UNIVERSITY OF SOUTHAMPTON

FACULTY OF ENGINEERING, SCIENCE & MATHEMATICS

School of Civil Engineering and the Environment

Predicting Suitable Areas for the Production of Tamarind (*Tamarindus indica* L.) - an Underutilised Fruit Tree Species

by

Colm Bowe

This thesis is submitted in fulfilment of the degree of Doctor of Philosophy

June 2007

UNIVERSITY OF SOUTHAMPTON
ABSTRACT
FACULTY OF ENGINEERING, SCIENCE & MATHEMATICS
SCHOOL OF CIVIL ENGINEERING &THE ENVIRONMENT

DOCTOR OF PHILOSOPHY

Predicting suitable areas for the production of tamarind (*Tamarindus indica* L.) an underutilised fruit tree species

By Colm Bowe B.Sc. (Hons) Env. Mgt.

Awareness is growing regarding the concerns of global climate change as populations continue to rise. The benefits provided by underutilised fruit tree species, such as tamarind (*Tamarindus indica* L.), in combating increasing malnutrition, hunger and poverty in a changing world become more apparent. Tamarind is a high value, multipurpose, underutilised fruit tree species which is drought tolerant and suitable for rain fed agriculture on marginal land. These characteristics make this species ideal for resource-poor farmers. The importance of identifying underutilised species such as tamarind's environmental adaptation and their ecogeographic distribution has been identified. However the lack of physiological or empirical yield or growth data for the species does not permit the use of more traditional methods of crop modelling.

In this study the statistical modelling technique, ENFA (Ecological Niche Factor Analysis), more commonly applied to wild species, was applied to occurrence data for tamarind derived from herbaria passport information records. This allowed the mapping of the niche of the entire species in ecological space and prediction of suitable production areas based on the species environmental requirements.

Tamarind was modelled across its entire global distribution and three continental regional subsets. Suitability maps were produced at global and regional extents. The models showed tamarind to respond differently on certain variables between the global and various regional extents, resulting in different patterns in the predicted distribution. Regional models were identified as providing a better prediction than global models. Due to the semi domesticated nature of the species, it was proposed that the likely effect of the regional variation may be due to evolutionary effects in the niche characteristics; resulting in geographical subpopulations. The global models are unable to incorporate this regional variation. Therefore for tamarind the regional models were identified as the "best" models and recommended for use in the prediction of potential production areas for the species.

Both regional and global model built in this study were built on a large scale. This provided information on the global and regional niche of the species and broad scale maps identifying potential production areas. Localised validation with field survey data indicated that the regional models, although performing well in the cross validation may not to be locally predicting very well in all areas. Caution should be taken when using these models to predict production areas at a local scale. In order to predict suitable locations at a practical scale for use by extension workers and farmers, models built at a more local scale may be more appropriate, incorporating the complex factors which may influence tamarind at this scale.

The implications of selection of global or regional models in regard to modelling underutilised species distribution under climate change scenarios are discussed. The use of geo-referenced herbarium passport data with statistical modelling is recommended as a relatively simple way of predicting the distribution/potential production areas for large numbers of underutilised crops, or any crop species with limited or no eco-physiological, empirical yield information. Future research is recommended to investigate niche requirements and niche variation of promising landraces or cultivars suitable to meet nutritional, medicinal and industrial requirements under current and future climates.

Table of Contents

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to give special thanks to my supervisor Dr. Nazmul Haq for his guidance, expertise, endless patience and his much appreciated mentoring over the period.

My sincere thanks also to Dr. Patrick Osborne, Dr Simon Bray, Mr John Jones and Dr Tony Lock, and all other staff within the Centre for Environmental Science and the School of Civil Engineering and the Environment, who have provided guidance, help and support

I want to thank all my friends and fellow PhD students in the Centre for Environmental Sciences and within the School of Civil Engineering and the Environment, University of Southampton, particulary Pedro Leitão, Toru Tsuzaki, Zoe Dunsiger and Clive Farquhar for their advice, support and friendship

I would also like to thank the administrators Barbara Hudson, Jacqui Holmes and Nicky Hancock for their support.

I would like to thank the Dr Daniel and his staff at BAIF, Pune and Dr Reddy and his staff at IIHR, Bangalore for help and support while carrying out my fieldwork in India. I would like to pay a particular thank you to Ganesh Karemane, for his advice, hard work and friendship during my visit.

My unending thanks also to my parents for their absolute love, support and encouragement and for their assistance with proof reading and formatting advice.

ACRONYMS AND ABBREVIATIONS

ACUC	Asian Centre for Underutilised Crops
AVH	Australian Virtual Herbarium
BRAHMS	Botanical Research and Herbarium management system
CABI	Commonwealth Agricultural Bureaux International
CGIAR	Consultative Group on International and Agricultural Research
DFID	Department of International Development - United Kingdom
DIVA	Freeware GIS software
ENFA	Ecological Niche Factor Analysis
ERIN	Environmental Resources Information Network, Australia
FAO	Food and Agriculture Organisation of United Nations
FNC	Forests Nations Corporation
ESRI	Environmental Systems and Research Institute California.
GBIF	Global Biodiversity Information facility
GFU	Global Facilitation Unit for Underutilised Species
GIS	Geographical Information System
GPS	Global Positioning System
ICUC	The International Centre for Underutilised Crops, established in Southampton
ICRAF	The World Agroforestry Centre established in Nairobi 1977
INFER	International Forum for Environmental Response
IPGRI	International Plant Genetic Resources Institute
ISRIC	International Soil Reference and Information Centre
MAI	Moisture Availability Index
MFP	Minor Forest Product
NARS	National Agricultural Research Systems
NAS	National Academy of Science
NHC	Natural History Collection
PCA	Principal Component Analysis
REMIB	World Information Network on Biodiversity
SEANUC	Southern and East Africa Network for Underutilised Crops
SOTER	Soil and Terrain Database
TERRSTAT	Database of Soil Properties
TKP	Tamarind seed Kernel Powder
UNCTAD	United Nations Conference on Trade and Development
UNEP	United Nations Environment Programme
USGS	Geographic Names Information Service (USA)
UTFANET	Underutilised Tropical Fruits in Asia Network
UTFT	Underutilised Tropical Fruit Tree
WHO	World Health Organisation
WISE	World Inventory of Soil Emissions

1 Chapter 1 Introduction

1.1 General Introduction

Over the past 60 years the introduction of technological advances has led to intensification of agriculture. Coined “the green revolution”; agricultural research has allowed the introduction of high yielding varieties of cereals (predominantly of maize, wheat, and rice), with the application of chemical fertilizers, pesticides and irrigation (Evenson and Gollin, 2003).

By the 1970s the new varieties of crops and production systems had replaced the traditional farming practices of millions of farmers in developing countries. By the 1990s, the new varieties accounted for almost 75% of the area under rice cultivation in Asia, approximately half the wheat planted in Africa, Latin America and Asia as well as more than 50% of the world's maize (Evenson and Gollin, 2003).

These developments have played key roles in boosting crop productivity which has brought about an increase in global food production (Evenson and Gollin, 2003). Although the global population has doubled since 1961; the food produced per person has risen by 24 percent (Evenson and Gollin, 2003). On average food calorific intake has risen with corresponding gains in health and life expectancy (Matson et al., 1997).

However in many countries this improvement has not been significant, and food shortage and malnutrition remain predominant. The aggregate global trends in statistics for the levels of poverty and food security, conceal that at regional and country levels, progress has not been satisfactory. In Sub-Saharan Africa the absolute numbers of malnourished continue to rise (von Braun et al., 2004; Desai, 2005). The FAO (2006) reported that there are 206 million hungry persons in South East Asia and East Asia , which is an increase of 40 million from the early 1990’s. The number of hungry people in the world is currently increasing at the rate of four million a year (FAO, 2006).

There are dramatic underlying disparities in the availability and impact of green revolution technologies across different agro-ecological zones and political regions (Evenson and Gollin, 2003). Many countries or regions failed to participate in the agricultural intensification, or gained little increase in production (Kydd1 et al., 2004). Regions which had access to irrigation or reliable rainfall, productive soil and good infrastructure (which allowed access to economic investment, agricultural inputs and markets) gained substantial improvement in agricultural productivity. Outside of these areas, implementation of new technologies was slower and more limited with new crop varieties performing poorly. Lack of economic stability and poor infrastructure limited farmers' ability to invest in and maintain intensive agriculture (Kydd et al., 2004).

Intensive agriculture's high dependence on chemical fertilizers, pesticides and irrigation led to incidences of disease, erosion of soil and loss of soil fertility. This resulted in reductions in yield with degradadation of the environment. Thus concerns were raised about intensive agriculture's long term environmental and economic sustainability (Tilman, 1998; Evenson and Gollin, 2003).

In Sub Saharan Africa the green revolution has had little positive impact. A number of countries achieved virtually no growth in food production (Evenson and Gollin, 2003). Agricultural advances have allowed India to go from being a chronic food importer to a massive grain exporter; however India still has the largest number of undernourished people in the world (FAO, 2006). Much of the population of India are too poor to buy the food produced in their own country.

The introduction of green revolution technologies has led to a homogenisation of farming systems. A small number of new varieties and intensive cropping systems replaced the multitude of traditional farming systems, species and cultivars adapted to specific ecotypes. This has dramatically reduced the diversity upon which global food security and agricultural income depend. For many, this has increased the vulnerability of agriculture; placed the future supply of food and rural income at risk, and impoverished the human diet.

The effects of climate change, the continued rise in global populations, and the spread of HIV/AIDS particularly in developing nations, are likely to have further impacts on crop production, agricultural income generation and availability of food. Changes such as globalization and rapid urbanization are issues to which agricultural production will have to adapt. Much of the further increase in food production will have to come from so-called "marginal" land which is not suitable for production of the major crops. Serious concerns exist about the ability of the narrow portfolio of today's agriculture to meet the needs of expanding populations in a changing world (Padulosi et al., 2003).

The contribution of major crops alone towards food and nutrition security, poverty alleviation, and ecosystem conservation, has been questioned. Particularly as awareness grows of the fact that diversification of crops at all levels and in all types of agro-systems is a crucial element for sustainability (Collins and Hawtin, 1998; Padulosi et al., 1999). The agricultural sector will need to respond in ways beyond the traditional focus on higher yields (Fresco, 2003)

1.2 Underutilised crops and species

Ethno-botanic surveys have identified hundreds of species found globally, which when grown in traditional farming systems represent an enormous wealth of agro-biodiversity with potential to contribute to improved incomes, food security and nutrition (Haq, 1995; Padulosi et al., 2003). Kuhnlien (2001) noted that in just five case studies of indigenous peoples in Asia, 716 types of traditional foods were reported (Kuhnlein, 2001). Studies conducted in the local markets of Malawi, Tanzania and eastern Zambia demonstrated the importance and popularity of indigenous fruits (Kwesiga and Mwanza, 1997). Over 7000 plant species have been known to be used for food and are either partly or fully domesticated. Approximately 30 species provide 95% of the world's food energy (Williams and Haq, 2002).

Many crops that have traditionally been important for feeding or providing income to the poorest sectors of society are now neglected, restricted to local production and subsistence use. These crops are underutilised in terms of their wider potential because emphasis has focused on a few species and varieties. Lack of investment, R&D and progressive policies, have meant that these underutilised species are often unable to compete with the major crops (which are supported by seed supply systems, production and post-harvest technologies and extension services) that have come to dominate the world food supply (Haq, 1995; Padulosi and Hoeschle-Zeledon, 2004). Food security programmes in many developing countries have been predominantly based on agricultural policies which favour the 'green revolution crops', focusing exclusively on maize, wheat or rice, and export crops (Anon, 2005). The National Agricultural Research Systems (NARS) have for a long time failed to include research and development of underutilised species in their policy development. The institutional frameworks in less developed countries do not have the capacity to extend research and experimentation to such crops. The frameworks suffer from lack of mechanisms to introduce, test and understand those species that are considered less economically important.

This has meant that the benefits of these species to human well-being and incomes is underutilised (Padulosi et al., 2003; Williams and Haq, 2003).
Many underutilised crops which were more widely grown are now in decline with erosion of their genetic base and knowledge of their production and utilisation (Padulosi et al., 2003; Williams and Haq, 2003). When grown in traditional subsistence farming systems, these underutilised species make a significant contribution particularly in marginal areas where poverty, food and nutritional security are significant problems (Campbell, 1987; Haq, 1995; Williams and Haq, 2003; ICUC, 2006). These species represent an enormous commodity resource that have traditionally provided food and nutrition, energy, medicine and industrial needs which will help to meet increasing demands in the future (Williams and Haq, 2003).

Many underutilised species having undergone generations of selection allowing them to withstand local, often stressful conditions. They produce high yields with minimal inputs. This can provide a comparative advantage (Padulosi et al., 2003), over more generic varieties of the major crops that may not be as suitable to such local environments (McNeely and Schutyser, 2003). Underutilised species are often grown in bio-diverse multi-crop systems such as home gardens or agro-forestry systems. They contribute to agro-ecosystem stability and ecosystem health and therefore may mitigate the effects of environmental changes (InWEnt/GFU, 2003).

These crops are often considered “minor crops” because of their status in terms of global production and market when compared with the major staple crops and agricultural commodities. However from the standpoint of the rural poor who depend on many of them for food security, nutrition, and incomes; they are not “minor”(Padulosi et al., 2003).

Global policy and strategy development are increasingly starting to focus on development and use of underutilised species. This has been encouraged by the establishment of organisations such as ICUC (International Centre for Underutilised Crops) and international networks such as UTFANET (Underutilised Tropical Fruits in Asia Network) (Rondolo, 2002), SEANUC (Southern and East Africa Network for Underutilised Crops)(Anthony et al., 1995) and ACUC (Asian Centre for Underutilised Crops)(ACUC, 2004). In 2002 GFU (Global Facilitation Unit for Underutilized Species)(Padulosi et al., 2003) was established. The GFU is a dedicated global hub created to promote international exchange and strengthen existing initiatives and networks on underutilised crops.

The FAO Global Plan of Action for the Conservation and Sustainable Use of Plant Genetic Resource for Food and Agriculture, which was adopted in 1996 by approximately 150 countries, identified the promotion and the development of underutilised species as one of its 20 main activities (FAO, 1996a). This program aims at: identifying, developing sustainable management practices, developing post-harvest and marketing methods, while promoting policies for the development and use of underutilized species. Increasingly NARS are starting to include underutilised crops in their national programmes (Williams and Haq, 2003).

Underutilised species provide a broad portfolio of crops to improve dietary diversity (Williams and Haq, 2003). Both the FAO and WHO identify dietary diversification as a key to combating malnutrition and micronutrient deficiencies (Anon, 2005). Many underutilised species are nutritionally rich, containing many vitamins and minerals often lacking in diets. They combat hidden hunger and thus have a direct impact on well-being and health. Many provide nutrients essential for maternal health and for child development (Anon, 2005; ICUC, 2006). They are accessible resources for the urban and rural poor (Verheij and Coronel, 1991; InWEnt/GFU, 2003), through home gardens, agroforestry systems or wild harvest. They often provide nutrition during famine periods and act as emergency foods. Many underutilised species are found to be richer in vitamins and nutrients than major crops grown for similar uses i.e. Oca (*Oxalis tuberosa*), ulluco (*Ullucus tuberosus*) and mashwa (*Tropaeolum tuberosum*). Three traditional Andean tuber plants are all richer in Vitamin A and Vitamin C than the well-known potato (*Solanum tuberosum*).

Underutilised species have been widly used in traditional medicine (Haq, 1995). Detailed nutritional and pharmaceutical studies are being conducted to identify and extract active compounds for a number of species such as *Ziziphus* (Azam-Ali et al., 2006)

New research with utilisation of traditional knowledge on processing of underutilised crops has lead to the development of technology procedures suitable for various production scales from the household to the large scale commercial level. Such technologies allow the development of a wide diversity of products, adding value and increasing the shelf life of the produce.
Product diversification has enhanced underutilised crop importance in both developing and developed country market places (Akinnisfsi *et. al.*, 2005). The nutritional and medicinal properties of many underutilised crops can be exploited through health foods or natural and organic products (Haq, 2007). Processing of crops has created new market opportunities with employment at various levels. This allows the opportunity for diversification of livelihoods, and the development of rural micro enterprises such as community participatory production and village

level processing and marketing (InWEnt/GFU, 2003; Haq, 2004). It is often the women that hold the knowledge for the production and utilisation of such crops (Williams and Haq, 2003; Haq, 2004). Therefore underutilised crops provide increase income and empowerment to potentally vulnerable members of society.

Underutilised crops also show potential for many industrial uses. They can address the needs of renewable and sustainable sources of energy, plastics, fibres and other materials (Haq, 1989; Pramanik, 2003; UNCTAD Biofuels Initiative team, 2006).

The growing world market for sustainable goods and environmental services is increasingly being exploited by use of current underutilised species in diverse agricultural systems (Kandji et al., 2006). Underutilised crops have great potential to contribute to the market for environmental services such as carbon sequestration, watershed management or bioremediation. Risk assessment and management options are of increasing importance for environmental insurers due the increased incidence of storm, flooding and droughts brought about by climate change (ICUC, 2006).

The true potential of underutilised species is related to the ways in which old and new uses are being re-addressed to meet current global issues (Padulosi et al., 1999). Emerging national and international attention on underutilized species has come about due to their ability to meet concerns of food and nutritional security, reduction of poverty, impacts of climate change and loss of biodiversity. They also meet the needs created by interest in new products and environmental markets (Padulosi et al., 2003; Williams and Haq, 2003). In fact underutilised crops have been identified to contribute to achieving 5 of the CGIAR (International Agricultural Research Centres) system priorities and towards to 7 of the 8 Millennium goals (ICUC, 2006)

1.3 Underutilised tropical fruit trees

Tropical fruit trees are important multipurpose species for small holders. Many underutilised fruit tree species are used in traditional farming systems and provide a diverse range of products (fruits, timber, fodder, resins and medicines). These products are often of high value in comparison to annual crops. The fruits provide essential vitamins and minerals often deficient in many diets (especially in urban areas), (Verheij and Coronel, 1991; Anon, 2005). Results from ethno-botanical studies conducted in Malawi, Tanzania and eastern Zambia demonstrated the importance and popularity of indigenous fruits (Kwesiga and Mwanza, 1997).

The diversification of the farming system with many crops is a risk avoidance strategy of many traditional farmers (Jarvis et al., 2006). One of the largest opportunities for farm diversification comes from small holders planting fruit trees alongside other crops (Sanchez *et al.*,1997). Underutilised fruit tree species can fill specific niches on farms, making the system ecologically stable and economically more rewarding, providing resilience against weather or price fluctuations. They have the potential to provide fruit throughout the year (Mateke et al 2002) with different species ripening at different times. They provide food security at strategic periods when conventional staple and vegetables crops are scarce (Okafor and Lamb, 1992).

Underutilised fruit tree species can grow in a wide range of climates and soil conditions including the hardiest of environments and in the most arid parts of the world. In regions where climate variability is common place and adverse impacts of climate change are expected; the role of trees in buffering against production risk can be of great importance (Ong and Leakey, 1999). Underutilised fruit tree species also provide environmental services such as reducing land degradation associated with rainfall variability and poor agricultural practices. They enhance resilience against the effects of adverse weather conditions, poor soils and pests (Verchot et al., 2005). They stabilise the soil, assist the cycling of nutrients and enhance biodiversity. They also provide improved infiltration of water, while reducing runoff and transportation of sediments. This improves water storage in the soil, buffering agricultural crops against water deficiencies (Verchot et al., 2005). The result is farm income is increased and diversified and food security improved (Sanchez et al., 1997)

Underutilised Tropical Fruit Tree species span almost the entire scope of the domestication process, from growing wild to some species having reached almost full domestication. Some are commercially cultivated and have undergone large amounts of selection. Few underutiliised species have distinguished cultivars or varieties. Many species are still found growing wild in forests or natural stands where they are gathered by local people, others are managed i.e. protected by local people but not actively cultivated. Some species have been semi or fully domesticated and are grown in home gardens or orchards.

The wide range of diversity shown by these species is important. The application of modern biotechnologies now allows identification of high quality planting material; consequently beneficial genetic traits can be introduced to cultivation through the selection and domestication process. Planned conservation and utilisation of local plant material is of prime importance for food security. This permits small-scale farmers to have access to seed and planting material of crop varieties adapted to their region.

Like all underutilised crops, fruit tree species have the ability to meet many of today's concerns, however for many reasons they do not meet their full potential. The major constraints to their production and utilisation include lack of availability of quality propagation material, availability of processing technologies and poor infrastructure. Survey results indicate local level markets exist for indigenous fruits but are largely constrained by poor promotion or low demand due to distorted prices, poor storage and transport facilities (Kwesiga and Mwanza, 1997)

There is a fear that much of the biodiversity provided by fruit tree species is being lost due to population growth, severe forest degradation, changes in farming systems and eating habits. This will affect the availability of genetic diversity they can provide which is a key element for sustainable agricultural development (Okafor and Lamb, 1992).

Research in conservation, utilisation and domestication of the underutilised tropical fruit tree species plays an important role in the diversification in agriculture and the conservation of genetic variation.

1.4 Tamarind

For this study the underutilised tropical fruit tree species tamarind (*Tamarindus indica* L.) is selected. Tamarind is a leguminous, multipurpose, tropical fruit tree species. It is a mainly subsistence based crop, which mostly meets local demands. Tamarind has broad geographic distribution across the tropics and subtropics. It is known to be adapted to a wide range of ecological conditions. Although tamarind does not substantially contribute to the economy, it is of major local importance in many areas where it is grown. Tamarind has a high social value and has numerous nutritional, medicinal and industrial uses. Women often play a major role in the gathering and processing of tamarind. It is grown on plantation scale in few countries (India, Mexico, Thailand), however a large proportion of its production comes from trees in home gardens and field borders. Therefore information on its area and quantity of production are either nonexistent or unreliable estimates (El-Siddig, 2006).

Burkina Faso, Cameroon, Chad, Cote d'Ivorie, Gambia, Guinea, Kenya, Mauritania, Nigeria and Senegal have prioritised tamarind for conservation based on utilisation and value. Market surveys of non – wood forest products carried out in Sudan by the FNC/FAO (1995) revealed that tamarind products used for home consumption ranked number one among species studied.

Tamarind is a smallholder tree crop especially suitable for rainfed and semi-arid regions. The trees produce well naturally with limited inputs. Tamarind can tolerate extended dry periods of weather and is known as a drought tolerant species. It is also known to tolerate high levels of salinity (Panchaban *et al.*, 1989). It is considered as a good candidate for diversification in smallholder production systems, in particular on marginal, i.e. arid and semi-arid, areas. The increasing integration of tamarind with other trees and crops on farmlands offer a strategy to minimise the risk of crop failure. These characteristics make this species ideal for resource-poor farmers and can thus contribute largely to income generation. Species with such characteristics are likely to have increased importantance due to the effects of climate change.

There exists a considerable land area where tamarind production areas could be expanded. Due to low priority allocation, many countries have not identified areas that could be used for expansion (Nyadoi, 2004). (For Further details on tamarind see section 2.1.)

1.5 Premise for research

Hall and O'Brien (2002) noted that the effective management of species populations retained in agricultural landscapes must take into account the ecology of the species. Williams and Haq (2002) reported that procedures for assessing the sustainability of underutilised crops requires focused research to evaluate the potential distribution and ecological requirements. This is important in advancing knowledge and avoiding limited or piecemeal research. The conferences on Neglected and Underutilized Crop Species Aleppo 9–11 Feb. 1998 and the International Workshop on Underutilized Plant Species, Leipzig, Federal Republic of Germany, 6 - 8 May, 2003 (InWEnt/GFU, 2003) both identified the need for investigation into the underutilised species' environmental adaptation and ecogeographic distribution. Such research would allow identification of where the species is locally available and accessible, adaptability to the local environment, identification and suitabilty of new areas for cultivation (Padulosi et al., 1999; InWEnt/GFU, 2003). To maintain larger markets as the economic importance of underutilised species continue to grow, it will become more important to identify sustainable sources of planting material supplies and areas for production.

1.6 Research thesis justification

As awareness grows regarding concerns of global climate change and rising human populations it is important that natural resources and associated ecosystems be used in a rational and sustainable way, optimising their benefits.

Serious concerns exist about the ability of the narrow portfolio of today's agriculture to meet the needs of expanding populations in a changing world (Padulosi et al., 2003). Underutilised crops provide a solution to the current major issues effecting agricultural production and food security. Much of the further increase in food production will have to come from so-called "marginal" land which is not suitable for production of the major crops. The exploitation of underutilised crops allows the opportunity to increase biodiversity and environmetal sustainability while providing food, nutrition and livelihoods to a rising human population in a changing environment.

Underutilised fruit tree species such as tamarind are adapted to local conditions and specific niches. They have the ability to survive and produce yields in conditions where the major crops may not survive. Therefore it is important that more research is undertaken on such crops in regard to their relationship with the environment. This will allow identification of suitable locations for their production under current and future climate senarios

Tamarind is a high value multipurpose underutilised fruit tree species which is a drought tolerant and suitable for rainfed agriculture on marginal land with minimum inputs. These characteristics make these species ideal for resource-poor farmers within a changing environment, contributing largely to food and nutritional security and income generation.

This study aims to model the environmental requirements of the underutilised fruit tree species tamarind (*Tamarindus indica*) and predict potential production areas.

Methods developed in this study can be applied to other underutilised speices allowing identification of their interactions with the environment and potential production areas under current climates.

2 Chapter 2 Literature review

2.1 Tamarind (Tamarindus indica L.)

Tamarind (*Tamarindus indica* L.) is a slow growing, long-lived, leguminous, evergreen or semi evergreen tree, which can grow up to 30m under favourable conditions (Morton, 1987; Gunasena and Hughes, 2000). *Tamarindus* belongs to the dicotyledonous family, Leguminosae which is the third largest family of flowering plants (Lewis et al., 2005). It belongs to the subfamily Caesalpinioideae which in turn has been divided into a number of tribes. Opinions are divided on how many tribes there are, or indeed to which one *Tamarindus* belongs. Leonard (1957) included it in the Amherstieae (Pettigrew and Watson, 1977) which contains 25 genera. More recently it was included in the tribe Detarieae thought to be close to Amherstieae. The genus *Heterosteman* was thought to resemble Tamarindus as do *Amherstia* and *Brownea*. The Tamarindus genus is monotypic, containing the sole species *T. indica* (Williams, 2006a).

2.1.1 Origin

Although the consensus is that tamarind originates in Africa (Coates-Palgrave, 1998), there is debate on this matter (El-Siddig et al., 2006). Some sources state the origin as India (Morton, 1987) or the Far East (Coates-Palgrave, 1998). Troup (1921) placed it in Ethiopia, but others considered it indigenous to the drier savannahs of tropical Africa, from Sudan, Ethiopia, Kenya and Tanzania, westward through sub-Sahelian Africa to Senegal (Brandis, 1921; Ridley, 1922)(Dalziel, 1937; Dale and Greenway, 1961; Irvine, 1961; NAS, 1979)).
If orginated in Africa, it is likely to have been introduced to South and Southeast Asia a very long time ago (Brenan, 1967; NAS, 1979) and has naturalised in many areas (Simmonds, 1984; Purseglove, 1987; Coronel, 1991).

2.1.2 Historical information

The spread of tamarind to Asia may have taken place in the first millennium BC. Farming of tamarind in Egypt by 400BC has been recorded and it was mentioned in the Indian Brahmasamhita Scriptures between 1200-200 BC. About 370-287 BC, Theophrastus wrote on plants and two descriptions refer to tamarind, which were probably from East Africa (Hort, 1916). Trade between the Mediterranean and the Orient flourished towards the end of the first millennium BC when spices were imported. By 1000 AD the Arabs dominated this trade. Marco Polo recorded that Arab traders made tamarind an important commercial item in Medieval Europe (Williams, 2006a).

2.1.3 Geographical distribution

Tamarind grows widely in most tropical/subtropical regions of the world. Table 1 and Table 2 give its native and exotic distribution range respectively.

Table 1 Tamarind Native Geographical Distribution (El-Siddig et al., 2006)

Burkina Faso	Chad	Guinea Bissau	Niger	Tanzania
Cameroon	Ethiopia	Kenya	Nigeria	Uganda
Cape Verde Is	Gambia	Madagascar	Senegal	
Central African Republic	Guinea	Mali	Sudan	

Table 2 Tamarind Exotic geographical distribution (ILDIS (International Legume database and Information Service), 2007)

Aldabra	Brazil	Dominican Republic	India	Malawi	Seychelles	Thailand
Andaman Is	Burundi	Ecuador	Indonesia	Mauritius	Sierra Leone	Togo
Angola	Cambodia	Egypt	Iran	Mexico	Singapore	Tonga
Argentina	Cayman Is	El Salvador	Iraq	Mozambique	Society Is	United States
Australia	China	Equatorial Guinea	Irian Jaya	Panama	Socotra	Venezuela
Bahamas	Christmas	Fiji	Ivory Coast	Papua New Guinea	Somalia	Vietnam
Bahrain	Colombia	French Guiana	Jamaica	Paraguay	South Africa	Yemen
Bangladesh	Comoro Is	Ghana	Java	Peru	South Yemen	Zaire
Belize	Cook Is	Gilbert Is	Laccadive Is	Philippines	Sri Lanka	Zambia
Benin	Costa Rica	Guatemala	Laos	Reunion	Sumatra	Zimbabwe
Bhutan	Djibouti	Haiti	Liberia	Rodrigues	Suriname	
Bolivia		Honduras	Libya	Saudi Arabia	Taiwan	

2.1.4 Production areas

Table 3 Major and Minor Production areas of tamarind

Major areas	Production (tons) and year	Minor Areas
Brazil	-	Bahamas
Costa Rica	221 (1997)	Bangladesh
Cuba	-	Burma
Egypt	-	Cambodia
Guatemala	-	Dominican Republic
India	250,000 (1964)	Fiji
Indonesia	-	Gambia
Mexico	37 (annually)	Kenya
Nicaragua	-	Pakistan
Puerto Rico	23 (1977)	Senegal
Philippines	-	Tanzania
Sri Lanka	-	Vietnam
Thailand	140,000	Zambia
Venezuela	-	Zanzibar

At present tamarind is cultivated in 54 countries of the world: 18 in its native range and 36 other countries where it has been introduced (El-Siddig et al., 2006).

