pointed my way back to my study areas in Finland and Africa.

In 1996 I went for the first time to Katima Mulilo (17°30'11"S, 24°16'26"E) in East Caprivi Namibia as a Master’s student in development geography, taking part in FINNIDA’s Forest Fire Control Pilot Project in the East Caprivi region. Later on, in 2004 I just passed East Caprivi on a bus trip from Windhoek Namibia to Livingstone in Zambia, but in 2007 I made my return to East Caprivi, this time not *in situ*, but from my working desk as my interest to understand the fires in Africa, and especially in East Caprivi, arose again. This time I was not using development geography methods but instead I used remote sensing, GIS and statistical techniques to predict fire occurrences in East Caprivi region. In 1998 I started working with *Melitaea cinxia* butterfly geospatial database at the Metapopulation Research group (MRG) lead by Professor Ilkka Hanski. Our study area was in Åland Islands, which meant that I made a return again. I returned to the landscape of my youth as I spent some summers in a small Island called Grytgrundet (19°59'22"E, 59°59'17"N) in southern Åland Islands. During my 3 ½ years in MRG I came very familiar with Åland Islands landscape, especially with those dry meadows habitat patches suitable for *Melitaea cinxia* butterfly that we were delineating using a Global Position System (GPS) method. My latest work position at the Department of Geosciences and Geography has taken me several times to Taita Hills in Kenya and again I made a kind of a return to my study area. In my sabbatical leave in 2004, we took a rough bus ride from Moshi in Tanzania to Mombasa in Kenya. During the trip our bus travelled through West Tsavo National Park and the first stop after crossing the West Tsavo plains was to be made in Mwatate. Before the short stop at the Mwatate tremendous rock wall of Bura Cliff (3°29'18"S, 38°20'42"E) could be seen and at that point I remembered that Finnish geographers, some of them my old fellow students, might be up on the Taita Hills making research. It took me less than two years from that incidence and I found myself working with Taita Hills geospatial data at the Department of Geosciences and Geography, University of Helsinki.

In my life I have had the great fortune to spend long periods in the tropics and subtropics from the heat of South East Asia to the colds of the High Altiplano and down to the burning savanna plains of Africa. On these trips over the last 25 years I have travelled in almost 90 countries and I have been able to witness the beauty of our Planet but unfortunately also the rapid environmental degradation that has happened in less than 25 years. Environmental degradation is proceeding with frightening speed and I deeply believe that something can, and has to be done, to stop environmental degradation. This thesis is my small contribution towards trying to protect our Planet as I believe that we need more information of the abiotic, biotic and anthropogenic factors and processes that are causing environmental degradation. We need more understanding of the diversity, distribution, and dynamics of species like Scopulini moths and the *Melitaea cinxia* butterfly and we need to develop methodologies for accurate estimation of forest cover change. We need more precise methods for up-to-date human population estimation and we need to create more explicit models of fire occurrence and burned area estimations

Geoinformatics and geospatial data sets have developed considerably in the last two decades and with the aid of GIS and remote sensing methods we can now create complex geospatial databases and build complex geospatial models. However, all too often geospatial data sets, study results and information we scientist are creating in our First World offices, unfortunately stay hidden in our premises. We are more interested to top up our publication list and to gain personal credits than actually sharing the knowledge we have created. The sharing culture is still poor, especially when the knowledge should flow from the “Rich North” to less developed countries in the south. It is therefore important to realise that if we create useful geospatial databases or geospatial models, we also have to share them. This is especially true if we have made our studies in Third World Countries. By saying this, I mean that, I don’t want my East Caprivi geospatial fire database or ARC/INFO AML macros to be buried with my crashing computer when the hard drive stops working and my external backup hard drive fails to function. I want to share my data and my ARC/INFO AML macros. For this reason, ARC/INFO AML macros and some other code to automate GIS processing can be freely copied from the Appendix section. I want also to share as much of my methodological geoinformatics knowledge as it is possible in an academic dissertation. My way of thinking might arise from my long GIS teaching experience and in this thesis I want to emphasize the importance of the geoinformatics methodologies and therefore methodological frameworks are presented in detail in section 4. By emphasising geoinformatics methodologies in this thesis, I also hope that whoever is reading this could use similar kinds of geoinformatics methodologies for his/her environmental modelling that are used in this thesis.

I strongly believe that we need to take into account abiotic, biotic and anthropogenic factors and processes behind environmental problems before we can resolve them. We need to have a broad knowledge for all these study fields, not just one. We need to have a holistic understanding for environmental problems and we need to combine more than just one entity to form something new — we need synthesis. And because geography is a discipline of synthesis; *“There has always been a view of geography as a discipline of synthesis. Holism has been there the whole time…”* (Holt–Jensen 1988), I therefore strongly believe that geographers can contribute fighting against environmental problems in a very special way as we geographers have the skill to see the “The Whole Picture”.

1. Introduction

1.1 The need for geospatial environmental modelling

We are on the edge – Earth is experiencing tremendous environmental problems at global, regional and local scales. According to the UN Intergovernmental Panel on Climate Change (IPCC) there is currently high scientific agreement that human-induced climate change is unequivocal and operating as an accelerator for combination of climate change associated environmental disturbances such as flooding, droughts, melting of snow and ice, rising sea level and wildfires. Other human actions such as land use change, landscape fragmentation and overexploitation of natural resources are acting as drivers for climate change. Therefore, there is a great need for action to save our environment from destruction. During the last few decades the development of remote sensing and GIS software has been fast, allowing for more sophisticated geospatial environmental applications to be developed. Geospatial environmental modelling using remote sensing, GIS and more recently spatial statistical techniques have deepened our understanding of the factors, mechanisms, and driving forces behind environmental destruction. However, we need to account for abiotic, biotic and anthropogenic factors and we need more broadminded and visionary approaches and modelling applications to understand environmental problems in a holistic way.

The need for accurate forest cover change analysis at a local scale

Deforestation continues at an alarming speed. About 13 million hectares were lost per year from 2000 to 2005, mainly to conversion for agricultural land. Globally ca. 36% of forest are primary forest, i.e. forests of native species, and about 6 million hectares of these forests has been lost or modified each year since 1990 (FAO 2006). However, during this same period from 2000 to 2005, natural expansion of forests and forest plantations affected 5.7 million hectares per year, resulting in an annual net loss of total forest cover estimated at 7.3 million hectares (FAO 2006). Furthermore, forests are acting as vital carbon sinks and it is estimated that the World's forests store 283 gigatonnes of carbon in their biomass and that this stock decreases by 1.1.gigatonnes annually because of deforestation and forest degradation. The largest net loss of forests occurs in South America (ca. 4.3 million hectares annually), followed by Africa (ca. 4.0 million hectares annually) (FAO 2006). It is therefore clear that deforestation, especially in the tropics, has become a major concern, not only because of the impact on the global carbon budget, but also due to the loss of natural habitats and biodiversity (Melillo *et al.* 1996).

The need to improve fire prediction and burned area estimation

Millions of hectares of forests are burned annually. Tansey *et al.* (2004) used SPOT VEGETATION satellite data and estimated vegetation fires to affect 350 million hectares in 2000, whereas Roy *et al.* (2008) used MODIS satellite data to assess the globally burned area to be 360 million hectares from July 2001 to July 2002. Warming climate may have profound impacts on global fires. Fires occurrence and area burned may increase and prolonged fire seasons and more extreme fire events are expected (Stocks *et al.* 1998; Hennessy *et al.* 2006; Flannigan *et al.* 2009). Greenhouse gases, such as CO_2 released by human activity are emitted into the atmosphere causing the earth to reach higher temperatures. It has been estimated that biomass burning contributes up to 40% of the annual CO_2 released into the atmosphere by human activities (Levine 1996; Dweyer *et al.* 2000) and therefore it is important to be able to have an accurate estimate of the global burned biomass. However, great uncertainty still exists in these calculations and one of the most critical inputs for these models is a precise estimation of burned areas. Different remote sensing methodologies and algorithms have been developed during the last few decades to achieve precise burned area estimation, but still there are some issues that have to be solved before truly accurate burned area information can be achieved. At the local and regional level, forest fire managers need more knowledge of active fire locations and areas that might be under fire risk. Fire prone areas can be identified by using remote sensing methods and preventive actions can be targeted to these core fire risk areas.

The need for species diversity and distribution models

Tropical forest loss, degradation and fragmentation is regarded as the major cause of the current biodiversity crisis (Andrén 1994; Achard *et al.* 2002). Tropical forest loss affects also climate change, which is also accelerating biodiversity loss. According to IPCC (2007) it is estimated if global average temperature increases exceed 1.5 to 2.5°C, there are projected to be major changes in ecosystem structure and function, e.g. in species ecological interaction and species geographical ranges. Approximately 20 to 30% of plant and animal species are likely to be at risk of extinction with the presented climate change scenario. Global warming also represents perhaps the most serious threat to biodiversity and according to Malcolm *et al.* (2006) rising temperatures could have a dramatic impact for the species richness of 25 "biodiversity hotspots". The worst affected areas include the Cape floristic region of South Africa, Southwest Australia, Caribbean, Indo-Burma, Mediterranean Basin and Tropical Andes. For these areas extinction of hundreds, even thousands of hotspot endemic plant and vertebrate species are expected under present climate scenarios. It is therefore important to understand the present global species diversity and distribution.

Habitat fragmentation and habitat loss can have profound effects on species at regional and local scale. In Åland Islands, south-western Finland habitat patches suitable for *Melitaea cinxia* butterfly have declined or degraded due to overgrowth of meadows and pastures, construction of roads and buildings. This has increased the extinction risk of *Melitaea cinxia* butterfly (Hanski 1998). It is therefore important to understand the effects of habitat fragmentation and loss to the local species. Sound spatio-temporal geospatial databases are needed to study species–environment interaction in real fragmented landscapes and for implementing realistic models and theories such as metapopulation theory using spatially explicit geospatial data.

The need for human population occurrence and abundance prediction and mapping

The rapid population growth and increasing need for natural resources are the main sources of environmental degradation (Cropper & Griffiths 1994; Preston 1996). According to the UN, world population reached 6.5 billion in 2005 and the population of the world is expected to increase from ca. 6.5 billion today to 9.1 billion in 2050. The largest population growth is expected in Third World countries having significant social, economic, and environmental consequences on these regions. Spatially explicit up-to-date information of population distribution and abundance are needed for these areas because the traditional nationwide population census is an error-prone, tedious operation and often undertaken only with a 10-year interval. In addition, demographic information is usually provided in national or administrative units and these sub-national reference units can be vastly different in size and shape (Li & Weng 2005; Mubareka *et al.* 2008). For spatial analysis it is often preferable to record human population estimates in regular analysis grids such as Landscan (Dobson *et* *al.* 2000). However, the problem for existing population models are their coarse scale e.g. Landscan 1 km^2, thus generalizing and obscuring the internal variability of population data. Therefore at a local scale cost-efficient applications to create spatially explicit human population geospatial databases and distribution maps at finer scales are in great demand.

1.2 Aims of the study

The aim of the study is to develop, describe, analyse and interpret GIS/RS based geospatial data modelling techniques. This thesis contributes to geospatial data processing and analysis by presenting advanced geospatial applications exploring novel methodologies and analysis techniques in the context of environmental modelling. This thesis covers five important study themes: (i) designing and developing environmental geospatial databases; (ii) studying species diversity and distribution using geospatial database and GIS techniques; (iii) analysing spatiotemporal forest cover change using remotely sensed data; (iv) developing a method for human population occurrence and abundance modelling using environmental geospatial data; and (v) developing predictive modelling techniques for improvement of fire occurrence and burned areas estimation. The constituting themes and Papers (I–VI) are summarized in Figure 1.

Paper (I) presents geospatial database development and analysis methodologies for ecological applications. Advanced database creation techniques were used to create a geospatial database for the *Melitaea cinxia* butterfly in the Åland Islands, south-western Finland. The database was implemented in MS Access database management system and the database was compiled from three main data sources: (i) from bi-annual field survey data for the *Melitaea cinxia* butterfly; (ii) from *Melitaea cinxia* habitat patch GPS-delineation data; and (iii) from climate data obtained from historical archives and from radar precipitation data. By using the geospatial database various spatial analyses was conducted.

Paper (II) examines temporal and spatial patterns of the species descriptions, diversity and geographical distribution for Scopulini moths at a world-wide scale using GIS. Species data was collected from multiple sources and statistical methods and GIS techniques were used for diversity and geographical distribution analyses

Paper (III) describes indigenous and exotic species spatiotemporal forest cover change analysis methodologies using remotely sensed airborne digital camera imagery for Taita Hills in Kenya. Airborne colour image mosaics for 2004, black and white aerial photographs for 1955 and 1994, and field survey data for 2007 and 2008, in addition with ancillary data from Taita Hills geospatial database were used for forest cover change detection analysis at a local scale.

Paper (IV) presents a predictive modelling methodology to estimate human population distribution and abundance in the mountainous rural area of Taita Hills, Kenya. Dwelling units were interpreted and digitized from airborne image mosaics and presence-absence of dwelling units in 100 m analysis grid was used as a response variable. Remote sensing-based (reflectance, texture and land cover) and geospatial (topography, climate and distance) data from the Taita Hills geospatial database were used as predictors. Generalized additive models (GAMs) were used to relate the dwelling units to remotely sensed and geospatial predictor data. Human population abundance models were compared with Kenyan population census data for 1999 and with two existing global coarse scale population estimate grids; Gridded Population of the World and LandScan 2005, respectively.

Paper (V) employs analysis techniques to improve knowledge for fire occurrence and burned area estimation at a regional scale in East Caprivi, Namibia. First fire hot spots occurrence was analyzed and spatial autocorrelation for fires investigated using Moran's *I* correlograms. Generalized linear models (GLMs) were used to relate MODIS hot spots fire data to geospatial predictors. Separate fire probability models were calibrated for abiotic, biotic, anthropogenic and combined predictors and autocovariate variable was tested for model improvement. Model performance was evaluated using area under the curve (AUC) from the receiver operating characteristic (ROC) plot and the hierarchical partitioning method was used to determine independent effects of explanatory variables. The predicted probability surfaces were translated into a burned area presence–absence classification maps for comparison analysis with Moderate Resolution Imaging Spectroradiometer MODIS burned area product (MCD45A1) data.

In Paper (VI) eight state-of-the-art predictive modelling techniques were tested for fire prediction, burned area estimation and fire risk mapping improvement relative to the existing MODIS burned area fire product (MCD45A1) in East Caprivi, Namibia. MODIS hot spots fire data was related to geospatial predictors using the BIOMOD modelling framework (Thuiller 2003; Thuiller 2009). Models were evaluated using AUC and Kappa statistics and the predicted probability surfaces were translated into a burned area presence-absence classification maps for comparison analysis with the MODIS (MCD45A1) burned area product.

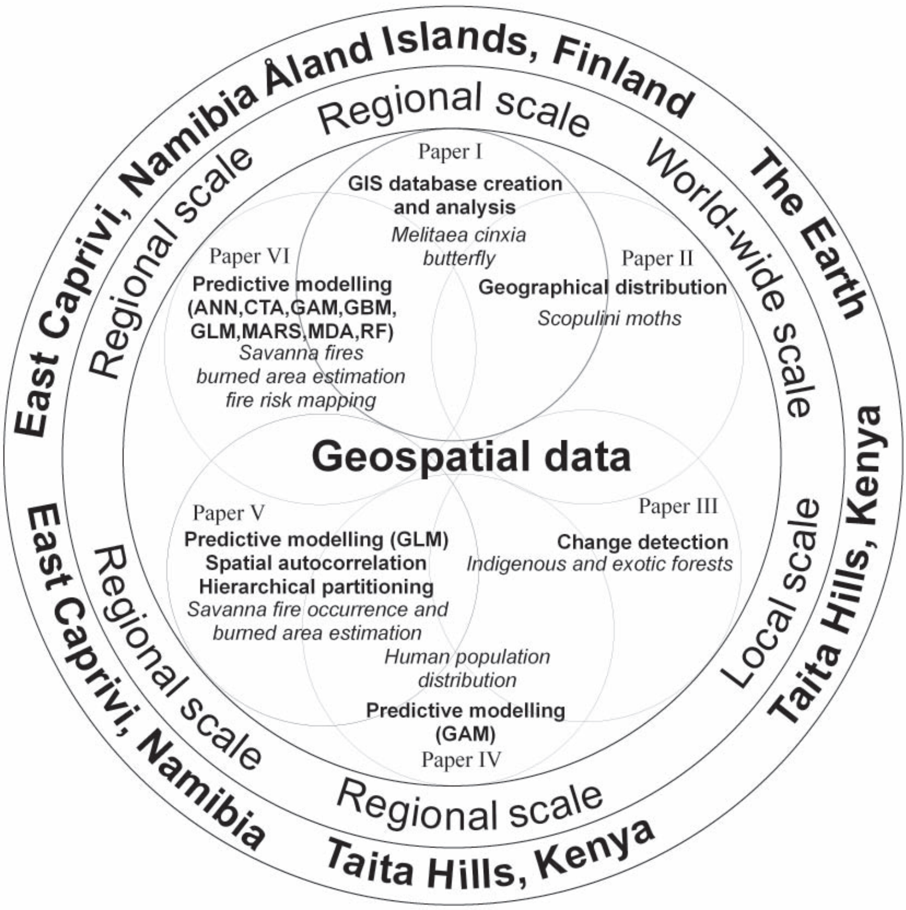


Figure 1. *Summary of study areas, analysis scales, analysis themes and methods for Papers I–VI.*

2. Theoretical and methodological background

2.1 Geospatial databases – design and development

Before any spatial analysis can be conducted a geospatial database has to be designed and developed. The first process is database design where the contents of the intended database are identified and described. According to Elmasri & Navathe (1994) database design can be divided into three major phases: (i) conceptual data modelling, where the data content is identified and described at an abstract, or conceptual, level; (ii) logical database design, where conceptual database design is transformed into the data model of a specific software system; and (iii) the physical design phase, where the data model is represented in the schema or semantic data model of the software. In this phase, issues such as data storage structures and indexing are determined. Laurini & Thompson (1992) named the final phase as *internal design*, a phase that provides a basis for data model implementation. In geospatial modelling the entity-relationship (E-R) modelling technique, developed by Chen in 1976, has gained popularity and it has been extremely effective over a wide range of applications. In this approach entities are the relevant objects for the database and in a GIS, an entity is any object that can be localised spatially. Attributes or physical characteristics of each object are attached to the entities. Between entities and attributes different kinds of spatially referenced relations or mechanisms can exist, e.g. 'located in' and 'contained in'.

All GIS software has in-built database management system (DBMS) capabilities such as structured query language (SQL) properties. However, GIS software cannot handle complex relationships and therefore present day geospatial databases are managed from relational database management systems (RDBMS). These commercially available or Open Source database management software programs can be programmed to perform these same tasks, e.g. database queries outside of the GIS or RDBMS can be connected through Open Database Connectivity (ODBC) to GIS software. The structure of geospatial databases can be very simple, consisting of only a few entities and attributes, or they can be very large systems dealing with highly structured geographic data and complex spatial relationships. In this thesis, six different geospatial databases were designed and developed but only one is described in details as it was the main aim in Paper (I) to design and develop a geospatial database for the *Melitaea cinxia* butterfly. The structure of databases ranged from very simple for the Scopulini moth Paper (II), to a rather complicated one for the *Melitaea cinxia* butterfly Paper (I), which consisted of many entities and attributes and relations. A logical flowchart of geospatial database analytical operations and data models in different levels of abstractions can be seen from Figure 2.

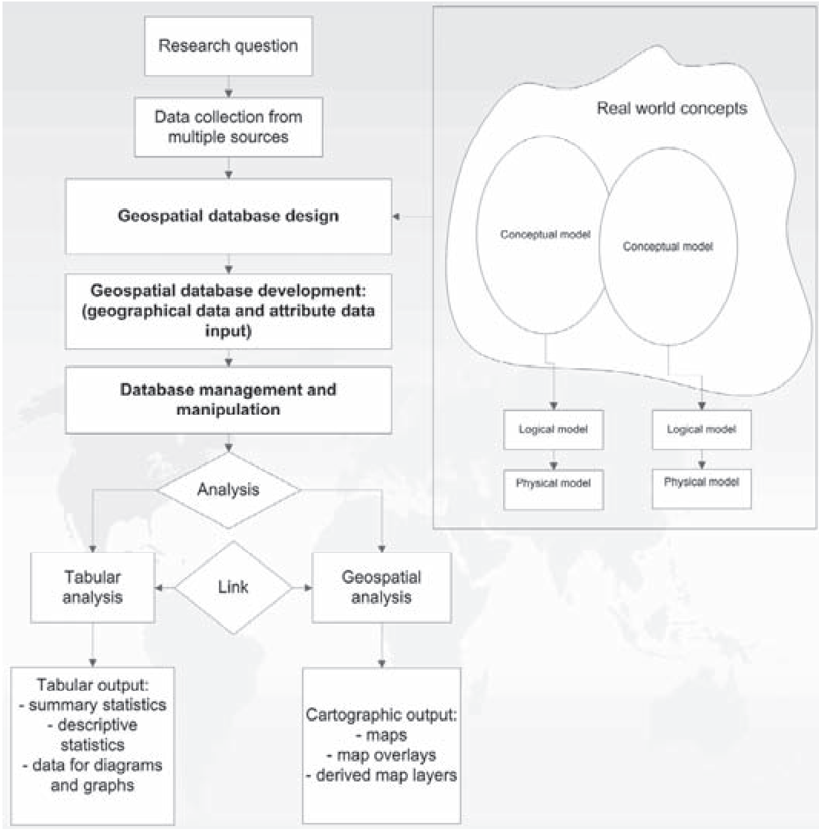


Figure 2. *A logical flowchart of geospatial database analytical operations (left hand side) and data models in different levels of abstraction (right hand side).*

2.2 Geographical diversity and distribution modelling and mapping for species data

For centuries, biologists and ecologists have been preoccupied with the question of species distribution. Early studies such as von Humboldt & Bonpland (1805) tried to explain why species are present in some places and absent from the others? In the 1960s, some iconoclastic theories have been developed such as the theory of island biogeography by MacArthur & Wilson (1967) explaining that the number of species increases with increasing area, or latitudinal gradient theory of diversity represented by Fischer (1961) where he showed that the number of species decline moving away from the tropical to extratropical areas. However, notable exceptions exist to the general patterns of species distribution and these anomalous representations may be dependent on such things as spatial scale or taxonomic hierarchy or elevational gradients (Beck & Chey 2008).

The term metapopulation was fist introduced in a paper by Richard Levins (1969) and it was a starting point for a gradual paradigm shift and emergence of a new metapopulation theory and empirical research. A 'metapopulation' is a "population of populations" (Levins 1969); in which distinct subpopulations (local populations) occupy spatially separated patches of habitat. The habitat patches exist within a matrix of unsuitable space, but organism movement among patches does occur, and interaction among subpopulations maintains the metapopulation. Levins's classical metapopulation theory considers that species persistence in the landscape depends on a turnover of extinction-(re)colonisation of suitable habitat patches at each generation and metapopulation dynamics are approximated by binary changes in the state of individual patches. However, real metapopulations do not consist of identical and equally connected populations, as is assumed in the basic models. Therefore, more recent empirical metapopulation models emphasize the importance of connectivity and isolation of populations in fragmented landscapes, ending up with the development of models like the Incidence Function Model (Hanski 1998).

The long-term research on the Glanville fritillary butterfly (*Melitaea cinxia*) in the Åland Islands south-western Finland was started in 1991 and soon it could be shown that *Melitaea cinxia* was an ideal modelling system for metapopulation studies. Since 1991 it has been the basis for a number of new empirical models in the context of metapopulation biology and ecology (see e.g; Hanski 1998; Hanski & Ovaskainen 2000). Creation of the *Melitaea cinxia* geospatial database started in 1998, which will be described in more detail in sections 3.1 and 4.1. and in Paper I. In short, the main idea was to incorporate existing *Melitaea cinxia* field survey data, climatic data and GPS delimited habitat patch data to geospatial database.

In the last decade, with the rise of advanced ecological applications using GIS and remote sensing, a new paradigm of quantitative, spatial species distribution modelling has emerged. At present, overwhelming numbers of predictive species distribution applications exist at global, regional and local scales. For a comprehensive review of the species distribution methodologies utilizing remote sensing and GIS techniques refer, for example, to Guisan & Zimmermann (2000) or Rushton *et al.* (2004). These new studies have given some cogent results determining the driving biotic and abiotic factors for species distribution at global, regional and local scales. Some of the most important biotic factors that affect diversity, distribution and abundance of species have been shown to be dispersal, habitat selection, competition, predation, parasitism and mutualistic interactions with other species. Many abiotic factors such as temperature, precipitation, potential evapotranspiration and incoming solar radiation, are limiting diversity, distribution and abundance of species. Species distribution models have become very sophisticated but at the same time very complex and hard to interpret. Therefore, in Paper II a more traditional species distribution modelling perspective was chosen to describe various temporal and spatial patterns of the species descriptions, diversity and geographic distribution.