Table 3 provides details of the major and minor production areas of tamarind The major areas of production are in the Asian and American continents (El-Siddig et al., 2006). Two main types of tamarind exist, those with a sweet pulp and those with a sour pulp..Most countries produce the sour type, comprising about 95% of the total world production.

India is the largest producer of sour tamarind in Asia and its annual production is in the range of 250,000 to 300,000 tonnes (NAS, 1979). Tamarind kernel powder (TKP) is in great demand and nearly 20,000 tonnes are produced annually in India . Annual returns from tamarind seed powder are estimated at 16,000,000-17,000,000 Indian rupees (US$ 346,400-368,050) (Hughes, in press).

Thailand is the second largest producer of tamarind in Asia and produces 30% of the sweet type, which is gaining status as a small-scale plantation crop. Although Thailand also produces the sour type, it is the sweet type which is gaining demand. The fresh pods are highly valued and large sweet pods reach high prices, particularly out of season.

In the America's tamarind is produced commercially in Mexico and Costa Rica. Mexico has over 4400 hectares producing over 37,000 tones of pulp. It exports a small amount of processed pulp to Central and South American countries and to the United States of America (El-Siddig et al., 2006). In recent years production in Costa Rica has risen, 1995 to 1997 figures show an increase from 192 to 221 tons. Mexico and Costa Rica show high potential for expansion of tamarind production, much of which remains unexploited.

In Africa and other minor production, tamarind is not produced on a large commercial scale; however it is used widely at the local scale. In these minor production areas (Table 3) tamarind occurs in small plantations on farm boundaries, roads and field edges. In these countries, production data are not available because the tree is thought to be unimportant for both domestic use and commerce (Hughes, in press).

2.1.5 Uses of tamarind

Tamarind is a multipurpose species and therefore has a wide range of uses. Virtually every part of the tree has some value in commerce and particularly in the subsistence of rural people (Williams, 2006b).

Fruit and Food Products: The sweet variety is eaten as a table fruit. The unique sweet/sour flavour of pulp is popular in cooking and flavouring. The acidic pulp is used as a favourite ingredient in culinary preparations such as curries, chutneys, sauces, ice cream and sherbet in countries where the tree grows naturally. Tamarind pulp is often made into juice, infusion or brine. Pulp is used commercially to prepare tamarind pickle jam, syrup, candy and champoy (balls rolled in sugar and wrapped in cellophane) .

Medical Uses: The medicinal value of tamarind is mentioned in traditional Sanskrit literature (Williams, 2006b). A number of recent surveys have listed local folk uses for tamarind remedies for ailments, which include anti inflamatories in North Africa (Rimbau et al., 1999), use for skin disorders in Gujarat (Punjani and Kumar, 2002). The laxative properties of the pulp and the diuretic properties of the leaf sap have been confirmed by modern medical sciences (Williams, 2006b).

Industrial uses: Tamarind pulp is used as a raw material for the manufacture of several industrial products, such as tamarind juice concentrate (TJC), tamarind pulp powder, tartaric acid, pectin, tartarates and alcohol. The major industrial use of the seeds is in the manufacture of Tamarind Kernel Powder. This can be used as a source of carbohydrate for the adhesive or binding agent in paper, textile sizing, weaving and jute products.

Agroforestry: Tamarind is used in agroforestry systems in many parts of the tropics due to its multiple uses (Relwani, 1993). Many farmers integrate tamarind and several other tree species such as *Annona* spp., with their agriculture crops (Yaacob and Subhadrabhandu, 1995). The increasing integration of tamarind with other trees and crops on farmlands offers a strategy to minimise the risk of crop failure. Tamarind acts to impove nutrient recycling and moisture storage in the soil.

Environmental benefits: Tamarind has been used in carbon sequestion projects by the ECCM (The Edinburgh Centre for Carbon Management) and the EETD (Environmental Energy Technologies Division). In such agroforestry systems tamarind is grown along other fruit species and forestry species. Farmers receive income for sale of carbon offset in addition to that generated by the sale of fruits (Satyanarayana 2004).

2.1.6 Tamarind eco-physical requirements

2.1.6.1 Climate

The tamarind ranges from Subtropical dry to Wet, through tropical very dry to Wet forest life zones. Tamarind tree is well adapted to the semi –arid tropics and subtropics (Chundawat, 1990; El-Siddig et al., 1999).

2.1.6.1.1 Rainfall

The is a large amount of variation for tamarinds rainfall requirement in the literature. Tamarind grows well in areas with more than 700mm (up to 3000) but is found throughout the Sahel in areas with 400 mm or more (FAO, 1988). The minimum annual rainfall was given by Gunasena and Hughes (2000) as 250 mm. However in areas where rainfall is low, the trees are usually located in areas with a shallow water table or along water course (Gunasena and Hughes, 2000)(FAO, 1988) (Vogt, 1995). In the Sahel regions of the Sudan where annual rainfall is 300 – 400mm tamarind is frequently found along seasonal streams (Khors) or in valleys (Wadis) (El-Siddig et al., 1999). Individual trees produce up to 50kg of fruit in West Africa where rainfall often totals less than 500mm per year (NFTA 1993). As per von Maydell (1986) it grows preferably in semi arid regions (with approx 400mm of rainfall) but also in the monsoon regions with more than 1500mm, it thrives in Southeast Asia where these conditions prevail (NFTA 1993). The maximum rainfall which tamarind can tolerate is up to 4000mm, provided the soil is well drained (Duke and Terrell, 1974) However where it grows in these conditions in the wet tropics, it has been reported not to flower (Allen and Allen, 1981; Coronel, 1991).

The pattern of rainfall and the amount of rainfall that seems to be important for flowering and fruit production in tamarind. A number of authors have emphasised the importance of a prolonged dry season (Allen and Allen, 1981; Von Maydell, 1986; Mahoney1, 1990), deeming it necessary for fruit production. The US National Academy of Science (NAS, 1979) states that tamarind is only suitable for growing in regions which have extended dry periods. In the humid tropics where rainfall is evenly distributed, the tree does not bear fruit and fails to grow unless the soil is well drained. However Coronrel (1991) states that tamarind was found where rainfall was evenly distributed as well as where the dry season was very dry and pronounced. He does however state that that wet conditions during the final stages of fruit development were detrimental. In some humid areas the tree will grows well, however too much rainfall may also affect growth due to water logging (FAO, 1988). A marked extended dry period seems necessary for the formation of fruit and regardless of total annual rainfall.

Tamarind is known as a drought tolerant plant and can be grown in very dry areas. With supplementary irrigation it can withstand up to six months without rainfall (Gunasena and Hughes, 2000). This is due in part to its deep and extensive root system (Coronel, 1991). It has been noted that pre-monsoonal drought can affect growth.

2.1.6.1.2 Temperature regime

Tamarind prefers a warm climate (Chundawat, 1990) and will survive in very high temperatures (Vogt, 1995). It is essentially a tree of tropical climates and thrives under a maximum annual temperature ranging from 33 -37 °C and a minimum of 9.5 to 20 °C (Gunasena and Hughes, 2000). FAO (1988) state that tamarind grows well where the mean monthly temperature does not drop below 21 °C. Mature trees are said to withstand temperatures as high as 47 °C (Gunasena and Hughes, 2000). Temperature has also been identified as important for fruit maturity and fruits are said not to ripen in cold conditions (Chundawat, 1990).

2.1.6.1.2.1 Absolute temperatures and frost

Tamarind is sensitive to frost (Troup, 1921; FAO, 1988; Chundawat, 1990; Mahoney1, 1990; Vogt, 1995). It is neither tolerant to persistent cold or brief frost and does not perform well in either cold temperatures or freezing conditions (Morton, 1987). However older trees are said to be more resistant to extreme temperatures (Coronel, 1991; Gunasena and Hughes, 2000) than young trees (Gunasena and Hughes, 2000) (Morton, 1987). Large specimens are said to be able to tolerate light frost and extreme temperatures as low as -3 °C without serious injury (Gunasena and Hughes, 2000) . Tamarind is said to be more cold tolerant than mango, avocado and lichee (Morton, 1987; Verheij and Coronel, 1991).

2.1.6.1.3 Light

Tamarind is a light demanding tree and grows very slowly. It is often isolated from other vegetation. The ground is usually bare around the tree due in part to the dense shading by the canopy (Gunasena and Hughes 2000). It is classed as a quantitative long day plant, since growth is enhanced by long days but not prevented by short days (El-Siddig *et al* 1999). Tropism is observed in tamarind, with leaflets folding after dark (Gunasena and Hughes 2000).

2.1.6.1.4 Wind

The tamarind tree is very resistant to strong winds (von Maydell 1986; von Carlowitz 1986) and can tolerate violent typhoons and cyclones (von Maydell 1986; von Carlowitz 1986). Often known as the hurricane resistant tree (NAS, 1979), it has long and strong plant branches with a deep and extensive root systems, which anchors it to the ground (Coronel 1986).

2.1.6.2 Edaphic

2.1.6.2.1 General

Tamarind can be grown in a wide range of soils (Chaturvedi 1985 and Sozolnoki 1985, Morton 1987) and it tolerates a wide range of physical site characteristics. It was suggested by Sozolnoki (1985) and Galang (1955 cited in Coronel 1986) to have no specific requirements. With little or no cultivation it can flourish in poor soils and on rocky terrain (Morton 1987; Coronel 1991; Gunasena and Hughes 2000). However it will not grow on swampy sites, where there is stagnant water (von Maydell 1986; FAO 1988). The tree thrives best in slightly acid (von Maydell 1986) loamy, deep well drained, alluvial soils (Sauls and Campbell, 1984; Vogt, 1985; Coronel, 1986; von Maydell,1986; Morton, 1987; FAO 1988; Coronel, 1991) which favours the development of a long taproot (Vogt 1995).

2.1.6.2.2 Soil pH

The tree prefers slightly acid soil (FAO, 1988, Sauls and Campbell, 1984, von Maydell, 1986) pH 5.5 – 6.8, (FAO, 1988) pH 5.5 (von Maydell, 1986), though it also grows well in neutral and alkaline soils (Singh et al., 1997 cited in Rao et al., 1999).

2.1.6.2.3 Salinity and sodicity

Tamarind tolerates saline soils (von Maydell, 1986; Hocking, 1993; El-Siddig et al., 1999). Dwivedi et al. (1996) found that tamarind could grow in soil containing up to 45% exchangeable sodium. In India it tolerates saline and sodic soils where it grows in ravines and on degraded land. Older plants are more resistant to salinity than seedlings and have been found growing in saline coastal regions (NAS, 1979; Pongskul et al., 1988; Anon 1991). In northeast Thailand tamarind has been reported to establish naturally in areas with recently salinised soils (Nemoto et al. 1987). Gebauer *et al.* (2001) concluded that tamarind seedlings tolerated a moderate salinity level of 80 mM (9.3 dS/m); probably due to an increase in leaf volume associated with succulence. No reduction in growth was seen at 40nM (5.1 dS/m) or 80 mM (9.3 dS/m) but a reduction in chlorophyll concentration and photosynthetic rate with an increase in moisture content was observed. This was most likely as a response to avoid excess in concentration in the leaf tissue. The seedlings were sensitive to high salinity and showed a reduction of 70% fresh and 75% dry weight at 160nM (18.0 dS/m) over a ten week period. Gebauer et al (2001) believed that at this salinity level the root system was unable to control the invasion of salt ions, which accumulated in the shoot tissue resulting in harmful salt concentration in the leaves. Panchaban *et al.* (1989) established that tamarind was one of the most tolerant trees to salinity among several fast growing trees tested in Thailand.

2.1.6.2.4 Soil depth

Tamarind thrives in deep soils (FAO 1988; Sauls and Campbell 1984; Coronel 1986; Vogt 1995) where a long tap root has room to develop. According to Coronel (1991), plantations in the central delta of Thailand showed charteristics of dwarfism and early bearing due to stress conditions brought about by a high water table.preventing growth of a tap root system.

2.1.6.2.5 Soil drainage

Tamarind prefers well drained sandy soils (Troup 1921, Sauls and Campbell 1984, Coronel 1986, von Maydell 1986 Vogt 1995, FAO 1988) and does not tolerate seasonal flooding or water logging (Relwani, 1993; Vogt 1995). Tamarind will not produce a deep tap root in poorly drained and compact soils (Chaturvedi 1985; Chaturvedi et al 1986). In Africa the tree is reported to grow near ant hills or termite mounds, which indicate its preference for well aerated soils (El-Siddig et al., 2006).

2.1.6.2.6 Soil Texture

Tamarind is said to prefer and produces best on loam soils (von Maydell 1986, Vogt 1995)

2.1.7 Genetic variation

Although there is a wealth of tamarind germplasm across the regions where it occurs, little systematic germplasm collection and evaluation has been attempted, in spite of current value and future potential of the species (Gunasena and Pushpakumara, 2006)

Within the species, there is evidence indicating the existence of a number of local types, differing in habitat, vigour, size and quality of pods, productivity as well as pest and disease resistance. Most types are of unknown origin and represent germplasm of each region. Thus the potential of selecting or breeding better quality, higher yielding cultivars exists (El-Siddig et al., 1999). The yield of tamarind varies considerably in different countries, depending on genetic and environmental factors (Gunasena and Hughes, 2000).

Variation has been reported for tolerance to drought, wind, poor soils, water logging, high and low pH and grazing (Gunasena and Hughes, 2000). Phenological diversity also exists and tree to tree variations are common in flowering and maturing fruits (Mahadeven, 1991)

Wide phenotypic variation in tamarind has been attributed to geographic isolation and gene mutation. One such example is the origin of sweet tamarind which has been attributed to a point mutation (Gunasena and Hughes 2000).

Germplasm collections undertaken in Thailand resulted in most of the best accessions being found in the provenances along the Mae Kong River, but the role of the river in all these cases has not been fully understood (Feungchan et al. 1996a). High morphological variation is found in the African Savannahs, thought to be the centre of origin for tamarind.

2.2 Identifying an appropriate method to predict the suitable production areas for tamarind – an underutilised fruit tree species

2.2.1 Matching crops with land

The increasing effects of global environmental change with the world's rapidly rising population, requires the best possible use of land resources for agriculture, horticulture, forestry and conservation. Being able to predict where and how well particular plants are likely to grow in different regions is vital for land use planning (Booth, 1995).

From the beginnings of agriculture, farmers have been deciding the best use for the land or to identify land suitable for the crops they wish to grow (Dent and Young 1981). Through the process of matching plants, crops or varities with land, it is possible to determine whether a crop will grow in a particular environment and how that crop will perform. When deciding which species will be most suitable for a particular location, (i.e. will achieve a high yield and meet the cultural and social needs of the population), a good knowledge of the physical, biological and socio-economic parameters is required (Miézan 1998).

Physical or environmental parameters include natural resources such as climatic (e.g. temperature, rainfall, photoperiod), soil characteristics (e.g. soil type, soil pH, salinity content, iron and aluminium ions content, soil fertility, water dynamics) and topographic characteristics (slope, aspect). Determination of optimal growth conditions is mostly based on the critical evaluation of some measure of productivity or success under different edaphic and climatic conditions (Sys et al., 1991). When species are introduced to a site, many lack the appropriate physiological traits to adapt in order to produce a feasible yield or even survive in the physical environment. In many cases species may survive but conditions may limit growth or reproduction.

Biological parameters include diseases, insect pests, nematodes, weeds and other plant parasites. If a species is introduced to an area, which is inhabited by potential pests or diseases it may grow and reproduce but the harvest may be destroyed or blighted.

Socio-economic parameters include government policies for agricultural food production, farming systems, agricultural practices and management options (e.g. amount of fertiliser, pest management), cultural practices, processing facilities, consumers preferences and market opportunities. Limitation due to socio-economic factors may be brought about by lack of infrastructure, transport, lack of work force. The crop may not fit with the present farming systems or there may be a more beneficial alternative. Lack of popularity due to cultural reasons or taste can also limit viability of a location for a species cultivation. If the plant species is not popular with the local people they will not be willing to invest their time and resources in its production (Miézan 1998).

This study will focus on the physical environmental factors associated with crop distribution modelling. The physical factors are a primary limiting factor for selection of appropriate production locations; this knowledge can then be combined with biological and socioeconomic information. Choice of species should reflect the farmers priorities, but a systematic system to suggest a good match between species and site is a useful management tool.

Traditional techniques for selecting particular species for a location include;

- Local appraisal of species near planting site; native and local exotics should be assessed for performance and potential

- Climatic matching; this technique involves comparing the climate of the planting area with other equivalent climatic areas around the world. Species are then selected from these areas with adjustments for soil types or special features, e.g. salt tolerance.

- Selection of provenance for planting; this is based on the concept that provenance has a genetic and evolutionary basis. It implies that genetic

> variation is associated closely with ecological conditions in which species evolved. Application of the concept involves recognition of intra-specific variation in particular characteristics and classification of forest reproductive material according to its geographical origin (Boland 1997).

More recently these principles have been applied with more sophisticated methods. Techniques have been developed to model a species response to the physical environment, which has allowed prediction of yield or suitability of a crop for a particular site.

- Mechanistic process based models have been developed to predict yield from location specific environmental information for many of the major crops and forestry species (Stape et al., 2004). Information gained from expensive studies are used to decipher the multiple interactions and develop process-based crop growth models, which are built using mathematical equations to model quantitatively plant–soil–atmospheric interactions.

- Empirical-statistical models relate crop yields (Drummond et al., 2000; Park et al., 2005) and tree growth (Louw and Scholes, 2006) to environmental variables using methods such as regression based techniques (Park et al., 2005; Louw and Scholes, 2006) and neural networks (Drummond et al., 2000; Schultz et al., 2000).

- Land Evaluation (FAO, 1984a, 1984b, 1985) was developed to assess the suitability of a unit of land for a particular crop or land use. Thus it allows for identification of suitable production areas for species for which a lack of empirical data is available to conduct growth modelling. Such models use informal information and expert knowledge to derive relationships between the plant and their environment (Bydekerke et al., 1998; Hackett and Vanclay, 1998).

2.2.2 Methods for modelling plant-environment interactions for crop species

The following section gives further details of methods which have been used in the modelling of crop species-environment interactions in order to identify potential productions areas. It suggests the limitations of using such methods for underutilised fruit species.

2.2.2.1 Mechanistic-Process crop productivity models

Process-based models describe plant productivity based on plant physiological processes that control growth (i.e., photosynthesis, allocation, respiration, transpiration and nutrition) (Stape et al., 2004). The aim of a process model is to simulate both temporal and spatial dynamics of crop yields. The ability to include temporal changes of crop yields and extrapolation potentials; means that unlike many of the alternative modelling techniques, there are few limitations when applied to applications such as climate change (Park et al., 2005).

Crop growth modelling and simulation have become accepted tools for agricultural (Rabbinge, 1986; Seligman, 1990) and forestry research (Constable and Friend, 2000; Porte and Bartelink, 2002; Stape et al., 2004). A wide variety of crop models have been developed to serve many different purposes, including yield prediction (Seligman, 1990; Horie et al., 1992; de Koning et al., 1993; Penning de Vries et al., 1995), investigating the effects of management options (Ungar, 1990; Aggarwal and Kalra, 1994; Rotter and Dreiser, 1994), to simulate the of environmental factors of morphological and physiological characteristics (Dingkuhn et al., 1993; Hunt, 1993; Kropff et al., 1995), and to explore of the effects of climate change on crop and forestry production (Wolf, 1993; Matthews et al., 1995; Constable and Friend, 2000). Demands for advisory systems for farmers and scenario studies for policy makers have resulted in the evolution of models geared towards tactical and strategic decision support (Bouman et al., 1996), examples include Rabbinge (1986) and van Keulen and Penning de Vries (1993).

Due to their ability to cope with changing environments over time, process-based modelling approaches are often preferred to empirical alternatives. However high demands of technological sophistication and demanding calibration–verification procedures are the main limiting factors for wider application (Park et al., 2005). Lack of time and resources can often restrict model input options relevant to the scientific or policy decisions in question (Stephens and Middleton, 2002). Calibration and verification issues can be problematic especially for developing countries, where the necessary technological and financial resources are not readily available (Bouman et al., 1996; Stephens and Middleton, 2002).Consequently, parameterization often comes from previous research conducted in different environmental conditions or from expert opinion (Park et al., 2005).

Park et al.(2005) noted that the uncertainty associated with such parameterization may greatly decrease the validity of model outputs and the reliability of model application (Penning de Vries et al., 1989; Aggarwal, 1995; Bouman et al., 1996; Stephens and Middleton, 2002).

The main process components for the major crops have already been developed, however they require careful calibration and verification for local crop varieties and different environmental conditions (Park et al., 2005). Process components have not been developed for underutilised species such as tamarind. For the majority of minor and perennial species, the experiments which allow development of such models would be too time consuming and costly.

2.2.2.2 Empirical agro-meteorological crop response models

Park et al. (2005) noted that empirical growth models may even be a more reliable method of investigating crop response than poorly calibrated process models when the necessary data are available. Empirical-statistical models relate a measure of productivity (i.e.crop yields (Drummond et al., 2000; 2005) and tree growth (Louw and Scholes, 2006)) to environmental and other variables using a number of alternative methods including regression based techniques (Park et al., 2005; Louw and Scholes, 2006) and neural networks (Drummond et al., 2000; Schultz et al., 2000).

Empirical modelling techniques have been used to model major crop species. Chen and Da Fonseca (1980) used regression analysis to model the effects of weather and technology on corn (*Zea mays* L.) yield used in the district of Ribeirão Preto, São Paulo State, Brazil. Landau et al. (2000) developed a multiple-regression model of wheat yield. Their intention was to build a model which was parsimonious (i.e., has the minimum number of parameters and maximum predictive power), but in which every parameter reflected a known climate effect on the UK crop-environment system to allow mechanistic interpretation.

There are a number of examples of regression analyses in forestry studies which use resources variables. Bateman and Lovett (1998) combine and analyse data from a variety of existing large area databases concerning tree growth, plantation management and the environmental characteristics of planted sites. Principal component analysis and regression techniques are employed to estimate a number of Yield Class models for Sitka spruce (Picea sitchensis (Bong). Carr.). Fruit crop yields have also been modelled using empirical techniques. Reddy (1997 cited in Yadav et al. 2002) developed a agro-meteorological yield model for Mango orchards using path coefficients and multiple regression analysis of yield with growth, meteorological and nutrient parameters.

Much of the empirical yield and distribution modelling in forestry, growth and yield are expressed as function of tree diameter, competition, and age. Vanclay (1994) notes that forestry growth models provide an efficient way to prepare resource forecasts, but a more important role may be their ability to explore management options and silvicultural alternatives. Hokka and Goot (1999) developed a basal area growth model to predict the growth of individual trees in second-growth black spruce (*Picea mariana* (Mill.) BSP) stands on north eastern Ontario peat lands. In the model, 5-year basal area growth of a tree was expressed as a function of tree diameter, stand-level competition, tree-level competition, and peat thickness. More recently the spatial heterogeneity of multivariate relationships between tree growth and diameter has been explored (Zang et al 2004). Porte and Bartelink (2002) give details of a large number of empirical growth models and discuss their application on mixed forests.

For many underutilised species, a lack of quantitative data on growth or yield means that the development of empirical growth models is not a possibility. For perennial tree crops, establishing yields under different environmental conditions would require a considerable amount of time and space.

2.2.2.3 Development of plant environment relationships from expert knowledge

Crop modelling techniques are often handicapped by a lack of suitable information on the performance of candidate species (or variety, provenance, etc.). Hackett and Vanclay (1998) argue that frequently, the problem is not the absence of information per se, but rather that decision support systems rely on models calibrated with empirical data, and unable to utilize alternative sources such as informal data and expert knowledge (Hackett and Vanclay, 1998).

Expert systems and other approaches developed enable such data to be incorporated into models compatible with prevailing planning systems (Schulze and Kunz, 1995; Hackett and Vanclay, 1998). Young (1980) produced a questionnaire to be used for the collection of information from researchers, extension workers and farmers on environmental requirements and limitations of individual crops. Hackett (1988) developed an expert system to prepare tabular description of plant requirements for a number of lesser-known species for a land suitability project in Papua New Guinea.

A limitation of expert knowledge is the subjectivity of the expert in defining the plant – environment response. Hackett (1996a), developed the system INFER which could be distributed to a large number of experts, entries could be made in a table which suggest how a species experiences a particular soil or climate. Simple rules were applied to this information to develop a functional relationship (Hackett and Vanclay 1999). Hackett and Vanclay (1999) stated that although these preliminary relationships will rarely be adequate at first, they could be improved and retested. However the initial step of turning raw data into a series of explicit and testable relationships had been achieved.

This information was used in the system PLANTGRO (Hackett 1991a; 1991b; Iris Media 1994; Hackett and Vanclay 1998). PLANTGRO uses simple notational relationships to express the plants response to environmental factors. The relationship expressed as spline curves are developed using informal data and expert knowledge along with experimental data (when available). PLANTGRO although originally designed for field crops, has received considerable attention from other disciplines, including forestry, entomology and plant pathology (Hackett and Vanclay, 1998).

There are a substantial number of literature resources, which give informal or expert knowledge on growth requirements of tree species. However most of these refer to timber species (FAO 1974; Web et al. 1980; National Academy of Sciences, 1980; Baumer 1983; Pandey 1983). They often suffer from broad generalisation and in some cases uncritical copying from one to another. As described by Young (1984) much data is in the form of “prefers deep soils” or “moderately drought tolerant”.

Database resources which contained environmental requirement information on tamarind and other underutilised tree species include Ecocrop 1 and 2 (FAO 1999), INSPIRE (INteractive SPecies Information REtrieval) (Web et al. 1984), Multipurpose Tree Species Computerised Database (von Carlwitz et al. 1991) TROPIS, Tree Growth and Permanent Plot Information System (a database which contains details about the objectives of experiments and plot systems) (CIFOR 1997), the Agroforestree Database (Salim et al. 2001), Forestry Compendium - a silvicultural reference (CAB International), MIRA (CATIE) and TREDAT a database of growth data accumulated from trials utilising Australian species by the ASTC (Australian Seed Centre; CIRSO 1996).

However for many of these databases, information on underutilised fruit tree species is very limited. Some give ranges in which the species are known to grow for a number of environmental factors. However there is a large amount of inconsistency in information given by the databases. There is little information on plant-environment relationships/responses or geographically referenced data on current distribution or production area's.

Some expert knowledge and informal information does exist on tamarind and other underutilised crops. However this type of data tends to be highly subjective, variable and highly descriptive and difficult to quantify. Variation found in environmental requirements and interactions between environmental variables are not easily incorporated into expert knowledge systems. Such limitations can lead to inaccurate outputs and in turn poor management decisions.

2.2.3 Modelling species distribution using herbarium passport data

For the majority of modelling techniques described above, it is lack of appropriate data that is the limiting factor when modelling underutilised crops such as tamarind. Therefore few models have been developed to predict the potential production areas for such species. The modelling of minor or underutilised crops up until now has been very limited (Azam-Ali et al., 2001).

However a large source of data does exist that has the potential to provide information which can be used to predict suitable production areas. The increasing availability of plant species location or presence data in the form of passport information from herbarium records and germplasm collections has unlocked a vast resource of useful data. For underutilised species which have an economic or social importance (i.e. tamarind), large collections of herbarium records exist stored in collections across the world. The use of such data allows the quantitative modelling of the potential distribution and potential production areas for plant species, for which limited or no empirical growth and yield data exists.

Although traditionally used in taxonomic studies, herbarium data has been used to study a broad range of topics, from aspects of ecological and evolutionary theory, to applications in conservation, agriculture and human health (Graham et al., 2004a). Much of the recent interest has focused on making use of this resource for biogeographical studies, which incorporate the technique of species distribution modelling (Pearce and Boyce, 2006).

Species distribution models often refered to as environmental niche models (Chapman et al., 2005), habitat distribution models (Binzenhofer et al., 2005) or resource select function's (Boyce et al., 2002), provide predictions of distributions by relating field observations of species' occurrence to environmental variables. This relationship can be described as a response surface, which in theory can be broken down into individualistic response curves for each variable (Guisan and Zimmermann, 2000). Using presence or presence-absence data empirical techniques are used to model these relationships and predict the species potential distribution.

Numerous statistical approaches have been applied to species distribution modelling including envelopes or distance-based measures (Busby, 1991; Walker and Cocks, 1991; Carpenter et al., 1993; Beaumont et al., 2005). Generalised regression analysis (Austin and Meyers, 1996; Barry and Welsh, 2002; Lehmann et al., 2002; Araujo et al., 2005) regression tree's , ordination (Pasinelli et al., 2001; De'Ath, 2002), factor analysis (Jones and Gladkov, 1999; Hirzel et al., 2002) and discriminant analysis (Manel et al., 1999a). Recently machine learning tools such as neural networks (Pearson et al., 2002), genetic algorithms (Stockwell and Peters, 1999) and maximum entropy (Phillips et al., 2006) have also been applied.

Species distribution models are more widely known for their use in the modelling of natural distribution of species rather than crops. Species distribution models have been used to model both plant (Austin et al., 2006) and animal species (Olivier and Wotherspoon, 2005), in both terrestrial (Hirzel et al., 2002) and aquatic environments (Wiley et al., 2003). They have been implemented as a management tool in order to identify conservation priorities (Jarvis et al., 2002; Loiselle1 et al., 2003; Tole, 2006), test biogeographic hypothesis (Leathwick, 1998), improve floristic and faunistic atlases (Hausser, 1995), study evolution (Graham et al., 2004) and as a tool to assess the impact of environmental change (Thuiller et al., 2003a). A review of recent examples of species distribution models can be found in (Guisan and Theurillat, 2000) and (Guisan and Thuiller, 2005).