2.3 Forest cover monitoring using satellite data and airborne imagery

Initially, remotely sensed forest cover maps were produced at a local scale using visual interpretation of black and white aerial photographs, but the commencement of the Landsat program in the 1970s enabled the first use of high- to moderate-resolution satellite imagery for forest cover mapping over large areas (Woodcock *et al.* 2002). Since its launch in 1972 Landsat satellite platforms (Landsat MSS, Landsat TM and Landsat ETM+), have been the most applied high- to moderate-resolution satellite imagery for forest cover monitoring (Fuller 2006). SPOT imagery has also been used to analyse land use change (e.g. Clark & Pellikka 2009). Coarse-resolution sensors have been utilized for global- and regional scale forest cover monitoring. At a coarse-scale, NOAA AVHRR (Advanced Very High Resolution Radiometer) satellite images with 1 km spatial resolution were used, for example, in the TREES project to create a map for tropical forest cover at scale of 1:5 000 000 (Malingreau *et al.* 1995; Mayaux & Lambin 1995, 1997). More recently other sensors such as Moderate Resolution Imaging Spectroradiometer (MODIS) and SPOT Vegetation sensor on board SPOT–4 and SPOT–5 satellites are replacing AVHRR to monitor forest cover in regional- and global scale because these sensors has improved geolocation and calibration capabilities relative to AVHRR sensors (Fuller 2006). Very high-resolution ≤ 5 m satellite imagery such as Ikonos, QuickBird and GeoEye–1 has gained some popularity recently for forest cover change mapping, but purchase costs are high and these images covers only a relatively small area, e.g. an Ikonos scene covers 11.3 ×11.3 km. Therefore, still at the present, Landsat is the most popular imagery for forest cover monitoring and change analysis at a regional scale.

A common feature for all the sensors described above is that they cannot penetrate clouds, thus hindering the acquisition of cloud free images, especially when monitoring tropical forests at low latitudes with persistent cloud cover. In addition, high- and moderate resolution satellites suffer from low temporal resolution. Therefore, radar imagery, for example from the Japanese Earth Resources Satellite (JERS–1) and European Remote Sensing Satellite (ERS), has been tested as an alternative method for forest monitoring (Sgrenzaroli *et al.* 2002; Podest & Saatchi 2002). However, it has been hard to interpret landscapes from radar images compared to imagery from the more traditional optical sensors, such as Landsat.

Due to the obstacles space-borne satellites might present, visual aerial photograph image interpretation can still be regarded as the most adequate and accurate remotely sensed method for forest cover change analysis at local and regional scale. Traditionally, aerial imagery has been the black and white photographs recorded on light-sensitive films but fairly recently the benefits of using digital aerial photographs has been recognized. There are some obvious advantages when digital airborne camera data is used: (i) the camera data is in a digital format allowing the automation of procedures during the whole processing chain from image data acquisition to image correction; and (ii) if needed, image manipulations are fairly easy to make for digital format images. However, digital camera image acquisition and processing is still a quite new field and therefore a consistent, repeatable methodological framework has to be developed suitable for forest cover change analysis. EnsoMOSAIC digital camera system represents a semi-automatic digital image processing technique which was used for a forest cover change study in Taita Hills (Paper III).

2.4 Human population estimation using geospatial data

Human population growth is one of the main threats to the world's environment. Since the start of the Industrial Revolution population has rapidly grown and environmental degradation has proceeded at an accelerated speed. Population growth has had profound social, economic, and environmental consequences and, at present, a great concern has emerged with the growing number of population increasing per capita greenhouse gases in to the atmosphere thus accelerating the global warming. To cope with population growth and the related environmental degradation it is therefore important to have precise estimates of the number of population and spatially explicit information of the human population distribution (Sutton *et al.* 1997). In developed countries the number of population and the distribution is well know and computerized, but in many Third World countries the traditional census methods are still used. Population censuses are found to be time-consuming, costly, error-prone, difficult to update and the census interval is often too long for many types of applications (Li & Weng 2005). Therefore, in addition to conventional census calculations other methods, such as aerial photo interpretation, have been used. Aerial photography has been the traditional airborne method to estimate population and to map the population distribution at local or regional scales. Large-scale aerial photographs have been used since the 1950s for dwelling unit counting, e.g. Porter (1956) used a rural dwelling unit count in Liberia, whereas Lo (1986) was able to estimate population for the city of Athens in Georgia U.S. using aerial photographs at the census tract level with high accuracy. However, the dwelling unit count method is a time-consuming, expensive process and it requires abundant high resolution aerial photographs to cover large areas (Lo 1989).

Orbital remote sensing has been considered as an adequate option for population estimation. Satisfactory population estimations results have been achieved despite the limitations of spatial resolutions of the sensors used, for example Landsat and SPOT (e.g. Forster 1983; Lo 1995). A number of studies have combined orbital remote sensing data with regression techniques for population estimation. Forster (1983) used Landsat Multispectral Scanner (MSS) data and developed a multi regression equation with standard deviation of separate bands and various reflectance ratios, to predict housing density and Lo (1995) linked spectral radiance values of image pixels with residential densities for the Kowloon metropolitan area of Hong Kong using a multispectral SPOT image. In addition to local and regional scale population estimations there have been a few attempts at global scale population estimation using geospatial data modelling to create gridded population models. The most used global population models are: Gridded Population of the World (GPWv3) at 5 km resolution (CIESIN 2005); LandScan at 1 km resolution (Dobson *et al.* 2000) and United Nations Environment Programme/Global Resource Information Database at 1–degree resolution (UNEP/GRID 2006), respectively. However, existing global population data have some serious issues, e.g. too coarse resolution generalize and obscure the internal variability of population when considered for use at a local or regional scale. Therefore, Paper IV concentrates on developing an enhanced probability based application that can be used for population distribution mapping and for population abundance estimation at a local or regional scale.

2.5 Fire detection and prediction methods using geospatial data

Fires burn million of hectares each year globally and fires play a very important role in ecosystems. There is scientific agreement that a significant fraction of trace gases and greenhouse gasses emitted into the atmosphere are originated from fires (Fuller 2000; Korontzi *et al.* 2004; Lentile *et al.* 2006). It has been estimated that biomass burning accounts up to 40% of annual anthropogenic CO_2 emissions and ca. one-quarter of anthropogenic particulate matter (Levine, 1996). Fire affects not only the atmosphere but also the terrestrial environment, as biomass burning is a major driving force in anthropogenic land-cover change, especially in the tropics. Moreover, burning has been proved to be a significant factor in altering vegetation cover which in turn may lead to land degradation and loss of biodiversity (Eva & Lambin, 2000; Fuller 2000; Lentile *et al.* 2006; Clark & Bobbe 2007). Therefore, there has been a great need to discover methodologies to analyse fire frequency and extent. Over the last few decades remote sensing has been successfully used for active fire detection and for fire-scar detection, thus increasing our understanding of fire frequency and extent at global to local scales. In addition, some studies have been conducted to predict fire occurrence. Table 1 summarizes some of the current satellite sensors that are used for active fire detection and fire-scar detection.

Table 1. *Characteristics of satellite sensors used for active fire and fire-scar detection (Modified after Fuller 2000).*

Sensor	Major applications	Spatial resolution	Swath width	Bands (µm range)	Major advantages	Major limitations
AVHRR	Active fires, Fire scars	1.1 km	2400 km	0.58–0.68 0.72–1.10 3.55–3.93 10.3–11.3 11.5–12.5	Widely available, low cost, high temporal Frequency	325 K saturation in channel 3
DMSP-OLS	Active fires	0.56 km 2.07 km	3000 km	0.58–0.91 10.3–12.9	High sensitivity; high temporal Frequency	Night-time use only during low lunar illumination
SPOT-4	Fire scars	10 m PAN from 0.61 to 0.68 µm 20 m MS	60 km	0.50–0.59 0.61–0.68 0.79–0.89 1.58–1.75	High spatial resolution, MIR band	Low temporal frequency, low area coverage, high costs
SPOT vegetation	Fire scars	1 km	2000 km	0.43–0.47 0.61–0.68 0.78–0.89 1.58–1.75	MIR band, large areas covered, high temporal Resolution	Unknown
Landsat (TM and ETM+)	Fire scars	15 m PAN 30 m MS	185 km	0.45–0.52 0.52–0.60 0.63–0.69 0.76–0.90 1.55–1.75 10.4–12.5 2.08–2.35	MIR band, high spatial resolution, well-known data source Used as fire scar reference data	Low temporal resolution, high cost
GOES-8	Active fires	1 km (visible) 4 km (infrared)	Hemisphere	0.52–0.72 3.78–4.03 6.47–7.2 10.2–11.2 11.5–12.5	Very high Temporal Resolution	Coarse spatial resolution, 3.9 µm band saturates at 335 K
ATSR	Fire scars	1 km	500 km	3.51–3.89 1.57–1.63 10.4–11.3 11.5–12.5	Good spectral Configuration for fire-scar Mapping	Unknown
MODIS	Active fires, Fire scars	250 m 500 m, 1km	2330 km	36 bands Including 3.9 and 11 µm	Saturation of 450 K at 4 µm and 400 K at 11 µm	Unknown
Ikonos	Fire scars	1 m PAN 4 m MS	11.3 km	0.440–0.516 0.506–0.595 0.632–0.698 0.757–0.853	Used as fire scar reference data	Low temporal resolution, high cost
QuickBird	Fire scars	61 cm PAN 2.4 m MS	16.5 km	0.450–0.520 0.520–0.600 0.630–0.690 0.760–0.900	Used as fire scar reference data	Low temporal resolution, high cost

Active fire detection, also known as hotspot detection, is based on the detection of the middle infrared (MIR) and thermal infrared (TIR) radiation emitted by fires (Qu *et al.* 2008). Advanced Very High Resolution Radiometer (AVHRR) on board (NOAA) satellites has been used in several studies since the 1980s for active fire detection even though the sensor was not designed for that purpose (see e.g. Setzer & Malingreau 1996; Eva & Lambin 1998a: Dwyer *et al.* 2000; Eva & Lambin, 2000; Fuller 2000). AVHRR has mainly been used for active fire detection at a broad-scale because of the coarse resolution (1 km). More recently, researchers have turned to using Moderate Resolution Imaging Spectroradiometer (MODIS) satellite data for active fire detection (see e.g. Justice *et al.* 2002; Roy *et al.* 2002, 2005, 2008; Miettinen *et al.* 2007). Since its launce in 1999 MODIS onboard Terra satellite, and MODIS onboard Aqua satellite since 2002, have been the most popular satellite platforms for fire scar studies. MODIS system extends the active fire detection and fire scar mapping capabilities of AVHRR and, unlike AVHRR, MODIS has specially designed channels (band 21 and 22) that are suitable for fire detection (Justice *et al.* 2002). At the present there are various other space-borne coarse scale satellites suitable for active fire detection, e.g. Geostationary Operational Environmental Satellite (GOES), the Defense Meteorological Satellite Program Operational Linescan System (DMSP-OLS), Visible and Infrared Scanner (VIRS), onboard the Tropical Rainfall Measuring Mission (TRMM) satellite, and European Remote Sensing Satellite Along Track Scanning Radiometer (ATSR) and its enhanced version (AATSR). For a full treatment of remotely sensed active fire detection issues, e.g. space-borne sensors, algorithms, products and fire detection methods, refer to Qu *et al.* (2008).

Fire scar detection (burn scar detection) is a method where burnt areas are identified and delineated using their spectral signature and/or fire-induced spectral changes. Over the last two decades a wide range of satellite data has been utilized for fire scar detection. At a coarse resolution (≥ 1 km), AVHRR has been used in the past for broad-scale burned area mapping and various algorithms and methods have been produced (see for example Kaufman *et al.* 1990; Barbosa *et al.* 1998; Barbosa *et al.* 1999; Roy *et al.* 1999; Fuller & Fulk 2001). Since 1999 MODIS has gained more attention and various burned area estimation studies at a moderate resolution (250–500 m) have been conducted since that time (see for example Roy *et al.* 2002, 2005, 2008; Sá *et al.* 2003; Chuvieco *et al.* 2005). Other satellite platforms, such as SPOT VEGETATION (1 km resolution), have also been utilized (see Stroppiana *et al*. 2002, 2003; Zhang *et al.* 2003; Silva *et al.* 2004). There are also a few radar based fire scar detection studies (e.g. Bourgeau-Chavez *et al.* 2002; Gimeno & San-Miguel-Ayanz 2004).

At a high resolution, Landsat TM and ETM+ images has been broadly used. Landsat has been used to derive reference burned area when comparison has been made with coarse or moderate resolution satellite imagery (see e.g. Eva & Lambin 1998a, 1998b). Landsat images allow for the measurement of the extent of burnt areas and the proportions of burnt surface and this information can used to estimate the effects of fires. The start of new millennium has brought several very high resolution space-borne satellites such as Ikonos and GeoEye–1 with resolution ranging from 0.5 meters to 4 m. With these images it is possible to have a very precise estimate of burned areas and validate the errors that occur in burned area estimations when using coarse or moderate resolution images. However, Landsat and very high resolution satellite data suffer from low temporal resolution and the costs for very high resolution images are still high and the areal coverage is very small.

Recently, various custom-made burned area products have been developed. The MODIS Burnt Area Product (MCD45A1) is a monthly Level 3 gridded 500 m product containing per-pixel burning and quality information, whereas GLOBCARBON (BAE) burned area data at 1 km resolution is based on three satellite sensors: ATSR2, AATSR and SPOT VEGETATION respectively. Another Global burned area product, L3JRC utilizes SPOT VEGETATION satellite images for burn area estimation using a single algorithm for burned area detection at 1 km resolution (Tansey *et al.* 2008).

However, the utilization of coarse resolution satellite images for fire detection raises some issues that need to be considered. Fire studies that have compared coarse resolution and high resolution reference fire data have found that apparent omission and commission errors exists (see Eva & Lambin 1998a; Boschetti *et al.* 2004; Laris 2005). Omission error occurs due to various reasons like: persistent cloud cover that hampers fire detection; difference between the time of fire occurrence and satellite overpass, and too coarse pixel size to detect small patchy fires (Eva & Lambin 1998a; Hawbaker *et al.* 2008; Roy *et al.* 2008). On the other hand, commission errors are caused by non-fire surfaces that are highly reflective, e.g. bare soil, water or cloud, or where sharp contrast exist, e.g. between desert and vegetation (Hawbaker *et al.* 2008). To overcome these problems, fire occurrence probability models have been developed. The few examples that exist mostly use regression techniques to relate fires to explanatory factors; for example Koutsias & Karteris (2000) used independent explanatory variables consisting of the spectral channels of the post-fire satellite image for logistic regression analysis to determine burned and unburned areas in Attica in Greece, whereas Lozano *et al.* (2007) used several spectral indices derived from multi-temporal Landsat data for fire occurrence probability modelling, and Preisler *et al.* (2004) utilized weather data as explanatory factors for probability based fire risk models. However, most of the regression models that have been calibrated for fire do not take into account the spatial autocorrelation effect of fires, even though it has been found to be a very important factor in probability based fire models (see Chou *et al.* 1993; Lynch *et al.* 2006). Papers V and VI explore methodologies to develop new applications for enhanced predictive fire occurrence modelling and burned area estimation in East Caprivi Namibia.

3. Studied areas and data sets

This thesis covers case studies from four different study areas ranging from the global scale to the local scale (Figure 3). Paper (I) concentrates on the Åland Islands in south-western Finland. Paper II presents a study made at a world-wide scale. Papers III and IV concentrate on Taita Hills, southeast Kenya and Papers V and VI presents case studies from East Caprivi, Namibia. The locations of these study areas can be seen from Figure 3.

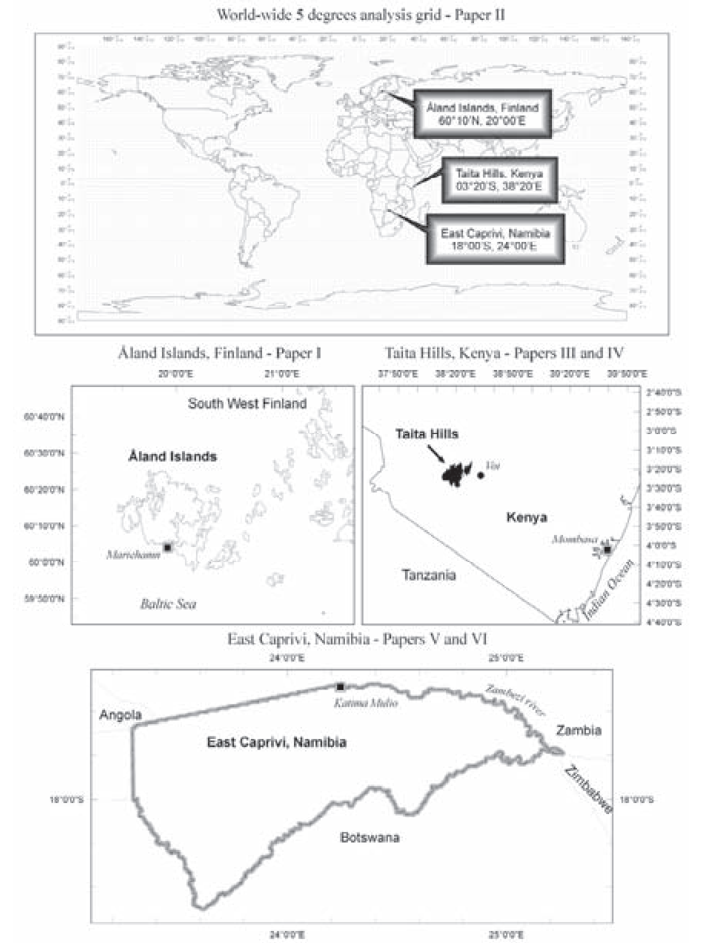


Figure 3. Locations of the studied areas. The roman numbers refer to the papers in the thesis.

3.1 Glanville fritillary *Melitaea cinxia* metapopulation data set, Åland Islands, Finland

The study area in Åland archipelago in south-western Finland covers approximately 50 × 70 km area (Hanski & Meyke 2005). The Landscape for the Åland mainland and other islands studied consists of heterogeneous landscape mosaics with forests, small farms, cultivated fields, pastures and meadows (Hanski *et al.* 1994, 1995, 1996). The long-term and large-scale metapopulation project on the Glanville fritillary butterfly (*Melitaea cinxia*) in the Åland Islands was started in 1991 to test of some metapopulation models (Hanski 1999). Since then, a large scale metapopulation data set for Glanville fritillary has been collected yielding to a database on a network of more than 4000 small habitat patches in the Åland Islands (Figure 4). The Glanville fritillary occurs in a highly fragmented landscape and has a classic metapopulation structure in the Åland Islands (Moilanen & Hanski 1998). The suitable habitat patches are dry meadows with one or both of the larval host plants, *Plantago lanceolata* and *Veronica spicata* (Hanski 1999). The knowledge about the occupancy of the habitat patches in the entire patch network is gathered bi-annually by surveying all patches with the aid of a large group of students. Of the roughly 4000 habitat patches in the entire study area, about 500–700 patches (small dry meadows) are occupied in any given year (Hanski 1999).

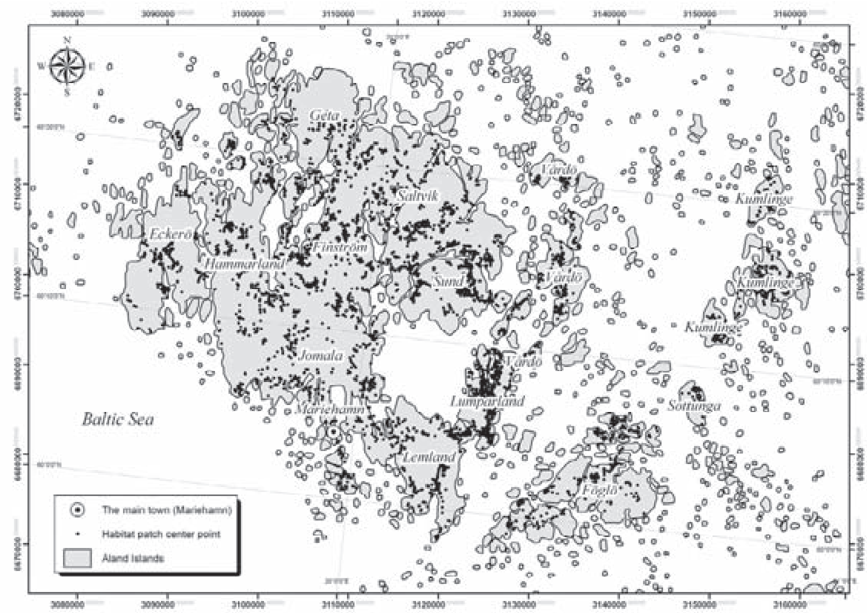


Figure 4. *Åland Islands and the distribution of Melitaea cinxia habitat patches presented as black dots.*

3.2 Scopulini moths data set at a world–wide analysis scale

The Scopulini moths were studied at a world-wide scale using a computerized database as the basis of the study. In ArcView 3.2 GIS an Avenue script (see Appendix 2) was used to divide the World into 5 degree equal-area grid squares in a geographic coordinate system (WGS84). Species data was geocoded using the information of type localities and further a geospatial database was created using the MS Access relational database management system to relate species information to 5 degree equal-area grid squares. MS Access was used because of the lack of suitable GIS software with support for internal geospatial database capabilities such as ArcGIS Geodatabase. The creation and content of Scopulini geospatial database is discussed in more detailed in methods section 4.2. The uppermost study area location map in Figure 3 shows the Scopulini moth analysis 5 degree equal-area grid squares at a world-wide scale.

3.3 Taita Hills highlands, Kenya

Taita Hills is located in the Taita-Taveta District of south-eastern Kenya at 03°25'S, 38°20'E. The study area covers ca. 327 km² of the Taita Hills highlands and the boundary to distinguish highland from surrounding semi-arid shrubland and dry savannah lowlands was set to 1100 m.a.s.l in Paper (IV) (see Figure 5). The average elevation of the Taita Hills is 1500 m and the highest peak is Vuria at 2208 m. The climate is influenced by the Inter-Tropical Convergence Zone (ITCZ) yielding a bi-modal rainfall incidence were long rainy season are expected during March to June and short rains in November–December. Annual precipitation in Taita Hills varies from 600 mm on the plains to over 1200 mm in the hills (Beentje 1988) and mist and cloud precipitation is usually a year-around phenomenon in the hills. Local climate is influenced by orographic rainfalls bringing more rain to south eastern slopes causing a distinctive 'rain shadow' effect on north western slopes of Taita Hills. *Euphorbia* *candelabrum* and more commonly *Euphorbia bussei var. kibwezensis* are growing in the drier conditions. Cloud forest fragments are found in areas receiving more than 900 mm of annual precipitation and being above 1400 m on the south eastern slopes, and above 1700 m on the north western slopes (Jaetzold & Schmidt 1983).

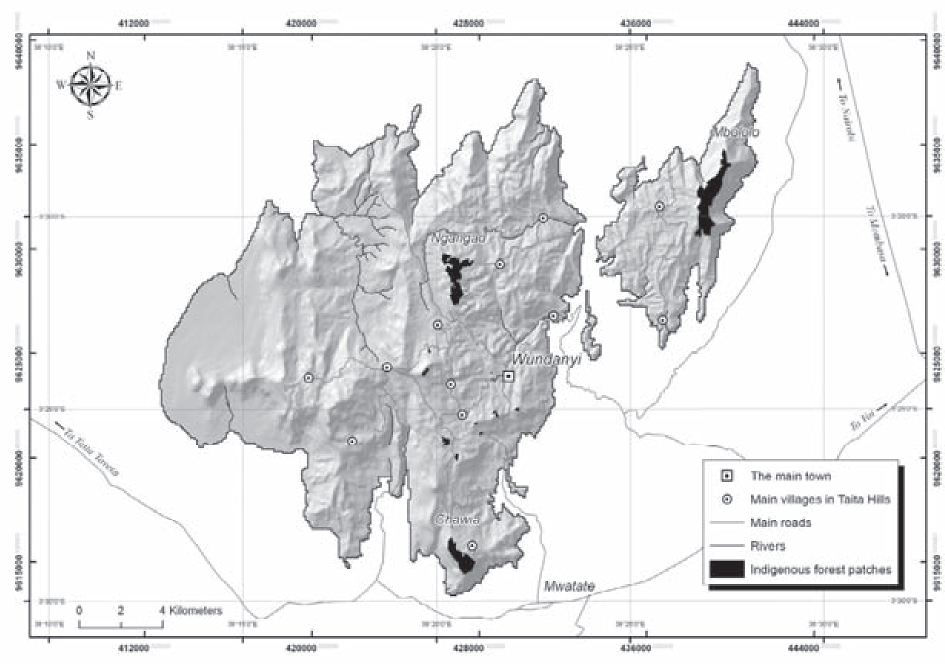


Figure 5. *Map of Taita Hills showing the main geographical features and indigenous forest patches.*

Due to the favourable climatic conditions Taita Hills are intensively cultivated and the landscape mosaic is constructed of abundant small-scale farms, exotic tree species patches such as *Cupressus lusitanica*, *Pinus spp*., *Eucalyptus spp*., and *Grevillea robusta* and indigenous forest patches. Only three larger hilltop indigenous forest remnants exist: Mbololo (ca. 179 ha), Ngangao (ca. 136 ha) and Chawia (ca. 94 ha) (Lens *et al.* 2002). In addition, much smaller indigenous forest patches are embedded in the landscape mosaic and they are scattered around the hills. The characteristic indigenous tree species include: *Newtonia buchananii*, *Tabernaemontana stapfiana*, *Macaranga conglomerata*, *Albizia gummifera*, *Phoenix reclinata*, *Strombosia scheffleri*, *Cola greenwayi*, *Podocarpus spp*., *Ochna holstii*, and *Millettia oblate* (Beentje 1988). Most people live on small farmlands in the hills at elevations between 1000 and 1700 m.a.s.l., where the rainfall is abundant and the temperature is more tolerable than in the lowlands. In the Taita Hills there is only one town – Wundanyi - with ca. 5000 inhabitants, and then smaller villages. Population in the Taita-Taveta district has grown from 60 000 in the 1960s to 246 671 in the last census of 1999 (Republic of Kenya 2001).