Although not widely used to model crop species distribution, the method has been used in biogeography studies of crop wild relatives. Jarvis et al. (2002; 2003) and Ferguson et al. (2005) modelled the distribution of the genus Arachis, wild relatives of the cultivated peanut *A. hypogaea* (Jones and Gladkov, 1999) in order to identify priority areas for collection of germplasm for ex-situ collection, as well as carrying out climatic adaptation analysis for each species. Other wild crop relatives modelled include wild rice species in East and Southern Africa (*Oryza* spp.) (Kiambi et al., 2000) and wild species of lima beans (*Phaseolus lunatus*) in the Central Valley of Costa Rica (Gonzalez et al., 2000).

Some techniques such as envelope modelling have been used to model the potential distribution of crop species and assist in identifying areas for plant introduction (Nix et al., 1977; Booth, 1999). Scheldeman (2002) modelled suitable cultivation zones of cherimoya (*Annona cherimola* Mill.) and highland papayas (*Vasconcellea* spp.). It was noted that the frequency of presence points collected under different edaphoclimatic conditions could be used instead of yields as a criterion to assess the range of edaphoclimatic parameters and their optimums. Azurdia (2006) modelled the potential cultivation area of 3 species of *Pouteria* spp. on 3 continents with passport data from accessions from Central America (Azurdia, 2004).

Herbarium passport data are significant because they provide both taxonomic and geographical information (Soberon, 1999). Of specific importance in terms of

modelling is the 'collecting event' a description of the time and place where a specimen was collected (Stein and Wieczorek, 2004). Natural history collection (NHC) passport information from herbarium, along with museum collections provide a vast resource of information (Suarez and Tsutusi, 2004).
NHC collections are thought to hold in the order of 3 x 10^9 specimens from all over the world (Soberon, 1999). Recent advances in web-enabled databases and biodiversity informatics have facilitated increased accessibility and information retrieval (Baker et al., 1998; Soberon, 1999; Bisby, 2000; Stein and Wieczorek, 2004; Soberon and Townsend Peterson, 2005; Elith et al., 2006).

There are however challenges when using such data to model species distribution, NHC data have often been collected over a long period of time in an ad-hoc (non-systematic) manner and therefore they often include taxonomic inaccuracies and biases in the spatial coverage (Graham et al., 2004a). However such issues can be accounted for (Chapman, 1999, 2004; Wieczorek et al., 2004) and passport occurance data has been successfully used in species distribution modelling (Estrada-Pena et al., 2006; Gaubert et al., 2006; Phillips et al., 2006). The processing of passport data to allow appropriate use in species distribution modelling is extensively covered in Chapter 3 and so will not be discussed here.

The major constraint with modelling occurrence data derived from herbarium records is that the intent and methods of collecting are rarely known, so that absences cannot be inferred with certainty (Elith et al., 2006). Many species distribution modelling approaches have focused on techniques which use both presence-absence data. The recent increased interest in making use of the vast resource of NHC collections has lead to an increased interest in the use of techniques for presence-only data.

Methods include adapting presence-absence techniques using samples from the background environment to create pseudo absences (Stockwell and Peters, 1999; Ferrier, 2002; Pearce and Boyce, 2006). Machine learning tool have also been applied to presence-only data (Phillips et al., 2006), although applying such methods to species distribution modelling is relatively new, the work has shown promise (Elith et al., 2006).

One type of presence only model is the envelope models. Conceptually envelope models are very close to the niche theory; as they try to delineate in environmental space, the hyper surface that circumscribes all suitable conditions (Hirzel1 and Arlettaz, 2003b). These models identify suitability habitat based on the environmental space and map this on to geographical space to produce suitability maps.

The majority of techniques developed to model presence-only data are based on envelope and distance based measures. Conceptually envelope models are very close to the niche theory (see section 2.2.3.1); as they try to delineate in environmental space, the hyper surface that circumscribes all suitable conditions (Hirzel1 and Arlettaz, 2003b). Examples include BIOCLIM (Busby, 1991), HABITAT (Walker and Cocks, 1991) and DOMAIN (Carpenter et al., 1993). A more recent development on such methods is ENFA (Ecological Niche Factor Analysis). ENFA overcomes the problem of other envelope models by taking into account interactions between environmental variables and taking account of observation density. Therefore it is not mislead by outliers (Hirzel and Arlettaz, 2003a). ENFA has been shown to perform well when compared to presence-absence models (Hirzel et al., 2001)

In this study species distribution models will be used to predict the distribution of the underutilised fruit crop species tamarind. Potential distribution models produced will be used to indicate high potential and new areas of production. This rest of this chapter introduces the use of species distribution models as a prudent alternative to modelling potential production area's of underutilised fruit crops (or any minor crop). Particularly when there is limited appropriate data (true of most cases), which limit use of the alternative methods. It will discuss the conceptual theory, the methodologies and data relevant to species distribution modelling with the implications, advantages and limitation in the use of such techniques to model the distribution of a minor or “underutilised” fruit crop species.

2.2.3.1 Niche theory

Guisan and Thuiller (2005) noted that species distribution models are not only useful tools for answering questions in applied ecology, they are also extremely relevant to fundamental science, because of the ecological and evolutionary theories underpinning them. However Austin (2002) commented that the ecological theory related to species distribution modelling has been sorely neglected in the literature resulting in a weakening in the overall approach. The concept of the Niche is central to ecological theory and thought to form the basis of species distribution modelling.

The term ‘Niche’ was coined by the naturalist Joseph Grinnell (Grinnel, 1924) and was later fully established as ‘the ultimate distributional unit within which each species is held by it structural and instinctive limitations’(Grinnel, 1928). Grinnel’s concept was geographical in nature and conceptually pre-interactive, it focused on the idealised distribution of individuals in the absence of their interactions with other species (Vandermeer, 1972). Around the same time Elton (1927) presented his notion of niche as ‘the species position in the community’ especially in regard to trophic interactions (Heglund, 2002; Morrison and Hall, 2002). His concept was post-interactive, integrating interactions with other species (Vandermeer, 1972). Arguably the greatest contribution to niche theory in relation to its use in distribution modelling was by Hutchinson (1957). His concept of the niche was conceptually non-geographic and quantitative; he described the niche by ‘the coordinates of the species with n-dimensional environmental axes’ or n-dimensional hypervolume; this hypervolume encloses ‘conditions that allow the species to exist indefinitely’ (Titeux, 2006). Hutchinson also attempted to amalgamate the pre-interactive and post-interactive concepts of niche described by Grinnel and Elton. It provided the valuable distinction between the fundamental niche, the full range of environmental conditions (biological and physical) under which an organism can exist indefinitely, and the realised niche a narrower range or reduced hyper-volume which the species is forced to occupy due to interactions with, other organisms within a community.

It is the Hutchinsonian niche theoretical framework that underpins most species distribution modelling techniques. They try to disentangle the complex population continuum and to understand the species-habitat relations in the environmental hyperspace (Titeux, 2006). The Geographic projection of the fundamental niche would identify all areas which environmental conditions were suitable. However a species is rarely found in all suitable areas (Titeux, 2006). A number of factors in addition to abiotic and biotic considerations will effect whether a species will occur at specific locations. These factors interact and have varying degrees of influence at various scales. Factors which influence species distribution include;

Environmental factors or abiotic factors; these impose physiological limits on species ability to exist (Soberon and Townsend Peterson, 2005) and their ability to provide resources for existence in an area (Austin, 2002; Morrison and Hall, 2002). Hutchinson (1957) considered abiotic factors to determine the fundamental niche.

Biotic factors are the set of interactions with other species that modify the species' ability to maintain populations (Soberon and Townsend Peterson, 2005) directly (i.e. interspecific territorial interactions) or indirectly (when two species exist together they change their own and the other species environment)(Pulliam, 2000). Interspecific competition for example may exclude individuals from some portions of their species fundamental niche (Austin, 2002). Such interactions are variable in time and space and vary with spatial scales (Wiens, 1989; Levin, 1992). Hutchinson (1957) considered biotic factors combined with abiotic factors to determine the realised niche.

Dispersal limitations (Cain et al., 1998; Clark, 1998); these limitations mean species cannot reach and therefore can be absent from suitable habitat (Pulliam, 2000). This factor is extremely useful in distinguishing a species' actual distribution from its potential distribution, based on landscape configuration and the species' dispersal abilities (Soberon and Townsend Peterson, 2005).

Historical influences and disturbance (Bolliger et al., 2000); Local extinctions can occur, due to stochastic effects; a plant or animal can be absent from a site which is highly suitable on the basis of climate and biota due to past geological or climatic events (e.g. glaciations), or disturbance by fire or human influence so that its range does not in the present time extend into this area (Guisan and Zimmermann, 2000).

The evolutionary capacity of populations of the species to adapt to new conditions; this factor, usually reserved from analysis or assumed negligible, is nevertheless an additional and important consideration in outlining the distributional possibilities of species. Ecological theory (Holt 1996a, b; Holt and Gaines 1992; Holt and Gomulkiewicz 1996; Kawecki 1995) and limited experiments carried out to date (Etterson and Shaw 2001), have indicated that the effects of evolution in niche parameters over short periods of time appear minor.

Geographical aspects and landscape pattern; this considers the composition and configuration of the landscape such as habitat fragmentation, connectivity (Wiens 1992, 1993, Forman 1995).

Meta-population dynamics (Pulliam, 2000; Baguette, 2004) and source-sink dynamics (Pulliam, 1988, 2000). These concepts further influence the relationship between niche concept and species distribution. In particular, the meta-population dynamics consists of a turnover of extinction and (re)colonisation of suitable habitat patches in the landscape (Pulliam, 2000; Baguette, 2004). This dynamics is governed by a variety of factors (not detailed here) and explains the existence of unoccupied suitable habitat patches.

Hutchinson accounted for competition through the development of the realised niche which describes the reduced hyper volume, brought about by the competition and interaction with other species. In a non-spatial context this does describe the conditions in which a species can survive based on its "structural or instinctive limitation". However the geographic distribution of a species is determined not only by factors considered in niche theory but is influenced by a great number of factors (as listed above), a number of which are geographical/spatial or temporal in nature. Therefore they are not easily described in ecological space and not easily incorporated into the niche concept. A considerable decoupling between habitat suitability as described by the niche and species distribution may therefore be observed in the real world and this distortion complicates any habitat suitability study (Pulliam, 2000).

2.2.3.2 Species distribution modelling and niche theory

The most frequent simplification found in species distribution model literature is to state that, because of the observed distribution is already being constrained by biotic interactions and limiting resources; species distribution models are *de facto* quantifying Hutchinson's realised niche of the species. Occurrence records however, being samples of the species geographic distribution include the effects of all factors that influence them (Phillips et al., 2006). In the majority of species distribution modelling studies, choice of model variables are often limited mainly to abiotic factors, those that determine the fundamental niche. Other variables which influence distribution are rarely used as they are generally difficult to measure, and therefore complex to interpret or unavailable (Thuiller et al., 2003; Soberon and Townsend Peterson, 2005). Therefore the multidimensional cloud/multidimensional space/model is developed and species environment relationship derived based on only abiotic factors but with distribution points which are influenced by all factors. A more accurate statement therefore might be that a model represents an approximation of the species' realized niche, in the study area with environmental dimensions being considered (Phillips et al., 2006).

In the prediction process the algorithm searches the map for regions resembling in abiotic terms the occurrence points, whose geographical position is determined by all processes that determine distribution. Therefore whether or not a model captures a species' full niche requirements; areas of predicted presence may include areas outside the current realised or actual distribution (Phillips et al., 2006). This projection of the model into geographic space is often termed the species' potential distribution (Phillips et al., 2006).

Models of species distribution usually focus on environmental (abiotic) predictors, however in a number of studies, processes such as dispersal, competition; succession and disturbance have attempted to be incorporated (Austin, 2002). When investigating the phenomena of 'beech gaps' is New Zealand Leathwick (1998) emphasised the need to combine ecological knowledge with statistical modelling. The gaps, which occurred on both North and South Island did not correspond with any environmental factors currently known to influence species distribution in New Zealand. It was found that variables used as surrogates for competition (Leathwick and Austin 2001; Leathwick 2002) disturbance and dispersal processes (Laethwick and Mitchell 2002) had more significant relationships than many of the environmental predictors. Austin (2002) noted that it is often difficult to distinguish whether the absence of the species is due to competition or an unidentified environmental variable.

2.2.3.3 Equilibrium between species distribution and the environment

As species and environmental data are usually sampled during a limited period of time or/and space, models fitted using these can only reflect a snapshot view of the expected relationship. Distribution models automatically assume equilibrium, or at least pseudo equilibrium (where change is slow relative to the life span of the biota (Austin, 2002)) between the environment and the observed species patterns (Guisan and Thuiller, 2005)(Guisan & Theurillat 2000). This will vary depending on the degree to which history and disturbance are important in the area under study (Austin, 2002). Often in reality the concept of non-equilibrium is more realistic (Johnston and Chapin, 2003). A non-equilibrium distribution occurs when the range is not static, this can result from history, disturbances, stochastic effects (Bolliger et al., 2000) such as geological or climatic events (Guisan and Zimmermann, 2000). Leathwick (1992) gave an example where disturbance by a volcano eruption meant that the slow dispersing Nothofagus species had not yet reached habitat it would normally dominate. Examples of tree species which appear not to be in equilibrium due to continued post glacial expansion in various continents have been discussed (Leathwick1, 1998; Johnston and Chapin, 2003; Svenning and Skov, 2004).

Non-equilibrium distribution can make predictions hazardous (Bolliger et al. 2000). Modelling of data which is not in equilibrium would lead to empirical estimates of the response surface of the species to be incomplete or inaccurate resulting in a biased representation of the full potential range of the species (Johnston and Chapin, 2003; Guisan and Thuiller, 2005). Invasive species are not in equilibrium with the environment within the introduced range, and thus should preferably be modelled using their distribution in the native range (Peterson 2003). Introduced crops species may also not be in equilibrium with the environment.

2.2.3.4 Species-environment response curves

In the majority of cases, the purpose of the statistical modelling is the prediction of species distribution. Underlying species distribution models is the premise that the predictable relations exist between species and certain features of the environment. The detection of relationships between species and environment, explanation of the response surface and response curves and the testing of ecological theory tend to be secondary considerations (Guisan and Zimmermann, 2000).

A common criticism of species distribution models is that most of them are based on correlations between species response and measured environmental variables. Correlation does not automatically infer causality or process, therefore providing little insight pertaining to the proximate mechanisms underlying such relations (e.g. Capen 1981;Van Horne 2002). It only provides a description of the functional relationship. Understanding the processes which create the observed patterns is critical and needed, if we are to devise or implement efficient management strategies (Heglund 2002).

Austin (1980, 1985), Austin et al. (1984), and Austin and Smith (1989) defined three types of ecological gradients, namely resource, direct, and indirect. Resource gradients address matter and energy consumed by plants or animals (nutrients, water, and light for plants, food, and water for animals). Direct gradients are environmental parameters that have physiological importance, but are not consumed (temperature, pH). Indirect gradients are variables that have no direct physiological relevance for a species' performance (slope, aspect, elevation, topographic position, habitat type, geology (Guisan and Theurillat, 2000).

The use of direct and resource gradients as predictive parameters means that predictions are based on what are thought to be more physiologically 'mechanistic' processes; this ensures that the model is more general and applicable over larger areas (Guisan and Theurillat, 2000). The more knowledge of ecological process that can be incorporated, the more explanatory power the model is likely to have (Austin, 2002). Prediction can be achieved without correlation having any particular ecological process, but the result is unlikely to be robust (Austin, 2002). Indirect variables can be used as surrogate, however as they are based on purely functional relationships, the response curve can take any shape (Austin, 2002) and the model may only be applicable within a limited geographical extent without significant errors (Guisan and Theurillat, 2000)

Assumptions about the shape of the response of species to an environmental variable are central to the predictive modelling environment. Most ecological textbooks present the response as a unimodal symmetrical bell shaped curve. Austin (1999) drew attention to the lack of evidence for this assumption. Mueller-Dombois and Ellenberg (1974) put forward the theory that assumes a more superior competitor can displace a species from the optimum of its fundamental niche. As a consequence the observed response may take a wide variety of shapes from skewed to bimodal (Austin, 2002). Effects such as biological interactions have the effect of decoupling systems from direct physical processes by introducing spatial lags in system dynamics or creating webs of indirect effects (Wiens, 1989).

2.2.3.5 Spatial nonstationarity in distribution modelling

As larger extents are modelled, it is highly likely that heterogeneity in the predictors variable increases and that areas are included where species respond to habitats in different ways because of different ecological status (Osborne and Suarez-Seoane, 2002; Estrada-Pena et al., 2006). Therefore the observed geographical patterns and relationships in ecology tend to be spatially variable (Jetz et al., 2005), this concept is termed spatial non-stationary.
Due to spatial nonstationarity, the parameters of the model describing the relationship may vary greatly in space. Models built on one small area may not apply to any other (Osborne and Suarez-Seoane, 2002). Global models built over a large area may have weak local predictive power because of differences in the habitats available or selected (Osborne and Suarez-Seoane, 2002; Foody, 2004). Even if the underlying ecological abiotic processes are universal, the observed patterns will vary with local conditions (Jetz et al., 2005). Osborne and Suarez-Seoane (2002) and Estrada-Pena et al. (2006) both found large scale models improved when data was geographically partitioned before analysis. Geographically-weighted regression has also been used as a solution to spatial nonstationarity (Foody, 2004).

Nonstationarity can be caused by variation in community structure. Due to the influence of inter-specific competition on the shape response curve, variation in community structure across the range may lead to variation in the observed response of the species (Peterson and Holt, 2003). Prinzing et al. (2002) suggests that although species susceptible to competitive displacement may be easily moved from their position along a niche axis, they may only be moved over a short distance.

Niche evolution can also cause spatial nonstationarity. Geographical variation within the species niche (Peterson and Holt, 2003; Estrada-Pena et al., 2006) brought about by processes such as adaptive radiation (Schluter, 2000) and genetic stochasticity or drift (Pulliam, 2000) leads to differentiated populations (on the basis niche characteristics) existing across the full species range. These populations are obscured when niche based models are applied to the entire area in which the species is distributed.

Ecological theory (Holt and Gaines, 1992; Holt, 1996; Peterson, 2003a) and evidence (Huntley et al., 1989; Beerling et al., 1995; Peterson et al., 1999; Peterson and Vieglas, 2001; Prinzing et al., 2001; Peterson and Holt, 2003; Peterson, 2003a; Thuiller et al., 2005; Estrada-Pena et al., 2006) have suggested the existence of phylogenetic conservation in niche characteristic's (niche conservation). This is thought to be caused by long term natural selection pressures which maintain the ecological (fundamental) niche without substantial modification. It is likely in at least some species over evolutionary time period. It must also be noted that apparent geographical variation within the niche which may be attributed to niche evolution may be caused by phenotypic plasticity (Schluter, 2000; Peterson and Holt, 2003).

2.2.3.6 Scale

Scale is best expressed independently as resolution (grain size) and extent of the study area (Guisan and Thuiller, 2005). Individual species and therefore their communities are influenced by factors at multiple scales (McPherson et al., 2006). The relationship between pattern and process closely relates to scale. These relationships can cause different spatial patterns of species distribution to be observed at different scales (Guisan and Thuiller, 2005). Processes may vary in importance at varying scales and patterns observed on one scale may not be apparent at another (Wiens, 1989; Guisan and Thuiller, 2005). For example inter-specific competition can only be detected at a resolution where organisms interact and compete for the same resources (Guisan and Thuiller, 2005). Patterns at large scale may be dependent on processes working at a local scale or visa versa (Heglund, 2002)

Evidence suggests that in broad-scale studies (continental or regional scales), abiotic factors (Pearson et al., 2002; Thuiller et al., 2003a; Pearson et al., 2004) and dispersal constraints prove particularly important (Soberon and Townsend Peterson, 2005). These large scale influences of habitat selection tend to override local effects of inter-specific competition. In fine-scale studies, at the level of local landscapes, issues of historical heterogeneity and accessibility are less important. The role of inter-specific interactions and meta-population (source-sink) dynamics have more influence (Peterson and Holt, 2003; Guisan and Thuiller, 2005; Soberon and Townsend Peterson, 2005). A gradual distribution observed over a large extent and at a coarse resolution is likely to be controlled by climatic regulators, whereas patchy distribution observed over a smaller area and at fine resolution is more likely to result from a patchy distribution of resources, driven by micro-topographic variation or habitat fragmentation (Guisan and Thuiller, 2005).

When modelling over larger scales the factors which affect the distribution pattern at local scales are less likely to have an influence over the shape of the species response in comparison to environmental predictors which act at large scales. It has also been argued that large scale patterns may be less likely to be influenced by chance events (Wiens, 1989) and are influenced by processes which change more slowly over time. Therefore patterns are more likely to be stable (in equilibrium) over longer periods. It is suggested that due to the greater number of factors effecting small scale distribution such as continuous natural or anthropogenic disturbance and succession species may never be in equilibrium with its environment at such scales (Bolliger et al., 2000). Therefore increasing resolution and extent, may reduce the influence of less predictable factors such as inter-specific competition (Beerling et al., 1995). Thuiller et al. (2003) gives an example of large scale models preforming more poorly in areas where the likely hood of human disturbance was high in comparison to areas less likely to be effected. This suggests that even large scale models may be influenced by factors such as disturbance.

The extent of the area included in a study can influence how much of the full species range is modelled. Most published papers on species distribution models describe models developed for sections of a species' range (Randin et al., 2006). Few models have been fitted at global scale (Prentice et al., 1992; Jeffree and

Jeffree, 1994; Box, 1996; Gevrey and Worner, 2006) and/or incorporate the full extent of a species' native or endemic geographical range (Peterson et al., 2000; Peterson, 2003a). When distributions span beyond the boundaries of study areas (Van Horn, 2002; Thuiller et al., 2003a), frequently only a part of the environmental gradient is sampled. The response curves derived are often incomplete descriptions of the species response to the environmental predictors, leading to nonrobust models, which are only applicable within that region (Thuiller et al., 2004; Guisan and Thuiller, 2005). A number of studies have attempted to calibrate models at a coarse resolution over a large extent to ensure capture of full climatic range of the species, before projecting the species distribution on finer scale grids or into new environments (Pearson et al., 2002; Pearson et al., 2004; Araujo et al., 2005; McPherson et al., 2006).

2.2.4 Modelling the global distribution of suitable production areas for the underutilised fruit tree species, tamarind

Tamarind is an ancient domesticate (Gunasena and Pushpakumara, 2006). Tamarind has been harvested, managed or cultivated by man for thousands of years and as a result has been subject to high levels of human disturbance. The current broad distribution of the species across the tropic is largely due to introduction by man. This may effect whether the species is in equilibrium with the environment (Thuiller et al., 2003; Thuiller et al., 2003a). While naturalised in a number of countries, tamarind is still considered an exotic in many. In locations which have seen relatively recent introduction it is less likely to be in equilibrium with the environment. The introduction and management of tamarind by man also means that it may exist outside its fundamental niche.

For most of the records collected from herbarium data it is very difficult to identify whether the plant in question was wild, managed or cultivated. Therefore a number of herbarium records' spatial-localities will be affected by those factors which influence the geographical location of a wild species. The mixed nature of the data (wild – cultivated) may influence the species environment relationship described by the models.

Tamarind is frequenty cultivated in low input, non-irrigated, agricultural systems. Conditions which limit the distribution of cultivated species will differ from those faced by wild species. Competition with native biota is often greatly reduced (although pests and disease could be important in some cases) (Wiens and Garham, 2005) . Natural barriers to dispersal may no longer apply. Issues of landscape pattern, habitat fragmentation and meta-population dynamic are likely to have less influence in regard to predicting potential production areas. In order to identify suitable conditions for production of these crops, it is important to focus on the factors which limit the distribution range in such agricultural systems. Factors which dictate suitable production areas are therefore likely to be abiotic, i.e. climate, soil etc. These were the factors considered by Hutchinson in the fundamental niche concept (James et al., 1984).

The high level of human introduction increases the likelihood of evolutionary effects in the form of geographic variation in niche characteristics (Peterson and Holt 2003). This leads to geographical subpopulations and therefore spatial nonstationarity. Introduction into new areas may have driven the evolutionary process as plants adapted and expanded into areas with different environmental conditions to natural habitats. They have also undergone human selection within various regions using different selection priorities for plant characteristics, based on utilisation and cultural difference.

Little work has been conducted to investigate the extent of niche evolution or niche conservatism occurring in domesticated or semi domesticated species. Miller and Knouft (2006) found significant differences between the distribution and environmental conditions experienced between wild and cultivated populations of *Spondias purpurea* (Anacardiaceae). This corresponded with the expansion of the species during the domestication process from its native habitat in the Mesoamerican tropical dry forests into "less seasonal" habitats. The Author's raised the testable hypothesis that these differences in the niche reflect artificial selection during domestication leading to differentiation of ecotypes. However, reciprocal transplant experiments are required to determine if there is a heritable basis for the habitat differences. Miller and Knouft (2006) did however find evidence to support conservation of the niche of wild populations within that of the cultivated populations. It is possible the expansion of realised niche (conditions experienced) may at least partly be due to the removal or reduction of natural dispersal barriers and competition, and the lack of need for conditions which allow self propagation. Vetaas (2002) suggested the removal of competition as reason for the differences in the ecological conditions experienced between populations of wild and exsitu *Rhododendron*.

2.3 Project aims and objectives

In chapter one it was noted that underutilised species have the ability to contribute to addressing the urgent issues in the world, in terms of agriculture, food and nutrition security. The increased need to make use of marginal land for agriculture and the occurrence of climate change, has resulted in the greater recognition of the importance of underutilised species. It is important to learn more about such species' relationship with the environment and to identify potential areas for production. In this study the underutilised tropical fruit tree tamarind as is selected as the target species.

Although there is limited data on the eco-physiology and environmental requirements of tamarind (as with most underutilised species); the existence of passport records from herbarium data provides information on the occurrence of underutilised species. By combining such data with environmental datasets and modern statistical techniques, it is possible to investigate species relationship with the environment and model of the potential distribution. This allows identification of potential production areas.

To incorporate the full environmental gradients encountered and delineation of the full response surface of the species; tamarind will be modelled over its full distribution range. It is assumed that at a global scale the local effects on the species distribution (such as competition or landscape pattern) which are responsible for skewing the response curve of the abiotic variables, will be overridden. This will allow the response surface to be closely fit the fundamental relationship between the species and the abiotic variables.

Tamarind is a specialised domesticate and is therefore likely to be subject to a high level of human introduction. This increases the likelihood of evolutionary effects in the form of geographic variation in niche characteristics (Peterson and Holt, 2003). This leads to geographical subpopulations.These populations can be obscured when models are applied to the whole study area. There is a need to identify whether global or regional models may be more suitabile for modelling underutuilised species such as tamarind.

The “presence-only” modelling technique ENFA will be used to model the distribution of the underutilised fruit tree species tamarind (*Tamarindus indica*).

This project aims to;

- Create a representative sample of the entire global range of the tropical fruit tree species tamarind.
- Aquire further information of the species relationship with the environment.
- Assess the benefits of using regional and global models to identify the potential distribution, production areas of tamarind (underutilised crop species).
- Map the potential production area of tamarind

The techniques used in this study should provide an example of how the potential production area can be modelled for all underutilised crops or any species for which limited ecphysiological or productivity (yield, growth etc.) data exist and would be too costly to aquire.

3 *Chapter 3 Developing a Species and Environmental database for modelling the potential production areas of* Tamarindus indica

3.1 Introduction

In order to carry out species distribution modelling, two types of data are required; i) data on where the species is known to occur and ii) data characterising the environmental condition for both locations of known occurance of the species and on the areas in which the species distribution will be predicted. Advances in computer technology have allowed increased accessibility and processing ability of such data. The digitisation of biological and passport data from herbaria, museums and germplasm surveys means that data on the species occurance has become widely accessible and available through the internet (Bisby, 2000; Graham et al., 2004a). Higher resolution and more accurate geographic environmental datasets continue to be created (Dobos et al., 2001; Hijmans1 et al., 2005). This has allowed modelling of the distribution of a greater number of species across a wide range of extents and scales (Hijmans et al., 2000; Scheldeman1, 2002; Jones and Thornton, 2003; Midgley et al., 2003; Reese et al., 2005).

However such data often inherently carries error, the data is often collected over a long period of time, collected through non stratified sampling design and not for the specific purpose of modelling. Data quality is an important issue (Chapman, 2004) which is often neglected in species distribution modelling (Wieczorek et al., 2004), leading to flawed outputs and poor management decisions.

This chapter discusses current best available global climate and soil datasets suitable for distribution modelling, the advantages and limitations of such data. It also covers the current growth in biodiversity informatics through the digitisation of specimen data from herbarium records and the importance of dealing with error and bias within such data when the intended use is in modelling of species distribution. In this chapter a species occurrence dataset which covers the entire global extent of the tamarind distribution will be created along with a database of appropriate geographical environmental datasets for modelling plant distributions at such a scale. This will be carried out through data processing, data quality evaluation and data cleaning of both environmental and species datasets. The clean species dataset will be mapped both in geographical and environmental space. Statistical analysis will be conducted in order to explore the data to gain information on the species environmental requirements and niche behaviour.

3.2 Geographic Environmental/Eco-geographical datasets

In geographical environmental datasets land is classified by specific characteristics such as soil type, annual temperature etc. This information is normally portrayed in the form of digital classification maps in which specific areas are classified as homogenous. Geographic digital data sets normally come in two formats, vector and raster. Datasets vary in detail/resolution, scale and extent (the size of the area covered). Geographic digitised environmental data can be readily combined, queried and displayed within a GIS. The type of environmental variables (i.e. temperature, rainfall, soil or topography) used in modelling are selected based on ecological assumption that they are likely to have a direct or indirect relationship with the species or are correlated with variables that do. Examples of Global geographic datasets include the CRU 1961-1990 Mean Monthly Climatology girded dataset (New et al., 1999) and the Digital Soil Map of the World (FAO, 1995).

3.2.1 Global climate datasets

There have been a number of significant developments in the production of global spatial raster climate datasets. These datasets are constructed from station climatological normals. The station data is interpolated as a function of latitude, longitude, and elevation using thin-plate spline (Hutchinson, 1995).
New et al (1999) created, a 0.5° latitude/longitude (55.6 km at the equator) surface climatology of global land areas, excluding Antarctica. A suite of 9 climatic

variables were constructed from station climatological normals (from 1961–1990), numbering between 19800 (precipitation) and 3615 (wind speed) records (New et al., 1999). The datasets were developed to fill a need by biophysical modellers for a suite of global high resolution long term mean climatology.

New et al. (2002) improved on this with a 10 minute latitude/longitude data set of mean monthly surface climate over global land areas, excluding Antarctica. The new datasets had an increased spatial resolution, the incorporation of additional station data and the inclusion of a description of precipitation variability, enabling the calculation of probability distributions of monthly precipitation.

More recently Hijmans et al. (2005) developed the WorldClim database, a 30 arc second spatial resolution (equivalent to about 0.86 km^2 at the equator) suite of climate surfaces. They compiled monthly averages of climate as measured at weather stations from a large number of global, regional, national, and local sources, mostly for the 1950–2000 periods. The data was interpolated using the thin-plate smoothing spline algorithm implemented in ANUSPLIN (Hutchinson, 2004).

3.2.2 Global soil datasets

Institutes and organizations involved in applied research on a global scale, for example climate change and the greenhouse effect, or studies such as "Agriculture Towards 2010" (Alexandratos, 1995), have a definite need for soil information. These data are a crucial input in models that simulate crop growth and calculate anticipated yields and water balance, or to assess the environmental impact of different land-use practices (Nachtergaele, 1999).

Soil factors are known to be important to plant species, and used in some plant distribution modelling studies (Bragazza and Gerdol, 1996; Pinto and Gegout, 2005). However they are often disregarded due to lack of accurate data. Coudun et al. (2006) found that the inclusion of soil variables in species distribution models significantly improved the quality of predictions *Acer campestre* (L.) in French forests.

Substrate data, both physical and chemical, can be one of the most difficult to obtain and the quality from one data source to another can be extremely variable. Soil mapping has been carried out in most regions of the world, but this is at varying scales and completeness (McBratney et al., 2003; Chapman, 2004).

Only two relatively large scale soil maps exist: a 1:10 million scale map prepared by Kovda and coworkers (Nachtergaele, 1999), and the 1:5 million scale FAO-Unesco Soil Map of the World (FAO, 1971-1981). It is generally accepted that the 1:5 million scale FAO-Unesco map is the most appropriate source of soil information for studies at a continental, regional or global nature (Nachtergaele, 1999).

The legend of the original Soil Map of the World (FAO, 1974) comprises an estimated 4,930 different map units, which consist of soil units or associations of soil units (FAO, 2003). It contains direct information on the composition of each mapping unit in terms of the soil type that is dominant, associated or included, the topsoil texture of the dominant soil type, the dominant slope class of the unit and the eventual soil phase present (Nachtergaele, 1999). Quantified soil information can be derived from this direct soil information on the basis of rules worked out for the interpretation of the Soil Map of the World.

These rules include pedotransfer function and taxotransfer function. A pedotransfer function is a mathematical relationship between two or more soil parameters which shows a reasonable high level of statistical confidence. This relationship is used to facilitate the estimation of a non-measured soil parameter from one or more measured ones (Nachtergaele, 1976). A taxotransfer function is the estimation of soil parameters based on modal soil characteristics of soil units, as derived from a combination of their classification name or taxon (which by definition often implies a certain range for a number of properties), expert knowledge and empirical rules, and a statistical analysis of a large number of soil profiles belonging to the same taxon. A large number of soil parameters have been derived in this way, particularly by FAO on the CD-ROM version of the Soil Map of the World (FAO, 1996) and by Batjes (1995; 1996; 1997) for the WISE data set (Nachtergaele, 1999).

In the early 1990's, FAO recognized that a rapid update of the Soil Map of the World would be a feasible option if the original map scale of 1:5 M were retained. Parallel programmes of ISRIC (International Soil Reference and Information Centre), UNEP (United Nation Environment Programme) and FAO (Food and Agricultural Organisation) merged together in mid-1995, when at a meeting in Rome the three major partners agreed to join all resources and work towards a common world SOTER-shell approach covering the globe at 1:5 M (Nachtergaele, 1999)

The Soil and Terrain Digital Database (SOTER) (UNEP/ISSS/ISRIC/FAO, 1995a) program provides natural resource data that can be readily accessed, combined and analyzed from the point of view of potential use and production, in relation to food requirements, environmental impact and conservation. Fundamental to the SOTER approach is the mapping of areas with a distinctive, often repetitive pattern of landform, morphology, slope, parent material and soils at 1:1 million scale (SOTER units). Each SOTER unit is linked through a Geographic Information System with a computerized database containing all available attributes on topography, landform and terrain, soils, climate, vegetation and land use. In this way, each type of information or each combination of attributes can be displayed spatially as a separate layer or overlay or in tabular form (Nachtergaele, 1999).

Staff at the ISRIC developed a uniform methodology for a global soil profile database in the framework of WISE (World Inventory of Soil Emission Potentials) (Batjes and Bridges, 1994). During this project a wide selection of soil profiles from all regions of the world were screened for completeness and incorporated into the WISE handling system. The central aim of the WISE database was to provide a basic set of uniform soil data for a wide range of global and regional studies. To avail this, all profiles have been classified according to both the original and revised legend of the soil map of the world. Thereby derived soil profile interpretations can be linked to, and spatially displayed geographically through the digital Soil Map of the World (FAO, 1995).

A number of efforts have been made at digitizing the Soil Map of the World. In 1993, FAO and ISRIC combined efforts to produce a raster map with a 30' x 30' cell size in the interest of the WISE (World Inventory of Soil Emissions) project (Batjes et al., 1995). This database contains the distribution of up to 10 different soil units and their percentages in each cell. In 1996, FAO produced its own raster version which had the resolution with a 5 x 5 minute cell size (9 km x 9 km at the equator) with a full database completely corresponding with the paper map in terms of soil units, topsoil texture, slope class and soil phase. Version (3.6) has recently been published (FAO, 2003).

In order to assign values to the cells, the FAO (1995) method was to use the Soil type found at the centre of each cell.This method ensures that each cell does not represent the soil type of largest polygon within the cell, but that of the soil polygon at its centre. The dataset developers believed this is a more accurate representation of the information, as this does not bias against the soil types that occur in smaller map units (FAO, 2003).When the digitised soil map is combined with derived soil parameters, another difficulty arises in assigning values to other than the dominant soil type. Consequently this disregards any of the associated types which may cover a significant part of the grid cell. This can lead to problem in distribution modelling with species being associated with soil characteristics values in which they would never be able to exist.

When a digitised version of the soil map of the world was presented as part of the TERRSTAT database (FAO, 2002) the values of the derived soil properties raster were presented in a range class format for both dominant and major associated soil type. Presenting data in class values and in terms of dominancy protects the sense of accuracy of the data provider and obliges the modeller to make a conscientious choice in selecting the most likely value for each cell (FAO, 2002).

With the completion of the Digital Soil Map of the world at FAO (FAO, 1995) and the World Inventory of Soil Emission Potentials (WISE) database at ISRIC (Batjes and Bridges, 1994; Batjes, 1997) it became possible to present updated, derived parameters for a number of soil units considered on the Soil Map of the World. This was delivered as ISRIC ver. 2.1 Derived Soil Parameters which contained data from both the WISE and SORTER database's (Batjes, 2002). The derived data set is considered appropriate for use in studies at a regional or global scale (Batjes, 2002).

Nachtergaele (1999) noted that the greatest disadvantage of soil maps in general, as perceived by modellers and geo-statisticians, is that boundaries drawn on a map are based on expert opinion and are highly dependent on the soil classification applied. Ideally, non-soil scientists would prefer remotely sensed datasets or thematic dataset developed from geo-statistically krigged randomly sampled data.

Nachtergaele (1999) believed that although some of this criticism is valid, the alternative as described above is an unworkable proposition on a worldwide scale. He believed the purely mechanistic approach was unsound, because it considers soils in isolation from their major pedogenetic factors, and requires a large (and uneconomical) sampling density to cater for soil variability. He noted that another weak point of this profile-based approach is that structural and morphological soil characteristics are often ignored and become highly dependent on laboratory analysis. This does not mean that soil profile information should not be collected, but it emphasizes that the information on its own and out of its landscape context is of little value. This is precisely why the SOTER approach, which marries the best of classical soil survey methodology with sophisticated modern technology approaches, is promoted by ISRIC, UNEP and FAO (Nachtergaele, 1999).

The original idea of SOTER was to develop the system worldwide at an equivalent scale of 1:1 M in order to replace the paper Soil Map of the World

(Sombroek, 1984). However, it soon became obvious that the resources were lacking to tackle and complete this huge task in a reasonable time frame. However, this still remains the long-term objective pursued on a country-by-country basis, mainly by ISRIC and UNEP (Nachtergaele, 1999)

Despite the continued focus on updating of current soil maps by the major actors; the production of digital soil maps based purely on soil profile information, as opposed to digitised (existing) soil maps, is moving inexorably from the research phase (McBratney et al., 2003), to production of maps for regions and catchments and whole countries. Notable examples include the map of the Murray–Darling basin of Australia (Bui and Moran, 2001) (Bui and Moran, 2003) comprising some 19 million 250_250 m pixels or cells and the digital Soil Map of Hungary (Dobos et al., 2000; Dobos et al., 2001). However development of such maps at a global scale has yet to be achieved.

3.2.3 Soil water balance models

The relationship encountered between plants, climate, water and soil are complex and many biological, physiological, physical and chemical processes are involved (FAO, 1979). The growing period for most crops continues beyond the rainy season and, to a greater or lesser extent, crops mature on moisture stored in the soil profile. However, the amount of soil moisture stored in the soil profile, and available to a crop, varies with a number of factors which include depth of the soil profile, the soil physical characteristics, and the rooting pattern and water management of the crop. A soil water balance equation is used to calculate the water inputs (rainfall/irrigation) and outputs (evapotranspiration) within the soil system and predict where there may be crop water shortages or overflow.

The practical procedure to estimate crop water requirements, which has become a widely accepted standard, was developed by the FAO in the 1970's. The methodology was published as FAO Irrigation and Drainage Paper No24 and "Crop water requirements" (FAO, 1977). This was followed by FAO Irrigation and Drainage No. 33 "Yield response to water"(FAO, 1979). In 1998 the Penman-Monteith combination was adopted as a new standard for reference evapotranspiration (ET_o). Published as Irrigation and Drainage Paper No 56: "Crop Evapotranspiration" (FAO, 1998), this document presented an updated procedure for calculating reference and crop evapotranspiration from meteorological data and crop coefficients.

The concept of the reference evapotranspiration (ET_o) was introduced to study the evaporative demand of the environment independently of crop type, crop development and management practices. Reference evapotranspiration is a representation of the environmental demand for evapotranspiration of a short green crop, completely shading the ground, of uniform height and with adequate water status. As water is abundantly available at the reference evapotranspiring surface, soil factors do not affect reference evapotranspiration. It obviates the need to define a separate evapotranspiration level for each crop and stage of growth. ET_o values calculated at different locations or in different seasons are comparable as they refer to the same reference surface (FAO, 1998).

Differences in the crop canopy and aerodynamic resistance relative to the reference crop are accounted for within the crop coefficient (K_c). The K_c serves as an aggregation of the physical and physiological differences between crops (i.e. resistance to transpiration, crop height, crop roughness, reflection, ground cover and crop rooting characteristics result in different evapotranspiration levels) (FAO, 1998). A Crop evapotranspiration (ET_c) was derived from a calculated reference evapotranspiration (Et_o) and crop coefficients (K_c). Precipitation records, soil water storage, crop evapotranspiration (ET_c) and a soil depletion factors are used to calculate actual Evapotranspiration (ET_a).

Soil water balance models compare moisture supply to crops from precipitation and storage in soils with reference evapotranspiration (ET_o). This allows estimations of actual evapotranspiration (ET_a), moisture availability, water excess (runoff) and deficits to be calculated at a particular location under particular conditions over a time period. The ET_a value can be used as an estimate of yield or to develop geographical datasets of moisture availability for use in land evaluation.

3.2.3.1 Crop coefficients for underutilised fruit crops

The crop coefficient (K_c) value for a number of the more major tropical fruit crops such as Citrus and Avocado are listed in FAO (FAO, 1998), however no K_c value's has been derived for underutilised fruit crops, such as tamarind. A K_c was derived for a generic "Fruit Tree Tropical" to be used in CATCHCROP an integrated crop model, which is capable of simulating yield response to water deficit and fertility depletion (Perez et al., 2002). A K_c of 0.9 was derived from the K_c values of Coffee, Cacao, Palm and Tea listed in FAO (1998) and field experiment estimates for Lichee and Mango (Perez pers. comms.).

CATCHCROP has also been used in order to simulate Javanese Gardens (containing banana, jackfruit, mango, rambutan and underlying annual crops) for which a $K_c = 1.0$ was selected (Perez pers comms.). Perez was aware of the limitations of using an average value but noted that CATCHCROP was not designed as a stand-alone crop model, but to be encapsulated into a more generic decision support model. K_c values for Mango and Lichee are almost certainly in the range of 0.9 and 1.1 nearly all year round (water availability not limiting), for mature trees. Perez suggests that stomata control in tamarind may be much more efficient in comparison with these "lush" tropical tree crops (Perez per comms.).

3.2.3.2 Water balance datasets

Water balance models have been used to characterise moisture availability conditions for use in land evaluation and species distribution modelling. The Global Agro-Ecological Zones (FAO/IIASA, 2000) developed a raster dataset based on the concept of Length of growing period (LGP) i.e. the period during the year when both moisture availability and temperature are conducive to crop growth. Thus, in a formal sense, LGP refers to the number of days within the period of temperatures above 5°C when moisture conditions are considered to be adequate. The FAO methodology (FAO, 1998) was used to develop a water balance using the reference evapotranspiration and a $K_c = 1$ in order to calculate a non-crop specific actual evapotranspiration. For established crops, an actual evapotranspiration of 0.4 - 0.5 times the level of reference evapotranspiration was considered sufficient to meet water requirements of dry land crops (FAO/IIASA, 2000).

Recent developments of the BIOCLIM modelling system (ANUCLIM) (Houlder et al., 2000) include moisture index values calculated from the weekly precipitation and evaporation values in conjunction with the soil type and maximum soil water availability.

3.2.4 Data quality and error in geographical environmental datasets

Errors can occur in geographical environmental datasets, these can be random, biased or spatially aggregated (Barry and Elith, 2006). All too often, environmental layers are selected without critical analysis (Chapman, 2004).

Errors often stems from inaccuracies in the raw data from which the dataset is created. Environmental dataset may have been derived from paper maps such as soil surveys, in which the location of boundary is uncertain and can often be drawn based on a human derived classification. This is often the case for soil datasets. Most digital climatic datasets are developed by elevation-sensitive spatial interpolations of climate station data (Hutchinson and Bischof, 1983; Hijmans et al., 2004; Hijmans1 et al., 2005).This can introduce spatial uncertainties because of (i) interpolation errors, (ii) lack of sufficient stations data, and (iii) the fact that standard climate stations do not reveal the biologically relevant microclimates (Guisan and Zimmermann, 2000).The dataset will have errors consistent with those of the original point data and interpolation algorithm used (Barry and Elith, 2006).

Positional accuracy of datasets will often depend on the accuracy of the underlying DEM (Digitail Elevation Model). A DEM prepared for South Africa at 10 arc-minutes, had a standard error of between 20 and 150 metres (Hutchinson, 2003a). The resulting climate maps had a standard error for temperature of approximately 0.5 °C and of rainfall grid between 5 and 15%, depending on data density and spatial variability of the actual monthly mean rainfall (Margules and Redhead, 1995; Hutchinson, 1996, 2003a)

The assigning of an incorrect datum or no datum to data can be another source of error. The same numerical coordinate in different Datum's can mean quite a large difference in position on the ground. The difference between the Australian Geodetic Datum (AGD66) and the World Geodetic System (WGS 84) in Australia can mean a shift of around 170m (Chapman, 2004). This can cause error if datasets with different Datum's are over-laid or data is extracted at a particular known coordinates recorded in a different datum. To avoid such error datum's of all datasets should be known and re-projection carried out to ensure all datasets are in the same datum (Chapman, 2004).

The scale at which datasets are created can lead to bias. If the scale at which data are recorded is coarse; units that exist at a finer resolution may be subsumed by more prevalent ones, leading to a bias against rarer classification units. When selecting which dataset set to use, the choice of scale of dataset is also important. Mixing of scales and the use of inappropriate scales can also lead to error. Too fine a scale (i.e. less than the known accuracy of the ecological data) will lead to errors due to mismatching with the ecological data being modelled against it. Too coarse a scale can lead to grids not covering all land surfaces especially off shore Islands (Chapman, 2004). Chapman (2004) believed that the ideal grid size for modelling continental basis is about 1/20th or 3 arc minute (about 5X5km).

When such erroneous datasets are used in modelling, any ecological assumption made about the ecological niche based on the derived relationship between the distribution data of the species and such environmental datasets may not be valid, models may not adequately delineate the appropriate environmental niche (Chapman, 2004; Barry and Elith, 2006)

3.3 Species data

Collections of plant and animal specimens have been in existence for centuries in museums in herbaria (Chapman, 1999). It is estimated that there up to 3 billion records are held in this form around the world (Chapman, 1999), each with an associated "collection event", describing the time and place where the specimen is found (Graham et al., 2004a). These records span a huge temporal and spatial extent. They are unique in that they cannot be recollected or obtained from any other source, and thus they provide a documented historical record of occurrence of species (Chapman, 1999).

Traditionally these collections have been used for taxonomic purposes (Chapman, 1999; Williams et al., 2002). However recently there has been renewed interest in using such records in biogeographic studies such as distribution modelling. This has opened the potential to provide valuable insight into the niche and current, historic and future distribution of many species which were previously considered not to have suitable data for such studies. In Recent studies such data has been applied to conservation planning, reserve selection, climate change studies (Loiselle1 et al., 2003).

Improved computer processing, data digitising techniques, digital storage capacities and internet accessibility has allowed this type of data to become increasingly available (Chapman, 1999; Graham et al., 2004a). Digitization of collections has processed slowly, but is now thought to include 5 – 10% of specimens in collections worldwide (Graham et al., 2004a).

Several initiatives have used innovate information technology to connect multiple data providers and users with a single internet site. Data is retained at the primary institution and distributed through network technology over the internet. Currently 20 – 40% (~ 60 million data records) of existing digitised specimen information is included in distributed networks (Graham et al., 2004a). Examples of such initiatives are listed in Table 4

Table 4 Herbaria data Information networks

Acronym	Name	URL
REMIB	World Information Network on Biodiversity	http://www.conabio.gob.mx/remib_ingles/doctos/remib_ing.html
BRAHMS	Botanical Research And Herbarium Management System	http://herbaria.plants.ox.ac.uk/bol/home/default.aspx?reset=-1
GBIF	Global Biodiversity Information Facility	http://www.secretariat.gbif.net/portal/index.jsp
AVH	Austraila Virtual Herbarium	http://www.anbg.gov.au/avh/

Germplasm records have also recently been used as a source of data to study the ecological requirements and distribution of species in addition to their well established role in conservation and crop improvement (Hijmans et al., 2000).

3.3.1 Data quality in species data

It is the use of specimen data in biogoegraphic studies, such as species modelling that has caused a focus on data quality issues which had previously been ignored (Chapman, 1999). Specimen data can be categorised broadly into three dimensions (i) identity (ii) space and (iii) time (Wieczorek et al., 2004). All three can create issues of data quality. Issues of data quality can be split broadly into error and bias although the two are interlinked. Error refers to a mistake in the data such as misidentification or human error in georeferencing while bias is due to problems intrinsic with the data such as not encompassing the full environmental niche. Error can lead to bias if not identified.

3.3.1.1 Error in species data

Error can occur in taxonomic identification; this can lead to an occurrence being recorded at a location incorrectly due to the mis-identification of an individual. Spatial error within species datasets includes georeferencing error, imprecision of location of a record or error in the original location of the record (Wieczorek et al., 2004). The usual view of errors and uncertainties is that they are bad, but a good understanding of error and error-propagation can lead to active quality control and managed improvement in the overall data quality (Burrough and McDonnell, 1998). Assessment of the accuracy of model input data is essential otherwise the predictions are meaningless. Correcting errors in data and weeding out bad records can be a time consuming and tedious process (Williams et al., 2002), but it cannot be ignored.

Herbarium collection localities have typically been recorded as textual descriptions, geographical coordinates are seldom given on specimen labels (Chapman, 1999, 2004; Wieczorek et al., 2004). Adding these geographic coordinates (georeferencing the data) after the collecting event can produce various kinds of error. Often these locations description are based on names and situations that can change over time. One place name may refer to several different localities and is thus easily misapplied (Chapman, 1999; Wieczorek et al., 2004). Human error may cause misreading of longitude and latitude or the accidental swapping or transposition of characters (Chapman, 1999).

In the relatively few cases in which localities have been assigned geographic coordinates by the collector or source institution there is seldom any record of the procedure used or further relevant information (meta data) regarding the georeferencing provided (such as meta data regarding the area encompassed by the localities). There is no record of any assumptions made and the uncertainties associated with generating the coordinates (Chapman, 2004). Thus, even where geo-referencing coordinates are assigned, they may be of limited utility since there is no knowledge of how they were assigned (Wieczorek et al., 2004).

It is important to have knowledge of both accuracy and precision with which a location record has been geo-referenced. Accuracy refers to whether the location has been correctly identified and precision to the resolution at which it has been geo-referenced. For pre-georeferenced, precision regarding a specimen's spatial localisation is generally both variable and often difficult to verify (Reutter et al., 2003; Chapman, 2004). The coordinates of some localities can be determined with great precision while others only include very broad locality information (Chapman, 1999) and can only be roughly estimated even under rigorous guidelines (Chapman, 2004; Wieczorek et al., 2004). If these differences in precision are not taken into account, uncertainties cannot be incorporated into analysis and decision making. It becomes impossible to determine whether a given record is appropriate for a particular application (Wieczorek et al., 2004). It should be noted that data of low precision is not necessarily of low quality, what is important for the users to be able to determine if the data is fit for purpose for which they wish to apply it (Chapman, 2004).

As with environmental data, the assigning of incorrect or no datum to data can be another source of error (see section 3.2.4. for further details).

Chapman (1999) and Williams (2002) both pointed out that errors in data are common and are to be expected. Chapman noted that species occurance record data received by Environmental Resources Information Network, Australia (ERIN) from a range of biological sources showed considerable error (up to 18% of records in some cases). In a case study on wild potatoes from Bolivia, in which a database consisted of records from 18 expeditions over a period of more than 40 years, more than 50% of the accessions were found to have an error of one kind or another (Hijmans et al., 1999).

The availability of Global Positioning System (GPS) has greatly facilitated taking geographical co-ordinates and it is hoped that this will reduce geographical error for such data. However Hijmans *et al* (1999) found that even the data of the accessions that were collected with a GPS had errors, both in the geographical names and in the co-ordinates, caused by typographical mistakes

3.3.2 Bias in species data

3.3.2.1 Spatial basis

Herbaria data was not collected with the intention for use in modelling. Datasets are often compiled from many different field collections. The details of sampling methods are often unrecoverable. Each collection might have its own particular basis, due to varying sample techniques and the original reason for data collection (Rich, 1997; Williams et al., 2002; Chapman, 2004). The collector of the specimens may have incorporated bias by collecting in places where they expect to find what they are looking for; in locations which are conveniently accessible, or specimens maybe collected opportunistically. Collectors have often been found to sample along roads and rivers, near town or biological stations or area's of high diversity for the species. (Engels et al., 1995; Rich, 1997; Hijmans et al., 1999; Hijmans et al., 2000; Williams et al., 2002; Chapman, 2004)

Unlike data from a well designed survey, which are based on a comprehensive and random design; herbaria data suffers from the common problem of *ad hoc* collections, non-systematic sampling and uneven sampling effort (Chapman, 1999; Stockwell and Peterson, 2001; Chapman, 2004). This can lead to a unknown but probable observation basis (Reutter et al., 2003), meaning a sample population may not provide a representative sampling in environmental or geographical space (Williams et al., 2002; Barry and Elith, 2006), and may violate assumptions of common multivariate statistical methods. Biases in data may cause the modeled relationships to be dominated by the patterns at sampled sites rather than the patterns across the entire study area, and this in turn is likely to lead to marked spatial variation in prediction uncertainty (Barry and Elith, 2006).

3.3.2.2 Temporal bias

Museum and herbarium data generally supply only information on the presence of the species at a particular time and say nothing about absences in any other place or time (Wieczorek et al., 2004). The species may no longer be present at a historical collection site. Presence locations may represent a demographic sink or source for the species (Wieczorek et al., 2004), include vagrants as well as breeding populations; and the pattern of occupancy may change with time (Williams et al., 2002) (see section 2.2.3.1 for further details on how this can effect modelling of the species niche). Data recorded over different times within close proximity can increase the temporal range of records, but surveys recorded over different times may not necessarily take records from the same place (Williams et al., 2002). In choosing and treating existing datasets, temporal bias and the dynamic of the population should be borne in mind (Williams et al., 2002).

Data sets may integrate observations that have been sampled over a very long period of time, thus possibly implying that different equilibria have been reached for each successive phase (Guisan and Theurillat, 2000)

3.3.3 Georeferencing of species data

Researchers interested in spatial analysis using museum and herbarium specimen data, face a daunting legacy of data without coordinates (Wieczorek et al., 2004). As of March 2003, 61.2% of the 3260453 specimens available through Lifemapper (University of Kansas Biodiversity Reserach Centre, 2002) did not have geo-referenced localities. This statistic is typical of natural history collections data that are in digital media today, and indicate the magnitude of the geo-referencing challenge. The georeferencing of data can be a long and difficult process and if not carried out correctly can cause error and bias in the data (see section 3.3.1.1).

A number of publications have been produced which provide reference for standardised guidelines to georeferencing for examples the MaNIS georeferencing guidelines (Wieczorek, 2001), MaPSTeDI guide to georeferencing (University of Colorado, 2003) and HISPID (Herbarium Information Standards and Protocols for Interchange of Data) (Conn, 1996, 2000).

3.3.3.1 Georeferencing information sources

Historical maps, atlases and gazetteers, and even travel books, can be useful sources of localities (Maxted et al., 1995). Gazetteers are lists of geographic names (usually in alphabetical order) and their co-ordinates. World gazetteers include The Times' Atlas of the World (Times Books, 1988) and the Geographic Names Data Base (GNS- National Geospatial-inteligence agency, 2005). The Geographic Names Data Base contains official standard names approved by the United States Board on Geographic Names and maintained by the National Geospatial-Intelligence Agency. Herbaria sometimes develop their own gazetteers, often providing geo-reference information for localities recorded in floras, such as Polhill (1988).

Geo-referencing coordinate sources are accompanied by rules governing the placement of the coordinates within a named place. For example, the US Geographic Names Information Service (USGS, 1981) places the coordinates of towns at the main post office unless the town is a county seat, in which case the coordinates refer to the county courthouse. Similarly, the same source places the coordinates of a river at its mouth. In the absence of one of these specific points of reference, the geographic centre of the named place is usually recorded. Hence the extent of the named place becomes an important consideration. Inconsistencies in assigning coordinates for named places can therefore greatly effect the precision of a geo-reference (Wieczorek et al., 2004).

New tools are being developed to assist institutions in the process of adding geo-code information to database collections (Chapman, 2004). Such tools include UTM converter (Knyazhnitskiy et al., 2000), Egaz (Shattuck 1997), BioGeoMancer (Beaman *et al* 2003) and Localidade (CIRA 2004a).

3.3.3.2 Identifying error and validating the quality of species data

Error detection methods designed to detect incorrect species identification, mistaken geo-references and other data problems have been developed to identify and flag data records that require inspection, assessment and perhaps correction (Wieczorek et al., 2004). Records of this type can often be detected because they represent outliers in geographical or environmental space (Wieczorek et al., 2004).

Checks can be made for logical consistency in relationships with and between records. For example checks can be made to check that a town cited in one field, is located within the correct state, district or country as cited in other fields of the same record (Chapman, 2004) i.e. the coordinates for a locality lie within the correct administrative boundaries (Hijmans et al., 1999). It is also possible to check against external reference, for example, if the passport data contains a value for altitude it can be checked against a DEM (Digital Elevation Map) (Chapman, 2004). Allen et al. (2001) tested the accuracy of the geo-references by comparing with the species known documented ranges and predictive habitat models generated using a database of known habitat associations for each species. When localities did not match either the documented range or predicted habitat, the specimen was tagged for investigation (Allen et al., 2001)

Mapping of records in a GIS based on their geographical coordinates with administrative unit map data, will also allows identification of records that occur within particular geographic regions. If a record has one country named and is located in another it will quickly be identified. Mapping of records can also be used to identify outlier in geographic space (Chapman, 2004). The most obvious of these is terrestrial records out at sea. Other methods that can be used to identify geographical outliers include statistical methods such as reverse jackknifing could be used with latitudinal or altitudinal data (Chapman, 2004). Mapping the collecting events of an expedition in temporal order have also been used to indentify errors; localities that lie outside of the normal patterns in the expedition should be flagged and checked (Wieczorek et al., 2004).