3.3.1 Data set for spatiotemporal forest cover change detection

Since the start of the Taita project in 2003 a large geospatial database has been collected. This data set consists mainly of digital maps, satellite images and aerial photographs acquired from different sources (Broberg & Keskinen 2004). In Paper (III) aerial photographs were used as the main data for analysing spatiotemporal forest cover change in Taita Hills. Black and white aerial photography for 1955 was obtained from Survey of Kenya and colour airborne digital camera data acquired during January 2004 (Pellikka *et al.* 2004). Airborne digital camera data was acquired using a custom true-colour NIKON D1X digital camera system along with a GPS navigation system and accompanying software (Holm *et al.* 1999). In addition, one black and white aerial photograph for year 1994 was obtained from the Survey of Kenya and used for Yale forest cover change analysis. GIS derived geospatial map layers such as road network, hydrography, administrative units and a digital elevation model (DEM) were used as ancillary data.

3.3.2 Data set for human population prediction modelling in Taita Hills

Taita Hills geospatial database was used as the main data source for the human population occurrence and abundance modelling (Paper IV). Dwelling units were mapped using on-screen digitization from airborne digital camera data acquired during January 2004 covering 30% of Taita Hills highlands. A SPOT 4 HRVIR 1 satellite image (15/10/2003, path & row 143-357) with 20 m pixel size was used to derive surface reflectance-, image texture- and land cover based predictors. A DEM with 20 m pixel size was derived from Survey of Kenya 1:50 000 scale topographic mapping (Clark & Pellikka 2005, 2009). The DEM was used to calculate the mean elevation, slope and aspect. The topographical wetness index (ω) was derived utilizing a custom-made ArcGIS geoprocessing model. In addition, irradiance (kWh/m^2/month) was calculated from the DEM using an ARC/INFO AML macro (shortwavc.aml) (Kumar *et al.* 1997; Zimmermann 2000). Long term mean precipitation grid layers were interpolated from monthly available meteorological data in Taita Hills and surrounding areas. ANUSPLINE software used the DEM and meteorological data for interpolation (Hutchinson 1995; Erdogan *et al.* in press). Vector map layers for main roads and rivers layers, digitized from the Kenya 1:50 000 scale topographic maps, were used as the source for Euclidean distance grid calculations in ArcGIS 9.3. Two existing global population datasets, Gridded Population of the World (GPWv3) at 5 km resolution and Landscan 2005 (Dobson *et al.* 2000) at 1 km resolution, and Kenyan 1999 census data (Republic of Kenya 2001) were used for human population abundance model comparison.

3.4 East Caprivi, Namibia

The East Caprivi region is situated in North-Eastern Namibia (18°30’-17°28’S, 23°18’- 25°22 E) surrounded by Angola, Zambia, Zimbabwe and Botswana. East Caprivi lies between the Kwando River in the west, and the Zambezi and Chobe Rivers in the east. The East Caprivi region covers an area of approximately 12,000 km^2, and has the highest rainfall in Namibia, receiving 600±700 mm of rain a year, at an altitude of 930±1020 m (Mendelsohn & Roberts 1997). According to von Breitenbach (1968) three main physiographic and vegetation regions can be defined: (i) an elevated upland region predominantly tree-bush in the northwest of East Caprivi with typical Zambezi Teak (*Baikiaea plurijuga*) forests, (ii) a lower lying southern and south-eastern region with Mopane (*Colophospermum mopane*) forest and savanna and (iii) a marsh and swamp region (Figure 6). Human population density in the East Caprivi is second highest in Namibia and the number of human population was estimated at 73,982 in the 1996 census (Mendelsohn & Roberts 1997). The people in this area depend on subsistence farming and cattle herding. The climate is sub-tropical with mild dry winters from April to August and hot wet summers from September to March. The first rains are expected in October starting with light showers and peaking in January. From the January peak season rains are gradually weakening so that they usually finish in the end of March. Approximately 90% of yearly precipitation falls between September and March (von Breitenbach 1968; Devereux 1993). Highest temperatures are reached in September and October and the lowest temperatures are recorded in June and July (von Breitenbach 1968).

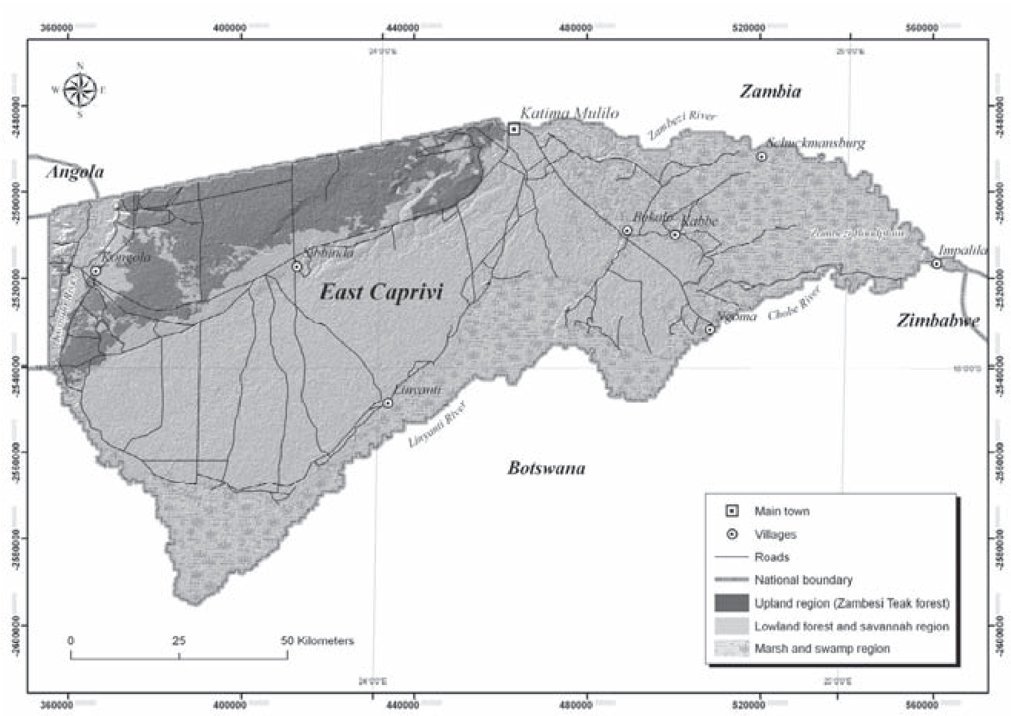


Figure 6. *A map of East Caprivi, Namibia showing the main geographic features and vegetation zones.*

3.4.1 East Caprivi geospatial data sets for fire prediction and burned area estimation

East Caprivi geospatial data sets used as predictors for fire occurrence and burned area estimation consist of vector GIS layers derived from the Caprivi database Internet page hosted by the Ministry of Environment and Tourism (Mendelsohn & Roberts 1997) and raster GIS layers derived from multiple sources. More detailed description of the geospatial data sets used in fire occurrence and burned area estimation is given in the original Papers V and IV. Moderate Resolution Imaging Spectroradiometer (MODIS) fire data were obtained from the Warehouse Inventory Search Tool (WIST). Two types of fire data were used: MODIS fire hot spots data (thermal anomalies/fire 8-day L3 global 1 km) for both Terra (MOD14A2) and Aqua (MYD14A2) satellites, and MODIS burned area product (MCD45A1). In addition a single Landsat ETM+ georeferenced satellite image (path 174, row 072; August 28, 2002) was obtained from Global Land Cover Facility and multiple MODIS 250 m images were obtained from the MODIS Rapid Response System Internet page. Landsat ETM+ and MODIS 250 m images were used in accuracy assessment for burned area estimation analysis as true *in situ* fire measurements were nonexistent for the studies.

4. Analysis methods and methodological frameworks

In this thesis, multiple analysis methods have been applied to various data sets and therefore all the details for geospatial database creation, GIS, remote sensing, and statistical methods are not described, except for Paper (I) because the details for database design and development are absent from the original paper. It is also appropriate to explain the database development in this thesis because the main emphasis in Paper I from a geoinformatics perspective was to describe the methodological aspects of the *Melitaea cinxia* GIS database. For each of the other studies a methodological framework is presented as flowcharts. The most important analysis methods used in this thesis can be seen from Table 2. The details for analysis methods are described in the original papers but, broadly, analysis methods used in this thesis can be divided into four main groups: (i) geospatial database creation (presented in detail for Paper I); (ii) spatial analysis for species diversity and distribution (Paper II); (iii) forest cover change detection methods (Paper III), and (iv) predictive modelling methods for occurrence and abundance mapping (Papers IV, V and VI).

Table 2. *The main methods used in this thesis. The Roman numbers refer to the original papers were a specific method was used.*

Analysis method	Main methodology	Used in paper
Areal change analysis	C	III, V, VI
Artificial neural networks (ANN)	D	VI
Change trajectory analysis	C	III
Classification tree analysis (CTA)	D	VI
Comparison analysis	C	IV, V, VI
Correlation analysis	D	IV, V, VI
Descriptive statistics	B, C, D	I, II, III, IV, V, VI
Distribution mapping	B, C, D	I, II, III, IV, V, VI
Exploratory data analysis	B, C, D	I, II, III, IV, V, VI
Generalized additive model (GAM)	D	IV, VI
Generalized boosting models (GBM)	D	VI
Generalized linear model (GLM)	D	V, VI
Geospatial RDBMS creation and maintenance	A	I (II, III, IV, V, VI)
GPS field work	A	I, III, IV
Hierarchical partitioning	D	V
Historical climate data retrieval and analysis	A	I
Image interpretation	C	III, IV
Image mosaicing	C	III, IV
Mixture discriminant analysis (MDA)	D	VI
Multiple adaptive regression splines (MARS)	D	VI
Multi-scale segmentation	C	IV
Precipitation estimation using weather radar data	A	I
Random forest (RF)	D	VI
Spatial autocorrelation	D	V
Spatial interpolation	D	IV, V, VI
Zonal statistics	D	IV, V, VI

A = Geographical databases; B = Geospatial analysis for distribution and diversity; C = Change detection; D = Predictive modelling

4.1 Geospatial database creation methodologies for the *Melitaea cinxia* butterfly

In Paper I, the geospatial database was created for the Glanville fritillary *Melitaea cinxia* butterfly. Data collecting for the species presence-absence and habitat patch characteristics for *Melitaea cinxia* has been ongoing since 1991. Field survey data are collected bi-annually in a way that each year ca. 4000 habitat patches are surveyed in autumn and those patches that were occupied in previous autumn are surveyed again in the next spring. Information from the field survey forms was manually entered in to the MS Access database. Each of the field survey forms had ca. 40 environmental entities and a hand written habitat patch map on the back of field form showing the outlining and locations of *Melitaea cinxia* larval nests, if any existed. The Access database consisted of data for bi-annual field surveys from 1993 to 2002. The most important part of the database consisted of presence-absence information for *Melitaea cinxia* butterfly larval nests and habitat characteristics for c.a. 4000 habitat patches and other data such as information of larval parasitoids. For modelling purposes, habitat patch central point (x-coordinate and y-coordinate in Finnish Uniform Coordinate System) was used as a spatial reference and the area for each habitat patch were estimated *in situ* during the field survey and from Basic maps.

To achieve more detailed information on habitat patches, a mapping project was started where GPS was used for delimiting the habitat patches with the information of surrounding habitat types. GPS field work for creating a geospatial database for *Melitaea cinxia* were conducted over three years. In each summer trained students outlined habitat patches in the field and the raw GPS data was later processed using PCGPS 3.6d2 software. First the raw GPS data in PC-GPS Feature File *.FTR format were imported from 3½-inch HD 1.44 MB floppy disks to PCGPS 3.6d2 software and for differential correction a correction data from Evo base station (http://gps-evo.hamk.fi/0001.htm) was used. Before the correction data could be used a conversion from the Pathfinder SSF format to the Receiver Independent Exchange Format (RINEX) format had to be made. Differential correction was necessary because of the intentional error signal of Selective Availability, until it was switched off in 2000. After the differential correction process, GPS data was exported to ESRI shapefile format. The necessary GIS operations to correct the data were processed in ArcView 3.2 software. Multiple GIS operations were used, e.g. connecting independent polylines to make closing lines in order to generate coherent polygons. Each of the 4000 patches was processed manually in GIS software.

The final database was created in the manner that first weekly outlined GPS patches were merged, then monthly and yearly patches were merged and finally all three years data were merged to create a full GIS database of suitable habitat patches for the *Melitaea cinxia* butterfly. This final polygon habitat database had the information of patch number, patch area and patch perimeter and landscape indices at patch level derived from area-perimeter ratio. In addition to the habitat polygon layer, another vector GIS layer was created as polylines. In this layer each of the lines had the information of habitat patch number and the surrounding land cover type, which were: forest, semi-open, field, other open, road and water, respectively (Figure 14). The GIS attribute tables for habitat patches were imported into the Access database and joined with the *Melitaea cinxia* field survey data using the Patch ID number as a common key field.

Two types of meteorological information were also collected. Firstly, historical meteorological data were collected from the Finnish Meteorological Institutes Meteorological yearbooks covering the years from 1885 to 2000 for the main weather stations in the Åland Islands. Both mean temperature and precipitation were recorded manually for June–August, first to notebooks and then to the Access database. Secondly, precipitation data was retrieved from the Finnish Meteorological Institute as raw radar data in ASCII text format for three summer months of 1998–2001. Each data file consisted of 12 hour precipitation values and x and y -coordinates in the Finnish Uniform Coordinate System for south-western Finland at a spatial resolution of 1 km^2. The data was first geocoded to ArcView 3.2 and further processed semi-automatically using Avenue scripts due to the vast information content (e.g. more than 256 columns and more than 65536 rows, which is the Excel limitations). One example of an Avenue script used in the precipitation radar data analysis can be seen in Appendix 1. The 1 km^2 spatial resolution radar precipitation data were masked to cover the entire Åland Islands and then imported into the Access database. The final *Melitaea cinxia* GIS database was then created by maintaining three types of information in the Access database: (i) patch level *Melitaea cinxia* field survey data; (ii) GPS information of habitat patches; and (iii) climatic data from yearbooks and radar. Using Access queries all the necessary information was derived from the database and imported through an Open Database Connectivity (ODBC) link to ArcView 3.2 software for spatial analysis and for map creation. For spatial analysis a 1 km^2 analysis grid was created using the Avenue script in ArcView 3.2 shown in Appendix 2. Habitat patch area information was summarized for each 1 km analysis square and also the monthly radar precipitation data were summed. The methodological framework for creating *Melitaea cinxia* geospatial database is presented as a flowchart in Figure 7.

Figure 7. *Flowchart for developing Melitaea cinxia geospatial database and the methodological steps of spatial analyses.*

4.2 Spatial analysis methods for diversity and distribution mapping

In Paper (II) various temporal, spatial, diversity and geographical distribution patterns of Scopulini moths were analysed. The methodological study design comprises patterns of species descriptions and relationships between the number of described species and their associated synonyms. The study design is similar to Gaston *et al.* (1995) and therefore enables a comparative analysis between these two studies. Eight variables, describing for example the rates of species description, synonymy rates, and geographical distribution of the Scopulini, were coded for all putatively valid species of Scopulini in 2003, and a database was created. The database is based on a preliminary world check-list of Scopulini, covering the world fauna, and it consists of 912 species. In addition, the *Geometrid Moths of the World – A Catalogue* (Scoble 1999) data was used for species that were not described in the world check-list of Scopulini.

More specifically, the following information was recorded in the database for each species: generic combination, author (multi-authored descriptions were counted under the senior author only), year of description, type locality (country), type locality (latitude and longitude), biogeographical region of type locality, number of synonyms, and type specimen depository. By following type localities, species were assigned to biogeographical regions after Gaston *et al.* (1995) to allow comparisons. Biogeographical regions are presented in Figure 8. All the eight variables were available for 792 species and type locality (country) was too inaccurate to be used in 11 cases, whereas the type locality (latitude and longitude) was too inaccurate in 25 cases. Furthermore, determining the biogeographical region for species was impossible in three occasions (ca. 0.3%) and type depository was unknown for 101 species. The database has ca. 9200 entries and in addition, numerous comments.

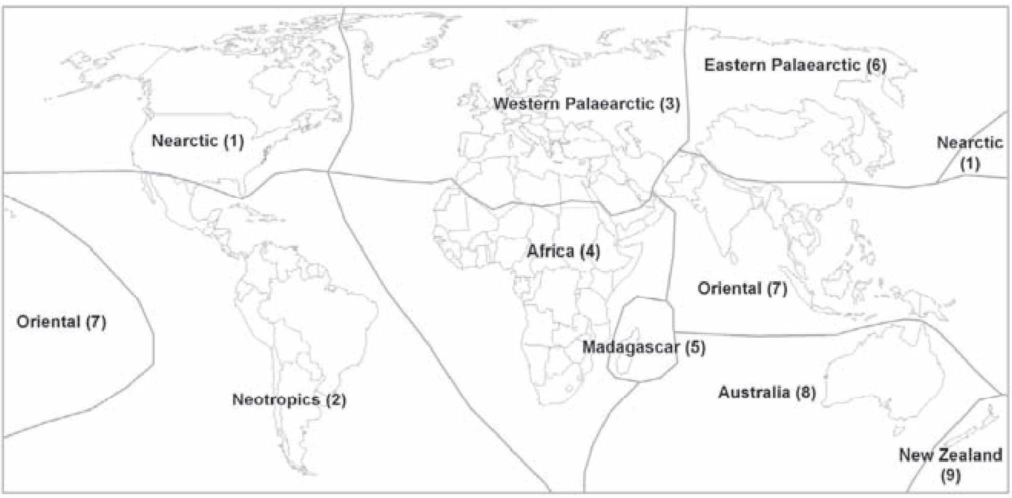


Figure 8. A map showing the boundaries of biogeographical regions (Modified after Gaston et al. 1995).

For geographical distribution analysis the species data included type localities information. i.e. latitude and longitude coordinates for species at an accuracy of 5 degrees. With the coordinate information of type localities, the species data could be imported into ArcView 3.2 software and geocoded to a geographic coordinate system with a WGS84 datum. A world-wide analysis grid with 5 degree analysis squares was created by using an ArcView

3.2 Avenue script (see Appendix 2). The frequency count of species in each analysis grid was calculated with the summarize function in ArcView 3.2 and the attribute data were exported to MS Access to develop a geospatial database. The data was then exported to a statistical program for statistical analyses and the species distribution maps were created in ArcView 3.2 GIS software. A flowchart of the methodological framework used to develop and analyse the Scopulini moth geographical database for diversity and distribution analysis can be seen from Figure 9.

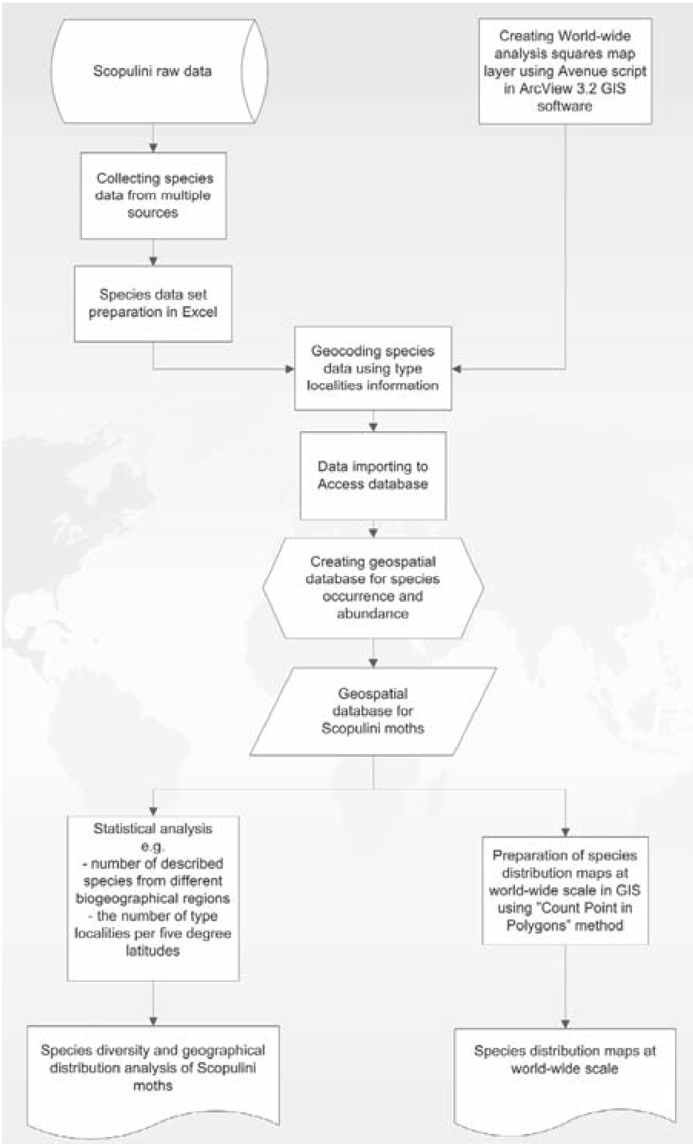


Figure 9. *A flowchart of methodological framework used for Scopulini moth diversity and distribution analysis.*

4.3 Analysis methods for spatiotemporal forest cover change detection

Paper (III) presents a spatiotemporal forest cover change analysis method for Taita Hills, Kenya. Airborne colour images and black and white aerial photography were used in analysis. Only the main analysis methods are presented here as the methodology section is described in detail in the original publication. Airborne colour images for 2004 were geometrically and spectrally corrected using EnsoMOSAIC software (StoraEnso 2003). This software covers the whole processing chain from flight planning to producing geo-referenced and ortho-rectified images and image mosaics (Holm *et al.* 1999). Black and white aerial photograph for years 1955 and 1994 was obtained from the Survey of Kenya. Images were co-registered using 2004 airborne imageries as reference data and co-referenced using the rubber sheeting method in ERDAS Imagine. GIS derived geospatial map layers were used as ancillary data.

For forest cover change analysis, a land cover model was created using on-screen digitizing in GIS software. Images for the year 2004 were first interpreted and digitized to four forest classes and into eight other land cover classes. Visual interpretation was based on image colour and texture and on the previous regional knowledge and it followed the LCCS class structure (Di Gregorio 2005). For black and white 1955 and 1994 images there was less forest classes interpreted. Table 1 and 2 in the original Paper III shows the forest classes and the land cover classes used in the analysis. Digitizing results for 2004 and 1994 images were verified by field survey in 2007 and 2008 and the digitizing was fine-tuned according to the field check. Three main parameters were used from the final land cover model: (i) area of the main indigenous forest fragment; (ii) the total area of the indigenous forest within the area analyzed; and (iii) total forest area including exotic forests and bushlands. For the change detection three calculations were conducted: (i) areal change; (ii) percentage change; and (iii) change trajectory analysis where the changes occurring from indigenous forest class to another classes, and from non-forest classes to indigenous or exotic forests were analyzed. Figure 10 presents the methodological steps used in the forest cover change analysis.

Figure 10. *A flowchart of the methodology adopted for spatiotemporal change analyses for indigenous and exotic forest cover between 1955–2004.*

4.4 Predictive modelling methods for geospatial data

Predictive modelling techniques were used for dwelling unit distribution and abundance analysis in Taita Hills (Paper IV) and for fire occurrence prediction and burned area estimation improvement for East Caprivi, Namibia in Papers V and VI.

4.4.1 Statistical predictive modelling calibration techniques

Linear regression techniques have been traditionally used in predictive modeling, e.g. in ecological and biogeographical research (Guisan & Zimmermann 2000). The basic linear regression model has the form:

$$Y = \alpha + X^T\beta + \varepsilon \qquad (1)$$

where Y is the response variable, α is a constant called the intercept and $X = (X_1, ..., X_p)$ is a vector of p predictor variables, $\beta = \{\beta_1, ..., \beta_p\}$ is the vector of p regression coefficients (one for each predictor), and ε is the error. However, when using linear regression four assumptions should be met: (i) linearity of the relationship between dependent and independent variables; (ii) independence of the errors; (iii) homoscedasticity (constant variance) of the errors and (iv) normality of the error distribution (Zar 1999). Violation of one or more of the multiple linear regression assumptions may lead to incorrect or misleading results. These assumptions are seldom achieved when using geospatial data sets. To overcome these violations new prediction methods have been introduced. These methods allow non-Gaussian error distributions and non-linear relationships between response and predictor variables (Guisan & Zimmermann 2000). In this thesis eight novel predictive modelling techniques were used: Generalized Linear Model (GLM) (McCullagh & Nelder 1989), Generalized Additive Model (GAM) (Hastie & Tibshirani 1990), Classification Tree Analysis (CTA) (Breiman *et al.* 1984), Artificial Neural Networks (ANN) (Ripley 1996), Multivariate Adaptive Regression Splines (MARS) (Friedman 1991), Mixture Discriminant Analysis (MDA) (Hastie *et al.* 1994), Generalized Boosting Models (GBM) (Friedman 2001), and Random Forest (RF) (Breiman 2001). Short descriptions of different predictive techniques are presented here as they are explained in more details in the original Papers IV, V and VI.