Suspect records can also be identified by searching for outliers in ecological or environmental space. By combining locality data for a given species with the corresponding environmental data at each location, data is mapped onto environmental space and an environmental profile is created. The environmental profile provides information on the statistical distribution of the species occurrence points on each of the environmental variables. This information can be used to reveal ecological or environmental 'outliers' that may have resulted from inaccuracies in the locality description or from the misidentification of the specimen (Williams et al., 2002; Chapman, 2004; Wieczorek et al., 2004). Any record that lies on or beyond the margin of the climate profile are flagged and rechecked (Williams et al., 2002). In the modelling system BIOCLIM possible outliers were removed by excluding records that fall outside the 90 percentile range of the environmental profile (Busby, 1991).

Cumulative frequency curves derived from environmental profile data (Lindenmayer et al., 1991) can be used to identify outliers on single variables (Busby, 1991; Williams et al., 2002). Using Principle Component Analysis (Flora map) (Jones and Gladkov 2001), Cluster analysis (Jones and Gladkov 2001) and reverse jack-knifing, it is possible to detect outliers in multivariate environmental space (Chapman, 1999). Such methods can also be used to detect bias within a dataset giving an indication of which regions have been poorly sampled (Wieczorek et al., 2004).

Hijmans et al. (1999) noted that caution should be taken when applying such procedures. To allow future interpretation of the data, one should avoid downgrading the database by creating artificially reinforced spatial relationships, by removal of outliers that are not erroneous. The exceptions to the general spatial patterns should not be changed/removed just because they are exceptions, but only when they are clearly wrong (Hijmans et al., 1999). Groups of outliers in environmental space may indicate ecotypes within the population.

3.4 Methodology

3.4.1 Environmental data

This section describes the collection, creation and processing of geographic environmental data in order to create an ecologically meaningful dataset suitable for modelling an underutilised tree species such as tamarind. Climate, soil and moisture availability raster data layers are sourced or created. A water balance model is created in order to create the soil moisture availability dataset.
All dataset were projected in geodetic coordinate system ('GEOGRAPHIC' or 'LATLONG' , datum WGS84)

3.4.1.1 Climate datasets

3.4.1.1.1 Temperature and Rainfall

Global (except Antarctica) temperature and rainfall raster dataset were acquired from the WorldClim data at a spatial resolution of 10 minutes (18.6 x 18.6 = 344 km^2 at the equator) (Hijmans et al., 2004). This dataset was compiled from monthly averages of climate as measured at weather stations from a large number of global, regional, national, and local sources, mostly for the 1950–2000 periods (Hijmans et al., 2004).

The bioclimatic variables represent annual trends (e.g., mean annual temperature, annual precipitation), seasonality (e.g., annual range in temperature and precipitation) and extreme or limiting environmental factors (e.g., temperature of the coldest and warmest month, and precipitation of the wet and dry quarters (a quarter is a period of three months) (Hijmans et al., 2004). These variables are considered to have ecological significance for the distribution of plant species and summarise annual and seasonal mean conditions, extreme values and intra-year seasonality (Busby, 1991).

The following variable were used

BIO1 = Annual Mean Temperature
BIO2 = Mean Diurnal Range (Mean of monthly (max temp - min temp))
BIO3 = Isothermality (P2/P7) (* 100)
BIO4 = Temperature Seasonality (standard deviation *100)
BIO5 = Max Temperature of Warmest Month
BIO6 = Min Temperature of Coldest Month
BIO7 = Temperature Annual Range (P5-P6)
BIO8 = Mean Temperature of Wettest Quarter
BIO9 = Mean Temperature of Driest Quarter
BIO10 = Mean Temperature of Warmest Quarter
BIO11 = Mean Temperature of Coldest Quarter
BIO12 = Annual Precipitation
BIO13 = Precipitation of Wettest Month

BIO14 = Precipitation of Driest Month
BIO15 = Precipitation Seasonality (Coefficient of Variation)
BIO16 = Precipitation of Wettest Quarter
BIO17 = Precipitation of Driest Quarter
BIO18 = Precipitation of Warmest Quarter
BIO19 = Precipitation of Coldest Quarter

3.4.1.1.2 Sunshine and Relative Humidity

Ten minute resolution Relative humidity (percent) and Sunshine (percent of maximum possible of day length) datasets were acquired from the Climate Research Unit (CRU Data) Distribution, University of East Anglia UK (New et al., 2000).

A C$^{\#}$ Program ProcessTemp was written in Microsoft Studio. Net (Microsoft, 2003) to produce the following dataset using CRU (New et al., 2000) and Worlclim (Hijmans1 et al., 2005) datasets;

Relative humidity coldest quarter
Relative humidity hottest quarter
Relative humidity driest quarter
Relative humidity wettest quarter

Sunshine coldest quarter
Sunshine hottest quarter
Sunshine driest quarter
Sunshine wettest quarter

A description of how the ProcessTemp program works can be found in the Appendix.

3.4.1.2 Soil data

A soil dataset with continuous values (as opposed to values being grouped into classes, see section 3.2.2) was produced by combining dominant soil type data and topsoil texture data from the TERRSTAT database (FAO, 2002) with ISRIC WISE Derived Soil Parameter Data Set (ver2.2)(Batjes, 2002).The IRSIC data is aggregated by soil unit, topsoil textural class and depth zone (i.e. topsoil (0-30cm) and subsoil (30 – 100cm)). The topsoil textural class is considered as a differentiating criterion, or flag, for the properties of the underlying profile (Batjes, 2002) in accordance with FAO conventions (FAO, 1995) and data requirements of the FAO-IIASA (see Fischer et al. (2000)). The TERRSTAT datasets was developed using The FAO/Unesco legend for the Soil Map of the World (FAO, 1971-1981) as an international correlation system to indicate the dominant soil unit in each cell.

The ISRIC ver. 2.2 derived soil parameters should be seen as the best possible estimates based on the present selection of soil profiles and adopted data cluster procedure. The derived data set is considered appropriate for use in studies at a regional or global scale (Batjes, 2002). Examples of crop simulation and agro-ecological zoning applications that used soil parameters derived from WISE, version 1.0 include, Fischer et al (2000; 2001) and Knox et al (2000).

The TERRSTAT dominant soil type dataset was converted from a grid to a shape file using Arc View. Each point was then assigned an individual ID number creating a table DOMALL.shp. The Texture Topsoil dataset from the TERRSTAT was reclassified so that course soil was reclassified as 1, medium soils as 2 and fine soils as 3, organic soils were reclassified as 5. The modeified Texture Topsoil grid file was then converted to a shape file and joined to DOMALL based on geographical location using Arc View (ESRI, 2000) “Geoprocessing extention” “Nearest” function. The outputted table was named DOMTEXTOP (Table 5)

Table 5 Section of DOMTEXTOP table

Id	Long	Lat	Dom_code	Toptex
74	-30.0417	83.5417	117	2
75	-29.9583	83.5417	117	2
76	-29.875	83.5417	117	2
77	-29.7917	83.5417	117	2
78	-29.7083	83.5417	117	2
79	-29.625	83.5417	117	2
80	-29.5417	83.5417	117	2
82	-37.875	83.4583	132	5
90	-36.625	83.4583	132	5
91	-36.5417	83.4583	132	5
……..	……….	………..	………..	……….

Dots represent continuing values, Id-Location ID, Long- Longditude, Lat-Latitude, Dom_code- Dominat Soil Code, Toptex- Topsoil Texture Code.

The SUMTAB74 table was acquired from the ISRIC WISE Derived Soil Parameter Data Set (ver2.2). This contains a letter coding derived for dominant soil type based on those from the FAO-UNESCO 1974 Legend, a texture code (1, 2, 3 and #) and the corresponding soil characteristic median values that correspond to each combination of dominant soil type and texture. A further column was added to the table with a code number for each dominant soil type which corresponded to the number assigned to the same dominant soil type in the TERRSTAT dominant soil dataset.

The modified SUMTAB74 and DOMALLTEXTOP datasets were then combined based on the corresponding Dominat soil codes and texture top soil code using specifically written program FAOSOIL. This created the CONSOILDATA (Table 6) dataset listing latitude and longitude for each grid location and its corresponding values for each of the soil characteristics included in the ISRIC WISE dataset. (Table 6 shows a section of CONSOILDATA showing the soil characteristics organic carbon topsoil, organic carbon subsoil, total nitrogen topsoil, total nitrogen subsoil. The full table contains all soil dataset).

Table 6 Section of CONSOILDATA table

Id	Long	Lat	Dom_code	Toptex	ORGC_TM	ORGC_BM	TOTN_TM	TOTN_BM
302426	-34.04	83.54	117	2	1.12	0.48	0.11	0.06
302427	-33.95	83.54	117	2	1.12	0.48	0.11	0.06
302428	-33.87	83.54	117	2	1.12	0.48	0.11	0.06
302429	-33.79	83.54	117	2	1.12	0.48	0.11	0.06
302430	-33.7	83.54	117	2	1.12	0.48	0.11	0.06
302438	-33.04	83.54	117	2	1.12	0.48	0.11	0.06
302439	-32.95	83.54	117	2	1.12	0.48	0.11	0.06
302440	-32.87	83.54	117	2	1.12	0.48	0.11	0.06
302441	-32.79	83.54	117	2	1.12	0.48	0.11	0.06
302442	-32.7	83.54	117	2	1.12	0.48	0.11	0.06
302443	-32.62	83.54	117	2	1.12	0.48	0.11	0.06
....	....	....	....	....	....	....	....	....

Dots represent continuing values, Id-Location ID, Long- Longditude, Lat-Latitude, Dom_code- Dominat Soil Code, Toptex- Topsoil Texture Code, ORGC_TM – Organic carbon topsoil, ORGC_BM - Organic carbon subsoil, TOTN_TM Total Nitrogen, TOTN_BM - Total Nitrogen Subsoil

The CONSOILDATA dataset was then converted to an ESRI shape file using Arc View and from this an ESRI grid file produced for each soil characteristic using the point shape file to grid converter in Arc View (ESRI, 2000). As this dataset only considers the dominant soil type within each grid cell it must be taken into consideration that it only captures characteristics that occupy 30-60% of the total grid area (Nachtergaele pers. comm.).

3.4.1.3 Water balance data

The computer program "WATBAL" was developed to calculate a monthly water balance based on the method used in Global Agro-ecological zones project(FAO/IIASA, 2000). Actual evapotranspiration (Eta) was calculated according to the FAO Irrigation and Drainage paper 33 (FAO, 1979), A monthly water storage/water-balance *(W)*, was calculated as follows

$$W_{j+1} = \min\left(W_j + P_j - ETa_j,\ Sa\right)$$

$$ETa_j = \begin{cases} ETo_j & \text{if } (W_j + P_j) \cdot d \geq Sa \cdot d \cdot (1-p) \\ \rho ETo_j & \text{else} \end{cases}$$

where,

$$\rho = \frac{ETa_j}{ETo_j} = \frac{W_j + P_j}{Sa \cdot (1-p)}$$

W_j	waterbalance
j	number of month in year
P	rainfall(mm)
d	rooting depth (m)
Sa	available soil moisture holding capacity (mm/m)
p	soil water depletion fraction below which $ETa < ETo$
ρ	actual evapotranspiration proportionality factor

A Monthly Penman-Montieth reference evapotranspiration dataset was acquired from the IMWI water Atlas website (New et al., 2000; IMWI, 2005). The dataset was created based on the FAO Penman-Monteith equation (FAO, 1998).

A "Maximum available soil moisture" (FAO, 2002) raster dataset was downloaded from the FAO Geonetwork and converted 10*10 arc minute. Monthly rainfall data data acquired from the WorldClim dataset (see section 3.4.1.1.1).

A monthly "Water Depletion Fraction" (p) was calculated based on FAO Irrigation and Drainage Paper 33 (FAO, 1979), The monthly reference evapotranspiration (Eto) dataset was reclassified based on figures from the "Soil Water Depletion Fraction (p) for Crop Groups and Maximum Evaoptranspiration (ETm)" ((FAO, 1979) (Table 8). This dataset was multiplied by 1000 to create integer values, resulting in a "p1000" dataset for each month of the year. All fruit tree species were assumed to be within group 3 of the "Crop Groups according to Soil Depletion" table (Table 7). Group 3 was selected due to tamarind being most commonly found growing in the same agro-ecological conditions under non-irrigated conditions as other crops within this group (see Table 19 in (FAO, 1979)).

Table 7 Crop Groups according to Soil Water Depletion (FAO, 1979)

Group	Crops
1	onion, pepper, potato
2	banana, cabbage, grape, pea, tomato
3	alfalfa, bean, citrus, groundnut, pineapple, sunflower, watermelon, wheat
4	cotton, maize, olive, safflower, sorghum, soybean, sugarbeet, sugarcane, tobacco

Table 8 Soil Water Depletion Fraction (p) for Crop Groups and Maximum Evapotranspiration (Etm) (FAO, 1979)

Crop Group	Etm mm/day								
	2	3	4	5	6	7	8	9	10
1	0.50	0.425	0.35	0.30	0.25	0.225	0.20	0.20	0.175
2	0.675	0.575	0.475	0.40	0.35	0.325	0.275	0.25	0.225
3	0.80	0.70	0.60	0.50	0.45	0.425	0.375	0.35	0.30
4	0.875	0.80	0.70	0.60	0.55	0.50	0.45	0.425	0.40

The Maximum available soil moisture dataset was combined with the 12 monthly p1000 datasets, the 12 monthly rainfall datasets and the 12 monthly evapotranspiration datasets to form maxmoist_pre_p1000_eto_1to12.txt which is entered into the "WATBAL" computer program. The WATBAL program was written with Visual Studio .Net (Microsoft, 2003) in $C^{\#}$. For a description of how the WATBAL program works please see the Appendix

The program outputs a monthly value for January to December for;

MAI Moisture availability
Index(Eta/Eto),
Deficit (Eto - Eta)
Runoff (W – Sa)

The average monthly MAI dataset was used to produce the following datasets using Arc View Grid calculator to produce the following datasets;

MAI Coefficient of Variance
MAI Maximum
MAI Minimum
MAI Mean

The average monthly MAI were also entered into the ProcessTemp program (see above) with the monthly temperature dataset to produce the following.

MAI Hottest quarter
MAI Coldest quarter
MAI Maximum MAI quarter
MAI Minimum MAI quarter

3.4.2 Species data

3.4.2.1 Collection and Recording of Species Distribution Information

A database of tamarind occurrence records was collected form a number of sources;

Herbarium Records: The majority of the data was collected from digitized herbarium collections from across the world, made available through the internet. Table 9 lists the digital herbarium sources from were tamarind occurrence records were found. Location descriptions were also aquired from herbarium passport records during a visit to the collection at the Royal Botanical Gardens, Kew.

Germplasm Passport data: A Masters thesis aquired from the University of Agricultural sciences Dharwad (Chanda, 2001), entilied "Evaluation of Tamarind (*Tamarindus indica* L.) Genotypes for Productivity and Grafting". This listed locations of mother tree specimens of the genotypes studied.

Field location data: The species distribution locations recorded using a GPS (Global positioning system) during a field survey in India conducted from June to September 2003 (see Figure 21)

Table 9 List of the digital herbarium sources for tamarind occurrence records used in this study

The Herbarium XAL of the Institute of Ecology, A.C., Mexico
Jararquia Taxonomica/The Herbarium of the National Institute of Biodiversity of Costa Rica
New York Botanical Garden
W3 Missouri botanical garden
Herbarium of the Yucatan Scientific Research Centre, Mexico
Native Trees and Shrubs for Restoration and Reforestation of Mexico
National Vegetable Germplasm Bank, Mexico
The Herbarium of the University of Texas - Austin, USA
Trees from the Yucatan Peninsula, Flora from Tehuantepec, Oaxaca and Asteraceae Family in Mexico
Herbarium of the Yucatan Scientific Research Centre, Mexico
The useful flora from two native communities of the Valley of Tehuacan-Cuicatlan
Australia Virtual herbarium
University of Florida Herbarium
Flora Zambesiaca
SysTax Biological Research Collections in Germany
Cameroon National Herbarium
Herbarium of the University of Aarhua
Oxford University Herbaria
Oxford Caribbean Specimens
South East Asia Botanical Collection Infromation Network

All location information was recorded in a “Location” table which was linked to “Collector Reference” and “Source” table, this allowed for cases where more than one species were found at a location and where a location record came from more than one source and/or was recorded by more than one collector.

3.4.2.2 Geo-referencing

All location descriptions were assigned an individual “location ID” and were geo-referenced. Latitude, longitude coordinates were assigned to each location using maps and digital gazetteers. Points were georeferenced based on the geodetic coordinate system ('GEOGRAPHIC LATLONG' , datum WGS84).
The Flora Zambesiaca – index of collecting localities (Pope and Pope, 1998), the Flora of Tropical East Africa - Index of collecting localities (Polhill, 1988) the Geographic Names Data Base (GNS- National Geospatial-inteligence agency, 2005) gazetters were used to identify georeferencing co-oridinates for the location descriptions.

Once Georeferencing was complete duplicate locations were identified. Where duplication in the location occurred, one record from the location table was removed. However the remaining location record was linked to information for both instances of the location in regard to the collector and source information in the “source” and “collector” tables.

3.4.2.3 Identifying and removing error from species occurrence records

In order to identify errorenous location records a number of methods were used in order to identify outliers in geographic and environmental space. Outliers were flagged and the georeference rechecked. Any records identified as erroroneous where possible were re-georeferenced in order to correct. If correction was not possible the records was removed from the dataset.

3.4.2.3.1 Geographical validation

All distribution points were mapped in ARC View (ESRI, 2000)and over laid onto the ESRI world administrative units dataset (ESRI, 2000) and GNS (GNS- National Geospatial-inteligence agency, 2005) country administrative unit datasets to identify geographical outliers. Initial identification was based on visual analysis of the map, points which appeared to be outside the geographical range of the species and those that appeared to fall in the sea rather than on land were flagged and checked.

Arc View (ESRI, 2000)‘Geo Processing wizard’ ‘Assign data by location’ function was used to join the species point distribution dataset to the GNS gazetteer dataset (GNS- National Geospatial-inteligence agency, 2005), the GNS Administrative units database (GNS- National Geospatial-inteligence agency, 2005) (from here called the DIVA administrative units as was acquired from the DIVA website (Hijmans et al., 2007)) and ESRI ‘administrative units’ (ESRI, 2000) dataset based on their spatial relationships. This combined dataset was then analysed to ensure that the location description, (i.e. place name, district, state) matched with the description assigned, from on the DIVA and ESRI datasets based on geographical location.

Distribution points for which no description was assigned by the ESRI or DIVA administrative databases or where the location description did not match that assigned by the DIVA or ESRI datasets were flagged as outliers and investigated. For the GNS gazetteer database the nearest place name and "distance to" was assigned to the coordinate point descriptions. Where these did not match the occurrence record was flagged and geo-reference checked.

In all cases where geographical errors were identified and correct coordinates could not be confidentially identified for a location description; the location description was discarded from the dataset and not used in any further analysis.

3.4.2.3.2 Environmental validation

Environmental validation of data quality involved the analysis of the records in environmental space and hence creation of an environment profile. The environment profile is a summary of the statistical distribution for tamarind occurrences on each of the environmental variables. In order to create the environment profile, the point distribution dataset was overlaid onto the environmental datasets, the values from the underlying environmental dataset grids were written to the corresponding point in the location dataset table, creating a dataset of the environmental conditions at each of the point distribution locations. This was carried out using the program 'grid sampler' (Zerger, 2004).

A statistical reverse jackknifing procedure was carried out(Barnet and Lewis 1978) and a cumulative frequency graph was produced for each variable in order to identify outliers in univarite environmental space. The reverse kackknifing procedure works by emphasizing the effect of marginal records in climate space, leading to critical values being obtained for each environmental variable (Chapman, 1999). The distance between each record and its neighbour is calculated. This is then manipulated by the distance between the mean and the outer record (i.e. for records less than the mean for all records, the lower of the two records is used; for values greater than the mean the higher of the two records is used). The result is divided by the standard deviation to give the critical value (C). If C is greater than a threshold value for that number of records then the record is regarded as an outlier and flagged as suspect (Chapman, 1999). This process was carried out in Microsoft excel (Microsoft, 2003).

Environmental profile data for each variable was used to create a cumulative frequency graph. Occurrence record points which created an elongation of the sigma curve were flagged and investigated.

Principle component analysis was carried out with only climate variables (temperature, rainfall, relative humidity, sunshine and moisture availability index data) in order to map the species records in multivariate environmental space. Outlier's were identified based visual analysis of component score plots. Points which appear to lie away from the main cluster of points were flagged and investigated

3.4.2.3.3 Accounting for bias in species occurrence dataset

Once all errorenous records had been removed, in order to reduce the effect of temporal bias, only occurrence records collected after 1950 were selected for use in modelling. It was was chosen due to tamarind being a long lived species. It is expected that most specimens recorded in 1950 would still be living, and this date allowed over 50% of the records to be kept. The Worldclim data was compiled from monthly averages mostly for the 1950–2000 periods (Hijmans et al., 2004), this matched realativily closely the period for which the species data was recorded.

The 1950 species dataset was converted to 10 X 10 minute (approx 18 X 18km) resolution, the same resolution as the environment datasets. This acted to reduce any sampling bias that may occur in the dataset. It also accounted for issues of accuracy of the georeference, allowing margin of error within the extent of the point.

The dataset was further reduced as not all environmental variables had the same extent. Only those points for which a value could be acquired for all dataset were used in modelling

3.4.2.3.4 Data Exploration of species occurrence dataset

The tamarind 10 minute resolution species distribution dataset was used to extract environment variable values for each presence record using the Grid Sampler software (Zerger, 2004). A statistical summary for each environmental variable was calculated from the data extracted using SPSS. Each presence record was categorised based on its Biogeographical region, using Udvardy's (1975) classification (Figure 1) . A shape file derived by the WCMC (2001) was used to classify the points in Arc view (ESRI, 2000). A Principle Component Analysis (PCA) (McGarigal et al., 2000) was run for all 359 data points using all environmental variables.

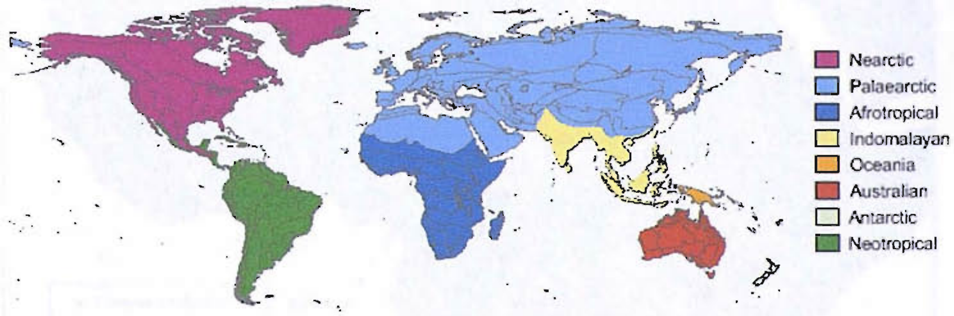


Figure 1 Udvardy's Biogeographical Realms

3.5 Results

3.5.1 Environmental data

The environmental dataset as described above were produced for use in modelling the potential production areas of tamarind.

3.5.2 Species data

3.5.2.1 Geographical validation of species occurrence data

Below are a number of examples of points which were flagged using the geographical validation method.

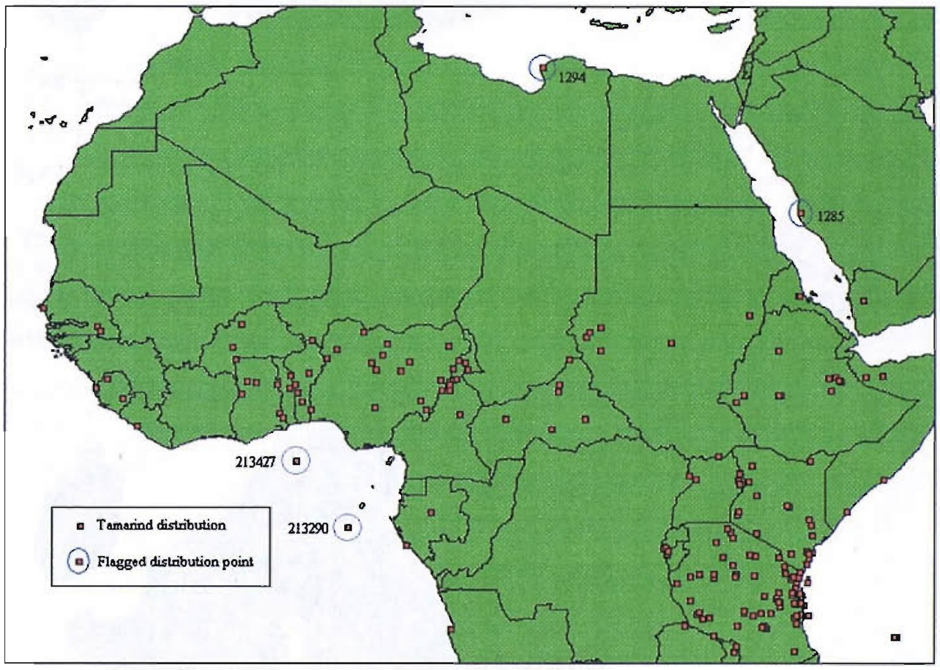


Figure 2 Flagged point occurrence records 1294, 1285, 213290, 21347 - geographical validation.

Figure 2 shows flagged points 1294 (El Feihat, Benghazi, Darnah, Lybia) and 1285 (Jedda, Makkah, Saudi Arabia) flagged based on visual analysis of distribution points. Both points appear to be at higher latitudes and therefore outside the geographical range indicated by the other distribution points for the species. Point 213290 (Pagalu Isle, Annobon Equatorial Guinea) flagged as identified as being in the sea. This was due to poor resolution of ESRI administration dataset, which is unable to show all islands in the island chain. On checking, the record was found to be correctly geo-referenced. Point 213427 Lougba Bante Zou Benin geo-referenced as 3.3833; 1.7833. It was found to be erroneous and was rectified to 8.3833; 1.7833. This shows the issue caused by misreading or mistyping of digits within coordinates.

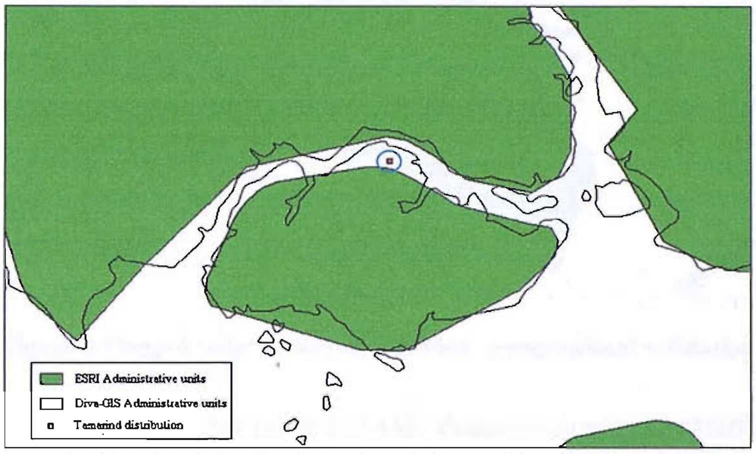


Figure 3 Occurrence point 213444 – geographical validation

Figure 3 shows point 213444 Sembawang, Singapore (1.450830; 103.82880); this indicates the difficulty in using shape file to identify geographic outliers due to differing resolution. The point would be flagged if only using ESRI administrative shape file to indetify outliers.

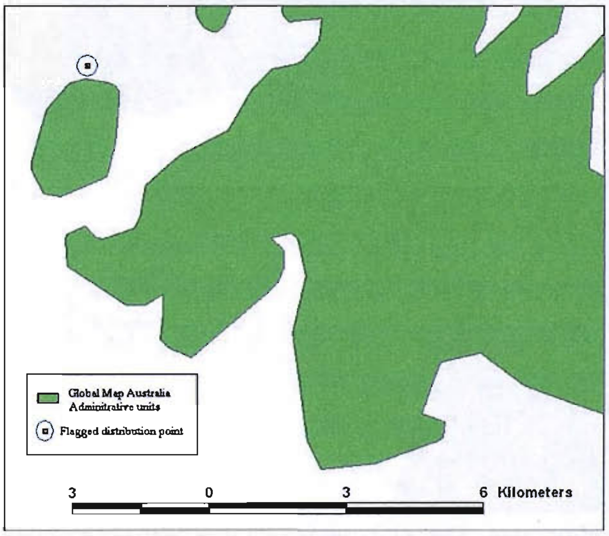


Figure 4 Flagged point occurrence 2080 - geographical validation

Figure 4 Shows point 2080 Queensland Bolton Hill flagged as appears to be in the sea

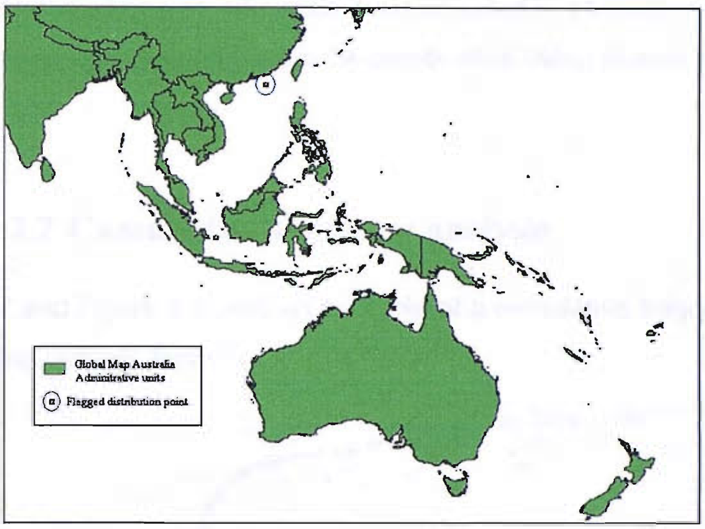


Figure 5 Flagged point occurrence 213448 - geographical validation

Figure 5 shows the point 213448, Western Australia, Mardie Station geo-referenced as 21.1833333; 115.98333. This record was flagged and corrected to -21.1833333; 115.983333 indicating the effect of an incorrect symbol.

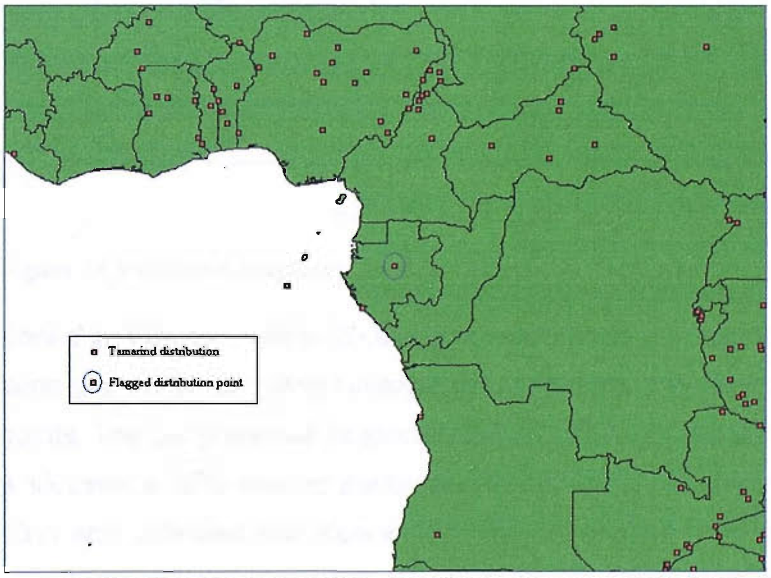


Figure 6 Flagged point occurrence 213448 - geographical validation

Figure 6 Point 213448 had the location description Burkino Faso 1 km E of Tenkodogo. It was geo-referenced as -0.3333; 11.7666 and flagged by geographic validation as these coordinates placed the point at Lope, Ogooue-Ivindo, Gabon. The geo-reference was corrected to 11.7666; -0.3333. This shows the effect of Latitude and Longitude coordinates being placed in the incorrect order.

3.5.2.2 Cumulative frequency analysis

Figure 7 and Figure 8 shows an example of a cumulative frequency graph used in identifying outliers in environmental space.

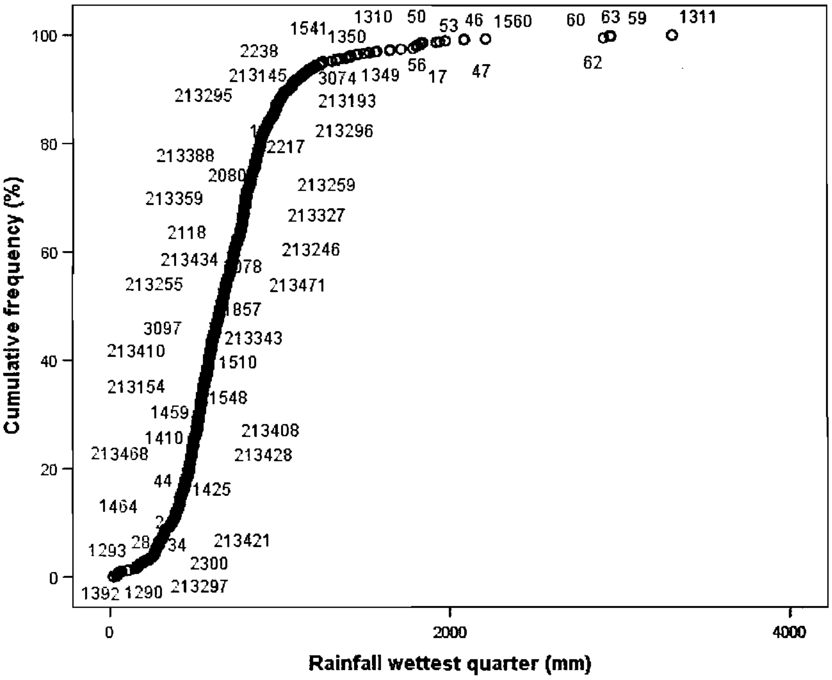


Figure 7 Cumulative frequency graph for Rainfall wettest quarter

Based on Figure 7, points 59,60,62,63 and 1311 are flagged as outliers due to the large gap in rainfall wettest quarter value between these and and the rest of the points. The geographical location for 59, 60, 62 and 63 were assigned during field work using a GPS making the chance these are erroneous less likely. Record 1311 was collected from Kew in 1959; any further details such as collector name of collector reference number are unknown. All points were rechecked and found to be correctly geo-referenced.

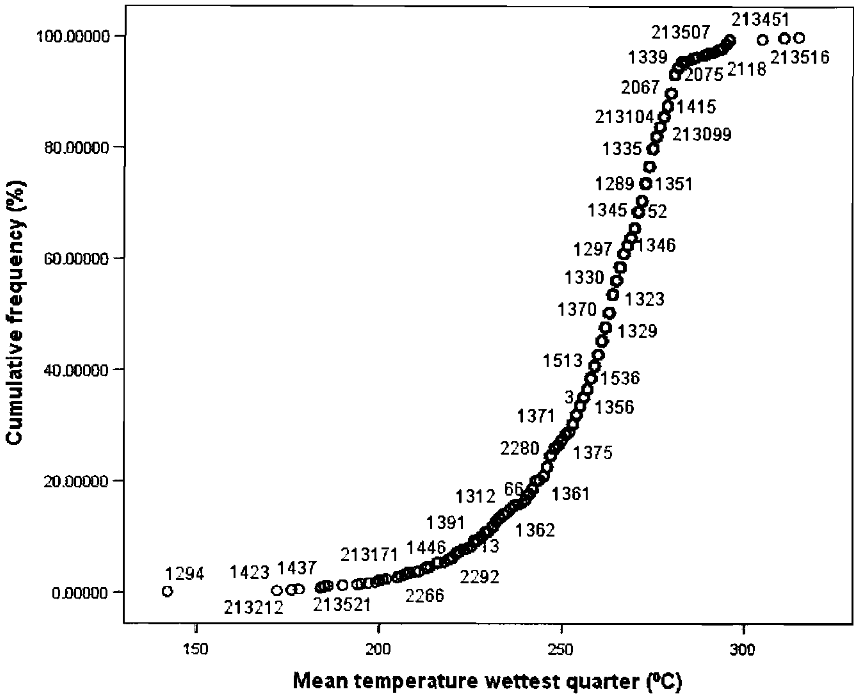


Figure 8 Cumulative frequency graph for Mean Temperature wettest quarter

Based on Figure 8 the record 1294 was flagged. This record 1294 was collected in 1959 and was sourced from RBG Kew. This record was also flagged by the Sunshine hottest quarter cumulative frequency graph and the MAI cumulative frequencies graph (Figure 10). It was also flagged by geographical validation methods.

Table 10 Number of occurrence records flagged by each variable type in the cumulative frequency analysis

Variable type	Number of records flagged
Temperature	17
Rainfall	16
Sunshine	11
Relative humidity	1
Moisture Availability Index	12
Soil	57

Table 10 lists the number of occurance records flagged by more by each variable type in the cumulative frequency analysis. Most variables were flagged by Soil, then Moisture Availability index (MAI), rainfall and temperature. Record was flagged for temperature did not tend to be flagged for rainfall and visa versa.

Table 11 Record flagged by two or more data types in the cumulative frequency analysis validation method

ID	Temperature data	Rainfall data	Sunshine data	Relative humidity	Moisture available Index data	Soil data	Total
1302	1	0	0	0	0	1	2
3094	1	0	0	0	0	1	2
1294	1	0	1	0	1	0	3
1417	1	0	0	0	2	0	3
1541	0	1	2	0	0	0	3
1285	1	0	0	0	3	0	4
1287	1	0	0	0	3	0	4
1391	0	0	1	0	3	0	4
1284	0	0	0	0	3	2	5
1288	1	0	0	0	4	0	5
1290	0	0	0	0	3	2	5
1327	0	5	1	0	0	0	6

Table 11 shows those occurrence points which have been flagged more than once in the cumulative frequency analysis. These points were remapped (Figure 9) to identify if there was a link between geographic position and being an outlier in environmental space.

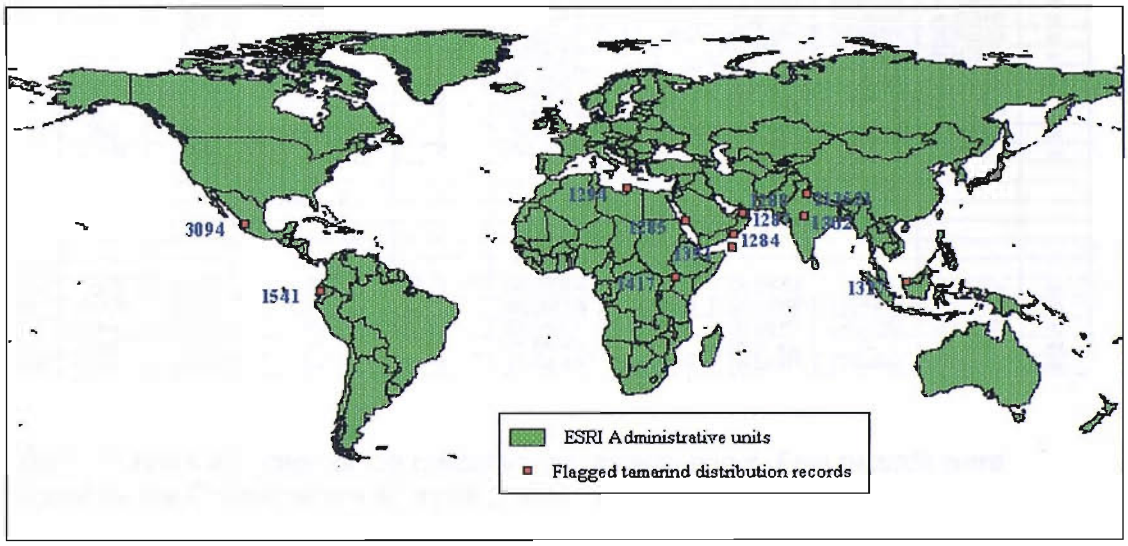


Figure 9 Geographical location of tamarind distribution flagged by 2 or more variables in the cumulative frequency analysis validation method

Figure 9 shows the geographical location of the records in Table 11, which were flagged by two or more data types. Occurance record 1294 was flagged by three different variable types, temperature sunshine and MAI. It was also flagged based on the visual analysis of mapped records. Records 1285, 1287, 1288 were flagged by both MAI and temperature.

3.5.2.3 Critical Value analysis

Table 12 Records flagged by Critical value analysis

ID	Variable
1294	Mean Temperature wettest quarter
213516	Relative humidity wettest quarter
1284	Calcium Carbonate topsoil/Calcium Carbonate subsoil
1289	Calcium Carbonate topsoil/Calcium Carbonate subsoil
1290	Calcium Carbonate Content topsoil/Calcium Carbonate Content subsoil
1339	Effective CEC subsoil
1531	Total nitrogen subsoil

Table 13 Critical value spreadsheet for mean temperature wettest quarter

ID	Xcoord	Ycoord	Mean_temp_wet_qrt	Neighbour distance	MEAN	Y	SD OF Y's	Critical value	Treshold value	FLAG
1294	20.1	32.08333	142	30	258.8385876	3505.157629	129.225452	27.1243596	26.9522429	yes
1423	35.9666	-1.8	172	4	258.8385876	347.3543506	129.225452	2.68797164	26.9522429	no
213212	-97.6165	18.4165	176	2	258.8385876	165.6771753	129.225452	1.28207851	26.9522429	no
1376	30.9833	2.3	178	6	258.8385876	485.0315259	129.225452	3.75337457	26.9522429	no
1398	37.75	8.2167	184	0	258.8385876	0	129.225452	0	26.9522429	no
1525	29.2166	-30.75	184	1	258.8385876	74.83858764	129.225452	0.57913195	26.9522429	no
1402	42.1166	9.3166	185	0	258.8385876	0	129.225452	0	26.9522429	no
1437	34.7	1.75	185	1	258.8385876	73.83858764	129.225452	0.57139353	26.9522429	no
1403	37.5833	11.4833	186	4	258.8385876	291.3543506	129.225452	2.25462048	26.9522429	no
213521	78.33333	30.36666	190	4	258.8385876	275.3543506	129.225452	2.13080586	26.9522429	no
1473	34.5	-9.66666	194	1	258.8385876	64.83858764	129.225452	0.50174781	26.9522429	no
1555	-100.15	18.9	195	2	258.8385876	127.6771753	129.225452	0.98801879	26.9522429	no
1475	35.7	-7.7667	197	2	258.8385876	123.6771753	129.225452	0.95706514	26.9522429	no
3069	-97.085	18.85	199	1	258.8385876	59.83858764	129.225452	0.46305574	26.9522429	no
1413	39.1341	15.4725	200	0	258.8385876	0	129.225452	0	26.9522429	no
1432	35.3333	1.9167	200	0	258.8385876	0	129.225452	0	26.9522429	no
2308	39.0865	15.4428	200	1	258.8385876	58.83858764	129.225452	0.45531733	26.9522429	no
1301	79.5	29.8333	201	1	258.8385876	57.83858764	129.225452	0.44757892	26.9522429	no
213135	-99.9	17.8	202	3	258.8385876	170.5157629	129.225452	1.3195215	26.9522429	no
2266	47.5166	-18.9166	205	0	258.8385876	0	129.225452	0	26.9522429	no
.										
.										
.										
.										
.										
.										
213496	136.85	-16.0167	296	0	258.8385876	0	129.225452	0	26.9522429	no
213509	135.4167	-15.5833	296	9	258.8385876	415.4527112	129.225452	3.21494493	26.9522429	no
213516	134.2	-19.65	305	6	258.8385876	312.9684741	129.225452	2.42187951	26.9522429	no
2074	115.9833	-21.1833	311	0	258.8385876	0	129.225452	0	26.9522429	no
213448	115.9833	-21.1833	311	4	258.8385876	224.6456494	129.225452	1.73840096	26.9522429	no

Table 13 shows a section of the critical values spread sheet. Few records were flagged by the Critical values analysis (Table 12)

3.5.2.4 Using geographical and environmental space to identify and explain outliers

The PCA analysis allows identification of outliers in mutlivarite environmental space. From investigation of the cumulative frequency plots combined with ploting the of occurnace points in environmental (the PCA component plots) and geographical space; it may be possible to identify why occurance records are appearing as outliers and make more informed judgements as to whether they are erroneous.

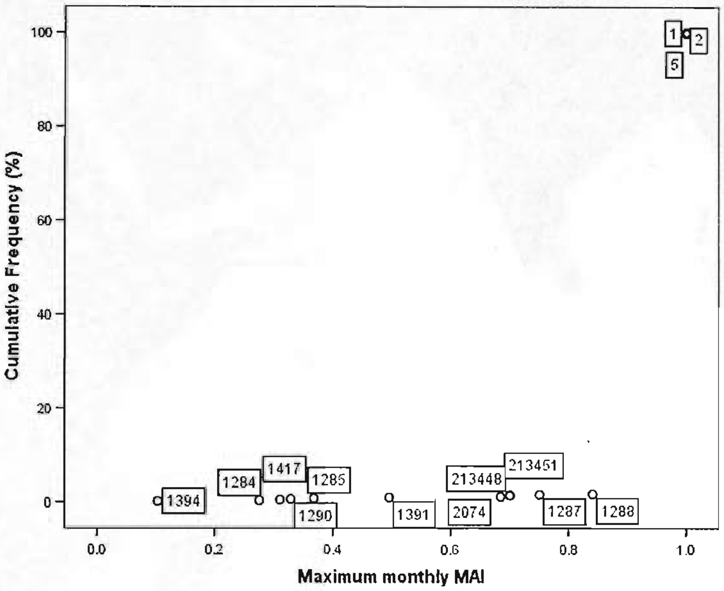


Figure 10 Cumulative frequency graph for Maximum Monthly MAI (points in top right corner represent 98.83533% of points which have an MAI of 1)

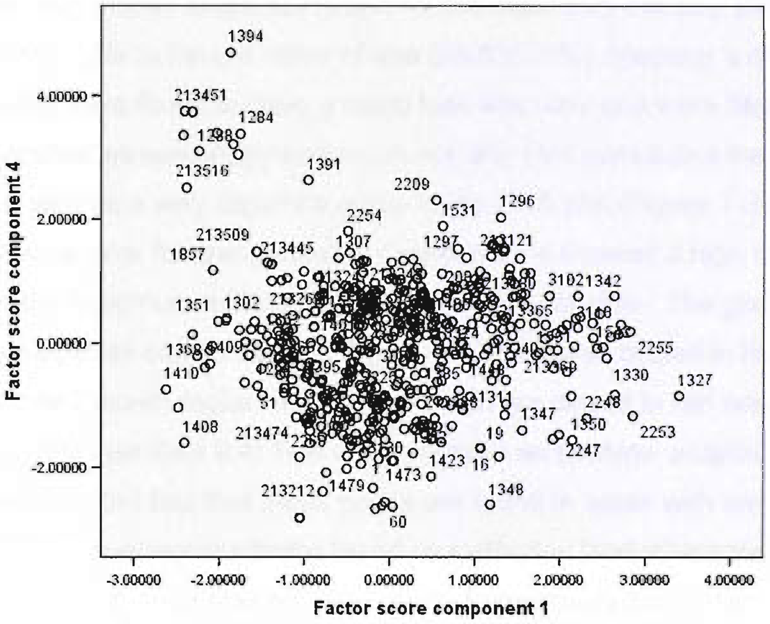


Figure 11 PCA component plot for climate only variables (component 1 and component 4)

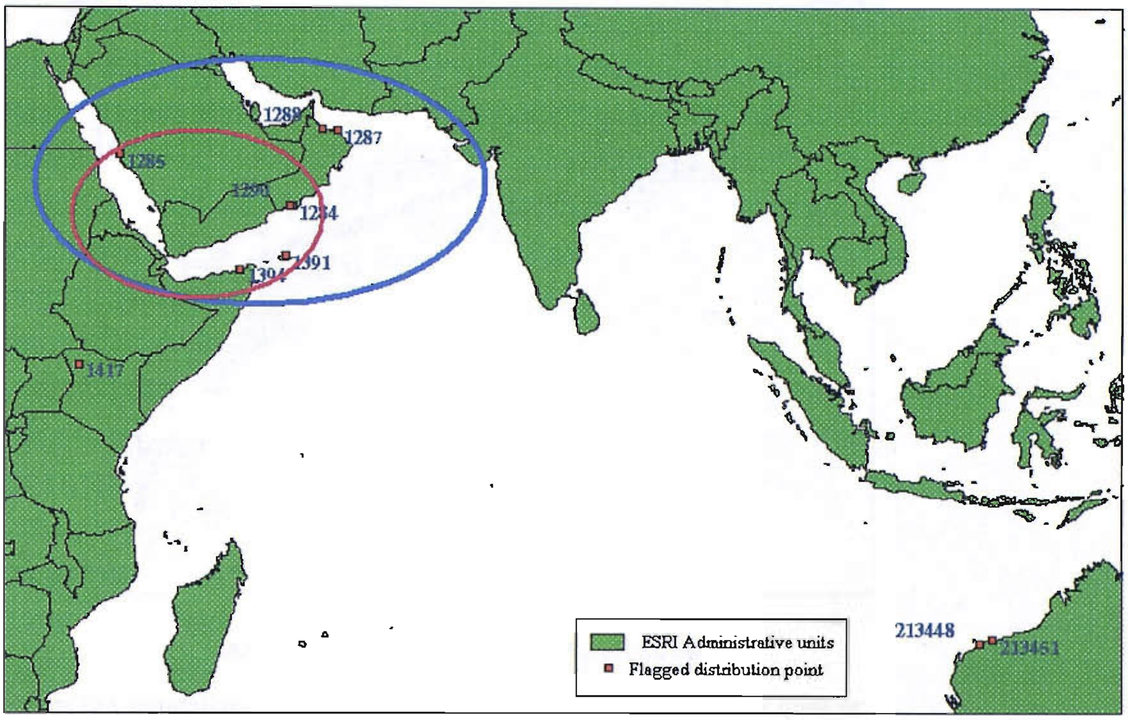


Figure 12 Geographical location of flagged points (cumulative frequency maximum monthly MAI)

The cumulative frequency graph for the maximum monthly MAI (Figure 10) show most points to have a value of one (98.83533%), however a small number of points were found to have a value less than one and were flagged. The points identified as outliers by maximum monthly MAI cumulative frequency graph can be seen as a very separate group in the PCA plot (Figure 11) due to high positive factor scores for component 4. Component 4 showed a high component loading for the maximum moisture availability index variable. The geographical position of these points can be seen in Figure 12. The points circled in blue are within the Middle Eastern region. The points which are circled in red area have particularly low MAI less than 0.4. This could indicate an ecotype adapted to drier conditions. However the fact that these points are found in areas with low MAI could be due to these speciemens being found on cultivated land where they are subject to irrigation. It could also be the case that specimens found near wadis or inhabit areas with a high water table allowing easy access to ground water (Tanton pers comm., 2007).

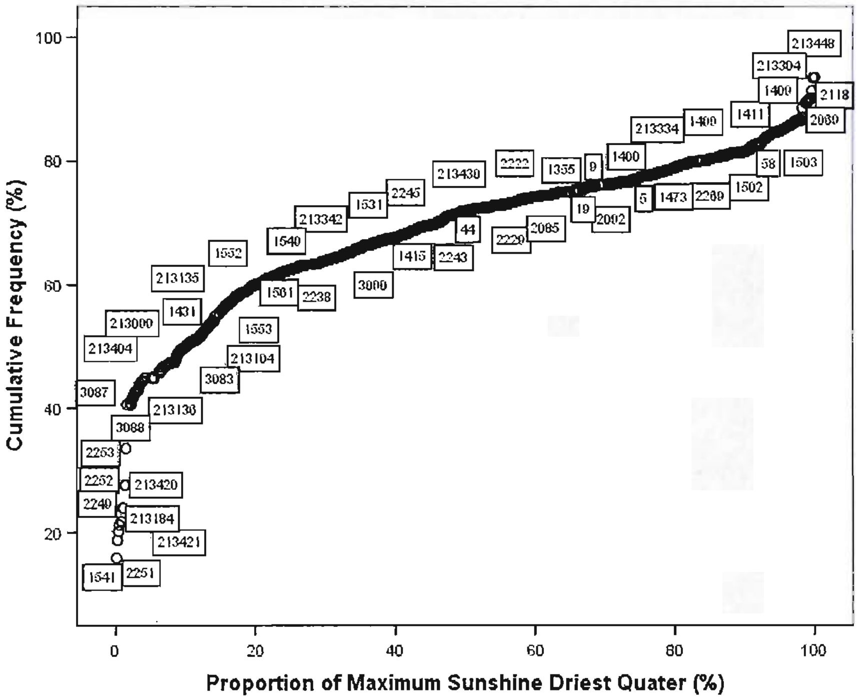


Figure 13 Cumulative frequency graph for Maximum sunshine driest quarter

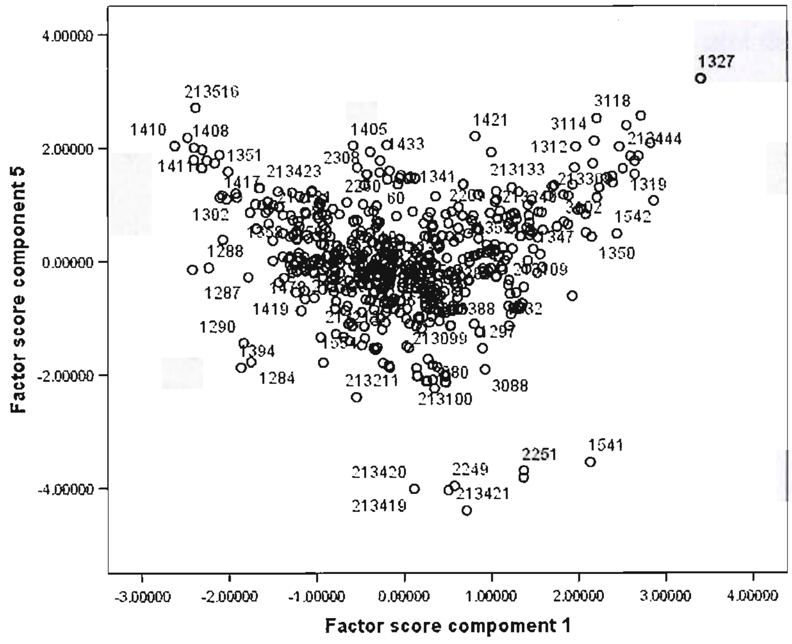


Figure 14 PCA component plot for climate only variables (component 1 and component 5)

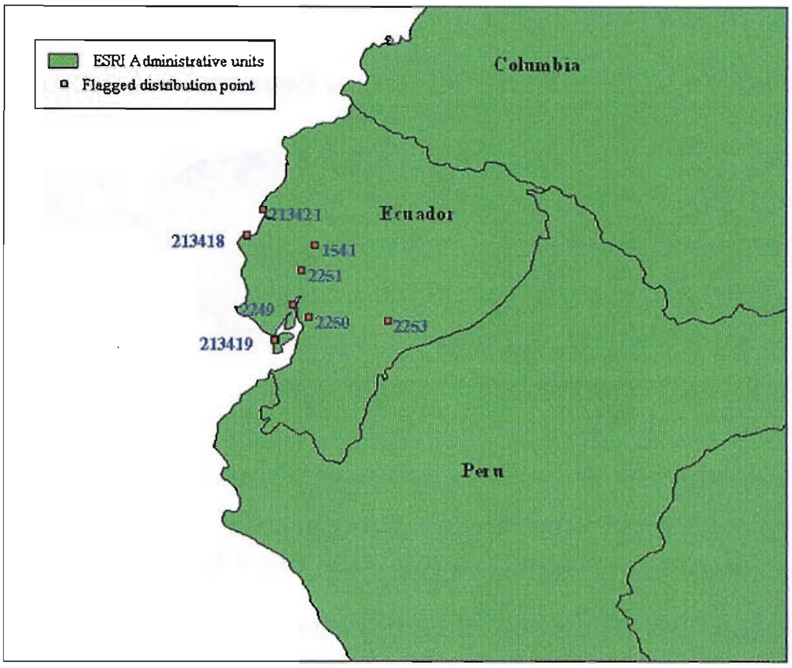


Figure 15 Geographical location of flagged occurrences records (cumulative frequency maximum sunshine driest quarter)

The points flagged as outliers (1541, 2249, 2250, 2251, 213184, 213419, 213421) by the proportion of maximum sunshine driest quarter cumulative frequency graph (Figure 13) can be seen as a very separate group in this plot due to very negative factor scores for component 5 (Figure 14). Sunshine driest quarter and sunshine both had high positive weightings for factor 5.

Figure 15 shows the points flagged in Figure 15 were found in Ecuador. The proportion of maximum sunshine at locations where tamarind is found in Ecuador appears to be significantly lower than any other location in which it is found. This is likely to be due to high levels cloud cover in this tropical region

3.5.2.5 Tamarind occurrence dataset for modelling

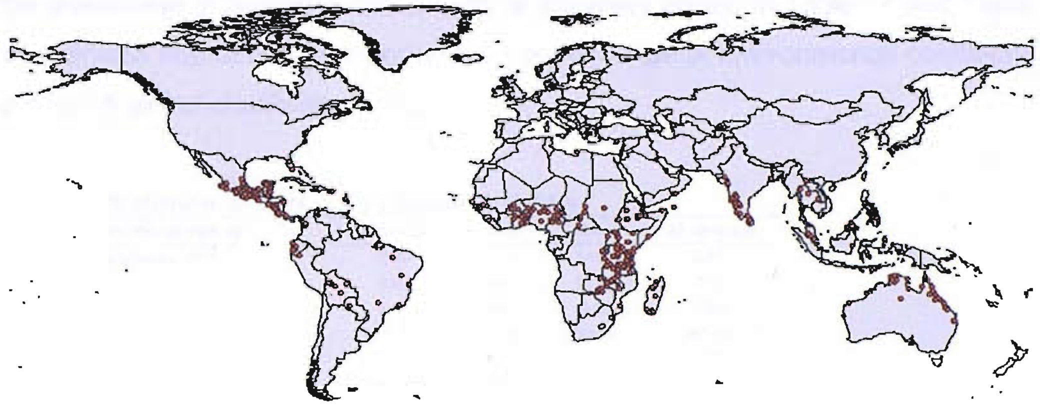


Figure 16 Tamarind dataset for modelling (359 occurrence records)

702 presence points remained after the removal of erroneous records. This was reduced to 543 point with removal of records collected before 1950. Once converted to 10X10 minute resolution this was reduced to 443. 359 of theses 443 presence points occured at locations where values were available for all environmental variables. This is due to the varying extent of environmental variable datasets. It is the 359 occurrence points that will be used in the distribution modelling process (Figure 16).

3.5.2.6 Data exploration of tamarind occurance dataset

Data exploration was carried out for the 359 data points to be used for modelling the distribution of tamarind. The statistical summary shown in Table 14 and Table 15 indicates that tamarind experiences a broad range of environmental conditions across its global distribution range.