- Generalized linear model (GLM)

Generalized linear models (GLMs) are an extension of classical multivariate linear regression, allowing non-normal response variables to be modelled (McCullagh & Nelder 1989). In GLMs, the predictor variables X_j (j = 1, ..., p) are combined to produce a linear predictor LP which is related to the expected value μ=$E(Y)$ of the response variable Y through a link function $g()$, such as:

$$g(E(Y)) = LP = \alpha + X^T\beta, \tag{2}$$

where α, X, β are those described in equation (1). The model is now written for the generic variables X and Y; the corresponding terms for the ith observation in the sample is:

$$g(\mu_i) = \alpha + \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_p x_{ip} \tag{3}$$

Unlike classical linear models, which presuppose a Gaussian (i.e., normal) distribution and an identity link, the distribution of Y in a GLMs may be any of the exponential family distributions (e.g., Gaussian, Poisson or binomial) and the link function may be any monotonic differentiable function (like logarithm or logit). GLMs do not force data into unnatural scales and therefore allow non-linearity and non-constant variance structure in the data (McCullagh & Nelder 1989; Collet, 2003). In Papers V and VI GLM models were built using a full stepwise approach, in which explanatory variables are included or excluded from the full model using Akaike Information Criterion (AIC) (Akaike 1974) and changes in scaled deviance (McCullagh & Nelder 1989; Venables & Ripley 2002).

- Generalized additive model (GAM)

Generalized additive models GAMs (Hastie & Tibshirani 1987; Yee & Mitchell 1991) support non-Gaussian error distributions and non-linear relationships between response and predictor variables. GAMs are non-parametric extensions of GLMs model regressions that apply nonparametric smoothers to each predictor and additively calculate the component response. GAMs are data-driven rather than model driven and allow consideration of more complex response shapes than those possible through GLMs (Yee & Mitchell 1991). GAMs were used in Papers IV and VI. A GAM model is expressed by

$$g(E(Y)) = \alpha + s_1(X_{1i}) + s2(X_{2i}) + \ldots s_p(X_{pi}) \quad (4)$$

where g is the link function that relates the linear predictor with the expected value of the response variable Y, X_{pi} is a predictor variable and s_p a smoothing function. The response variable was linked to the set of predictor variables through a logit link function for dwelling unit occurrence and log link function for dwelling unit abundance in Paper IV. In Paper VI a logit link function was used for fire occurrence prediction.

- Classification Tree Analysis (CTA)

Classification and regression trees explain variation of a single response variable by one or more explanatory variables providing an alternative to regression techniques (Thuiller *et al.* 2003). The CTA method consists of recursive partitions of the dimensional space defined by the predictors into groups that are as homogeneous as possible in terms of response. The partition divides the data in an iterative manner into smaller groups with binary split. A tree is built by repeatedly splitting the data based on thresholds for individual explanatory variables (Breiman *et al.* 1984).

- Artificial Neural Networks (ANN)

Feed forward neural networks belong to machine learning techniques and provide a flexible way to achieve generalize linear regression functions (Venables & Ripley 2002). The backpropagation algorithm by Rumelhart *et al.* (1986) is used in layered feed-forward ANN models. In feed forward ANN models, artificial neurons are organized in layers, and send their signals "forward", and then the errors are propagated backwards. Then the network receives inputs by neurons in the input layer, and the output of the network is given by the neurons on an output layer. In the model one or more intermediate hidden layers can parameterized. Cross-validation is normally used to avoid the model overfitting and as different runs can provide different results, a common procedure is to make N-fold model runs and average the results in the final model.

- Multivariate Adaptive Regression Splines (MARS)

Multivariate adaptive regression splines (MARS) is a multivariate non-linear regression method that combines linear regression, mathematical construction of splines and binary recursive partitioning to produce a local model in which relationships between response and predictors that are either linear or nonlinear (Friedman 1991). In MARS models, the amount of smoothing required for each predictor as well as the interaction order of the predictors are automatically selected.

- Mixture Discriminant Analysis (MDA)

MDA is a classification method and can be viewed as an extension of linear discriminant analysis (LDA) (Venables & Ripley 2002). MDA allows the classifier to handle different prototype classes such as a mixture of Gaussians. The mixture of normals is used to obtain a density of estimation for each class (Hastie *et al.* 1994; Fraley & Raftery 2002). For optimal scaling process different regression methods can be used. R-BIOMOD modelling framework uses MARS to increase the predictive power of the models (Thuiller 2009).

- Generalized Boosting Models (GBM)

GBM is a non-parametric, highly efficient modelling method based on the Gradient Boosting Machine algorithm developed by Friedman (2001). A boosted regression tree (BRT) algorithm fits a large number of relatively simple models whose predictions are then combined to give more robust estimates of the response that classifies binary responses (e.g. fire presence-absence) where each of the individual models consists of a simple classification or regression tree; i.e. a rule based classifier that consists of recursive partitions of the dimensional space defined by the predictors into groups that are as homogeneous as possible in terms of response. The tree is built iteratively by repeatedly splitting the data, defined by a simple rule based on a single predictor. At each split, the data are partitioned into two exclusive groups, each of which is as homogeneous as possible (Ridgeway 1999; Friedman 2001). In the first step predictors are the input to the first regression tree and thereafter for each step the focus is on the residuals. At the second step a tree is fitted to the residuals of the first tree and the model is then updated to contain two trees, and the residuals from these two trees are calculated and the sequence is repeated for as long as necessary. In the modelling the maximum number of trees can be set, e.g. to 2000–3000 (Elith *et al.* 2008).

- Random Forest (RF)

Random forest (RF) is an ensemble machine-learning algorithm (Breiman 2001). RF generates hundreds of random classification trees by using both bagging and random variable selection for tree building. Rather than using all predictors and all individual data points to make a single tree, RF makes a forest of many trees, each one based on a random selection of predictors and individuals (Breiman 1996). Each tree is grown with a randomized subset of predictors and fitted using a bootstrap sample of data and grown utilising CART methodology to the largest extent possible. Each node is then split using the best among a subset of predictors randomly chosen at that node. Prediction is then made from the complete forest based on a majority vote of the prediction of all random variable trees (Breiman, 2001).

4.4.2 Evaluation of predictive models

In this thesis predictive models were evaluated as follows: (i) by using the percentage of explained deviance as an indicator of model explanatory power (D^2) (Paper IV). This is obtained by dividing the difference between null and residual deviance by the null deviance (Guisan & Zimmerman 2000); (ii) using the area under the curve (AUC) from the receiver operating characteristic plot to indicate the model predictive power (ROC, Fielding & Bell 1997). As a general rule, an AUC between 0.5 and 0.7 indicates a poor discriminate capacity; 0.7-0.9 indicates reasonable capacity; and 0.9 or higher indicates a very good capacity (Swets 1998). AUC were used in Papers IV, V and VI; (iii) conducting a five-fold cross-validation of area under the curve (AUC) from ROC plot (CVROC) Paper (IV); (iv) calculating cross-validated Cohen's Kappa (Papers IV and VI); (v) calculating the contribution for each predictor, giving an indication of the contribution of the variable within the selected model and corresponding to the possible range of variation on the scale of the linear predictor (Paper IV); (vi) visual interpretation of prediction maps (Papers IV, V, VI).

4.4.3 Predictive modelling for human population studies in Taita Hills

In Paper (IV) human population distribution and abundance in Taita Hills were modelled using Generalized Regression Analysis and the Spatial Prediction (GRASP) modelling framework which uses generalized additive models (GAMs) for model calibration (Lehmann *et al.* 2002). The response variable (dwelling unit presence–absence) was derived from airborne imagery covering ca. 30% of Taita Hills using on-screen digitizing in GIS. Geospatial GIS- and remote sensing-based map layers were used as predictors. Prior to modelling, the full data set (n = 10488, 100 m analysis squares) was randomly divided into model calibration 70% (n = 7342) and model evaluation 30% (n = 3146) datasets following the split sample approach (Guisan & Zimmermann 2000). The outcome of dwelling unit prediction models were extrapolated to cover the whole Taita Hills area (n = 34143 100 m analysis squares) and a human population distribution map was created for Taita Hills semi-automatically. In addition, human population abundance model was compared with two existing global population datasets, GPWv3 and LandScan 2005, and Kenyan census data for 1999. Only the main steps for predictive modelling are presented here. For more detailed description for the methodologies used in Taita Hills dwelling unit occurrence and abundance prediction refer to the original Paper (IV). Figure 11 shows the methodological flowchart of the modelling steps.

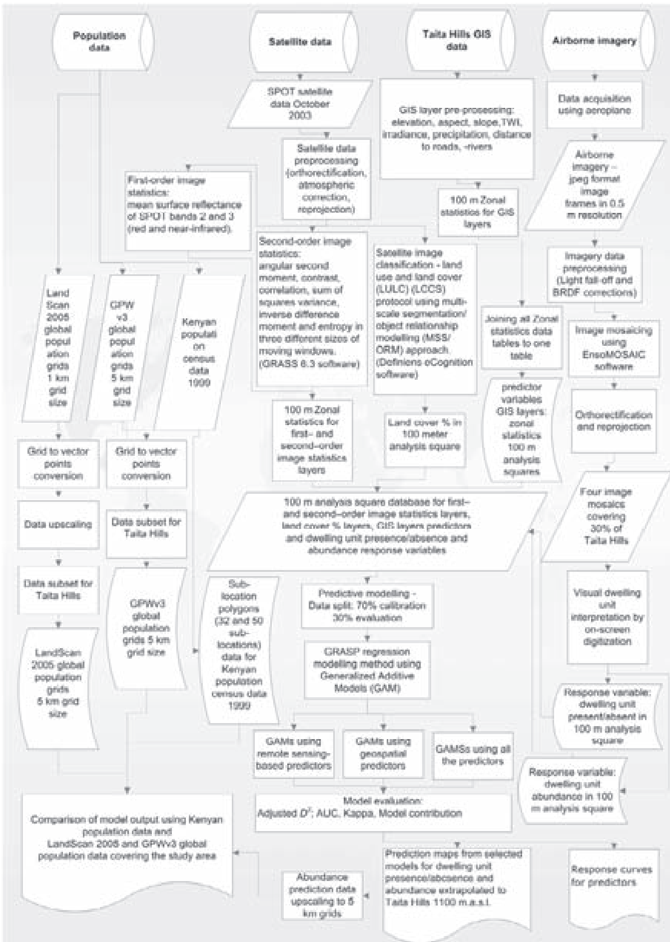


Figure 11. *Methodological flowchart of the modelling steps for human population distribution and abundance modelling.*

4.4.4 Fire prediction and burned area estimation in East Caprivi

An application for fire prediction and burned area estimation was developed for East Caprivi, Namibia. A methodological flowchart to create the application can be seen from Figure 12.

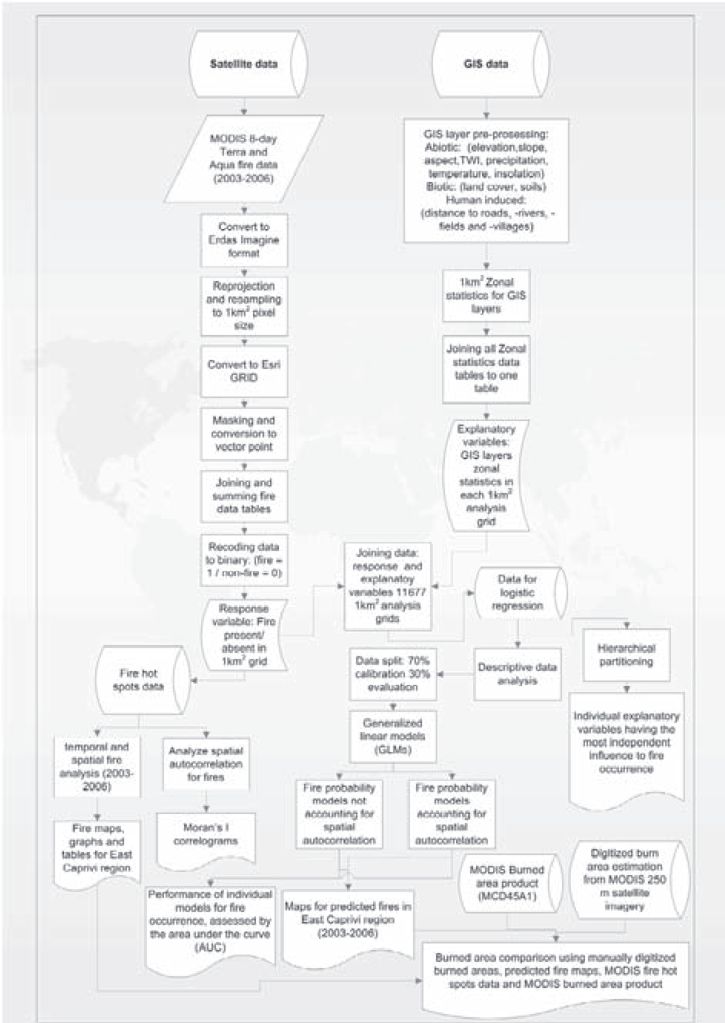


Figure 12. *Methodological flowchart for fire occurrence prediction and burned area estimation application for East Caprivi, Namibia.*

The first step was to create the geospatial database and prior predictive fire occurrence model calibration, spatiotemporal characteristics of fires were analysed using fire hot spots frequency counts. In addition, spatial autocorrelation of fires was tested with Moran's *I* correlograms. In Paper (V) fire occurrence was predicted using generalized linear models (GLMs) with binominal error distribution and logit link functions. Fire probability models were calibrated by using MODIS fire hot spots data as binary (0/1) response variable data and abiotic, biotic and anthropogenic factors as exploratory variable data. Prior to modelling, data was randomly divided into a model calibration set (n = 8167) and a model evaluation set (n =

3510). GLM models were built using a full stepwise approach, where variables were included or excluded from the full model using Akaike's Information Criterion (AIC) (Akaike, 1974) and changes in scaled deviance (McCullagh & Nelder 1989; Venables & Ripley 2002). Models were developed for all four years 2003–2006 and for an aggregated fire occurrence (2003–2006). Four different explanatory variable sets (abiotic, biotic, anthropogenic and combined) were tested and two types of models were constructed: models with and without consideration of the spatial autocorrelation of fires. Models were evaluated using area under the curve (AUC) from the receiver operating characteristic (ROC) plot. Hierarchical partitioning (HP) method was used as a complementary analysis to GLMs to identify individual explanatory variables having the most independent influence on response variable, in this case, fire occurrence (MacNally 1996, 2000; Walsh & MacNally 2003). Predicted probability surfaces were translated into burned area presence-absence classification maps using threshold cut-off values determined with PresenceAbsence R package function *MinROCdist* (Freeman 2007; Freeman & Moisen 2008). Burned area classification maps were then compared with the MODIS burned area product (MCD45A1) and MODIS fire hot spots data.

4.4.5 Fire prediction and burned area estimation in East Caprivi using eight techniques

In Paper (VI) the BIOMOD computation framework (Thuiller 2003; Thuiller *et al.* 2009) was used for fire occurrence prediction. BIOMOD enables the use of up to eight state-of-the-art predictive modelling techniques; GLM, GAM, MARS, CTA, MDA, ANN, GBM and RF, respectively (Thuiller *et al.* 2009). Fire hot spots presence–absence in 1 km analysis squares was used as a response variable and geospatial environmental data were used as predictors. Data was first randomly divided into calibration (70%) and evaluation (30%) data sets and final prediction models were built for full data set (100%). Models were evaluated using cross-validated area under the curve AUC and with cross-validated Kappa values. Predicted probability surfaces were translated into burned area presence-absence classification maps using optimized threshold cut-off values determined with Kappa statistics. Classified burned area maps were compared with the MODIS burned area product (MCD45A1) and MODIS fire hot spots data. In addition to fire occurrence prediction and burned area estimation, GIS map overlay technique was used for fire risk mapping. Figure 13 presents the methodological steps that were used in the analysis.

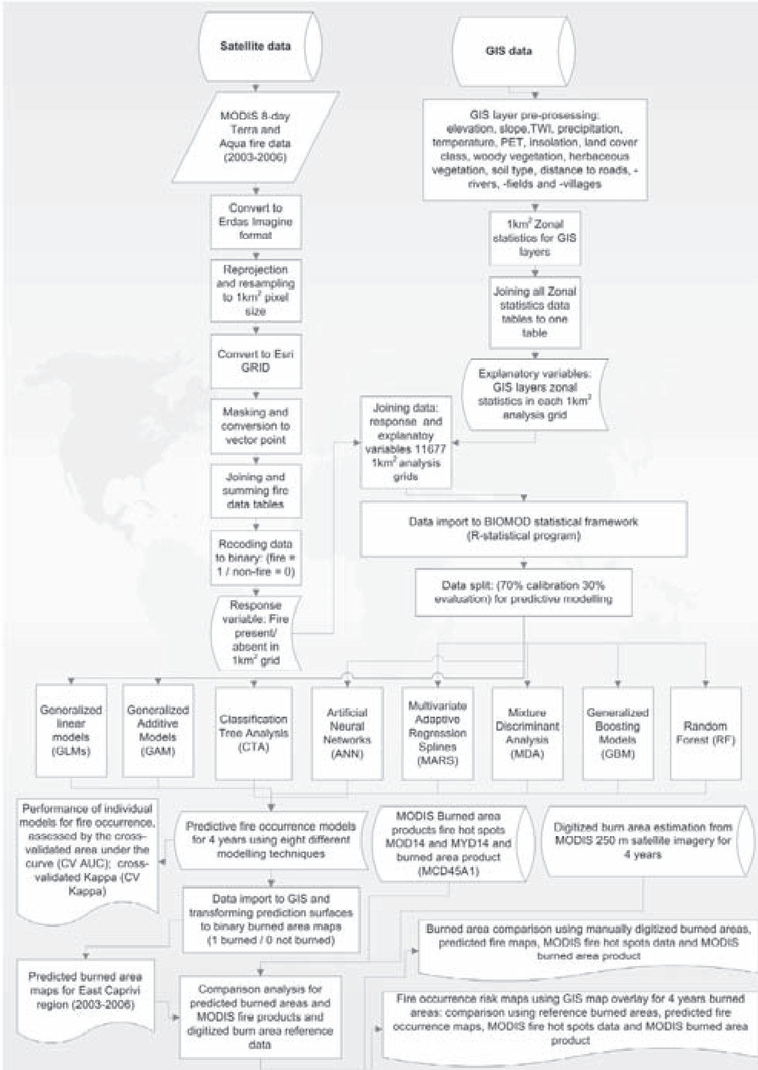


Figure 13. *Methodological flowchart for fire occurrence prediction in East Caprivi using eight state-of-the-art modelling techniques*

5. Results and discussion

5.1 Geodatabase creation for the *Melitaea cinxia* butterfly (Paper I)

The main aim in Paper (I) was to design and develop a GIS database for the *Melitaea cinxia* butterfly and to conduct geospatial analyses using the database. The database consisted of bi-annual field survey data of *Melitaea cinxia*, GPS delimited habitat patch information and meteorological data for temperature and precipitation. A relational database was created for all the entities and relations were created between different data sets. An ODBC link was then established to ArcView 3.2 GIS software to enable spatial analysis and cartographic presentations (seen from I: Figures 4.3, 4.11, 4.12 and 4.14). These images present some of the spatial analysis capabilities for relational geospatial databases. The advance of a geospatial database is that tedious consistently repeatable manual operations can be replaced with semi-automatic data handling and processing. With ODBC linking, database queries could be conducted first in Access and then visually presented in ArcView 3.2 GIS software. Figure 4.13 in the original paper shows an analysis for temporal changes in habitat patch occupancy and population size for *Melitaea cinxia*. With the geospatial database this type of analysis could be made in several spatial scales and not only one. The habitat patch GPS work for *Melitaea cinxia* improved significantly the habitat patch information. Now the exact location of the patch is known and the patch area and patch perimeter calculations are more precise and the knowledge on the habitat surrounding the patch is also known. Thus, now the spatial modelling of *Melitaea cinxia* butterfly metapopulation is more realistic than it was before the mapping work. Figure 14 shows and example of one habitat patch information before and after the GPS work. It can be clearly seen that the true location, patch shape, area and perimeter, and habitat types surrounding the patch can now be modelled more realistically.

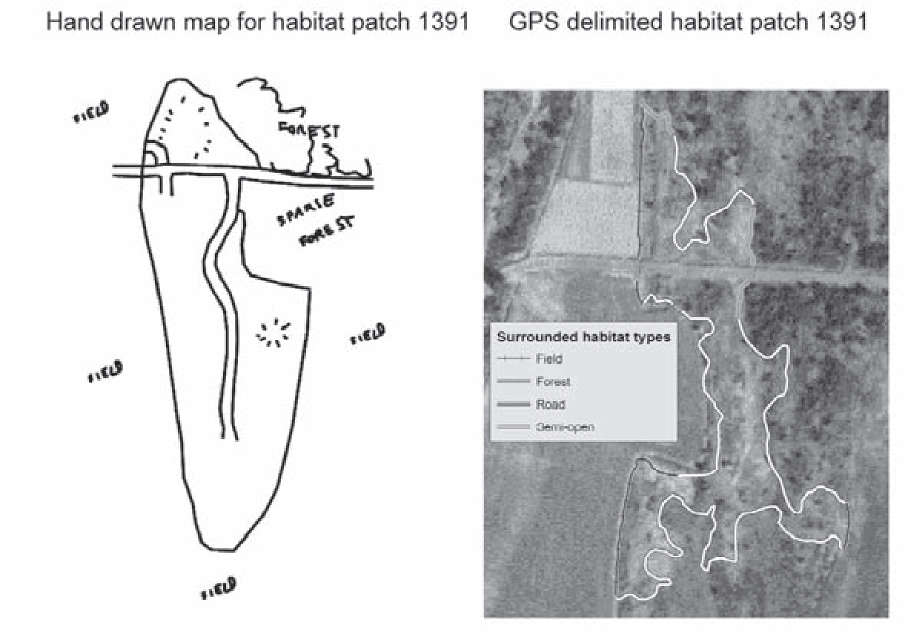


Figure 14. *A presentation of a habitat patch before and after the GPS work. The "before" map on the left presents one example of the habitat patch drawings which was drawn during the field survey. The "after" image on the right shows the same habitat patch with GPS delimited surrounded areas overlaid on an aerial photograph.*

Some critical notes of designing and developing the *Melitaea cinxia* GIS database should be pointed out. The geospatial database was implemented to be used with ArcView 3.2 software, which does not support Unified Modelling Language (UML) conceptual modelling schema nor the new object-relational geodatabase capabilities seen in the current ArcGIS 9.3 software. Contemporary GIS databases are often designed using UMLs as a conceptual modelling tool, e.g. Lambers & Sauerbier (2003) used UML to create an object-oriented UML data model to organize and integrate geoglyph data for Nasca lines in Peru. There are now overwhelming numbers of UML designed geodatabases and it is therefore recommended that UML conceptual modelling should be used for database design with present day GIS software, such as ArcGIS 9.3. However, ESRI has announced in their very recent deprecation plan for ArcGIS 9.3.1 (January 7, 2010) that they will not support CASE tools for ULM modelling schema any more in their forthcoming ArcGIS 10.0 software version. Therefore it is recommended to closely follow which conceptual modelling technique will be used in the future GIS software. Another major shortcoming of using MS Access 97 as a basis for a geospatial database is that it is not possible to manage geographic or location-data in a native format in the way it is possible for example within an Oracle Spatial database system. Oracle Spatial provides functions that facilitate the storage, retrieval, update, and query of collections of spatial features in an Oracle database. Moreover, the *Melitaea cinxia* GIS database was designed as a closed system using proprietary data structures and file formats which normally causes interpretational problems between GIS and other software. GPS work was partly carried out before May 1, 2000, which means that the feature called Selective Availability (SA) was still on. Before SA was switched off it added intentional time varying errors up to 100 m to the publicly available GPS navigation signal. This error was clearly seen in raw GPS data before the differential correction was made and even after differential correction some errors still existed for a number of habitat patches. This added some uncertainty to the accuracy of GPS delimited patches.

Some uncertainty also existed in the radar precipitation data. There was some clear "noise" in the data and mostly it could be eliminated but in some locations it could not be removed. However, a correlation analysis where radar precipitation data was plotted with weather station precipitation data shows statistically significant correlation (Hanski & Meyke 2005). Moreover, 1 km^2 precipitation radar data revealed that in Åland Islands there exists high spatio-temporal variation in precipitation patterns, which was not so evident when using only meteorological data from weather stations. On average, there was a clear decreasing precipitation trend from west to east and high variation existed in weekly, monthly and yearly precipitation in different areas. This high spatio-temporal variation may have serious consequences to local *Melitaea cinxia* populations. For example in dry summer months host plants may dry out or during wet condition larvae groups may drown and therefore local population may go extinct. It can be concluded that it appears evident that at a large-scale spatially correlated weather conditions are one of the primary causes of spatially correlated changes in *Melitaea cinxia* population sizes.