Table 14 Statistical Summary for climatic variables

Environmental Variable	Minimum	Maximum	Mean	Std. Deviation
Mean Annual Temperature (°C)	15.80	28.60	24.66	2.17
Mean Diurnal Range (°C)	5.80	19.60	11.24	2.17
Isothermality (°C)	4.20	8.90	6.48	0.93
Temperature Seasonality	26.90	471.70	181.44	87.18
Maximum Temperature of Warmest Month(°C)	25.90	41.20	33.33	2.85
Minimum Temperature of Coldest Month(°C)	4.20	24.10	15.78	3.35
Temperature Range(°C)	7.50	30.50	17.55	4.00
Mean Temperature Wettest Quarter(°C)	14.20	31.10	25.33	2.27
Mean Temperature Driest Quarter(°C)	13.60	29.00	23.41	2.96
Mean Temperature Warmest Quarter (°C)	18.00	32.60	26.83	2.41
Mean Temperature Coldest Quarter (°C)	13.50	27.00	22.30	2.65
Annual Precipitation (mm)	32.00	3817.00	1333.25	635.70
Precipitation of Wettest Month (mm)	9.00	1373.00	257.28	135.44
Precipitation of Driest Month (mm)	0.00	190.00	18.49	30.66
Precipitation Seasonality (mm)	10.00	151.00	78.99	25.65
Precipitation of Wettest Quarter (mm)	15.00	2902.00	650.43	315.79
Precipitation of Driest Quarter (mm)	0.00	599.00	68.40	102.96
Precipitation of Warmest Quarter (mm)	3.00	1240.00	323.25	200.38
Precipitation of Coldest Quarter (mm)	0.00	1829.00	180.49	244.01
Relative Humidity of Coldest Quarter (%)	19.63	90.33	69.02	12.32
Relative Humidity of Driest Quarter (%)	17.70	89.40	62.99	14.27
Relative Humidity Warmest Quarter (%)	30.83	90.70	66.12	11.75
Relative Humidity Wettest quarter (%)	36.23	90.70	76.71	6.22
Mean Sunshine Duration Coldest Quarter (%)	18.33	91.37	62.56	14.21
Mean Sunshine Duration Driest Quarter (%)	18.87	93.57	67.32	12.73
Mean Sunshine Duration Hottest Quarter (%)	21.80	85.43	59.63	9.96
Mean Sunshine Duration Wettest Quarter (%)	21.80	80.67	49.82	9.23

Table 15 Statistical Summary for edaphic variables

Environmental Variable	Minimum	Maximum	Mean	Std. Deviation
Mean Moisture Availability Index of Coldest Quarter	0.00	1.00	0.62	0.38
Moisture Availability Index Seasonality	0.00	122.00	47.05	30.28
Mean Moisture Availability Index of Hottest Quarter	0.01	1.00	0.74	0.25
Maximum Monthly Moisture Availability Index	0.10	1.00	0.99	0.08
Mean Moisture Availability Index of Highest Quarter	0.06	1.00	0.99	0.09
Mean Annual Moisture Availability Index	0.03	1.00	0.74	0.19
Minimum Monthly Moisture Availability Index	0.00	1.00	0.27	0.35
Mean Moisture Availability Index Lowest Quarter	0.00	1.00	0.33	0.36
Bulk Density subsoil (g/cm^{-3})	0.80	1.76	1.40	0.14
Bulk Density topsoil (g/cm^{-3})	0.76	1.65	1.36	0.14
Percentage Clay subsoil (%)	3.75	68.00	31.31	13.57
Percentage Clay topsoil (%)	4.00	60.00	26.25	11.63
Percentage Sand subsoil (%)	3.00	89.00	39.06	15.96
Percentage Sand topsoil (%)	7.00	90.00	45.68	15.15
Percentage Silt subsoil (%)	4.00	53.00	23.71	8.47
Percentage Silt topsoil (%)	5.00	55.00	28.18	8.57
Effective CEC subsoil (cmolc kg^{-1})	2.00	59.00	17.83	10.46
Effective CEC topsoil (cmolc kg^{-1})	3.00	55.00	16.94	9.84
Organic Carbon content subsoil (% by weight)	0.20	1.91	0.50	0.27
Organic Carbon content topsoil (% by weight)	0.35	7.00	1.20	0.82
pH subsoil	4.60	8.90	6.33	0.93
pH topsoil	4.40	9.00	6.25	0.83
Total nitrogen subsoil (% by weight)	0.02	0.18	0.06	0.02
Total nitrogen topsoil (% by weight)	0.03	0.34	0.11	0.06

3.5.2.6.1 Principle Component Analysis

Investigation of the variable weighting scores of the first three components in the PCA analysis shows that component 1 (24% of the variance), 2 (14% of the variance) and 3 (10% of the variance) relate highly to water availability (i.e. moisture availability index seasonality, precipitation driest quarter and relative humidity hottest quarter), Soil characteristics (i.e. organic carbon, total nitrogen, Percentage sand, Effective CEC) and temperature (i.e. mean annual temperature, mean temperature driest quarter) respectively.

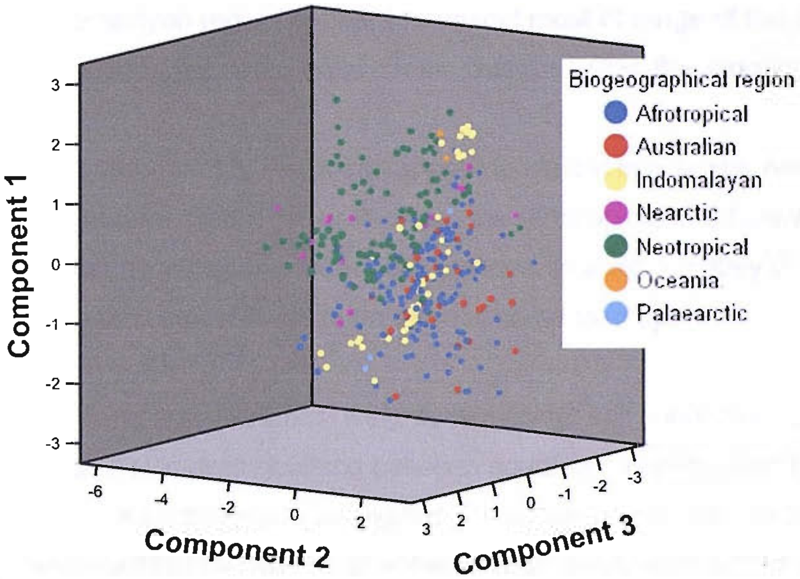


Figure 17 PCA Plot of component scores for Component 1, 2 and 3 classified by geographical region.

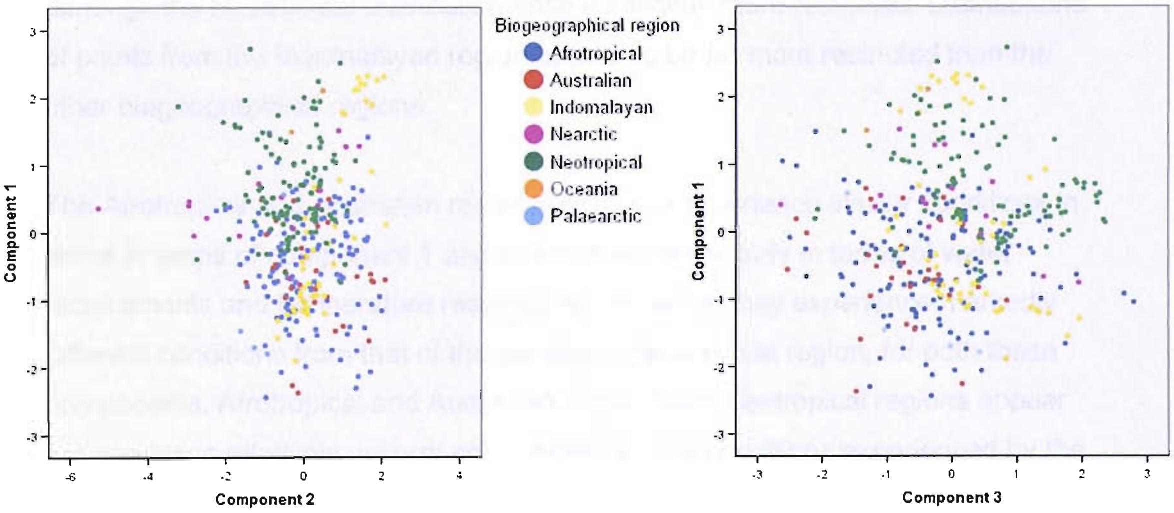


Figure 18 PCA component plots of component scores for Component 1- 2 and 1-3, Points are classified biogeographical region

Figure 17 shows a clustering of tamarind presence points based on biogeographical region within the first 3 components. Figure 18 shows the presence records within each region are distributed across each component. For Component 1 there is a clear gradient from points within the Neotropical region range from positive values to approximately zero, to points found in Afrotropical and Australian regions which range from approximately zero to negative values, with some over lap between the groups/clusters. Points in the Neoarctic region are found mainly within the middle of the range while the points found in the Indomalayan region appear to transect most of range of the component, although there does appear to be two main clusters within this biogeographical region.

For component 3, the points appear to cluster in a similar pattern to that of component 1, with the points from the Afrotropical and Australian regions clustering in the area of negative values, to approximately 0, and the Neotropical from zero onwards although there is some overlap.

For component 2, which weights heavily for soil conditions, appears to be less influential in distinguishing between conditions experienced by tamarind in different biogeographical regions. Presence points from Afrotropical, Australian and Neotropical regions all show a broad distribution across the component, although the Neotropical distribution appears slightly more restricted. Distributions of points from the Indomalayan region appear to be far more restricted than the other biogeographical regions.

The Afrotropical and Australian regions appear to experience similar conditions in terms in terms of component 1 and 3, which weight heavily in terms of water requirements and temperature respectively. However they experience markedly different conditions from that of the species in Neotropical region, for both these components. Afrotropical and Australian regions and Neotropical regions appear to experience relatively uniform soil conditions. Soil conditions experienced by the species in the Indomalyan region do not show the same variation at those experienced in the other regions.

3.6 Discussion

A number of validation techniques which identify outliers in both geographic and environmental space were used to flag potentially erroneous data points.
Most erroneous records were identified through the process of geographical validation. Though a large number of outliers were flagged by the cumulative frequency analysis; a recheck of geo-referencing seemed to show few to contain erroneous records. Few records were flagged by the critical value analysis method in comparison to the cumulative frequency graphs. This was due to the critical value analysis method requiring a very high threshold score to be exceeded before records could be flagged.

A number of processes were also carried out to account for bias which may occur in such datasets. By using a resolution of 10 x 10 minutes (approx 19 x 19 km) both in term of species and environmental data both sampling bias and spatial precision are addressed. In converting the species occurance dataset to 10 x 10 minute resolution; this ensured only one occurrence point in each 10 x 10 minute grid cell of the environmental datasets. This had the effect of reducing the effects of sampling bias within the species dataset which may occur due differing sampling effort across the species distribution. Stockwell and Peterson (2001) noted that the significance of the spatial accuracy of the data is often dependent on the scale of the environmental data used. Although precision in georeferencing dataset could not always be consistant, a standard procedure was always used. By working at a resolution of 10x10 minute resolution a realtivily large scale in terms of grain size, the majority of occurrence records are highly likely to be precise and accurate to this resolution.

Temporal bias was addressed by only using occurance records collected after 1950. Although this will not fully account for any change which may have occurred in distribution over time, all trees sampled are still likely to be alive. This period also matches the time duration over which the climate averages for worldclim climate datasets were recorded and allows a workable number of occurance points to be used in the model.

The end result of treatment of existing data will always be limited, particularly if the data was collected for other purposes, such as taxonomic description. If existing data contain a strong bias of unknown direction, no amount of treatment can remove it (Williams et al., 2002). However it is important to note that although errors can occur, herbaria collections in which the deficiencies are recognised , can offer by sheer number invaluable insights into a species spatial distribution and ecology (Reutter et al., 2003). The challenge is using existing data to devise appropriate treatments so the effects of these sampling biases are minimised (Williams et al., 2002).

In this study tamarind occurrence data has been acquired from a broad number of sources from around the world. Occurrences have been collected from the full global range of the species. A number of error identifying procedures have been carried out on the data. Sampling and temporal bias and spatial accuracy has been accounted for. Therefore it is assumed that this dataset is appropriate for the purpose of modelling tamarinds worldwide potential distribution. It should also be noted that at such an extent the process of systematic field surveys would be impractical and very costly.

During the error identification process, cumulative frequency analysis identified a number of groups of outliers where the species are experiencing different conditions, indicating the possibilities of ecotypes. However as erroneous points had not been removed and the PCA's only included climate data, these were not ecologically investigated. However these findings did provide reason to carry out exploratory analysis on the "clean" error free occurance dataset.

The exploratory analysis indicated that across its global distribution tamarind is subject to a broad range of both climatic and edaphic conditions. However this is characterised by a difference in environmental conditions being encountered between biogeographical regions (particulary in terms of water availability and temperature). This could be explained by variation of environmental conditions available to the species within the biogeographical regions it inhabits. A transitional change across the range of the species between biogeographical regions (indicated by the overlap in clusters) endorses this.

However the different conditions could indicate the possibility of regional subpopulations within the main population - a differentiation of the niche (niche evolution) within the population. This is the effect of geographical barriers or selection of material by farmers adapted to local environmental conditions. Caution however should be taken when suggesting such explanations without substantiation from analysis of morphological characteristic, genetic analysis, or reciprocal transplant experiments.

Due to the semi managed and domesticated nature of tamarind the likelihood of selection of material and development of ecotypes is high. Genetic variation is likely to exist between plant populations for plant species that cover large and/or environmentally heterogeneous distribution areas, both for survival and growth rate and quantity and quality of products (Lillesø et al., 2001).

4 *Chapter 4 Modelling the potential production areas of* Tamarindus indica *using species occurrence data*

4.1 Introduction

An extensive dataset representive of the full world distribution of tamarind was created mainly from herbaria records, (see chapter 3). The data shows variation between biogeographical regions for the environmental conditions experienced by the species. This indicates that it may be appropriate to produce regional as well as global models to allow analysis in more detail for the relationship between the species and the environment. As discussed previously the methods of collection of such data are rarely documented and so it is therefore not possible to infer absences (Elith et al., 2006). Due to the broad geographical extent of the study area, it is not possible to acquire reliable absence data personally.

However, species distribution models have been developed for use with presence only data (Busby, 1991; Walker and Cocks, 1991; Carpenter et al., 1993; Peterson et al., 1999; Stockwell and Peters, 1999; Hirzel et al., 2001)(Hirzel et al., 2002) (see section 2.2.3). It is now possible to obtain electronic access to vast resources of museum and herbaria data. For this reason such methods have become increasingly important (Elith et al., 2006).

One type of presence only model is envelope models. Conceptually envelope models are very close to the niche theory; as they try to delineate in environmental space, the hyper surface that circumscribes all suitable conditions (Hirzel1 and Arlettaz, 2003b). The presence-only modelling technique applied in this study is Ecological Niche Factor Analysis (ENFA) and this is implemented in the software, Biomapper (Hirzel et al., 2002). ENFA has been shown to perform well when compared to presence/absence models (Hirzel et al., 2001)

ENFA is an envelope model and so conceptally similar to Hutchinson niche. ENFA uses species presence to compute a number of ecologically meaningful orthogonal factors from a number of environmental predictor variables. ENFA uses these factors to quantify a hyper volume defined in the multidimensional space of ecological variables within which a species can maintain a viable population. The models identify suitable habitat based on the ecological/environmental space and map this on to geographical space to produce suitability maps.

Since factors are built to maximise the discrimination between the areas where the species is present compared to the rest of the study area. These factors may be seen as the most important gradients that the species is responding to in the study area. It is assumed that the response of the species along the principle axes constitutes a description of its observed niche (Cassinello et al., 2006).

ENFA has been used to model the potential habitat of a number of species; examples including reptiles and amphibians (Tole, 2006), bird species (Hirzel1 and Arlettaz, 2003a; Hirzel1 and Arlettaz, 2003b; Hirzel et al., 2004), mammals (Hirzel et al., 2002; Cassinello et al., 2006), insects (Gallego et al., 2004; Chefaoui et al., 2005; Estrada-Pena et al., 2006) marine (Bryan and Mataxas, 2007) and plants species (Feria et al., 2007; Ortega and Vackavik, 2007)

In this study ENFA is used to characterise the niche of tamarind across its entire distribution range and regional subsets, based on a global dataset of the species distribution. This will be used to produce predictive maps of potential production areas.

4.2 *Methodology*

4.2.1 Sources of species and environmental variable data

The tamarind species presence dataset containing 359 presence records was converted to 10x10 minute boolean (0 or 1) raster grid for use in the biomapper software (Hirzel et al., 2006).

Raster based grids for environmental data (climate, soil and moisture availability) at 10X10 minute resolution (as described in Chapter 3) were used in this study.

A list of these variables can be seen in Table 16. Environmental variables were normalised as far as possible through Box-Cox transformation (Sokal and Rohlf, 1981); with the biomapper software (Hirzel et al., 2006) .The Box-Cox transformation was implemented to ensure multivariate normality (see section 4.2.4).

Table 16 Environmental variables used in analysis and their abbreviations

Climate		**Moisture**	
C-TAn	Mean Annual Temperature (°C)	M-AICdQ	Mean Moisture Availability Index of Coldest Quarter
C-TDR	Mean Diurnal Range (°C)	M-AISn	Moisture Availability Index Seasonality
C-TIso	Isothermality (°C)	M-AIWmQ	Mean Moisture Availability Index of Warmest Quarter
C-TSn	Temperature Seasonality	M-AIMxM	Maximum Monthly Moisture Availability Index
C-TMxWmM	Maximum Temperature of Warmest Month(°C)	M-AIMxQ	Mean Moisture Availability Index of Maximum Quarter
C-TMnCdM	Minimum Temperature of Coldest Month(°C)	M-AIAn	Mean Annual Moisture Availability Index
C-TRn	Temperature Range(°C)	M-AIMn	Minimum Monthly Moisture Availability Index
C-TWtQ	Mean Temperature Wettest Quarter(°C)	M-AIMnQ	Mean Moisture Availability Index Minimum Quarter
C-TDyQ	Mean Temperature Driest Quarter(°C)	**Soil**	
C-TWmQ	Mean Temperature Warmest Quarter (°C)	S-BDsub	Bulk Density subsoil (g/cm^{-3})
C-TCdQ	Mean Temperature Coldest Quarter (°C)	S-Bdtop	Bulk Density topsoil (g/cm^{-3})
C-PAn	Annual Precipitation (mm)	S-ClyPsub	Percentage Clay subsoil (%)
C-PWtM	Precipitation of Wettest Month (mm)	S-ClyPtop	Percentage Clay topsoil (%)
C-PDyM	Precipitation of Driest Month (mm)	S-SndPsub	Percentage Sand subsoil (%)
C-PSn	Precipitation Seasonality (mm)	S-SndPtop	Percentage Sand topsoil (%)
C-PWtQ	Precipitation of Wettest Quarter (mm)	S-SltPsub	Percentage Silt subsoil (%)
C-PdyQ	Precipitation of Driest Quarter (mm)	S-SltPtop	Percentage Silt topsoil (%)
C-PWmQ	Precipitation of Warmest Quarter (mm)	S-GrPtop	Percentage Grave topsoil (%)
C-PCdQ	Precipitation of Coldest Quarter (mm)	S-GrPsub	Percentage Gravel subsoil (%)
C-RHCdQ	Relative Humidity of Coldest Quarter (%)	S-ECECsub	Effective CEC subsoil (cmolc kg^{-1})
C-RHDyQ	Relative Humidity of Driest Quarter (%)	S-ECECtop	Effective CEC topsoil (cmolc kg^{-1})
C-RHWmQ	Relative Humidity Warmest Quarter (%)	S-OCsub	Organic Carbon content subsoil (% by weight)
C-RHWtQ	Relative Humidity Wettest quarter (%)	S-OCtop	Organic Carbon content topsoil (% by weight)
C-SDCdQ	Mean Sunshine Duration Coldest Quarter (%)	S-pHsub	pH subsoil
C-SDDyQ	Mean Sunshine Duration Driest Quarter (%)	S-pHtop	pH topsoil
C-SDWmQ	Mean Sunshine Duration Warmest Quarter (%)	S-TNsub	Total nitrogen subsoil (% by weight)
C-SDWtQ	Mean Sunshine Duration Wettest Quarter (%)	S-TNtop	Total nitrogen topsoil (% by weight)

4.2.2 Global and regional study areas

Tamarind distribution, ranges through the tropic, subtropics and equatorial regions and is found in both dry and wet zones from arid, semi arid to monsoon regions (El-Siddig et al., 2006) (please see chapter 2 and 3 for further details on tamarind distribution).

For this reason the tropics, subtropics and equatorial regions was used for this study. All dataset were clipped to latitude 35° north to 35° south to include the regions above (see Figure 19). By reducing the size of the background dataset it is likely to further increase the sensitivity of the model in relation to marginality and specialisation (see below – section 5.2.3) (Hirzel et al., 2002) .

Exploratory statistical analysis of the tamarind species dataset in relation to the environmental variables, (See Chapter 3) was applied using PCA (Principle Component Analysis). This indicated that in different biogeographical regions, tamarind appeared to be experiencing varying environmental conditions across its global range. This is likely to have an impact on the result of a global model due to spatial nonstationarity within niche across the study area(Osborne4 and Suarez-Seoane, 2002f). This could result in biasing the species environment relationship and an inaccurate prediction. Osborne and Suarez-Seoane (2002) and Estrada-Pena et al. (2006) found that large scale models were improved when data was partitioned before analysis.

For this reason the full global species dataset was divided into three regions; based on geographic separation of occurance data points across 3 continents (Figure 19); the “Americas” (126 presence points), “Africa” (149 presence points) and “Australasia” (84 presence points).

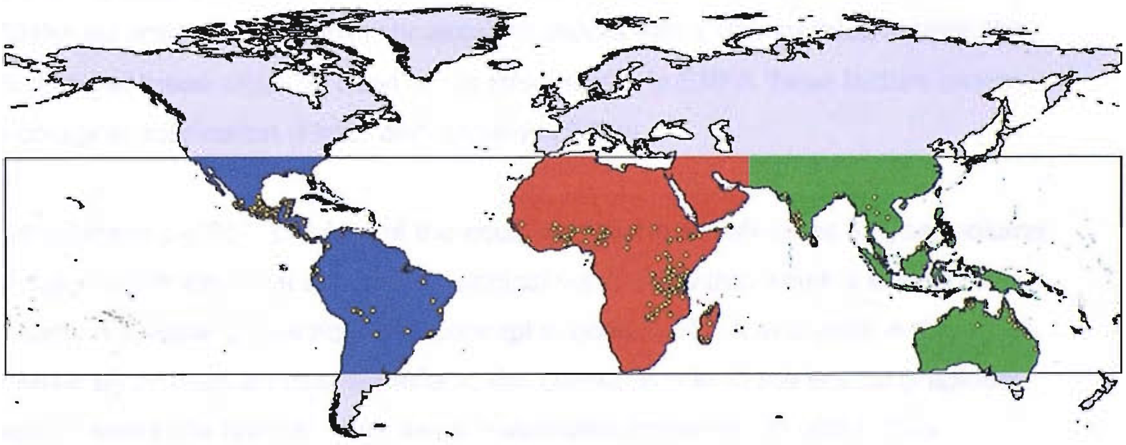


Figure 19 Delineation of global and regional study areas
The rectangle shows extent for the global model, colours identify "regional" areas, Americas (blue); Africa (red); Australia (green), yellow circles represent tamarind occurance points.

4.2.3 Variables selection

It has been stated that for ENFA, no variable selection was required as highly correlated variables do not affect the model output(Hirzel et al., 2002). ENFA extracts all relevant information from variables while discarding their correlations and the background noise(Hirzel et al., 2004). However preliminary runs of ENFA found both validation score and suitability map output to be sensitive to variable selection. Two sets of models are used; one run with all variables and the other with a subset of variables, each set includes global and regional models. The reduced variable set was created by removal of variables that were co-correlated above 0.7.

4.2.4 Ecological Niche Factor Analysis

Environmental predictor variables (Table 16), and species presence dataset (consisting of locations where the species has been detected) were entered into the Biomapper software and ENFA analyses conducted to characterise the niche of the species and produce predictive maps of potential production areas.

ENFA compares the environmental predictor variable distribution for the whole study area **(global dataset)** to a subset of it where the species occurs **(species dataset)**. Like Principle Component Analysis (PCA)(McGarigal et al., 2000), ENFA summarises all X environmental predictors into X uncorrelated factors, the first few of these retaining most of the information. In ENFA these factors convey ecological information (Hirzel and Arlettaz, 2003a).

Hutchinson's (1957) concept of the ecological niche is defined as a hyper-volume in the multidimensional space of ecological variables within which a species can maintain a viable population. The concept is used in ENFA in exactly in the same sense; by ecological niche we refer to the subset of cells in the ecogeographical space where the focal species has a reasonable probability to occur. This multivariate niche can be quantified on any of its axes by an index of marginality and specialization. ENFA uses these factors to quantify a hyper volume in multivariate space by an index of marginality and specialisation (Hirzel et al., 2002). Each cell within the dataset is thus associated to a factor whose components are the values of the environmental variables in the underlying area. This can be represented by a point in the multidimensional space of the environmental variables. If distributions show multivariate normality, the scatter plot will have the shape of a hyper ellipsoid. The cells of the species dataset constitute a subset of the global distribution and are plotted within the global one (see Figure 20a)(Hirzel et al., 2002).

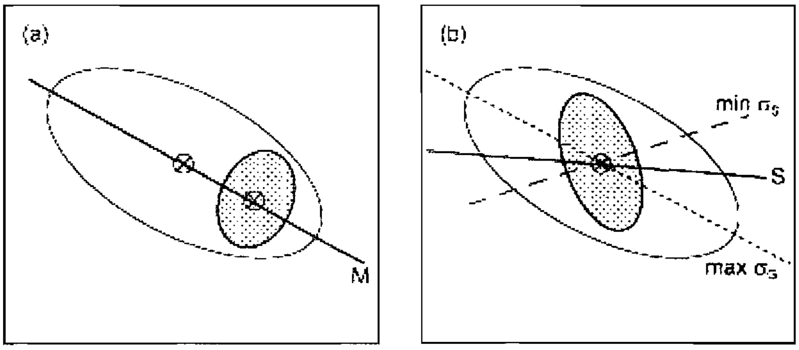


Figure 20 a and b, Geometrical interpretation of marginality and specialisation factors.
The two dimensional distribution of the global and species sites are symbolised by the white and dotted ellipses, with a crossed circle marking the centroid of each. (a)The marginality factor (M) is the axis through both centroid. (b) The specialisation factor (S) axis maximise the ratio of global variance σ_G to species variance σ_S; it's the intermediary between the axes of maximal global variance (dotted line) and the minimum species variance (dashed line). Adapted from Hirzel et al. (2004)

The first factor accounts for marginality of the species niche, this passes through both centroids of the two ellipsoids. The remaining factors account for specialisation, the first specialisation factor maximises the ratio between global variance and the species variance (see Figure 20b). The remianing specialisation factors are then extracted in turn, each step removing one dimension from space until all X factors are extracted (Hirzel et al., 2002)

Marginality is defined as the absolute differences between the global mean (that of the whole study area) and the species mean (the locations where the species have been detected) for the environmental variables (1).

$$M = \frac{|m_G - m_S|}{1.96\sigma_G} \qquad (1)$$

Where the global mean is defined as (m_G) and the species mean as (m_S) and (σ_G) the standard deviation of the global distribution.

The multivariate extension of (1) when applied to more than one environmental variable and as implemented in the Biomapper software (Hirzel et al., 2006) is provided by equation (2). Where (M) represents the overall marginality, (m_i) represents the coefficients of the marginality factor (expresses the marginality of the focal species on each environmental variable), and (V) represents the number of variables.

$$M = \frac{\sqrt{\sum_{i=1}^{V} m_i^2}}{1.96} \qquad (2)$$

A species is more marginal the further the species-mean differs from the global mean. The marginality factor also reflects the direction in which the species niche mostly differs from the available conditions in the global area (Hirzel et al., 2001; Hirzel et al., 2002). If the species statistical distribution along the variables is bimodal or asymmetrical it will be reflected by this factor (Hirzel and Arlettaz, 2003a).

The first factor is chosen to account for 100 percent of the marginality of the species, it will also account for some proportion of specialisation.

The remaining factors are then extracted orthogonally, chosen to maximise specialisation. The specialisation indicates how restricted or selective the species niche is in relation to the study areas, defined as the ratio of the variance in the global distribution to that in the species distribution.

Specialisation is calculated as;

$$S = \frac{\sigma_G}{\sigma_S} \qquad (3)$$

Where (σ_G) represents the standard deviation of the global distribution and (σ_S) that of the species distribution

A Global specialisation can be calculated as;

$$S = \frac{\sqrt{\sum_{i=1}^{V} \lambda_i}}{V} \qquad (4)$$

Where (λ_i) represents the eigenvalue associated to each factor and expresses the amount of specialisation it accounts for (Hirzel et al., 2002).

A more specialised species will have a much lower variance in comparison to the global data (Hirzel et al., 2002). Specialisation is expected to depend on interactions among variables. For instance a species preference in temperature may vary with changing humidity. Species may thus specialise on a combination of variables, rather than on every variable independently (Hirzel and Arlettaz, 2003a).

It is important to note that both the marginality and specialisation factors are highly dependent on the extent of the global dataset (Hirzel et al., 2002). The position of the factor axes and hence values of marginality and specialisation depend on the global set chosen as a reference. A species may appear very marginal or specialised on the scale of a whole country, but much less so on a subset of it (Hirzel et al., 2002). By reducing the size of the study area to latitude 35° north to 35° south (see section 4.2.2), this should remove large amounts of information from the global cloud in environmental space that are far from the ecological niche of tamarind thus increasing the sensitivity of the model.

The first few factors normally account for most of the information and are necessary to calculate habitat suitability. Marginality and specialisation factors were computed and the Mc-Authur's broken stick method (MacArthur, 1960) was used to select the number of factors to produce the suitability map.