The making of the *Melitaea cinxia* GIS database was a challenging task and all the unexpected problems occurring during the development of database cannot be explained in this thesis. Furthermore, we have to put the project to the right time perspective. At the time the database was created, about ten years ago, the GPS receivers were not at such an advanced and accurate level as they are now. As an example, we tried to use a commercial differential correction signal obtained from FOKUS-service but most of the time we could not receive it in a proper way. In addition, GIS- and database software had limited capabilities compared to the contemporary software with all the sophisticated geospatial data handling possibilities. However, it can be concluded here that the geospatial database has greatly improved the knowledge of true *Melitaea cinxia* metapopulation structure and it has been used successfully in a number of studies since it was built (see e.g. Hanski & Meyke 2005).

5.2 Scopulini moth spatial analysis for diversity and distribution (Paper II)

Spatiotemporal analyses were conducted using a computerized geospatial database of Scopulini moths (Lepidoptera: Geometridae, Sterrhinae). Various aspects related to species description, diversity and geographical distribution were analysed. In general, the analyses showed that diversity patterns for Scopulini moths reflected a similar pattern as those reported earlier for the whole Geometridae family (see Gaston *et al.* 1995). The numbers of described species of Scopulini by decade showed that it peaked between 1900–1940, during which period 528 species were described. Since the 1940s there has been a lower rate of descriptions c.a. 15 species per decade but the cumulative numbers of species descriptions has not yet reach the asymptote and inform a biogeographical perspective, species descriptions has been very uneven (II: Figures 1 and 2). Generally, the most species rich biogeographical area for the Geometridae has been reported to be Neotropics. However, for the Scopulini it was found to be Africa (II: Table 1). The synonymy rates were found to be highest in the New Zealand and Nearctic regions and lowest in Madagascar.

The Scopulini was found to be cosmopolitan in distribution but the distribution of type localities is uneven. The most species-rich analysis squares were located in sub-Saharan Africa and in northern India (II: Figure 6). Many of the most speciose squares of the Scopulini have also been identified earlier as biodiversity hotspots, based on non-invertebrate taxa (Myers *et al.* 2000). In other areas, virtually no species have been described from the interior parts of the Nearctic and Neotropics. When latitudinal patterns were investigated, it was found that the majority of the species have been described from low latitudes and numbers decrease steadily towards higher latitudes, following broadly the latitudinal gradient theory of diversity by Fischer (1961) (II: Figure 6). This finding contradicts the earlier finding of Holloway (1997) where he stated that Scopulini have been successful also in temperate zones.

5.3 Spatiotemporal forest cover change detection (Paper III)

The main aim in Paper (III) was to analyse forest cover spatiotemporal change between 1955 and 2004 in Taita Hills, Kenya. Forest cover change was analysed in the study area quantitatively using black and white aerial photographs (1955 and 1994) and airborne digital camera mosaics (2004), and field survey data (2007 and 2008). Another aim was to create a consistent methodological framework for forest cover change detection using digital imagery techniques. Analysis methodologies were described in detail in the Paper III because previous forest cover studies in Taita Hills have lacked profound explanations of the used techniques and therefore hampered the validation of earlier results. Moreover, there exists a great variation in the forest cover areas given by other authors due to the different principles used for delineating indigenous forest from other land cover, due to the definition for indigenous, native or original forest or due to the interpretation methods used.

The main results indicated that the total forest cover decreased in the study area only by 2% due to the planting of exotic trees. However, indigenous forests decreased by 50% (260.2 ha) and based on change detection results they were mainly changed to agricultural areas but also to exotic tree plantations. This finding is in concordance with a land use study by Imberon (1999) were he used SPOT satellite image and aerial photographs to analyse land-use over the past 40 years (1958 to 1995) at Embu highland on the slopes of Mount Kenya. He found that the total tree cover including forest, woodlots, tree lines and tree plantations, had not changed. However, the composition of tree cover had changes so that exotic tree species, such as eucalyptus, dominated over natural trees where in 1958 the composition was opposite. Moreover, the area under annual food crops increased significantly from 1958 to 1985, probably because of the population growth in the area. A typical pattern in Sub–Saharan land cover change is that natural tropical forest cover has changed to mainly agricultural land or to exotic trees.

This study presents the most accurate evaluation so far of the forest cover change for the studied forest patches in Taita Hills. However, there are still some important considerations about the accuracy that should be pointed out here. These accuracy issues have to be taken in to account when using historical black and white aerial photographs and airborne digital imagery. Firstly, for the black and white photographs the metadata was very poor, only information on focal length of the camera was available and the aerial photographs had no fiducial marks, hindering their ortho-rectification correction. Therefore, an image-to-image co-registering method was employed. Black and white aerial photographs were co-registered using 2004 airborne imageries as reference data. By doing so it has to be recognized the fact that uncorrected black and white aerial photographs will inherit all the geometric distortions and location errors of the reference images and positional inconsistency between different map sets will become apparent. Fortunately in this study the 2004 airborne imagery had only minor geometric distortions (ca. 2 meters around main forest areas) validated by using differentially corrected GPS reference points.

Black and white aerial photographs were co-referenced using rubber sheeting, a piecewise polynomial method for geometric correction of digital imagery. Rubber sheeting attempts to correct errors by stretching a map to fit a known set of ground control points (GCPs) by forming a triangulated irregular network (TIN) over all the GCPs. The image area covered by each triangle in the network is rectified by the first (linear) or fifth (nonlinear) order polynomials (see e.g. White & Griffin 1985; Doytsher 2000; Doytsher & Hall 1997) therefore the rubber sheeted image is suffering for positional inconsistency related to referenced image. It is desirable that ample amount of evenly distributed GCPs are collected covering the whole image but in this study it was exceptionally hard to find matching control points between 1955 and 2004 images because of significant landscape changes. Fortunately some matching GCPs could be found from unchangeable places such as road crossings. But obviously the quality of GCPs might have some effect on the accuracy of co–registration.

A critical notice has to be placed also over on-screen digitization of forest patches using GIS. On screen digitizing is solely a subjective way of interpreting and delineating the forest patches and therefore errors may occur. This fact is showed in the original Paper III were some forest patches was fine-tuned by field check and the comparison between the visual interpretation before and after field work showed some differences (III: Figure 5). These probable interpretation errors are even more significant if considering that the interpretations of forests from 2004 colour images is fairly easy. This cannot be said about digitizing forests from 1955 and 1994 black and white images. From black and white images the interpretation is based only on the different colours of gray levels and image texture, it lacks the assistance of colour. As an example here, from 2004 airborne images it was fairly easy to interpret eucalyptus forests due to the distinctive greenish-brown crown cover.

These two possible sources for inaccuracy discussed above have to be taken in to account when analysing the exact forest cover change numbers. However, the advantages of digital airborne camera data compared to traditional airborne camera data were showed to be obvious. Firstly, when using digital imagery the whole process is digital from the start to the end. In addition, automation and integration of the GPS-system to the camera gave more accuracy to the aerial photographing. Moreover, the digital airborne imaging does not need as much light as traditional aerial photography and has important advantages such as digital data storage, manipulation, transmission and easier display. The quality of the data also increases when steps like chemical film, processing and scanning become unnecessary. On the other hand, the disadvantages of digital aerial photographs include unstable geometry and so called dead pixels.

It can therefore be concluded here that forest cover change at local and regional scale can be analysed by using historical images and airborne digital camera images. Nonetheless, it should be noted that spatiotemporal change trajectory analysis using historical data (photography or maps) is not a new method and over recent years it has been applied in various studies. Lung & Schaab (2004) used remote sensing imagery and aerial photographs to study forest cover change in Kakamega forest in west Kenya. Vuorela & Toivonen (2003) analysed historical land cover and woodland cover change between 1690 and 1998 on the island of Ruissalo south-western Finland. Fuse & Shimizu (2004) used historical maps to construct a model for ancient Tokyo city (Edo). Imberon (1999) studied land cover change at Embu highland on the slopes of Mount Kenya. He used successfully aerial photographs and SPOT satellite image for change detection.

In Paper III the forest cover change was analysed with high accuracy for the selected forest patches in Taita Hills by using airborne colour imagery and historical black and white aerial photographs. The accuracy of co-registering of historical black and white photographs and the delineations of forest patches from these images using on-screen digitizing with GIS, in addition to the lack of historical ground-truth data, could lead however to some errors in the forest cover change calculations. As a final statement, according to Hording (2004) five significant advantages of visual interpretation of photo products over satellite image based land cover models can be found: *"Less time required to create a usable product; Little, if any, expense incurred beyond the acquisition of the image; Image illumination "problems" (such as shadows and brightly illuminated surfaces) can be used as an interpretation aid; Minimal expertise required to interpret the image; and uses the power of the brain"*.

5.4 Human population prediction in Taita Hills (Paper IV)

Population distribution and abundance were modelled for the rural mountainous area of the Taita Hills, Kenya, using dwelling unit data (presence-absence) and population count data (abundance) as the response variable and geospatial data as predictors. Prediction models were created using the GRASP method that utilizes the generalized additive model (GAM) regression technique. The results showed that population distribution models explained 19 to 31% of variation in the dwelling unit occurrence data indicating a fair explanatory power and the predictive power for population distribution models was good (AUC 0.80 to 0.86). The abundance models explained 28% to 47% of the variation in human population abundance in the study area. Combined geospatial- and remote sensing-based predictors gave the overall best modelling results when compared with only remote sensed and GIS predictor models. The best single predictors for modelling the variability in human population distribution using combined predictors were: angular second moment image-texture measurement, precipitation, mean elevation, surface reflectance for SPOT red- and near-infrared bands, correlation image-texture measurement and distance to roads, respectively (IV: Tables 3, 4, 5 and 6). The fairly poor performance of land cover classes used as predictors in human population predictive modelling indicates that it is not necessary to use classified land cover, instead first and second order image-texture measurement derived from satellite image should be used. Second-order image texture factors have been shown to be important factors in earlier urban population density analysis (see for example Shaban & Dikshit 2001; Li & Weng 2005). The second benefit to use the first and second order image-texture measurements is the fact that they can be used to quantify the variability of vegetation as a continuous variable in statistical modelling. These findings are important, especially when considering the laborious satellite image classification work to derive an accurate land cover map is not necessarily needed. This study also revealed that models using solely geospatial predictors had by far the lowest model performance in population models and therefore it is suggested that they should not be used alone as a predictors for dwelling unit distribution and abundance modelling. However, availability of a more precise geospatial predictors, e.g. a road network that also includes footpaths, might have improved the model performance. This is true especially in mountainous rural areas, such as the Taita Hills, were the majority of houses are accessed only by footpaths.

Generalized additive model (GAM) were chosen for modelling because in various species distribution studies GAM models have outperformed conventional linear regression techniques (Yee & Mitchell 1991; Thuiller *et al.* 2003). Moreover, GAMs are more suitable for geospatial data modelling as environmental predictors are often non-Gaussian with non-constant variance. However, in earlier human population modelling studies using geospatial data, mainly linear regression techniques have been used (see, e.g. Lo 1995; Schnaiberg *et al.* 2002; Gustafson *et al.* 2005; Li & Weng 2005). To our knowledge, this study was the first time that GAM models were used for human occurrence and abundance prediction using geospatial predictors and the good modelling results encourage other predictive human population and abundance studies to consider the GAM modelling technique to be used alongside or alone with more traditional regression methods.

The GAM abundance model was extrapolated for the whole study area and the model was capable of discriminating between inhabited and uninhabited areas (VI: Figure 4). For example, there are no dwelling units in Ngangao forest and the model predicting presence of human population concentrations in and around the villages and absence of dwellings on cultivated fields. When abundance models were compared with two existing global population data sets, GPWv3 and LandScan 2005, the results affirmed that there was statistically significant correlation between combined and remote sensing based models and the GPWv3 product ($r > 0.8$) but the correlation was non-significant with geospatial model ($r = 0.19$). For LandScan 2005 the correlations were lower (VI: Table 7). The correlation between Kenyan census data for 1999 and predicted population abundance models are high for remote sensing data ($r = 0.71$) and combined models ($r = 0.51$) when solely sub-locations over 1100 m.a.s.l. (n = 32) were used. For geospatial models the correlation was non-significant. There was low correlation ($r = 0.34$) between remotely sensed population abundance models and Kenyan census data for 1999 for the sub-locations also extending into the lowlands (n = 50) and no correlation for combined- and geospatial models (VI: Table 8). As a final conclusion it can be stated that the predictive models using predictors from remote sensing and geospatial data were found to be more accurate than global datasets and correlated well with the Kenyan 1999 census data too. However, it must be kept in mind that the modelling performance can be affected by different factors such as: analysis scale, spatial autocorrelation, chosen predictors and modelling technique and model parameterization.

5.5 Fire prediction and burned area estimation in East Caprivi using GLM (Paper V)

Generalized linear models (GLMs) were used for predictive fire modelling in East Caprivi, Namibia. MODIS hot spots fire data for four years 2002–2006 were used as a response variable and geospatial data was used as predictors. Prior to predictive modelling, spatiotemporal characteristics of fires were investigated. Firstly, fire frequency counts on a monthly basis were conducted and the fire frequency analysis showed a typical fire trend in Southern Africa were most of the fires are occurring during a few months in the dry season. In East Caprivi the peak season for fires is September–October (V: Table 4). The extent of burned areas was roughly estimated from fire hot spots data assuming that each hot spot represents a 1 km^2 of burned area, fully aware of the possible misinterpretation that may occur with this type of assumption. For example Roy *et al.* (2008) have estimated that hot spots based burned area calculations underestimated the burned area by ca. 24% due to high omission error. The burned area in East Caprivi was estimated from hot spots data to vary between 19.2–24.4% (V: Table 4). The spatial characteristics of fires were investigated using Moran's *I* correlograms and the results showed that high positive autocorrelation was present for fire occurrence at short lag distances for all four years and aggregated four years fire data (V: Figure 3). This finding was in concordance with an earlier study made in Mali by Laris (2005) showing that the people set fires in regular annual pattern and the patches that burn every year are approximately in the same place at the same time.

Due to the high spatial autocorrelation of fires, two types of predictive models were conducted: models with and without consideration of the spatial autocorrelation of fires. Moreover, separate model were developed for abiotic, biotic, anthropogenic and combined explanatory variables, respectively. The modelling results showed that models accounting for spatial autocorrelation, i.e. autologistic models, had by far the best performance. This result is similar to Chou *et al.* (1993) and Lynch *et al.* (2006) where model improvement were significant when accounting for spatial autocorrelation in logistic fire models. Moreover, there was practically no difference in model performance between autologistic models using abiotic, biotic, anthropogenic or combined explanatory variables. Hierarchical partitioning (HP) was used as a complementary analysis method in order to reveal individual explanatory variables having the most independent influence on fire occurrence. The result showed that autocovariate (a mean value of total number of fires in the eight, where possible, nearest neighbour analysis squares) had by far the strongest independent influence on fire occurrence and as much as 24% of the variation in fire occurrence was explained by the autocovariate variable.

Fire prediction surfaces were transformed to classified burned area maps and compared with MODIS burned area product (MCD45A1) and reference burned area data that were derived from MODIS 250 m images. The overall result was that in this study, autologistic fire probability models estimated burned area more accurately than MODIS fire products (V: Table 9). Therefore, based on the clear results of superior performance for predictive fire occurrence modelling relative to MODIS burned area product, the future MODIS burned area product should investigate the possibilities to incorporate predictive fire occurrence modelling and improve the present burned area estimations.

However, some critical issues have yet to be raised even though the findings in this study were found to be satisfactory for both fire occurrence prediction and burned area estimation. Firstly, in this study there was no ground truth reference data available for burned areas, instead MODIS 250 m imagery were used to derive the reference burned area for each year. Thus, it has to be recognized that burned area estimation derived from coarse scale satellite imagery may also have errors due to incorrect georectification and boundary differences during on-screen digitizing (Felderhof & Gillieson 2006; Verlinden & Laamanen 2006). This will result that small fires are overlooked and burned areas are simplified as seen from (V: Figure 6). Secondly, the predictive modelling was conducted at 1 km analysis square due to the MODIS hot spots data resolution (1 km). The model performance could thereby be affected by the relatively coarse scale and downscaling would have been an option to consider for model performance improvement (Kidson & Thompson 1998), especially because the original spatial resolution for predictor map layers were less than 1 km, e.g. for the DEM the resolution was 90 m. On the other hand, a smaller analysis grid side would increase the sample size and risk pseudoreplication occurring in the statistical analysis (Hurlbert 1984). Thirdly, predictive models are scale and place -dependent and the results should be interpreted somewhat cautiously particularly when fire models are extrapolated to other areas or to different analysis scales.

5.6 Fire prediction and burned area estimation using eight modelling methods (Paper VI)

Eight state-of-the-art modelling methods (GLM, GAM, MARS, CTA, MDA, ANN, GBM and RF) were used for predictive fire occurrence modelling, burned area estimation and fire risk mapping in East Caprivi, Namibia. MODIS hot spots fire data for four years 2002–2006 were used as a response variable and geospatial data was used as predictors. Two types of predictive models were tested: models accounting for spatial autocorrelation, i.e. models including the autocovariate predictor and models without the autocovariate predictor. Predictive accuracy was evaluated using cross-validated AUC and Kappa values. The results showed that distinctive variation existed between different modelling methods (VI: Figure 3 and 4). The best overall model performance was achieved using generalized boosted models (GBM). The GBM models had good predictive power and good accuracy for burned area estimation. For fire occurrence prediction random forest (RF) had the best predictive power. However, when RF models were transformed to classified burned area maps, the performance was very poor. The model performance results and burned area estimations for all the models can be seen from the original paper VI (VI: Tables 4, 5 and 6). Furthermore, each of the modelling method results are discussed separately in detail in the original Paper VI section 5.

To outline the main findings: firstly, to the latest knowledge, GBM MARS, GAM modelling techniques, which had good modelling performance in this study, have never been used before in fire occurrence, burned area estimation or fire risk mapping. This is surprising considering the good results for these novel modelling techniques in, for example, ecological studies (Prasad *et al.* 2006; Elith & Leathwick 2007; Elith *et al.* 2008; Morin & Thuiller 2009). Artificial neural networks (ANNs) and classification tree analysis (CTA) have been used successfully earlier for fire occurrence prediction and burned area estimation. ANNs were used in a study by Maeda *et al.* (2009) to predicted fire occurrence in Brazilian Amazon and Al-Rawi *et al.* (2001) estimated burned areas using NOAA-AVHRR imagery for the eastern part of Spain. However, in this study ANN models had very low accuracy for fire prediction and they overestimated burned areas (VI: Tables 4, 5, 6 and Figures 3 and 4). Stroppiana *et al.* (2003) used SPOT VEGETATION data and CTA modelling for burned area mapping in the Australian savanna with acceptable accuracy. In this study CTA models had also fairly good prediction accuracy and burned area estimation. This study also shows that GLM models had fair modelling performance for fire prediction and burned area estimation. The advantage of GLM models relative to more novel statistic techniques is that GLMs are fairly simple to understand and they present clear regression equations whereas, for example, RF and CTA models do not. The random forest regression technique has been used before for burned area estimation at coarse scales. Archibald *et al.* (2009) predict fires at a sub-continental scale in Africa using 100 km analysis window size with fairly good results and the RF technique has out-performed more traditional modelling techniques (see Prasad *et al.* 2006; Peters *et al.* 2007). However, in this study RF models had low accuracy for burned area estimation. Mixture Discriminant Analysis (MDA) models had generally poor performance for fire prediction and burned area estimation. This can be clearly seen from (VI: Figure 3) where burned area estimation for year 2003 is unrealistic as the model has detected some odd artificial boundaries for burned and unburned areas. For models including autocovariate predictor, i.e. for the models accounting for spatial autocorrelation of fires, MDA models had better model performance (VI: Figure 4). Figure 5 from the original paper VI presents a fire risk map created using GIS map overlay techniques. This figure and Table 7 shows that predictive GBM modelling technique estimated fairly good areas that were burnt in four years time.

The results of this study showed that there existed some noticeable variation between different modelling techniques for fire occurrence prediction and burned area estimation. Variation existed between different prediction techniques but there was also some intra-model variation within single modelling techniques. Moreover, this study indicated that no single modelling method had absolutely the best performance for fire occurrence prediction and for burned area estimation, as there was some variation between different years. It is therefore important that a combination of modelling techniques and statistical algorithms using models with and without autocovariate should be utilized for obtaining the best modelling results for predictive fire occurrence modelling and burned area estimation. This study also highlighted the need to improve the current MODIS (MCD45A1) burned area product as several of the predictive modelling techniques used in this study had superior performance for burned area estimation relative to that of the MODIS product.

5.7 The challenge to improve active fire detection and burn-scar detection

Throughout the World fires threaten human lives, property and natural resources. In recent years devastating fires in California, Australia, Indonesia and Greece have gained a lot of media attention but the majority of fires burning around the World are not reported. For example, annually thousands of square kilometres are burned in Africa. In the past, fire has been an essential part of ecosystem dynamics in African savannas but at the present fires have increased and the majority of these fires are man-made (Laris 2005). Figure 15 shows the extent of present day fires in Africa. In the African continent and Madagascar there were ca 2.5 million active fires detected with MODIS Terra and Aqua satellites in 2008 (data set obtained from MODIS Rapid Response Project). A picture on the right shows an example of burned savanna environment in East Caprivi in 1996.

Figure 15. *Active fires detected in African continent and Madagascar during 2008 with MODIS Terra and Aqua satellites and recently burned savanna in East Caprivi (photo by Mika Siljander 1996).*

Remote sensing methods have been used for three decades for active fire and fire-scar detection. A multitude of sensors, algorithms and applications have been used during these decades. However, uncertainties in active fire detection and burned area estimation exist and the accuracy of fire models is still far from perfect. A lot of work has to be done to improve present fire models and researchers are persistently seeking for new algorithms and modelling techniques to enhance active fire detection and burn scar estimation. New novel predictive modelling techniques are one possible solution to improve currently available fire occurrence and burned area models, as there are still great challenges before local fire managers can obtain precise fire occurrence information at local and regional scales, and before global climate modellers receive accurate burned area estimation at global scale derived from remote sensing or GIS originated geospatial data.

6. Conclusions and future prospects

The main purpose of this dissertation was to develop geospatial environmental data modelling applications using remote sensing, GIS and spatial statistics and this thesis presented six studies for environmental geospatial data analysis covering five themes: (i) design and development of an environmental geospatial database; (ii) species diversity and geographical distribution analysis; (iii) forest cover change detection; (iv) predictive modelling for human population distribution; and (v) predictive modelling for fire occurrence and burned area estimation. The main conclusions of the thesis are as follows:

- A cogent geospatial database is an essential part of spatial analysis and functional RDBMS should be created for each study that utilizes geospatial data. Contemporary GIS has some inbuilt database capabilities such as Geodatabase in ArcGIS but to work effectively with geospatial data, true RDBMS software should be used. Present day database software such as commercial Oracle Spatial and Open Source PostgreSQL/PostGIS adds support for geographic objects. For creating the *Melitaea cinxia* geospatial database, MS Access 97 software was used which does not support geographical objects. However, with Access RDBMS it was possible to make complex database SQL queries and then link the database to ArcView 3.2 GIS software by using Open Database Connectivity (ODBC) thus providing semi-automatic functionality to spatial analyses. The challenges to manage geospatial databases have increased at the same time as the data availability and the sizes of the data files have increased and therefore it is essential for scientists and for environmental modellers also to understand the potentialities and constraints of geospatial databases.

- The species data derived from the world check-list and from species catalogues can be computerized and incorporated into GIS for species description, diversity and geographic distribution analyses. A geospatial database was developed for Scopulini moths and it was found that Scopulini moths have a cosmopolitan distribution. The majority of the species have been described from the low latitudes, sub-Saharan Africa being the hot spot of species diversity. However, the taxonomical effort has been uneven among biogeographical regions.

- Tropical forest cover measurements at regional and local scales using orbital satellite imagery suffer from errors. Airborne digital camera imagery is more suitable to gain accurate forest cover measurements. However, when spatiotemporal forest cover change is studied care has to be taken in co-registration and image interpretation when historical black and white aerial photography is used. Furthermore, visually interpreted GIS-based land cover models should be fine-tuned with field survey to enhance the accuracy.

- Human population occurrence and abundance can be modelled with satisfactory results using GIS and remote sensing based data and non-Gaussian predictive modelling techniques. Land cover layer is not necessary needed as a predictor because first and second-order image texture measurements derived from satellite imagery had more power to explain the variation in dwelling unit occurrence and abundance. Predictive, local and regional scale human population abundance models were more suitable than existing more coarse scale global population data sets GPWv3 and LandScan 2005 to estimate the number and showing the distribution of population in the rural mountainous area of Taita Hills, Kenya.

- Generalized linear model (GLM) is a suitable technique for fire occurrence prediction and burned area estimation. When predictive models were transformed to classified burned area maps they outperformed the MODIS (MCD45A1) burned area product in estimating the burned areas at a regional scale in East Caprivi. Namibia. However, spatial autocorrelation of fires has to be taken into account when using the GLM technique for fire occurrence prediction.

- Novel statistical predictive modelling techniques can be used to improve fire prediction, burned area estimation and fire risk mapping at a regional scale. Predictive fire modelling can be used as a tool to provide communities and fire managers an early warning system to identify areas at a risk of likely fires. The burned area product MODIS (MCD45A1) is not accurate for estimating burned areas at a regional scale and predictive modelling based burned area estimations outperformed the MODIS burned area product. However, some noticeable variation between different predictive modelling techniques for fire occurrence prediction and burned area estimation existed. The random forest (RF) had the best predictive accuracy but performed very poorly for burned area estimation. Generalized boosted models (GBM) had both good predictive accuracy and burned area estimation.

All the major problems that we are facing throughout the world today – overpopulation, food shortages, reduced agricultural production, poverty, deforestation, land degradation, land cover change, biodiversity loss, flooding, droughts, wildfires and climate change, - these are all fundamentally geographic problems. Geoinformatics is the science that develops and uses information science technologies and tools, combining geospatial analysis and modelling and development of geospatial databases to resolve geographic problems. Geoinformatics tools and techniques have been successfully used in contemporary environmental research and without these tools and techniques environmental modelling would not be the same at is it now.

We have come a long way from the 1960s map overlay analysis where transparent sheets of different map layers e.g. geology, land cover and infrastructure were overlaid one on top of the other to achieve most suitable areas of interest for a specified analysis. At present, GIS technology allows us to do e.g. map overlay analysis in a “push the button” manner and do things just in minutes instead of struggling for hours with a pen and paper. However, at the same time as GIS and remote sensing software will get easier to use there lurks a danger that users with a single afternoon GIS course will think themselves to be GIS experts. But this is not true as to become a GIS/RS professional takes years of hard learning, and mastering GIS/RS seems to be a “mission impossible” – a lifelong learning process with no ending in sight. Before understanding GIS/RS one has to understand geography as a complete discipline; one has to understand the fundamentals of geography e.g. principles of projections and mapping scale, spatial autocorrelation, map interpretation, topology, and one has to understand the limitations and problems of spatial data; problems with data standards, problems of data definition, data accuracy, and data exchange. And most of all, one has to understand that GIS is only a tool, though a powerful tool in the right hands.

There is no doubt that the use of remote sensing and GIS in environmental research are increasing day by day. Governmental organizations and private companies are rapidly increasing the use and applications in various fields of vital importance. Increasing use of these techniques in environmental research will show also increasing specialization in GIS/RS applications in the future. Therefore there is a great challenge ahead to develop and implement innovative applications for environmental research. But if we are going to use these tools and techniques we must integrate physical, cultural, economic, and political geographical knowledge to understand the "Whole Picture". And by doing that, we can understand more deeply the factors and actors that are behind environmental degradation and we might have a change to save our Planet for the future generations, so that they also have the chance to see our Planet as beautiful as I have seen it during my travels.

Acknowledgements

Firstly, I gratefully acknowledge my supervisor professor Petri Pellikka as he took me as a very unsure "non-traditional" student, returning from two-year sabbatical trip, to study for a Ph.D. Petri, I also have to thank you that you tenaciously insisted for me to take part in Taita Hills field trip and to return to Africa in winter 2006. At that point I was not too keen to return to Africa because of my mugging at knife-point incident in Windhoek in 2004 that was still too clear in my mind. But thanks Petri, the trip to Taita Hills returned my trust and love for Africa, and the "late night board meetings" that we shared with a glass or two of lukewarm Whisky or Cointreau at the Hebron Guest House in Wundanyi, these moments have melted deeply in to my consciousness. I am also grateful to you Petri for giving me free hands to study the subjects that I have a great interest in, like forest fires in East Caprivi. I know that most of the papers presented in this thesis are not about Taita Hills but I believe that you can cope with this. My interests have always been to understand things broadly as I am more a "Jack of all trades and master of none" kind of person, than a precise scientist knowing so much about so little. This might be the explanation why I also cover some other study areas, such as East Caprivi, in this thesis. Secondly, warm thanks go to Dr. P.K. Joshi and Dr. Tarmo Virtanen for pre-examining this thesis.

During my Ph.D. work Barnaby Clark and Alemu Gonsamo have been as supporting and encouraging as one can expect true working colleagues could be. I highly respect you both and it has been a fun ride with both of you "Boyz". Barnaby, we have had some extraordinary moments talking deeply about the real life subjects, not about anything related to either Ph.D. thesis, remote sensing or GIS. These talks have given me a great pleasure and fun and more trust to believe in the wisdom of humankind. We had some tremendous time at Safari Park Hotel outside Nairobi when we were giving a remote sensing and GIS course at the Regional Centre for Mapping Resources for Development (RCMRD). And you Alemu, we have shared some three different rooms at the Department of Geosciences and Geography and there have been some outstanding moments with you. You truly reflect the best spirit of Africa.

This thesis was mainly made at the Department of Geosciences and Geography at the University of Helsinki premises. I therefore want to thank all the staff members and I also want to express gratitude to some particular persons. I would firstly like to thank the Head of the Department, John Westerholm, for giving me the opportunity to work at the Department of Geosciences and Geography. Secondly, as I am mostly spending my working days with GIS it means computers and computers usually mean troubles! Tom Blom, who is now working at the IT-Services has been "the guy who gets you out of the trouble" at least when computers issues are concerned. Thanks Tom, you have helped me out so many times, I really appreciate it. This same goes also to you Hilkka Ailio; you have also helped me a lot with all sorts of computer issues. There would not be a Geographical Department without an efficient administration – "The Office". And from "The Office" I especially want to thank Airi Töyrymäki to make most of my paperwork feel "smooth as a silk". The thanks goes also to Johanna Jaako helping me out with some tricky administrative issues. Three persons from the Department have had some influence in my career. In 1996 Ritva Kivikkokangas-Sandgren was the leader for a Master's student trip to East Caprivi in Namibia and, as I still have some interest to East Caprivi fires, I thank you Ritva for the trip to East Caprivi. In 1997 Mari Vaattovaara was the person who had the courage to select me at *Programme for Maintenance and Applications of Geographical Information Systems* course. Without this course I doubt I would be working with GIS, remote sensing and databases issues. So Mari, I would like to thank you for having such an impact on my working career. Tuuli Toivonen you are the person who gave me the first GIS practical and that practical was with MapInfo. It took me 20 minutes before I left the room with my classmate and we had to go to clear our heads with couple of beers in Tube Bar. I just could not understand anything about MapInfo – I pretty much still don't. So I thank you Tuuli for giving me the first push towards the "ESRI World", but also for many other things. I would also like to thank you also Nina, Tino, Jan, Jari–Pekka and Eduardo for assistance and great working companion and Miska Luoto for his early influence.

There are some persons not working at Department of Geosciences and Geography that I want also to acknowledge. Jaakko Suikkanen and Juha Oksanen were giving me the first GIS course after my MapInfo experience. I thank you both for taking my MapInfo GIS trauma away with the more sophisticated software ARC/INFO Workstation. Probably my final words will be "`mape meku`". Jaakko we have had some memorable moments also outside the GIS world down in the Galaxy cellar jamming the night away so let's keep it that way. I want to express gratitude to Prof. Ilkka Hanski to take me "On the wings of butterfly" at the MRG and I also want to thank Marko Nieminen and Pasi Sihvonen who were the main authors in Paper I and Paper II and I also wish to thank every one of the other co-authors in my thesis for advice, comments and support. And Janne Heiskanen, we shared some valuable talks while you were still working at the Department of Geosciences and Geography.

I also want to thank my family, especially my deceased mother and father as you let me live my life the way I wanted. My great love goes to my sisters Susanna and Heidi. You both have always supported and encouraged me. I want also to thank for encouragement and support from some of my relatives and friends: Petteri, Taru, Kati, Jussi, Jukka, Taina, Kimmo, Veijo, Tuula, and Marco.

Finally and above all, I am eternally grateful to my wife Tuija. Tuija, you are the force behind my existence and you were the first to encourage me to finish my high school and continue to University. We have soon shared 20 wonderful years and we have made countless trips together and witnessed together the beauty of this Planet. It has been a magnificent journey and I must admit that I am still very much in love with you. **ผมรักคุณ**

Huvilakatu, Helsinki, February 2010

Mika Siljander 

References

Akaike, H. (1974). A new look at statistical model identification. IEEE *Transactions on Automatic Control* 19, 716–722.

Al-Rawi, K.R., Casanova, J.L. & Calle, A. (2001). Burned area mapping system and fire detection system, based on neural networks and NOAA-AVHRR imagery. *International Journal of Remote Sensing* 22, 2015–2032.

Achard, F., Eva, H.D., Stibig, H.J., Mayaux, P., Gallego, J., Richards, T. & Malingreau, J.P. (2002). Determination of deforestation rates of the world's humid tropical forests. *Science* 297, 999–1002.

Andrén, H. (1994). Effects of habitat fragmentation on birds and mammals in landscapes with different proportions of suitable habitat—a review. *Oikos* 71, 355–366.

Archibald, S., Roy, D., van Wilgen, B. & Scholes, R.J. (2009). What limits fire? An examination of drivers of burnt area in Southern Africa. *Global Change Biology* 15, 613–630.

Barbosa, P.M., Pereira, J.M.C. & Grégoire, J.M. (1998). Compositing criteria for burned area assessment using multitemporal low resolution satellite data. *Remote Sensing of Environment* 65, 38–49.

Barbosa, P.M., Grégoire, J.M. & Pereira, J.M.C. (1999). An algorithm for extracting burned areas from time series of AVHRR GAC data applied at continental scale. *Remote Sensing of Environment* 69, 253–263.

Beck, J & Chey, V.K. (2008). Explaining the elevational diversity pattern of geometrid moths from Borneo: a test of five hypotheses. *Journal of Biogeography* 35, 1452–1464.

Beentje, H.J. (1988). An ecological and floristic study of the Taita Hills, Kenya, *Utafiti* 1, 23–66.

Bourgeau-Chavez, L.L., Kasischke, E.S., Brunzell, S. & Mudd, J.P. (2002). Mapping fire scars in global boreal forests using imaging radar data. *International Journal of Remote Sensing* 23, 4211–4234.

Boschetti, L., Flasse, S.P. & Brivio, P.A. (2004). Analysis of the conflict between omission and commission in low spatial dichotomic thematic products: the Pareto boundary. *Remote Sensing of Environment* 91, 280–292.

Breiman, L., Friedman, J.H., Olshen, R.A. & Stone, C.J. (1984). *Classification and Regression Trees*. Wadsworth and Brooks/Cole, Monterey, CA, 358 pp.

Breiman, L. (1996). Bagging predictors. *Machine Learning* 26, 123–140.

Breiman, L. (2001). Random forests. *Machine Learning* 45, 15–32.

Broberg, A. & Keskinen, A. (2004). Geodatabase over Taita Hills, Kenya. In: Pellikka, P., J. Ylhäisi & B. Clark (eds.). *Taita Hills and Kenya, 2004 – seminar, reports and journal of a field excursion to Kenya*. Expedition Reports of Department of Geography, University of Helsinki 40. 148 pp. Department of Geography, University of Helsinki.

Chou, Y.H., Minnich, R.A. & Chase, R.A. (1993). Mapping probability of fire occurrence in San Jacinto Mountains, California, USA. *Environmental Management* 17, 129–140.

Chuvieco, E., Ventura, G., Pilar Martín, M. & Gomez, I. (2005). Assessment of multitemporal compositing techniques of MODIS and AVHRR images for burned land mapping. *Remote Sensing of Environment* 94, 450–462.

CIESIN (Center for International Earth Science Information Network), Columbia University, & Centro Internacional de Agricultura Tropical (CIAT). (2005). *Gridded population of the world version 3 (GPWv3): Population grids*. Palisades, NY: Socioeconomic Data and Applications Center (SEDAC), Columbia University. Available at http://sedac.ciesin.columbia.edu/gpw.

Clark, B. & Pellikka, P. (2005). The Development of a Land Use Change Detection Methodology for Mapping the Taita Hills, South-East Kenya: Radiometric Corrections. *Proceedings of the 31st International Symposium on Remote Sensing of Environment (ISRSE)*, 20-24 June, 2005, St Petersburg, Russian Federation. CD-Publication, no page numbers. (available from: http://www.helsinki.fi/geography/bclark_p-pellikka-ISRSE-paper.pdf).

Clark, B.J.F. & Pellikka, P.K.E. (2009). Landscape analysis using multiscale segmentation and object orientated classification. In: Röder, A. & Hill, J. (Eds.). *Recent Advances in Remote Sensing and Geoinformation Processing for Land Degradation Assessment*, pp. 323–342 (Taylor & Francis).

Clark, J. & Bobbe, T. (2007).Using Remote Sensing to Map and Monitor Fire Damage in Forest Ecosystems. In: Wulder, M. A. & Steven E. Franklin, S. E. (Eds.). *Understanding forest disturbance and spatial pattern: remote sensing and GIS approaches* pp. 246. CRC Press

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement* 20, 37–46.

Collett, D. (2003). *Modelling Binary Data*, 2nd ed. Chapman and Hall/CRC, Boca Raton, 387 pp.

Cropper, M. & Griffiths, C. (1994). The Interaction of Population Growth and Environmental Quality. *The American Economic Review* 84, No. 2, Papers and Proceedings of the Hundred and Sixth Annual Meeting of the American Economic Association. 250–254.

Devereux, S. (1993). *Water and Sanitation in East Caprivi: A Knowledge Attitudes and Practices Survey*. Social Sciences Division. Multi-Disiplinary Research Centre. University of Namibia.

Dobson, J.E., Bright, E.A., Coleman, P.R., Durfee, R.C. & Worley, B.A. (2000). LandScan: a global population database for estimating populations at risk. *Photogrammetric Engineering and Remote Sensing* 66, 849–857.

Doytsher, Y. (2000). A rubber sheeting algorithm for non-rectangular maps. *Computers & Geosciences* 26, 1001–1010.

Doytsher, Y. & Hall, J.K. (1997). Gridded affine transformation and rubber sheeting algorithm with Fortran program for calibrating scanned hydrographic survey maps. *Computers & Geosciences* 23, 785–791.

Dwyer, E., Pereira, J.M.C., Grégoire, J.M. & Da Camara, C.C. (2000). Characterization of the spatio-temporal patterns of global fire activity using satellite imagery for the period April 1992 to March 1993. *Journal of Biogeography* 27, 57–69.

Elith, J. & Leathwick. J. (2007). Predicting species distributions from museum and herbarium records using multiresponse models fitted with multivariate adaptive regression splines. *Diversity and Distributions* 13, 265–275.

Elith, J., Leathwick, J.R.& Hastie, T. (2008). A working guide to boosted regression trees. *Journal of Animal Ecology* 77, 802–813.

Elmasri, R. & Navathe, S.B. (1994). *Fundamentals of Database Systems*. 2ed. Redwood City, California: The Benjamin/Cummings Publishing Company, Inc. 873 pp.

Erdogan, E.H., Pellikka, P. & Clark, B. Impact of land cover change on soil loss in the Taita Hills, Kenya between 1987 and 2003. *International Journal of Remote Sensing*, accepted for publication.

Eva, H. & Lambin, E.F. (1998a). Remote sensing of biomass burning in tropical regions: sampling issues and multisensor approach. *Remote Sensing of the Environment* 64, 292–315.

Eva, H. & Lambin, E.F. (1998b). Burnt area mapping in Central Africa using ATSR data. *International Journal of Remote Sensing* 19, 3473–3497.

Eva, H. & Lambin, E.F. (2000). Fires and land-cover change in the tropics: A remote sensing analysis at the landscape scale. *Journal of Biogeography* 27, 765–776.

FAO, (2006). *Global Forest Resources Assessment 2005: Progress towards sustainable forest management*. FAO Forestry Paper 147, Rome.

Felderhof, L. & Gillieson, D. (2006). Comparison of fire patterns and fire frequency in two tropical savanna bioregions. *Austral Ecology* 31, 736–746.

Forester, B.C. (1985). An examination of some problems and solutions in monitoring urban areas from satellite platforms. *International Journal of Remote Sensing* 6, 39–151.

Flannigan, M.D., Krawchuk, M.A., De Groot, W.J., Wotton, B.M. & Gowman, L.M. (2009). Implications of changing climate for global wildland fire. *International Journal of Wildland Fire* 18, 483–507.

Freeman, E. (2007). *PresenceAbsence: an R Package for Presence–Absence Model Evaluation*. USDA Forest Service, Rocky Mountain Research Station, 507 25th street, Ogden, UT, USA. http://cran.r-project.org/

Freeman, E.A. & Moisen, G.G. (2008). A comparison of the performance of threshold criteria for binary classification in terms of predicted prevalence and Kappa. *Ecological Model*ling 217, 48–58.

Friedman, J. (1991). Multivariate adaptive regression splines. *The Annals of Statistics* 19, 1–67.

Friedman, J.H. (2001). Greedy function approximation: a gradient boosting machine. *The Annals of Statistics* 29, 1189–1232.

Fraley, C. & Raftery, A.E. (2002). Model-based clustering, discriminant analysis, and density estimation. *Journal of the American Statistical Association* 97, 611–631.

Fuller, D.O. (2000). Satellite remote sensing of biomass burning with optical and thermal sensors. *Progress in Physical Geography* 24, 543–561.

Fuller, D.O. & Fulk, M. (2001). Burned area in Kalimantan, Indonesia mapped with NOAA-AVHRR and Landsat TM imagery. *International Journal of Remote Sensing* 22, 691–697.

Fuller, D.O. (2006). Tropical forest monitoring and remote sensing: A new era of transparency in forest governance? *Singapore Journal of Tropical Geography* 27, 15–29. doi:10.1111/j.1467-9493.2006.00237.x

Fuse, T. & Shimizu, E. (2004). *Visualizing the landscape of old-time tokyo (Edo city)*. Proceedings of the ISPRS working group v/6 International archives of photogrammetry, Remote sensing and spatial information sciences Volume xxxvi, part 5/w1 Editors: Gruen, A., Murai, Sh., Fuse, T. , Remondino, F. Thursday 18 November 2004.

Gaston K.J., Scoble M.J. & Crook A. (1995). Patterns in species description: a case study using the Geometridae (Lepidoptera). *Biological Journal of the Linnean Society* 55, 225–237.

Guisan, A., & Zimmermann, N.E. (2000). Predictive habitat distribution models in ecology. *Ecological Modelling* 135, 147–186.

Gustafson, E.J., Hammer, R.B., Radeloff, V.C. & Potts, R.S. (2005). The relationship between environmental amenities and changing human settlement patterns between 1980 and 2000 in the Midwestern USA. *Landscape Ecology* 20, 773–789.

Hanski, I. (1998). Metapopulation dynamics. *Nature* 396, 41–49.

Hanski, I. (1999). *Metapopulation Ecology*. Oxford University Press Inc.: New York.

Hanski, I., Kuussaari, M. & Nieminen, M. (1994). Metapopulation structure and migration in the butterfly *Melitaea cinxia*. *Ecology* 75, 747–762.

Hanski, I., Pöyry, J., Pakkala, T. & Kuussaari, M. (1995). Multiple equilibria in metapopulation dynamics. *Nature* 377, 618–621.

Hanski, I., Moilanen, A., Pakkala, T. & Kuussaari, M. (1996). Metapopulation persistence of an endangered butterfly: a test of the quantitative incidence function model. *Conservation Biology* 10, 578–590.

Hanski, I. & Ovaskainen, O. (2000). The metapopulation capacity of a fragmented landscape. *Nature* 404, 755–758.

Hanski, I. & Meyke, E. (2005). Large-scale dynamics of the Glanville fritillary butterfly: landscape structure, population processes, and weather. *Annales Zoologici Fennici* 42, 379–395.

Hastie, T., & Tibshirani, R. (1990). *Generalized Additive Models*. Chapman and Hall, London. 335 pp.

Hastie, T., Tibshirani, R, & Buja, A. (1994). Flexible Disriminant Analysis by Optimal Scoring. *Journal of the American Statistical Association* 89, 1255–1270.

Hawbaker, T.J., Radeloff, V.C., Syphard, A.D., Zhu, Z. & Stewart, S.I. (2008). Detection rates of the MODIS active fire product in the United States. *Remote Sensing of Environment* 112, 2656–2664.

Hennessy, K.J., Lucas, C., Nicholls, N., Bathols, J.M., Suppiah, R. & Ricketts, J.R. (2005). *Climate change impacts on fire-weather in southeast Australia*. Consultancy report by CSIRO Marine and Atmospheric Research, Bureau of Meteorology and Bushfire CRC. 88. http://www.cmar.csiro.au/e-print/open/hennessykj_2005b.pdf.

Holloway, J.D. (1997). The moths of Borneo: Family Geometridae, subfamilies Sterrhinae and Larentiinae. *Malayan Nature Journal* 51, 1–242.

Holm, M., Lohi, A., Rantasuo, M., Väätäinen, S., Höyhtyä , T., Puumalainen, J., Sarkeala, J. & Sedano, F. (1999). *Creation of large image mosaics of airborne digital camera imagery.* In: Proceedings of the 4th International Airborne Remote Sensing Conference and Exhibition, vol. II, Ottawa, Canada, 21–24 June, pp. 520–526.

Holt-Jensen, A. (1988). *Geography. History and concepts* 2nd ed. Chapman, London. 162 pp.

Hording, N. (2004). *Justification for using photo interpretation methods to interpret satellite imagery*. (unpublished). Available at: http://nilerak.hatfieldgroup.com/French/NRAK/EO/10._Using_photo_interp_methods_to_interpret_sat_imagery_AMNH.pdf (accessed 9 October 2009).

Hurlbert, S.H. (1984). Pseudoreplication and the design of ecological field experiments. *Ecological Monographs* 54, 187–211.

Imbernon, J. (1999). Pattern and development of land use changes in Kenyan highland since 1950s. *Agriculture, Ecosystems and Environment* 76, 67–73.

Intergovernmental Panel on Climate Change. (2007). *Climate change 2007: Synthesis report. summary for policymakers*. Tech. rep. URL http://www.ipcc.ch/ipccreports/ar4-syr.htm

Jaetzold, R. & Schmidt, H. (1983). *Farm Management Handbook of Kenya*, vol.II. East Kenya. Ministry of Agriculture, Kenya.

Justice, C.O., Giglio, L., Korontzi, S., Owens, J., Morisette, J.T., Roy, D.P., Descloitres, J., Alleaume, S., Petitcolin, F. & Kaufman, Y. (2002). The MODIS fire products. *Remote Sensing of Environment* 83, 244–262.

Kaufman, Y.J., C., Tucker, C.J. & Fung, I. (1990). Remote sensing of biomass burning in the tropics, *Journal of Geophysical Research* 95, 9927–9939.

Kidson, J.W & Thompson C.S. (1998). A comparison of statistical and model-based downscaling techniques for estimating local climate variations. *Journal of Climate* 11, 735–753.

Korontzi, S., Roy, D.P., Justice, C.O. & Ward, D.E. (2004). Modeling and sensitivity of fire emissions in southern Africa during SAFARI (2000). *Remote Sensing of Environment* 92, 376–396.

Koutsias, N. & Karteris, M. (2000). Burned area mapping using logistic regression modelling of a single post-fire landsat-5 thematic mapper image. *International Journal of Remote Sensing* 21, 673–687.

Kumar, L., Skidmore, A.K. & Knowles, E. (1997). Modelling topographic variation in solar radiation in a GIS environment. *International Journal for Geographical Information Science* 11, 475–497.

Lambers, K. & Sauerbier, M. (2003) *A data model for a GIS-based analysis of the Nasca lines at Palpa (Peru)*. Institute of Geodesy and Photogrammetry, ETH Hoenggerberg. doi:10.3929/ethz-a-004599145.

Laris, P. (2005). Spatiotemporal problems with detecting and mapping mosaic fire regimes with coarse-resolution satellite data in savannah environments. *Remote Sensing of Environment* 99, 412–424.

Laurini, R. & Thompson, D. (1992). *Fundamentals of Spatial Information Systems*. Academic Press, Février, 680 pp.

Lehmann, A., Overton, J.M. & Leathwick, J.R. (2002). GRASP: generalized regression analysis and spatial predictions. *Ecological Modelling* 157, 189–207.

Lens, L., Van Dongen, S., Norris, K., Githiru, M. & Matthysen, E. (2002). Avian persistence in fragmented rainforest. *Science* 298, 1236–1238.

Lentile, L.B., Holden, Z.A., Smith, A. M.S., Falkowski, M.J., Hudak, A.T., Morgan, P., Lewis, S.A., Gessler, P.E. & Benson, N.C. (2006). Remote sensing techniques to assess active fire characteristics and post-fire effects. *International Journal of Wildland Fire* 15, 319−345.

Levine, J.S. (1996) Introduction. In: Levine, J. (Ed.). *Biomass burning and global change*, Vol. 1. MIT Press, Cambridge, Massachusetts. p. xxxv-xliii. 646 pp.

Levins, R. (1969). Some demographic and genetic consequences of environmental heterogeneity for biological control. *Bulletin of Entomological Society of America* 15, 237–240.

Li, G. & Weng, Q. (2005). Using Landsat ETM+ Imagery to Measure Population Density in Indianapolis, Indiana, USA. *Photogrammetric Engineering & Remote Sensing* 71, 947–958.

Lo, C.P. (1986). Accuracy of population estimation from medium-scale aerial photography. *Photogrammetric Engineering and Remote Sensing* 52, 1859–1869.

Lo, C.P. (1989). A Raster Approach to Population Estimation Using High-Altitude Aerial and Space Photographs. *Remote Sensing of Environment* 27, 59–71.

Lo, C.P. (1995). Automated Population and Dwelling Unit Estimation from High-Resolution Satellite Images: A GIS Approach. *International Journal of Remote Sensing* 16, 17–34.

Lozano, F.J., Suárez-Seoane, S. & de Luis, E. (2007). Assessment of several spectral indices derived from multi-temporal Landsat data for fire occurrence probability modelling. *Remote Sensing of Environment* 107, 533–544.

Lung, T. & Schaab, G. (2004). Change-detection in Western Kenya. The documentation of fragmentation and disturbance for Kakamega Forest and associated forest areas by means of remotely-sensed imagery. In: *ISPRS Archives, Vol. XXXV, Part B* (DVD), Proceedings of the ISPRS XXth Congress, 12-23 July 2004, Istanbul.

Lynch, H.J., Renkin, R.A., Crabtree, R.L. & Moorcroft, R.R. (2006). The influence of previous mountain pine beetle (Dendroctonus ponderosae) activity on the 1988 Yellowstone fires. *Ecosystems* 9, 1318–1327.

MacArthur, R.H. & Wilson, E.O. (1967). *The theory of island biogeography*. Princeton University Press, Princeton. 203 pp.

MacNally, R. (1996). Hierarchical partitioning as an interpretative tool in multivariate inference. *Australian Journal of Ecology* 21, 224–228.

MacNally, R. (2000). Regression and model-building in conservation biology, biogeography and ecology: the distinction between – and reconciliation of – 'predictive'and 'explanatory' models. *Biodiversity and Conservation* 9, 655–671.

Maeda, E.E., Formaggio,A.R., Shimabukuro,Y.E., Arcoverde, G.F.B. & Hansen, M.C. (2009). Predicting forest fire in the Brazilian Amazon using MODIS imagery and artificial neural networks. *International Journal of Applied Earth Observation and Geoinformation* 11, 265–272.

Malcolm, J.R., Liu, C., Ronald P. Neilson, R.P., Lara Hansen, L. & Hannah, L. (2006). Global Warming and Extinctions of Endemic Species from Biodiversity Hotspots. *Conservation Biology* 20, 538–548.

Malingreau, J.P., Achard, F., D'Souza, G., Stibig, H.J., D'Souza, J., Estreguil, C. & Eva, H. (1995). AVHRR for global tropical forest monitoring: The lessons of the TREES project. *Remote Sensing Reviews* 12, 29–40.

Mayaux, P. & Lambin, E.F. (1995). Estimation of tropical forest area from coarse spatial resolution data: a two-step correction function for proportional errors due to spatial aggregation, *Remote Sensing of Environment* 53, 1–16.

Mayaux, P. & Lambin, E.F. (1997). Tropical forest area measured from global land-cover classifications : inverse calibration models based on spatial textures. *Remote Sensing of Environment* 59, 29-43.

McCullagh, P. & Nelder, J.A. (1989). *Generalized Linear Models*. 2nd ed. Chapman and Hall, New York. 511 pp.

Melillo, J.M., Houghton, R.A., Kicklighter, D.W. & McGuire, A.D. (1996). Tropical deforestation and the global carbon budget. *Annual Review of Energy & Environment*. 21, 293–310.

Mendolsohn, J. & Roberts, C. (1997). *An environmental profile and atlas of Caprivi*. Directorate of Environmental Affairs, Namibia.

Miettinen, J., Langner, A. & Siegert, F. (2007). Burnt area estimation for the year 2005 in Borneo using multi-resolution satellite imagery. *International Journal of Wildland Fire* 16, 45–53.

Moilanen, A. & Hanski, I. (1998). Metapopulation Dynamics: Effects of Habitat Quality and Landscape Structure. *Ecology* 79, 2503–2515.

Moisen, G.G. & Frescino, T.S. (2002). Comparing five modelling techniques for predicting forest characteristics. *Ecological Modelling* 157, 209–225.

Morin, X. & Thuiller, W. (2009). Comparing niche- and process-based models to reduce prediction uncertainty in species range shifts under climate change. *Ecology* 90, 1301–1313.

Mubareka, S., Ehrlich, D., Bonn, F. & Kayitakire, F. (2008). Settlement location and population density estimation in rugged terrain using information derived from Landsat ETM and SRTM data. *International journal of remote sensing* 29, 2339–2357.

Myers N., Mittermeier R.A., da Fonseca G.A.B. & Kent, J. (2000). Biodiversity hotspots for conservation priorities. *Nature* 403, 853–858.

Pellikka, P., Clark, B., Hurskainen, P., Keskinen, A., Lanne, M., Masalin, K., Nyman-Ghezelbash, P. & Sirviö, T. (2004). Land use change monitoring applying geographic information systems in the Taita Hills. *Proceedings of the 5th AARSE conference* (African Association of Remote Sensing of the Environment), 18–21 October, 2004, Nairobi, Kenya. CD-Publication.

Peters, J., Baets, B.D., Verhoest, N.E., Samson, R., Degroeve, S., Becker, P.D. & Huybrechts, W. (2007). Random forests as a tool for ecohydrological distribution modelling. *Ecological Modelling* 207, 304–318.

Podest, E., & Saatchi, S. (2002). Application of multiscale texture in classifying JERS-1 radar data over tropical vegetation. *International Journal of Remote Sensing* 23, 1487–1506.

Porter, P.W. (1956). *Population Distribution and Land Use in Liberia*. PhD thesis, London School of Economics and Political Science, London.

Prasad, A.M., Iverson, L.R. & Liaw, A. (2006). Newer classification and regression tree techniques: bagging and random forests for ecological prediction. *Ecosystems* 9, 181–199.

Preisler, H.K., Brillinger, D.R., Burgan, R.C. & Benoit, J.W. (2004). Probability based models for estimation of wildfire risk. *International Journal of Wildland Fire* 13, 133–142.

Preston, S. (1996). The effect of population growth on environmental quality. *Population Research and Policy Review* 15, 95–108.

Qu, J.J., Wang, W., Dasgupta, S. & Hao, X. (2008). Active fire monitoring and fire danger potential detection from space: A review. *Frontiers of Earth Science in China* 2, 479–486

Republic of Kenya (2001). *The 1999 Population & Housing Census*. Central Bureau of Statistics, Ministry of Planning and National Development, Kenya.

Ripley, B.D. (1996). *Pattern Recognition and Neural Networks*. Cambridge University Press.

Roy, D.P., Giglio, L., Kendall, J.D. & Justice, C.O. (1999). Multi-temporal active-fire based burn scar detection algorithm. *International Journal of Remote Sensing* 20, 1031–1038.

Roy, D.P., Lewis, P.E. & Justice, C.O. (2002). Burned area mapping using multi-temporal moderate spatial resolution data – a bi-directional reflectance model-based expectation approach. *Remote Sensing of Environment* 83, 263–286.

Roy, D.P., Jin, Y., Lewis, P.E. & Justice, C.O. (2005). Prototyping a global algorithm for systematic fire-affected area mapping using MODIS time series data. *Remote Sensing of Environment* 97, 137–162.

Roy, D.P., Boschetti, L., Justice, C.O. & Ju, J. (2008). The collection 5 MODIS burned area Product–global evaluation by comparison with the MODIS active fire product. *Remote Sensing of Environment* 112, 3690–3707.

Rumelhart, D.E., Hinton, G.E. & Williams, R.J. (1986). Learning internal representations by error propagation. *Nature* 323, 533–536.

Rumelhart, D. & McClelland, J. (1986). *Parallel Distributed Processing*. MIT Press, Cambridge.

Rushton S.P., Ormerod, S.J. & Kerby, G. (2004). New paradigms for modelling species distributions? *Journal of Applied Ecology* 41, 193–200.

Sá, A.C.L., Pereira, J.M.C., Vasconcelos, M.J.P., Silva, J.M.N., Ribeiro, N. & Awasse, A. (2003). Assessing the feasibility of sub-pixel burned area mapping in miombo woodlands of Northern Mozambique using MODIS imagery. *International Journal of Remote Sensing* 24, 1783–1796.

Sawada, M. (1999). ROOKCASE: an Excel 97/2000 Visual Basic (VB) add-inn for exploring global and local spatial autocorrelation. *Bulletin of Ecological Society of America* 80, 231–234.

Schaab, G., Lung, T. & Mitchell, N. (2005). Land use/cover change analyses based on remotely-sensed imagery and old maps as means to document fragmentation and disturbance for East-African rainforests over the last ca. 100 years. In: CD-ROM *Proceedings of the International Cartographic Conference 2005*, 9-16 July 2005, A Coruña (Spain).

Schnaiberg, J., Riera, J., Turner, M.G. & Voss, P.R. (2002). Explaining human settlement patterns in a recreational Lake District: Vilas County, Wisconsin, USA. *Environmental Management* 30, 24–34.

Scoble, M.J. (1999) (Ed.). *Geometrid Moths of the World: A Catalogue* (Lepidoptera, Geometridae). Apollo Books, Stenstrup, Denmark. 1280 pp.

Sgrenzaroli M, de Grandi G.F, Eva, H., & Achard, F. (2002). Tropical forest cover monitoring: estimates from the GRFM JERS-1 radar mosaics using wavelet zooming techniques and validation. *International Journal of Remote Sensing* 7, 1329–55.

Shaban, M.A. & Dikshit, O. (2001). Improvement of classification in urban areas by the use of textural features: the case study of Lucknow city, Uttar Pradesh. *International Journal of Remote Sensing* 22, 565–593.

Silva J.M.N, Cadima, J.F.C.L., Pereira, J.M.C. & Grégoire J.M. (2004). Assessing the feasibility of a global model for multitemporal burned area mapping using SPOT-VEGETATION data. *International Journal of Remote Sensing* 25, 4889–4913.

Setzer, A.W. & Malingreau, J.P. (1996). AVHRR monitoring of vegetation fires in the tropics: towards a global product. In: Levine, J.S. (Ed.). *Biomass burning and global change*, Cambridge: MIT Press, 48–81.

Stocks, B.J., Fosberg, M.A., Lynham, T.J., Mearns, L., Wotton, B.M., Yang, Q., Jin, J.Z., Lawrence, K., Hartley, G.R., Mason, J.A. & McKenney, D.W. (1998). Climate change and forest fire potential in Russian and Canadian Boreal Forests. *Climatic Change* 38, 1–13.

StoraEnso (2003). *EnsoMOSAIC, Mosaicking user's guide 5.00*. StoraEnso Forest Consulting Oy Ltd.

Stroppiana, D., Pinnock, S., Pereira, J.M.C. & Grégoire, J.M. (2002). Radiometric analysis of SPOT-VEGETATION images for burnt area detection in Northern Australia. *Remote Sensing of Environment* 82, 21–37.

Stroppiana, D., Grégoire, J.M. & Pereira, J.M.C. (2003). The use of SPOT VEGETATION data in a classification tree approach for burnt area mapping in Australian savanna. *International Journal of Remote Sensing* 24, 2131–2151.

Sutton, P., Roberts, C., Elvidge, C. & Meij, H. (1997). A comparison of nighttime satellite imagery and population density for the continental united states. *Photogrammetric Engineering and Remote Sensing* 63, 1303–1313.

Swets, K. (1988). Measuring the accuracy of diagnostic systems. *Science* 240, 1285–1293.

Syphard, A.D., Volker, C., Radeloff, V.C., Keeley, J.E., Hawbaker, T.J., Clayton, M.K., Stewart, S.I. & Hammer, R.B. (2007). Human influence on California fire regimes. *Ecological Applications* 17, 1388–1402.

Tansey, K., Grégoire, J.M., Stroppiana, D., Sousa, A., Silva, J., Pereira, J.M.C., Boschetti, L., Maggi, M., Brivio, P.A., Fraser, R., Flasse, S., Ershov, D., Binaghi, E., Graetz, D. & Peduzzi, P. (2004). Vegetation burning in the year 2000: Global burned area estimates from SPOT VEGETATION data. *Journal of Geophysical Research* 109. D14S03.

Tansey, K., Grégoire, J.M.C., Defourny, P., Leigh, R., Pekel, van Bogaert, E., Bartholomé, E. & Bontemps, S. (2008). A new, global, multi-annual (2000-2007) burned area product at 1 km resolution and daily intervals. *Geophysical Research Letters*, VOL. 35, L01401, doi:10.1029/ 2007 GL031567

Thuiller, W. (2003). BIOMOD—optimizing predictions of species distributions and projecting potential future shifts under global change. *Global Change Biology* 9, 1353–1362.

Thuiller, W., Araújo, M.B. & Lavorel, S. (2003). Generalized models vs. classification tree analysis: Predicting spatial distributions of plant species at different scales. *Journal of Vegetation Science* 14, 669–680.

Thuiller, W., Lafourcade, B., Engler, R. & Araújo, M.B. (2009). BIOMOD – a platform for ensemble forecasting of species distributions (Version 1.1-0). *Ecography*, 32, 369–373.

UNEP/GRID, 2006, *UNEP/GRID spatial data clearinghouse*. Available at: http://grid2.cr.usgs.gov/datasets/datalist.php (accessed 9 October 2009).

Venables, W.N. & Ripley, B.D. (2002). *Modern Applied Statistics with S*. 4th ed. Springer–Verlag, New York. 495 pp.

Verlinden, A. & Laamanen, R. (2006). Long term fire scar monitoring with remote sensing in northern Namibia: relations between fire frequency, rainfall, land cover, fire management and trees. *Environmental Monitoring and Assessment* 112, 231–253.

von Breitenbach, F. (1968). Long-term Plan of Forestry Development in the Eastern Caprivi Zipfel. Department of Forestry, Saasveld, South Africa.

von Humboldt, A. & Bonpland, A. (1805). *Essai sur la géographie des plantes : accompagné d'un tableau physique des régions équinoxiales, fondé sur des mesures exécutées, depuis le dixième degré de latitude boréale jusqu'au dixième degré de latitude australe, pendant les années 1799, 1800, 1801, 1802 et 1803*. Chez Levrault, Schoell et compagnie, libraires, XIII–1805. Paris.

Vuorela, N. & Toivonen, T. (2003). Can the past suggest the present day biodiversity? – detecting and classifying change transitions in the landscape. In: Mander, Antrop, M. (eds.): *Multifunctional Landscapes - Vol III: Continuity and Change*, 135-166. WIT Publications.

Walsh, C., MacNally, R. (2003). *Hierarchical Partitioning: R Project for Statistical Computing*. http://cran.r-project.org/

White, M.S. & Griffin, P. (1985). Piecewise linear rubber-sheet map transformations, *The American Cartographer* 12, 123–131.

Willig, M.R., Kauffman D.M. & Stevens R.D. (2003). Latitudinal gradients of biodiversity: Pattern, process, scale, and synthesis. *Annual review of ecology, evolution, and systematics* 34, 273–309.

Woodcock, C.E., Scott, A., Macomber, S.A. & Kumar, L. (2003). Vegetation mapping and monitoring In: Skidmore, A. (Ed.) *Environmental Modelling with GIS and Remote Sensing*. pp. 268 Taylor & Francis.

Yee, T.W. & Mitchell, N.D. (1991). Generalized Additive Models in Plant Ecology. Journal of Vegetation Science 2, 587-602.

Zar, J.H. (1999). *Biostatistical Analysis*, 4th ed. Prentice-Hall, London. 929 pp.

Zhang, Y.H., Wooster, M.J., Tutubalina, O. & Perry, G.L.W. (2003). Monthly burned area and forest fire carbon emission estimates for the Russian Federation from SPOT VGT. *Remote Sensing of Environment* 87, 1–15.

Zimmermann, N.E. (2000). Shortwavc.aml, 9.08.2000. http://www.wsl.ch/staff/niklaus.zimmermann/programs/aml1_1.html.

Appendices 1–6

Examples of ArcView 3.2 Avenue scripts, ARC/INFO AML macros, and ArcGIS geoprocessing model used for GIS analysis and process automation in different studies.

Appendix 1. ArcView 3.2 Avenue script to calculate subtotals for attribute table (Paper I).

```
' valisumma.ave
' made by M.H SYKE and M.S HY 2002
' ensin avaa aktiivinen attribuuttitaulu
taulu = av.GetActiveDoc
if (taulu.Is(Table)) then
  theVTab = taulu.GetVTab
  ' etsitään kentät
  k_paiva = theVTab.FindField("paiva")
  k_arvo = theVTab.FindField("arvo")
  uusi_k = theVTab.FindField("uusi_arvo")
  ' mikäli uusi_arvo kenttää ei ole, niin tehdään se
  '
  if (uusi_k = NIL) then
    uusi_k = Field.Make ("uusi_arvo", #FIELD_DECIMAL , 4, 0)
    theVTab.SetEditable(true)
    if (theVTab.IsEditable) then
      theVTab.addfields({uusi_k})
    end
    theVTab.SetEditable(false)
  end
  uusi_arvo = 0
  ed_arvo = "999"
  ' tehdään dictionary, jonne valisumma tiedot tallennetaan
  '
  uusi_Dic = Dictionary.Make(11)
  ' käydään loopilla läpi taulun arvot
  '
  for each rec in theVTab
    paiva = theVTab.ReturnValue(k_paiva,rec)
    arvo = theVTab.ReturnValue(k_arvo,rec)
    if (arvo > 0) then
      uusi_arvo = uusi_arvo + arvo
    else
      uusi_arvo = 0
    end
    if (ed_arvo.asstring <> arvo.asstring) then
      uusi_Dic.Empty
    end
    onko = uusi_Dic.Add(paiva,rec.asstring)
    ' käydään läpi dictionaryyn tallennetut arvot, viimeisellä
    ' kierroksella välisumma tallentuu oikein
    '
    for each a in uusi_dic.returnkeys.clone
        r = uusi_dic.Get(a).AsNumber
        theVTab.SetEditable(true)
        if (theVTab.IsEditable) then
           theVTab.SetValue(uusi_k,r,uusi_arvo.AsString)
        end
        theVTab.SetEditable(false)
     end
    ed_arvo = arvo
  end
```

```
end
' scriptin loppu
```

Appendix 2. ArcView 3.2 Avenue script to create Åland Islands 1 km^2 analysis squares (Paper I) and world-wide analysis grid for Scopulini moths (Paper II).

```
'*********************************************************************
' Nimi:  Analyysi_Gridi.ave
' Title:  Creates Analysis Grid (vector polygon) shapefile
' Modified by Mika Siljander in 2002 original by Francisco Olivera
' The View Map units and distance units have to be UNKNOWN
'********************************************************************
theView=av.GetActiveDoc
viewUnits=theView.GetDisplay.GetUnits
viewProj=theView.GetProjection
' Defining the rectangle
' *** theRect=theView.ReturnUserRect
theLabels ={"ALA VASEN X","ALA VASEN Y","YLÄ OIKEA X","YLÄ OIKEA Y"}
theDefaults = {"","","",""}
theParameters = MsgBox.MultiInput("Valitse gridin Parametetrit",
"Gridin Parameterit",theLabels,theDefaults)
x1 = theParameters.Get(0).AsNumber
y1 = theParameters.Get(1).AsNumber
x2 = theParameters.Get(2).AsNumber
y2 = theParameters.Get(3).AsNumber
therect = Rect.MakeXY(x1,y1,x2,y2)
if (theRect=nil) then
  exit
end
llx=therect.getleft
lly=therect.getBottom
theWidth=therect.getwidth
theHeight=therect.getheight
rows=Msgbox.Input("rivien määrä (ylös/alas)","Analyysi Gridi","10")
if ((rows = nil) or (rows.asNumber < 1)) then
  MsgBox.Error("Must use more than one row.","Aborting")
  exit
end
cols=Msgbox.Input("sarakkeiden määrä (vasen/oikea):","Analyysi Gridi","10")
if ((cols =nil) or (cols.asNumber < 1)) then
  MsgBox.Error("Must use more than one column.","Aborting")
  exit
end
rows=rows.asnumber
cols=cols.asnumber
rsize=theheight/rows
csize=thewidth/cols
letters="ABCDEFGHIJKLMNOPQRSTUVWXYZ"
newshpname=filedialog.Put("AnalysisGrid.shp".AsFilename,"*.shp","Analysis
Grid file")
if (newshpname =nil) then
  exit
end
gridftab=Ftab.MakeNew(newshpname,POLYGON)
labelfield=Field.Make("Label",#FIELD_CHAR,6,0)
gridftab.SetEditable(true)
gridftab.Addfields({labelfield})
gridftab.SetEditable(True)
shpfield=gridftab.FindField("shape")
labfield=gridftab.FindField("label")
```

```
countup=0
sizer=rows*cols
av.setstatus(0)
Av.ShowMsg("Analyysi ruudukkoa luodaan...")
for each r in 1..rows
  row_id=(rows-r+1).AsString
  for each c in 1..cols
    countup=countup+1
    av.setstatus(countup/sizer*100)
    if ((c/26) > 1) then
      col_prefix=letters.middle((c/26).floor-1,1)
    else
      col_prefix=""
    end
    if (c.Mod(26) = 0) then
      col_name=letters.Middle(25,1)
    else
      col_name=letters.Middle(c.Mod(26)-1,1)
    end
    col_id=col_prefix+col_name
   originx=llx+((c-1)*csize)
   originy=lly+((r-1)*rsize)
   size=csize@rsize
   theOrigin=originx@originy
    rct=rect.make(theOrigin,size)
    if (viewunits=#UNITS_LINEAR_METERS) then
      rct=rct.ReturnUnprojected(viewproj)
    end
    newrec=gridftab.AddRecord
    gridftab.SetValue(shpfield,newrec,rct.aspolygon)
    gridftab.SetValue(labfield,newrec,col_id+row_id)
  end
end
gridftab.SetEditable(false)
av.clearMsg
mytheme=ftheme.Make(gridftab)
mysym=av.GetSymbolWin.GetPalette.GetList(#PALETTE_LIST_FILL).Get(0)
mysym.setolwidth(0.1)
mytheme.GetLegend.GetSymbols.Set(0,mysym)
mytheme.SetName("AnalysisGrid")
mytheme.SetActive(true)
myTheme.SetVisible(true)
theview.AddTheme(mytheme)
mytheme.Invalidate(true)
theview.GetDisplay.Flush
yn=MsgBox.YesNo("Tekstitä analyysi Gridi?","Analyysi Gridi",true)
if (yn) then
  for each thm in theView.GetThemes
    if (thm <> mytheme) then
      thm.SetActive(False)
    end
  end
  theView.LabelThemes(false)
end
```

Appendix 3. ARC/INFO AML macro to automatically create raster GRIDS and shapefile point layers from MODIS fire hot spots data. (Papers V and VI)

```
/**********************************************************************
/*
/* AML macro for creating Grids and pointcoverages from MODIS fire data
/* Looping for: imagegrid,projectgrid and gridpoint commands by Mika
/* Siljander /* 31.5.2006 21.00
/*
/**********************************************************************
/* -----------------------------------------------------------------
/* input images are ERDAS IMAGINE *.img named m1.img to m46.img
/* output will be grids named caprivi1 to caprivi46 and point shapefiles
/* shpnt1.shp to shpnt46.shp
/* Input Projection is sinusoidal from MODIS (HDF) fire hot spot data
/* Output coordinate system is Lambert azimuthal equal area
/* projection file is project.txt in working directory (look below)
/* -----------------------------------------------------------------
/* projection.txt file
/* --------------------------------------------------------------
/* INPUT
/* PROJECTION SINUSOIDAL
/* UNITS meters
/* DATUM WGS84
/* PARAMETERS
/* 6371007.1809184756 /* Radius of the sphere of reference
/* 0 0  0.000 /* Longitude of center of projection
/* 0.0 /* False easting (meters)
/* 0.0 /* False northing (meters)
/* OUTPUT
/* Projection    LAMBERT_AZIMUTHAL
/* Zunits        NO
/* Units         METERS
/* Xshift        0.0000000000
/* Yshift        0.0000000000
/* Parameters 6370997.0000000000  0.0000000000
/* 6370997.00000 /* radius of the sphere of reference
/* 20 0 0.000 /* longitude of center of projection
/* 5 0 0.000 /* latitude of center of projection
/* 0.00000 /* false easting (meters)
/* 0.00000 /* false northing (meters)
/* END
/* --------------------------------------------------------------
/* Macro starts from here
&type  *************************************************************
&type  * Macro for creating Grids and point coverages from MODIS fire data
&type  * Looping for: imagegrid,projectgrid, gridpoint, arcshape commands
&type  *          made by Mika Siljander 31.5.2006 21.00
&type  *
&type  *************************************************************
&pause &seconds 2
&type macro is starting to run
&pause &seconds 2
/* imagegrid loop
&do i := 1 &to 46
  &type image to grid converting %i% image
  imagegrid m%i%.img m%i%
 &end
grid
/* projectgrid loop
&do i := 1 &to 46
```

```
  &type projecting %i% image
 capritmp%i% = PROJECTGRID(m%i%, project.txt, NEAREST)
&end
setwindow 324585.375  -2595160.250 568285.375 -2482560.250
setmask capri_mask
setcell 1000
&type SET WINDOW, SETMASK and SETCELL set to Caprivi Mask
&pause &seconds 2
/* caprivi tmp loop after setwindow
&do i := 1 &to 46
  &type  %i% GRID coping
 caprivi%i% = capritmp%i%
&end
q
/* deleting capritmp grids
&do i := 1 &to 46
  &type DELETING temporary projected m%i% GRID
 kill capritmp%i% all
 &end
/* deleting temporary m grids
 &do i := 1 &to 46
   &type DELETING temporary projected m%i% GRID
  kill m%i% all
  &end
grid
/* gridpoint loop
&do i := 1 &to 46
  &type GRIDPOINT Processing %i% GRID
 capnt%i% = gridpoint(caprivi%i%, Grid_code)
&end
setmask off
q
/* looping coverages to shapefiles
&do i := 1 &to 46
  &type converting to shapefile %i% processing
  arcshape capnt%i%  points  shpnt%i%
&end
&do i := 1 &to 46
  &type DELETING GRIDS m%i% GRID
 kill caprivi%i% all
 &end
&do i := 1 &to 46
  &type DELETING GRIDS m%i% GRID
 kill capnt%i% all
 &end
```

Appendix 4. Arc/Info AML macro to interpolate rainfall data using IDW interpolation method (Papers V and VI).

```
/*******************************************************************
/* IDW interpolation of rainfall
/* weather stations (rstation coverage) from Namibia, Botswana and Zambia
/* created by Mika Siljander 24.6.2006 21.00
/* capri_mask GRID to masking
/*
/*******************************************************************
grid
/* mask grids needs to be in the same directory
setcell rmaskgrd
setwindow rmaskgrd
setmask grd_mask
```

```
&type SET WINDOW, SETMASK and SETCELL done
&pause &seconds 2
&do i := 1 &to 12
&type Interpolate %i% day
r_%i%_temp = idw(rstation, month%i%, #, #, sample, #, #, 2000)
r%i%month = r_%i%_temp
kill r_%i%_temp all
&end
&type %i% month IDW interpolation completed
&pause &seconds 1
rgrd_ave = int(idw(rstation, month%i%, #, #, sample, #, #, 2000)
&type IDW interpolations done
&pause &seconds 1
/* gridpoint loop
&do i := 1 &to 12
  &type GRIDPOINT Processing %i% GRID
 rain_%i%p = gridpoint( r%i%month, value)
&end
r_ave = gridpoint(rgrd_ave, value)
&type Converting GRIDS to points done
&pause &seconds 2
q
/* looping coverages to shapefiles
&do i := 1 &to 12
  &type converting to shapefile %i% processing
  arcshape rain_%i%p  points  r_pnt%i%
&end
&type Converting to shapefiles done
&pause &seconds 2
arcshape r_ave points rain_ave
setmask off
q
```

Appendix 5. ARC/INFO AML macro to calculate monthly Thornthwaite's potential evapotranspiration (PET) for East Caprivi study area. (Paper VI)

```
&type ***************************************
&type * Macro to calculate Thornthwaite PET
&type *   (Potential evapotranspiration)
&type *      using Worldclim data
&type * copyright© Mika Siljander 24 May 2009
&type ***************************************
/* &type PET calculation in brief:
/*&type Potential evapo-transpiration (mm/day),
/*&type   PET = if Ta > 0 then dl*16*(10*Ta/I)^a
/*&type   if Ta > 26.5 then b = -415.85 + 32.24 * t%i% - 0.43 * ta^2
/*&type  else 0
/*&type where
/*&type   Ta is mean monthly temperature (Celsius)
/*&type   a = 0.49+0.0179*I-7.71*10^ -5*I^2+6.75*10^ -7*I^3
/*&type   dl = daylength in hours / 12
/*&type   I = sum(i)
/* &type   i is a monthly heat index given by
/* &type     i = if Ta>0 then (Ta/5)^1.5
/* &type         else 0
/* &type start now. For large grid the calculation can take long time
/* &type For one Worldclim area calculation takes up to one hour
/* Macro to calculate Thornthwaite PET
/* created by Mika Siljander 24 May 2009
/* Use the mean value GRIDS tiles from Worldclim Internet site
/* http://www.worldclim.org/tiles.php
```

```
/* Macro will first convert *.bil images to GRIDS before calculations
/* Tarvitaan lämpötilan keskiarvo gridit:
/* joka kuukaudelle lasketaan heat index gridit:
/* temp is the average temperature of the i:th month in degrees C.
/* start from Arc promt
/* If you want to calculate PET for the whole tile
/* then remove the mask part of the code
/* This macro uses mask GRID layer for East Caprivi
/* this macro change coord. system from geographic to Lambert-Azimuthal
/* project.txt is needed in working directory
/* If other projection is needed adjust output parameters.
/* create a separate project.txt file as below
/* INPUT
/* Projection    GEOGRAPHIC
/* Datum         WGS84
/* Zunits        NO
/* Units         DD
/* Spheroid      WGS84
/* Xshift        0.0000000000
/* Yshift        0.0000000000
/* Parameters
/* OUTPUT
/* Projection    LAMBERT_AZIMUTHAL
/* Zunits        NO
/* Units         METERS
/* Xshift        0.0000000000
/* Yshift        0.0000000000
/* Parameters 6370997.0000000000  0.0000000000
/* 6370997.00000 /* radius of the sphere of reference
/* 20  0  0.000 /* longitude of center of projection
/* 5  0  0.000 /* latitude of center of projection
/* 0.00000 /* false easting (meters)
/* 0.00000 /* false northing (meters)
/* END
/* Caprivi mask grid copied from
/* remove copy part of code if no mask grid needed
copy C:\Temp\xCapclim\caprivi\capri_mask  C:\Temp\xCapclim\capri_mask
&type EAST CAPRIVI MASK GRID copied
/* this part of code is from Worldclim Internet pages
&TERMINAL 9999
&s program [locase [show program]]
&if %program% ^= grid &then grid
/* &if [exists yyy -grid] &then kill yyy
/*&if [exists yyy2 -grid] &then kill yyy2
/*&if [exists zzz -grid] &then kill zzz
/*&if [exists zzz2 -grid] &then kill zzz2
/*&if [exists zzz3 -grid] &then kill zzz3
/*&do i := 1 &to 60
/* &if [exists z%i% -grid] &then kill z%i%
/*&end
&do climvar &list tmean tmin tmax prec bio alt
  &type Starting import of %climvar%
  &sv nvars = 12
  &if %climvar% = bio &then &sv nvars = 19
  &if %climvar% = alt &then &sv nvars = 1
  &do m := 1 &to %nvars%
    &if %climvar% = alt &then
    &do
      &if [exists %climvar% -grid] &then kill %climvar%
      &s bils := [listfile %climvar%*.bil -image]
      &if [null %bils%] &then &type There are no tiles for %climvar% month %m%
```

```
    &end
    &else
    &do
      &if [exists %climvar%%m% -grid] &then kill %climvar%%m%
      &s bils := [listfile %climvar%%m%_*.bil -image]
      &if [null %bils%] &then &type There are no tiles for %climvar% month %m%
    &end
    &if [null %bils%] &then &type next
    &else
    &do
      &s num := [token %bils% -count]
      &do i := 1 &to %num%
        &sv name = [extract %i% %bils%]
        &if %i% < 31 &then
        &do
          &if %i% = 1 &then &sv listnames = z%i%
          &else &sv listnames = %listnames%, z%i%
        &end
        &else
        &do
          &if %i% = 31 &then &sv listnames2 = z%i%
          &else &sv listnames2 = %listnames2%, z%i%
        &end
        arc imagegrid %name% z%i%
      &end
      &type merge
      yyy = merge(%listnames%)
      zzz = con(yyy >= 32768, yyy - 65536, yyy)
      kill yyy
      &if %num% > 30 &then
      &do
        yyy2 = merge(%listnames2%)
        zzz2 = con(yyy2 >= 32768, yyy2 - 65536, yyy2)
        kill yyy2
        zzz3 = merge(zzz, zzz2)
        kill zzz
        kill zzz2
        %climvar%%m% = setnull(zzz3 == -9999, zzz3)
        kill zzz3
      &end
      &else
      &do
        %climvar%%m% = setnull(zzz == -9999, zzz)
        kill zzz
      &end
      &do i := 1 &to %num%
        kill z%i%
      &end
      arc projectdefine grid %climvar%%m%
        projection geographic
        units dd
        datum WGS84
        parameters
      &if %climvar% = alt &then rename alt1 alt
      &type %climvar%%m% done
    &end
  &end
&end
&type coversion of BIL files to GRIDS finished successfully!
/* Code from Worldclim Internet pages ends
/* Laskenta alkaa - NOW STARTS CALCULATION!
```

```
&type START temperature adjusting
&do i := 1 &to 12
   &type temperature float GRID number %i% is processing now
tf%i% = float(tmean%i%)
   &end
&do i := 1 &to 12
   &type devide by 10 all temperature float grids. GRID number %i% is processing
t%i% = tf%i% / 10
   &end
   /* projectgrid loop
&do i := 1 &to 12
  &type projecting t%i% GRID
 capritmp%i% = PROJECTGRID(t%i%, project.txt, NEAREST)
&end
/*mask part of the code
setwindow 324585.375  -2595160.250 568285.375 -2482560.250
setmask capri_mask
setcell 1000
&type SET WINDOW, SETMASK and SETCELL set to Caprivi Mask
/* caprivi tmp loop after setwindow
&do i := 1 &to 12
  &type  %i% GRID coping
 caprivi%i% = capritmp%i%
&end
/* deleting capritmp grids
&do i := 1 &to 12
  &type DELETING temporary projected m%i% GRID
 kill capritmp%i% all
 &end
&do i := 1 &to 12
  &type DELETING temporary projected t%i% GRID
 kill t%i% all
 &end
 &do i := 1 &to 12
  &type  %i% GRID coping
 t%i% = caprivi%i%
&end
/* deleting temporary m grids
 &do i := 1 &to 12
   &type DELETING temporary projected m%i% GRID
  kill caprivi%i% all
  &end
  /* ----------------------------------------------------------
&do i := 1 &to 12
   &type delete all float grids. Delete GRID %i% now.
kill tf%i% all
   &end
/* temperature devide by 5
 &do i := 1 &to 12
   &type devide by 5 tempeature GRID %i% now.
 tdiv%i% = t%i% / 5
   &end
/* tdevide GRID POW by 1.514
&do i := 1 &to 12
 &type TEMP devide POW by 1.514 GRID %i% ready now.
tpow%i% = pow(tdiv%i% , 1.514)
 &end
&type ALL TEMP/5 POW by 1.514 are ready now.
&do i := 1 &to 12
   &type delete all tdiv grids. Delete GRID %i% now.
kill tdiv%i% all
```

```
   &end
/* outgrid = pow(ingrid1
/* con([t1]  > , ([t1] / 5)Pow(1.514) , [t1])
&type calculate conditional statements for 12 temperature GRIDS
&type if pixel in t1 GRID > 0, then give value tpow1, otherwise 0.
ht1 = con(t1 > 0, tpow1, 0)
ht2 = con(t2 > 0, tpow2, 0)
ht3 = con(t3 > 0, tpow3, 0)
ht4 = con(t4 > 0, tpow4, 0)
ht5 = con(t5 > 0, tpow5, 0)
ht6 = con(t6 > 0, tpow6, 0)
ht7 = con(t7 > 0, tpow7, 0)
ht8 = con(t8 > 0, tpow8, 0)
ht9 = con(t9 > 0, tpow9, 0)
ht10 = con(t10 > 0, tpow10, 0)
ht11 = con(t11 > 0, tpow11, 0)
ht12 = con(t12 > 0, tpow12, 0)
&do i := 1 &to 12
   &type delete all tpow grids. Delete GRID %i% now.
kill tpow%i% all
   &end
/* con(tempgrd%i% > 0, (tempgrg%i% / 5)^1.514,tempgrd%i%
&type Monthly Heat index grids ready
/* sum ht grids to get annual heat index grid
I = ht1 + ht2 + ht3 + ht4 + ht5 + ht6 + ht7 + ht8 + ht9 + ht10 + ht11 + ht12
&type Annual Heat index SUM GRID (I) is now ready
I2 = I * I
&type I^2 GRID ready
I3 = I * I * I
&type I^3 GRID ready
&type Annual heat index sum GRIDS (I) is ready
&do i := 1 &to 12
   &type delete all ht grids. Delete GRID %i% now.
kill ht%i% all
   &end
/* kirjan mukaan m = (6.75x10–7) I^3–(7.71x10–5) I^2+(1.79x10-2)I+0.492
mgrd1 = (I3 * 0.000000675)
mgrd2 = I2 * 0.0000771
mgrd3 = I * 0.0179
m4 = mgrd1 - mgrd2 + mgrd3
mgrd = m4 + 0.492
&do i := 1 &to 3
   &type delete all mgrd grids. Delete GRID %i% now.
kill mgrd%i% all
   &end
kill m4 all
/* m = (6.75 x 10 –7) I3 – (7.71 x 10–5) I2 + (1.79 x 10-2) I + 0.492
&type TESTI TESTING
&do i := 1 &to 12
 &type temp * 10 calculation GRID %i% ready now.
xtok%i% = t%i% * 10
&end
&do i := 1 &to 12
 &type temperature * 10 divided by I calculation GRID %i% ready now.
xtdiv_I%i% = xtok%i% / I
 &end
kill I all
kill I2 all
kill I3 all
&do i := 1 &to 12
   &type delete temp * 10 grids. Delete GRID %i% now.
```

```
kill xtok%i% all
   &end
&do i := 1 &to 12
 &type tdiv_I POW mgrd calculation GRID %i% ready now.
xpow16%i% = pow(xtdiv_I%i% , mgrd)
 &end
&do i := 1 &to 12
   &type delete xtdiv_I grids. Delete GRID %i% now.
kill xtdiv_I%i% all
   &end
&do i := 1 &to 12
 &type tdiv_I calculation GRID %i% ready now.
xg16div%i% = xpow16%i% * 16
 &end
&do i := 1 &to 12
   &type delete all xpow16 grids. Delete GRID %i% now.
kill xpow16%i% all
   &end
&do i := 1 &to 12
 &type temp times temp (t*t) calculation GRID %i% ready now.
xtt%i% = t%i% * t%i%
&end
&type Calculate PARAMETER a = -415.85 + 32.24 * t%i% - 0.43 * pow(t%i%, 2)
&do i := 1 &to 12
 &type -415.85 + 32.24 *t1 - 0.43 * t1 * t1 calculation GRID %i% ready now.
xt%i%_26 = -415.85 + 32.24 * t%i% - 0.43 * pow(t%i%, 2)
&end
&type calculate conditional statements for 12 temperature GRIDS
&type if pixel in t1 GRID > 0, then give value tpow1, otherwise 0.
&type calculate if statements for 12 temperature GRIDS
&type if t1 > 0 then t1 <= 26.5, g16div1, xt1_26 or 0
if (t1 > 0 && t1 <= 26.5) then xPET1 = xg16div1
    else if (t1 > 26.5) then xPET1 = xt1_26
  else xPET1 = 0
endif
&type 1 ends
&type calculated if statements for 1 temperature GRID
if (t2 > 0 && t2 <= 26.5) then xPET2 = xg16div2
    else if (t2 > 26.5) then xPET2 = xt2_26
  else xPET2 = 0
endif
&type 2 ends
&type calculated if statements for 2 temperature GRID
if (t3 > 0 && t3 <= 26.5) then xPET3 = xg16div3
    else if (t3 > 26.5) then xPET3 = xt3_26
  else xPET3 = 0
endif
&type 3 ends
&type calculated if statements for 3 temperature GRID
if (t4 > 0 && t4 <= 26.5) then xPET4 = xg16div4
    else if (t4 > 26.5) then xPET4 = xt4_26
  else xPET4 = 0
endif
&type 4 ends
&type calculated if statements for 4 temperature GRID
if (t5 > 0 && t5 <= 26.5) then xPET5 = xg16div5
    else if (t5 > 26.5) then xPET5 = xt5_26
  else xPET5 = 0
endif
&type 5 ends
&type calculated if statements for 5 temperature GRID
```

```
if (t6 > 0 && t6 <= 26.5) then xPET6 = xg16div6
    else if (t6 > 26.5) then xPET6 = xt6_26
  else xPET6 = 0
endif
&type 6 ends
&type calculated if statements for 6 temperature GRID
if (t7 > 0 && t7 <= 26.5) then xPET7 = xg16div7
    else if (t7 > 26.5) then xPET7 = xt7_26
  else xPET7 = 0
endif
&type 7 ends
&type calculated if statements for 7 temperature GRID
if (t8 > 0 && t8 <= 26.5) then xPET8 = xg16div8
    else if (t8 > 26.5) then xPET8 = xt8_26
  else xPET8 = 0
endif
&type 8 ends
&type calculated if statements for 8 temperature GRID
if (t9 > 0 && t9 <= 26.5) then xPET9 = xg16div9
    else if (t9 > 26.5) then xPET9 = xt9_26
  else xPET9 = 0
endif
&type 9 ends
&type calculated if statements for 9 temperature GRID
if (t10 > 0 && t10 <= 26.5) then xPET10 = xg16div10
    else if (t10 > 26.5) then xPET10 = xt10_26
  else xPET10 = 0
endif
&type 10 ends
&type calculated if statements for 10 temperature GRID
if (t11 > 0 && t11 <= 26.5) then xPET11 = xg16div11
    else if (t11 > 26.5) then xPET11 = xt11_26
  else xPET11 = 0
endif
&type 11 ends
&type calculated if statements for 11 temperature GRID
if (t12 > 0 && t12 <= 26.5) then xPET12 = xg16div12
    else if (t12 > 26.5) then xPET12 = xt12_26
  else xPET12 = 0
endif
&type 12 ends
&type calculated if statements for 12 temperature GRID
&do i := 1 &to 12
   &type delete xg16div 1 to 12 GRIDS. Delete GRID %i% now.
kill xg16div%i% all
   &end
&do i := 1 &to 12
   &type delete xtt%i%_26 GRIDS. Delete GRID %i% now.
kill xtt%i% all
   &end
&do i := 1 &to 12
 &type -415.85 + 32.24 *t1 - 0.43 * t1 * t1 GRIDs %i% ready now.
kill xt%i%_26 all
&end
 &do i := 1 &to 12
    &type delete tmean%i% GRIDS. Delete GRID %i% now.
 kill tmean%i% all
    &end
/* &do i := 1 &to 12
/*   &type delete xt%i%_26 GRIDS. Delete GRID %i% now.
/* kill t%i% all
```

```
/*   &end
kill mgrd all
&type YOU have calculated unadjusted Potential evapotranspiration (mm)
&type for each month and as annual sum using Thornthwaite method.
&type Macro is proceeding to calculate adjusted PET values using daylight
&type correction
&type for each month in East Caprivi (latitude is S -18)
&type correction factors are presented now
&type PET01 = xPET1 * 1.14
&type PET02 = xPET2 * 1.0
&type PET03 = xPET3 * 1.05
&type PET04 = xPET4 * 0.97
&type PET05 = xPET5 * 0.96
&type PET06 = xPET6 * 0.91
&type PET07 = xPET7 * 0.95
&type PET08 = xPET8 * 0.99
&type PET09 = xPET9 * 1.0
&type PET10 = xPET10 * 1.08
&type PET11 = xPET11 * 1.09
&type PET12 = xPET12 * 1.15
PET1 = xPET1 * 1.14
PET2 = xPET2 * 1.0
PET3 = xPET3 * 1.05
PET4 = xPET4 * 0.97
PET5 = xPET5 * 0.96
PET6 = xPET6 * 0.91
PET7 = xPET7 * 0.95
PET8 = xPET8 * 0.99
PET9 = xPET9 * 1.0
PET10 = xPET10 * 1.08
PET11 = xPET11 * 1.09
PET12 = xPET12 * 1.15
PETann = PET1 + PET2 + PET3 + PET4 + PET5 + PET6 + PET7 + PET8 + PET9 + PET10 + PET11 + PET12
PETannmean = PETann / 12
 &do i := 1 &to 12
    &type delete xPET%i% GRIDS. Delete GRID %i% now.
 kill xPET%i% all
    &end
&do i := 1 &to 12
  &type GRIDPOINT Processing PET%i% GRID
 capPET%i% = gridpoint(PET%i%, Grid_code)
&end
&do i := 1 &to 12
  &type GRIDPOINT Processing t%i% GRID
 capt%i% = gridpoint(t%i%, Grid_code)
&end
petan = gridpoint(petann, Grid_code)
petannme = gridpoint(petannmean, Grid_code)
q
/* looping coverages to shapefiles
&do i := 1 &to 12
  &type converting to shapefile PET%i% processing
  arcshape capPET%i%  points  PETshp%i%
&end
&do i := 1 &to 12
  &type converting to shapefile t%i% processing
  arcshape capt%i%  points  tshp%i%
&end
arcshape petan  points  petannshp
arcshape petannme  points  petannmeanshp
 &do i := 1 &to 12
```

```
    &type delete capPET%i% GRIDS. Delete GRID %i% now.
 kill capPET%i% all
    &end
 &do i := 1 &to 12
    &type delete capt%i% GRIDS. Delete GRID %i% now.
 kill capt%i% all
    &end
kill petan all
kill petannme all
&return
```

Appendix 6. ArcGIS 9.2 Geoprocessing model (Java script) to calculate topographical wetness index (TWI). (Papers IV, V and VI)

```
// ---------------------------------------------------------------
// EastCaprivi_TWI_index.js
// Created on: to elo 14 2008 09:50:34
//   (generated by ArcGIS/ModelBuilder)
// Description:
// Calculates Topographical wetness index (TWI-index) for DEM.
formula: Wi = ln(As / tanB)
where:
As = drainage area in m2
B = slope
For final TWI-index grid model fills voids using focalmean function and
conditional statement (CON-function) with focalmean function until no voids
detected
CON function used:
con(IsNull([twi_temp]), focalmean([twi_temp], RECTANGLE, 4, 4, DATA),
[twi_temp])
// ---------------------------------------------------------------
// Create the Geoprocessor object
var gp = WScript.CreateObject("esriGeoprocessing.GPDispatch.1");
// Check out any necessary licenses
gp.CheckOutExtension("spatial");
// Load required toolboxes...
gp.AddToolbox("C:/Program Files/ArcGIS/ArcToolbox/Toolboxes/Spatial Analyst
Tools.tbx");
// Local variables...
var dem_fill20m = "C:\\WorkSpace\\dem_fill20m";
var FlowAcc = "C:\\WorkSpace\\grdpois\\flowacc";
var flowdir = "C:\\WorkSpace\\grdpois\\flowdir";
var Output_drop_raster = "";
var drainarea = "C:\\WorkSpace\\grdpois\\drainarea";
var Raster_cell_size_20m = "20";
var slopedeg = "C:\\WorkSpace\\grdpois\\slopedeg";
var slope__2_ = "C:\\WorkSpace\\grdpois\\slope";
var Tanslope = "C:\\WorkSpace\\grdpois\\tanslope";
var drainslope = "C:\\WorkSpace\\grdpois\\drainslope";
var twi_temp = "C:\\WorkSpace\\grdpois\\twi_temp";
var contemp1 = "C:\\WorkSpace\\grdpois\\contemp1";
var Focal1 = "C:\\WorkSpace\\grdpois\\focal1";
var Focal2 = "C:\\WorkSpace\\grdpois\\focal2";
var focal3 = "C:\\WorkSpace\\grdpois\\focal3";
var TWI_index = "C:\\WorkSpace\\grdpois\\twi_index";
var streams = "C:\\WorkSpace\\grdpois\\streams";
var Streamlines_shp = "C:\\WorkSpace\\grdpois\\streamlines.shp";
// Process: Flow Direction...
gp.FlowDirection_sa(dem_fill20m, flowdir, "NORMAL", Output_drop_raster);
// Process: Flow Accumulation...
```

```
gp.FlowAccumulation_sa(flowdir, FlowAcc, "", "FLOAT");
// Process: Times...
gp.Times_sa(FlowAcc, Raster_cell_size_20m, drainarea);
// Process: Slope...
gp.Slope_sa(dem_fill20m, slope__2_, "DEGREE", "1");
// Process: Tan...
gp.Tan_sa(slope__2_, Tanslope);
// Process: Single Output Map Algebra...
gp.SingleOutputMapAlgebra_sa("tanslope / 57.2957795
", slopedeg, "C:\\WorkSpace\\grdpois\\tanslope");
// Process: Single Output Map Algebra (2)...
gp.SingleOutputMapAlgebra_sa("drainarea / slopedeg", drainslope,
"C:\\WorkSpace\\grdpois\\drainarea;C:\\WorkSpace\\grdpois\\slopedeg");
// Process: Single Output Map Algebra (3)...
gp.SingleOutputMapAlgebra_sa("ln(drainslope)", twi_temp,
"C:\\WorkSpace\\grdpois\\drainslope");
// Process: Single Output Map Algebra (4)...
gp.SingleOutputMapAlgebra_sa("con(IsNull([twi_temp]), focalmean([twi_temp],
RECTANGLE, 4, 4, DATA), [twi_temp])", contemp1,
"C:\\WorkSpace\\grdpois\\twi_temp");
// Process: Focal Statistics...
gp.FocalStatistics_sa(contemp1, Focal1, "Rectangle 5 5 CELL", "MEAN",
"DATA");
// Process: Focal Statistics (2)...
gp.FocalStatistics_sa(Focal1, Focal2, "Rectangle 5 5 CELL", "MEAN",
"DATA");
// Process: Focal Statistics (3)...
gp.FocalStatistics_sa(Focal2, focal3, "Rectangle 5 5 CELL", "MEAN",
"DATA");
// Process: Single Output Map Algebra (5)...
gp.SingleOutputMapAlgebra_sa("con(IsNull([focal3]), focalmean([focal3],
RECTANGLE, 12, 12, DATA), [focal3])", TWI_index,
"C:\\WorkSpace\\grdpois\\focal3");
// Process: Single Output Map Algebra (6)...
gp.SingleOutputMapAlgebra_sa("setnull(FlowAcc <100,1)", streams,
"C:\\WorkSpace\\grdpois\\flowacc");
// Process: Stream to Feature...
gp.StreamToFeature_sa(streams, flowdir, Streamlines_shp, "SIMPLIFY");
```