4.2.5 Algorithm selection

The biomapper software allows selection of one of 4 algorithms to calculate habitat suitability (Hirzel1 and Arlettaz, 2003b). These include the **median algorithm** which is based on the frequency distribution of the observations in environmental space; **distance geometric mean**, **harmonic mean** and **minimum distance algorithms**, which are functions of the distance to the observation points in the environmental space (Hirzel1 and Arlettaz, 2003b).

The distribution of tamarind has been influenced to great extent by human activity through introduction and selection of plant material. This may have lead to the adoption to two or more kinds of environment, or inhabitation of suboptimal environments. This leads to a unsymmetrical or bimodal distribution of populations with environmental factors (Austin, 2002). The **median algorithm** was not selected as it assumes that the factor distributions are symmetric and unimodal (Hirzel and Arlettaz, 2003a). Although the **minimum distance algorithm** makes no assumption about the distribution of the species points, it does not consider density of presence points and therefore can be easily influenced by outliers.

Therefore the **geometric mean** and **harmonic mean** were the algorithms selected to compute the suitability score for all grid cells in the study area.
The geometric mean and harmonic mean make no assumption about distribution and take the density of points into account. The distance geometric mean algorithm was shown to be a good trade-off between two contradictory constraints, precision and generality (Hirzel1 and Arlettaz, 2003a; Titeux, 2006).The harmonic mean places high weightings on individual points this can lead to very close fitting of the data (Hirzel, per comms). However it may include information from points in areas which have been poorly sampled, but not considered by the geometric mean algorithm.

4.2.6 Creation of suitability maps

The algorithms selected in this study were the **distance geometric mean** and **distance harmonic mean**. The principle of geometric mean algorithm is to draw in the environmental space, the influence field or suitability field of each species point in such a way that when they are close together, their attraction powers are mutually reinforced. For any point of the environmental space, this is done by computing the geometric mean of the distance to all species points. The distance between each point on each dimension is weighted based on the eigenvalue for that factor. Thus the higher the density of species presence points in environmental space around a point in the global space, the higher is its habitat suitability (Hirzel and Arlettaz, 2003a; Hirzel1 and Arlettaz, 2003b). The harmonic mean algorithm is similar to the geometric mean but uses the harmonic mean of the distances instead. The effect of this mean is to give a high weight to all observations while keeping the information of observation density in the factor space.

Let O_i be the *N* species presence points given by their coordinates in *D*-dimensional environmental space. In this space, the weighted Euclidian distance between two points, A and B – whose coordinates are $A_1, A_2, ..., A_D$ and $B_1, B_2 ..., B_D$, is given by equation 5:

$$\delta(A,B) = \sqrt{\sum_{i=1}^{D} w_i (B_i - A_i)^2} \qquad (5)$$

Where w_i are the weights given by the amount of information explained by each factor (Hirzel and Arlettaz, 2003a; Hirzel1 and Arlettaz, 2003b).

Equal weights are given to marginality and specialisation, all the marginality component goes to the first factor, the specialisation component is apportioned among all factors proportional to their eigenvalue. The marginality factor will often account for more than half the weight as it normally accounts for some specialisation (Hirzel et al., 2002)

For any point P of the environmental space, the geometric mean of its distances to all observations points O_i is given by:

$$\mu_G(P) = N\sqrt{\pi_{i=1}^{N}\delta(P,O_i)} \qquad (6)$$

These means range from zero to infinity. High values of μ_G reflect low occurrence density in the hyperspace.

For any point P of the environmental, space the harmonic mean of its distances to all observations points O_i is given by

$$\mu_H(P) = \frac{1}{\frac{1}{N}\sum_{\substack{i=1\\P\neq O_i}}^{N}\frac{1}{\delta(P,O_i)}} \qquad (7)$$

In the case of both algorithms, envelopes are delineated in factor space, encompassing hyper volumes below a fixed threshold; thereby delineating the niche. Envelopes can then be delineated by circumscribing all points in the hyperspace that have a value lower than a certain threshold. Several envelopes can thus be defined enclosing different proportions of occurrences, from the innermost to the outermost part of the occurrence cloud. A Habitat Suitability (ENFA-HS) value is associated to each envelope by counting the proportion of occurrences they encompass. Classifying suitability of habitat on a continuous scale from 0 – 100, so that the class with a habitat suitability score of 99 contains 1% of the species points, the class 98, 2%, the class 50, 50% of the points, and so on. Those classes are used as the final HS values (Hirzel and Arlettaz, 2003a; Hirzel1 and Arlettaz, 2003b; Titeux, 2006). These points in ecological space are then mapped on to geographical space to produce a suitability map (see section 4.3.4).

4.2.7 Model Validation

4.2.7.1 Continous Boyce Cross validation Index and Cross validation Contrast Validation Index

Two measures of model quality were used to validate the model. The Continuous Boyce cross validation Index (Hirzel et al., 2006) and a cross validated Contrast Validation Index (CVI). The Continuous Boyce cross validation Index measures the suitability index's proportionality to probability of presence, in other words, the models ability to distinguish different classes of suitability. The CVI gives a measure of how much the model differs from chance expectation or randomness.

The cross validation process partitions into k (in this case 10) mutually exclusive sets. K -1 partitions are used to compute a model, while the left out partition is used to validate it, as independent data. The partitions are set so as they do not geographically overlap, this makes the cross validation more robust to autocorrelation attempts to account for spatial nonstationarity within the validation process. This process is repeated (k) times each time leaving out a different partition. The process result is 10 different habitat suitability maps.

For the continuous Boyce index (Hirzel et al., 2006) a moving window classification is applied to each habitat suitability map of width W (in this case W = 20 (Hirzel pers comm.)). Computation starts at the first class covering the suitability range [0,W(20)]. For each class a predicted-to-expected ratio (P/E) is calculated.

The predicted-to-expected ratio is given by

$$F_i = \frac{P_i}{E_i}$$

P_i is the predicted frequency of evaluation points, calculated as;

$$P_i = \frac{p_i}{\sum_{j=1}^{b} p_j}$$

Where p_i is the number of evaluation points predicted by the model to fall into habitat suitability class *i* and $\sum p_j$ is the total number of evaluation points, i.e. the frequency expected from a random distribution across the study area.

E_i is the expected frequency of points, i.e. the frequency expected from random across the study area. This is given by the relative area given by each class;

$$Ei = \frac{a_i}{\sum_{j=1}^{b} a_j}$$

Where a_i is the number of grid cells belonging to habitat suitability class *i*, and $\sum a_j$ is the overall number of cell in the study area.

The predicted-to-expected ratio (P/E) is plotted against the average suitability value of the class. The window is then moved a little forward and the process repeated until it reaches the last possible range 100 – W. This provides a smooth P/E curve. The continuous Boyce index is computed as the Spearman's rank correlation coefficient of the curve. The results from all (k) habitat suitability maps from the cross validation are used to calculate an average Boyce index.

4.2.7.2 Field data validation

Due to the unknown reliability and accuracy of the data records used in the study, a further validation was conducted in addition to the cross validation. Presence data (23 records) collected during a field survey in India from June – September 2003 using a GPS (therefore thought to be more reliable and accurate) were excluded from the occurrence dataset. All global models and Australasia models were rerun with the remaining 336 and 61 occurrence points' respectivily. The 23 field survey data occurrence records were used to validate the predicted distribution maps. (Figure 1 show the location of records collected during the field survey). The Contrast Validation Index is calculated for the field survey data.

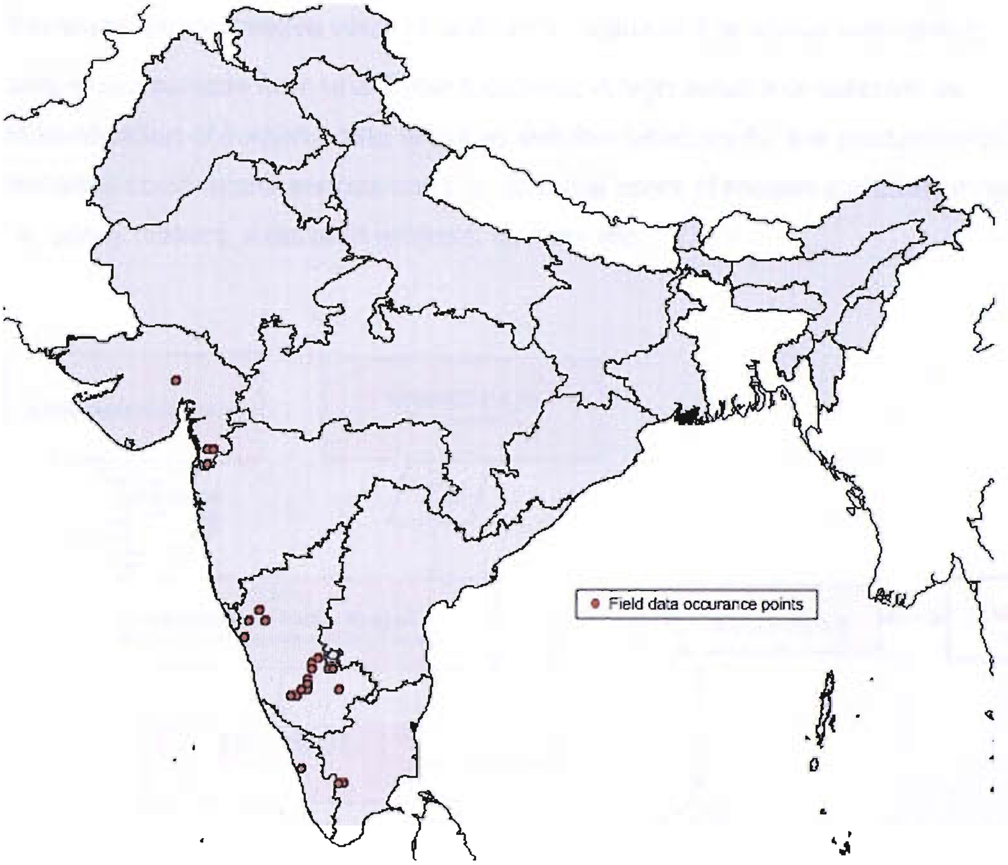


Figure 21 India field survey occurrence records

4.2.8 Suitability threshold selection

Even good predictive models suffer from uncertainty; this makes the use of the full continuous suitability scale inappropriate. A model showing only a few classes may be "more honest" about its actual content (Hirzel et al., 2006). Often suitability thresholds are selected on an arbitrary basis.

In this study suitability thresholds were selected for the models based on the shape of the mean F_i curve from the continous Boyce cross validation index. The method involves placing thresholds at natural boundaries of the curve (Hirzel et al., 2006). Hirzel et al. (2006) suggested using as the boundary a mean F_i value of above 1, to distinguish the first class of suitable habitat from unsuitable habitat; a F_i value of 1 or above shows that the model differs from random. However in this study a conservative value of a mean F_i value of 2 or above was used to distinguish suitable from unsuitable locations. A high value was selected as indentification of inappropriate areas as suitable locations for the production of tamarind could incour serious costs to potential users of models suitability maps, i.e. policy makers, extension workers, farmers etc .

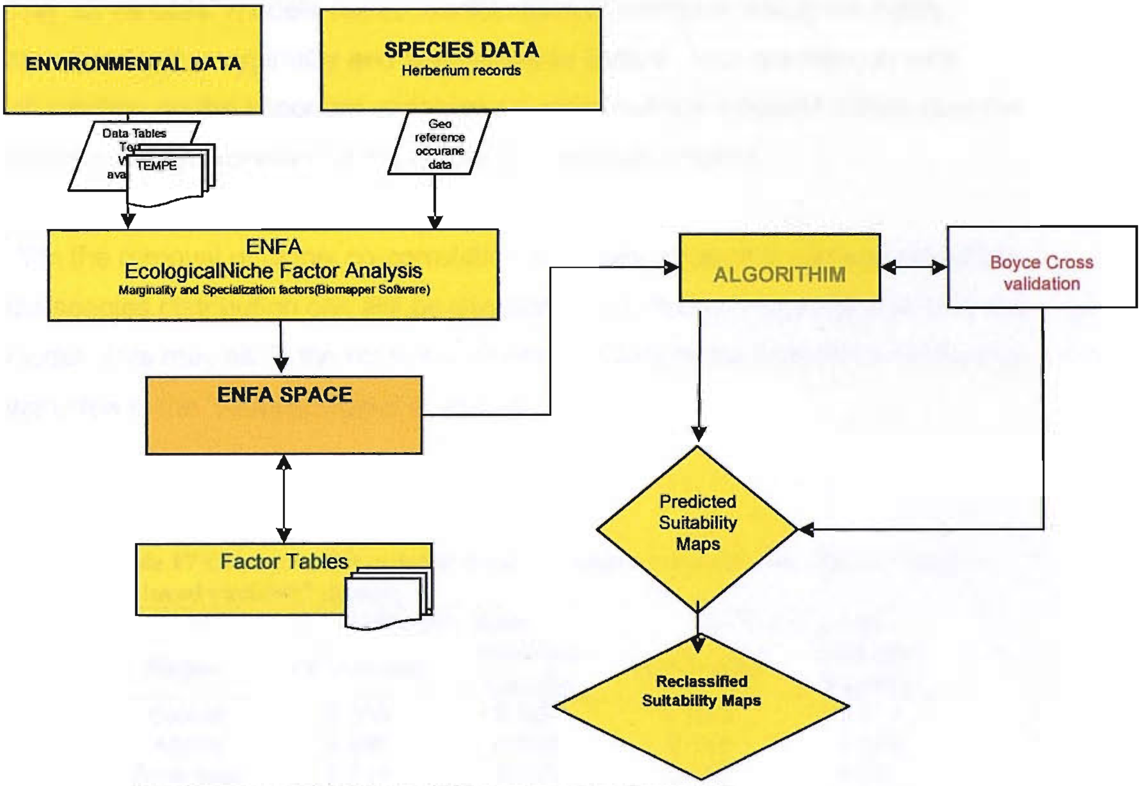


Figure 22 Flow diagram of ENFA modelling process and outputs

4.3 Results

Two sets of models are presented; the first set have been run with all variables ("all variable" models), the second with a subset of variables ("reduced variable" models). The reduced subset of variables was created by removal of variables that were co-correlated above 0.7. Each set includes global and regional models.

The "all variable" models allows identification of variables which are highly correlated with marginality and specialisation factors. This provides us with information on the important variables on which tamarind habitat differs from the background environment at the global and regional extents.

With the removal of highly co-correlation variables, most of the information from the species distribution can still be explained in by those remaining in use by the model. This may allow the factors to fit more closely to the axis of the remaining variables in the "reduced variable" model

4.3.1 Algorithm selection

Table 17 Continous Boyce index cross validation score for "all variable" and "reduced variable" models

	Harmonic mean		Geometric mean	
Region	All Variable	Reduced Variable	All Variable	Reduced Variable
Global	0.903	0.852	0.663	0.811
Africa	0.599	0.588	0.166	0.244
Americas	0.734	0.702	0.312	0.681
Australasia	0.667	0.354	0.068	0.398

The highest cross validation continous Boyce index validation score was received for the harmonic mean when with "all variable" selection was applied. The highest cross validation continous Boyce index validation score was received for the geomertic mean when the "reduced variable" selection was applied (Table 17). The outputs for these models are presented below.

4.3.2 ENFA factor analysis

Marginality and specialisation coefficients listed in ENFA factor tables (Table 18 - Table 25) were interpreted as described by Hirzel and colleagues (Hirzel et al., 2002; Hirzel and Arlettaz, 2003a; Hirzel et al., 2004); a brief description follows of this interpretation. The higher the absolute value of a marginality coefficient, the further the species departs from the mean available habitat (global mean) in relation to that variable. The signs (+/-) of the variable coefficient indicate if the mean for that environmental variable is higher or lower than the global mean. The higher the absolute value of a specialisation coefficient, the further the species variance differs from the global variance and the more restricted the range of the species is on the corresponding variable in relation to the study area.

In order to account for any effect caused by the Box-Cox transformation of variables on the coefficient values, coefficients are only interpreted to one decimal place (Hirzel et al., 2004). When the absolute marginality or specialisation coefficient is below 0.1, the species data was not considered to be significantly different from the global data (Hirzel pers comm.).

An overall (global) value for marginality and specialisation value are given for each model. As the overall (global) specialisation goes from 0 to infinity; tolerance (defined as the inverse of specialisation), is usually preferred as it ranges from 0 to 1(Hirzel et al., 2004). The overall marginality indicates how far (with all environmental variables being accounted for), the species optimum is from the average conditions in the study area. The specialisation or tolerance is indication of niche width. It must be noted that marginality and specialisation are relative to the study area and can only be compared between models when the same variables are used (Hirzel et al., 2004)

4.3.2.1 All variable models

4.3.2.1.1 Global "All variable" model

The global model for tamarind showed a marginality of 1.226, a specialisation 2.189 and a therefore a tolerance (1/Specialisation) of 0.457. This indicates that tamarind habitat differed dramatically from the mean conditions across the study area; however it does not have a highly restricted niche width.

The seven factors retained to produce the suitability map for the global model explained 83% of the information (the total sum of the eigenvalues), 100 % of marginality and 66% of the specialisation.

Table 18 (below) gives details of the specialisation explained and information on marginality and specialisation coefficients of variables for the first 5 of the 7 ecological factors retained to produce the suitability map for the global "all variables" model. From these factor correlation coefficients (Table 18) it is possible to identify the variables on which tamarind showed high marginality and specialisation (the variables which are highly correlated with marginality and specialisation factors). These are the variables that the model uses to discriminate suitable from non-suitable habitat.

Temperature: A high positive coefficient on the marginality factor showed tamarind habitat to have higher averages than the study areas for minimum temperatures both over a season and for the extreme conditions (Minimum temperature coldest quarter, minimum temperature coldest month). This was also the case for mean annual temperature and temperature during the driest quarter, although these variables had lower marginality coefficients. Highly negative marginality coefficients in regard to variation in temperature both in terms of the daily range (diurnal range) and variation between months (temperature seasonality, temperature range) identified tamarind habitat as having averages considerably lower than of the study area for these variables. This indicated that tamarind prefers areas with less temperature variation than found across the study area. Minimum temperature coldest month and temperature diurnal range showed high specialisation coefficients which indicated a restricted range width in the study area for these variables. Tamarind habitat showed no correlation with marginality for maximum temperatures and temperatures during warm periods.

Water availability: Precipitation over the whole year for tamarind habitat has a higher average than that of the study area. Highly positive marginality coefficients are given for precipitation during the wettest and warmest periods of the year (precipitation wettest month, precipitation wettest quarter, precipitation warmest quarter). However precipitations during dry and cold periods show conditions similar to the average for the study area. Variation in precipitation in tamarind habitat is higher than that in the study area indicated by high marginality coefficient for precipitation seasonality. Most precipitation variables have low coefficients on the specialization factors, indicating that tamarind shows a wide niche breadth in terms of precipitation variables. A similar pattern is reflected by with moisture availability. Areas inhabited by tamarind have a higher average mean annual moisture availability index (MAI) than that of the study area. Tamarind also shows high positive marginality coefficients for MAI during the wet periods; though MAI driest quarter and driest month is not different to the global average. As with precipitation variables, moisture availability variables show little specialization. So although tamarind seems to prefer greater than average water availability during the wet periods it does not appear to be restricted to these regions.

Soil: Relatively high marginality coefficients relating to soil with high organic material (total nitrogen, organic carbon) and medium soil texture (percentage silt). However a low level of specialization in the soil variables indicates that it can tolerate a broad range of conditions. Soil pH show little marginality or specialization indicating that tamarind distribution is not greatly limited by this soil characteristic.

Table 18 Specialisation explained by the first 5[$] of the 7 ecological factors retained and correlation coefficient symbols for marginality and specialistion for the most important 27[$] variables, for the Global "All variable" model

Environmental Variable	Factor 1 (18%) Marginality§	Factor 2 (14%) Spec.1#	Factor 3 (14%) Spec. 2#	Factor 4 (7%) Spec. 3#	Factor 5 (6%) Spec.4#
C-TMnCdM	++	****	**	****	0
C-TRn	- -	*******	******	*****	0
C-TCdQ	++	0	***	0	**
C-PWtM	++	0	0	0	*
M-AIMxQ	++	0	0	0	*
C-PWtQ	++	0	*	**	0
C-TSn	- -	0	0	*	0
C-TDR	- -	*	**	0	*
C-RHWtQ	++	0	0	0	0
M-AISn	++	0	0	0	0
S-OCtop	+	0	0	0	0
C-TAn	+	0	**	**	******
S-TNsub	+	0	0	0	0
S-OCsub	+	0	0	0	*
C-TIso	+	0	*	0	0
S-TNtop	+	0	*	0	0
C-PAn	+	0	**	****	*
C-RHWmQ	+	0	*	0	0
C-PWmQ	+	0	*	0	0
C-RHCdQ	+	0	0	0	0
C-RHDyQ	+	0	*	0	*
S-ECECsub	+	0	0	0	0
S-ECECtop	+	0	0	0	0
C-PSn	+	0	*	**	**
C-TDyQ	+	0	*	0	*
C-SDWtQ	-	0	0	0	0
S-SltPtop	+	0	0	0	0

§ The symbol + means that the species was found in locations with higher values than average in the study area. The symbol – means the reverse. The greater the number of symbols, the higher the correlation (the more the species mean differs from the global mean). 0 indicates a very weak correlation. Each symbol is representive of a unit of 0.1 for the coefficient value (e.g. ++ = 0.20 – 0.29).

#The symbol * means the species was found occupying a narrower range of values than available. The greater the number of asterix, the narrower the range (the more the species variance differs from the global variance). 0 indicates a very low specialization. Each symbol is representative of a unit of 0.1 for the coefficient value, (** = (+/- 0.20 – 0.29)).

$See appendix for table with all variables used in this model and all 7 factors retained to produce suitability map

Climate: C-TAn Mean Annual Temperature(°C), C-TDR Mean Diurnal Range(°C), C-TIso Isothermality(°C), C-TSn Temperature Seasonality, C-TMnCdM Minimum Temperature of Coldest Month(°C), C-TRn Temperature Range(°C), C-TDyQ Mean Temperature Driest Quarter(°C), C-TCdQ Mean Temperature Coldest Quarter(°C), C-PAn Annual Precipitation(mm), C-PSn - Precipitation Seasonality(mm), C-PWtQ Precipitation of Wettest Quarter(mm), C-PWmQ Precipitaton of Warmest Quarter(mm), C-RHCdQ Relative Humidity of Coldest Quarter(%), C-RHDyQ Relative Humidity of Driest Quarter, C-RHWmQ Relative Humidity Warmest Quarter (%), C-RHWtQ Relative Humidity Wettest quarter(%), C-SDWtQ Mean Sunshine Duration Wettest Quarter(%). ***Moisture Availability:*** M-AISn Moisture Availability Index Seasonality, M-AIMxQ Mean Moisture Availability Index of Maximum Quarter. ***Soil:*** S-SltPtop Percentage Silt topsoil(%),S-ECECtop Effective CEC topsoil (cmolc kg^{-1}), S-ECECsub Effective CEC subsoil(cmolc kg^{-1}), S-OCsub Organic Carbon content subsoil(% by weight), S-OCtop Organic Carbon content topsoil(% by weight), S-TNsub Total nitrogen subsoil(% by weight), S-TNtop Total nitrogen topsoil(% by weight).

4.3.2.1.2 Evidence for carrying out Regional models

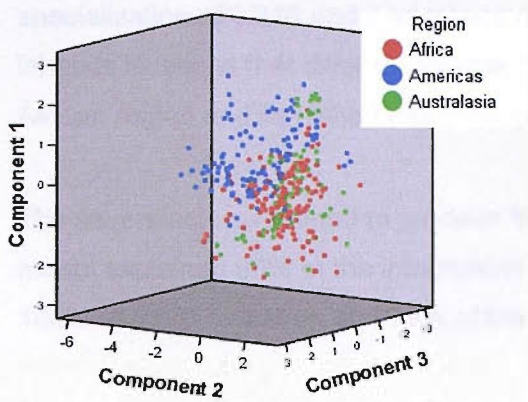


Figure 23 PCA Component plot for the 359 presence points, and all variables (see chapter 3), classified by the geographical regions defined in this chapter (see section 4.2.2.).

Figure 23 shows the component plot of the PCA analysis for the 359 presence points and all variables, classified by the geographical regions defined in this chapter (see section 4.2.2). The presence points from the 3 regions form clusters within component space. Figure 24 (a) shows the same points plotted in ecological spaces described by the factors produced from the ENFA analysis (described above, section 4.3.2.1.1). Clustering between groups can be seen although less defined than the PCA. Figure 24(b) shows the density of presence points plotted on the marginality factor. These plots (Figure 23 and Figure 24a and b) illustrate further evidence that the species are occurring in different realised or experienced niche within the different regions of study area and provides a further case for producing regional models.

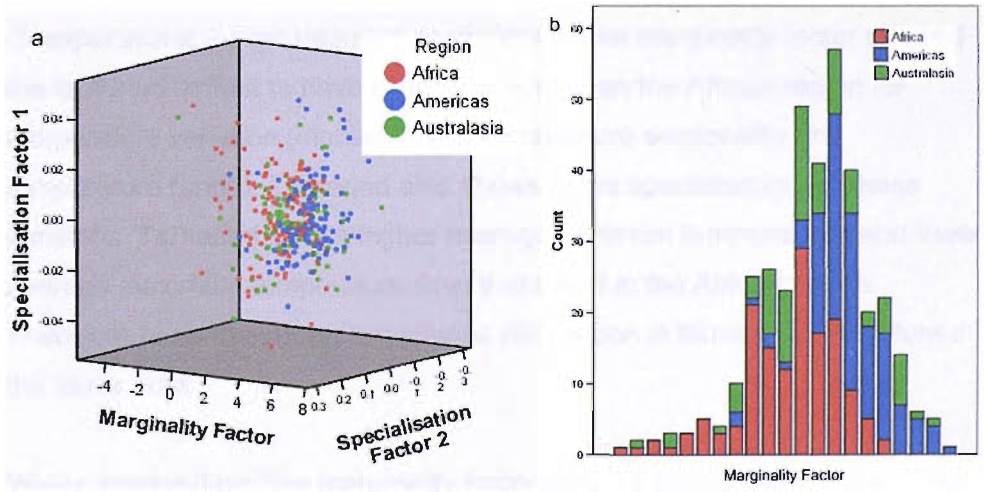


Figure 24 a and b, Presence points in ENFA space, (a) tamarind points mapped in three dimensional ENFA space (b) tamarinds presence point density plotted on marginality factor.

4.3.2.1.3 Africa region"All variable" model

The "Africa" regional model for tamarind showed a marginality of 1.629, specialization of 3.216 and a tolerance of 0.311. This indicates that tamarind inhabits locations that differ dramatically from the mean conditions across the African region and its niche breadth is quite restrictive.

The seven factors retained to produce the suitability map for the global model explained 90% of the information (the total sum of the eigenvalues) 100% of marginalisation and 79% of the specialisation.

Table 19 (below) gives details of the specialisation explained and the marginality and specialisation coefficients for variables for the first 5 of the 7 ecological factors retained to produce the suitability map. From these factor correlation coefficients (Table 19) it is possible to identify the variables on which tamarind showed high marginality and specialisation (the variables which are highly correlated with marginality and specialisation factors). These are the variables that the model is using to discriminate suitable from non-suitable habitat. The identified variables are listed below.

Temperature: A high negative coefficient on the marginality factor showed the tamarind habitat to have lower averages than the African region for temperature variation (diurnal range, temperature seasonality and temperature range). Tamarind also shows some specialisation for these variables. Tamarind shows higher average minimum temperatures and lower average maximum temperature than that found in the African region. Therefore tamarind shows a restricted distribution in terms of temperature in the study area.

Water availability: The marginality factor showed that tamarind on average was associated with areas of higher water availability over the whole year; (mean annual water availability index - species mean 0.67 *vs* global mean 0.48, annual precipitation - species mean 1058 mm *vs* global mean 715 mm);

in comparison with the African region. A high coefficient in the first and second specialisation factor for annual precipitation showed that for this variable tamarind has a narrow niche width. Tamarind also favoured areas where water availability was higher than the average found in the study area during the wettest periods of the year both in terms of precipitation (precipitation wettest month -212mm vs 122mm, precipitation wettest quarter 536 mm *vs* 339mm) and moisture availability (moisture availability index maximum quarter - 0.97 *vs* 0.75). During the dry season tamarind favoured areas where the water availability was found to be higher than the average in the African region (precipitation driest quarter - 31mm *vs* 26mm and moisture availability index minimum quarter - 0.20 vs 0.18). Tamarind also favoured area's were relative humidity was higher than average (relative humidity wettest quarter - 75% *vs* 65%, relative humidity coldest - 67% vs 57%, warmest - 61% *vs* 51% and driest -57% *vs* 48% quarters), particularly during the wettest periods.

Soil: Tamarind was found in higher than average conditions for variables which can influence soil nutritional variables such as organic carbon content (top and subsoil) and total nitrogen. Although tamarind is discriminating based on these soil characteristics; the low level of specialization indicates that although tamarind may prefer such conditions, it can tolerate a broad range in this region. Tamarind showed a preference for more acidic soils than found on average in the study area. Tamarind did not show preference for soil structure characteristics (see appendix).

Table 19 Specialisation explained by the first 5$^{\$}$ of the 7 ecological factors retained and correlation coefficient symbols for marginality and specialistion for the most important 31$^{\$}$ variables, for the Africa region "All variable" model

Environmental Variable	Factor 1 (34%) Marginality§	Factor 2 (17%) Spec.1#	Factor 3 (10%) Spec. 2#	Factor 4 (8%) Spec. 3#	Factor 5 (4%) Spec.4#
C-TRn	++	***	**	*******	****
C-PWtM	++	*	*	0	0
C-PWtQ	++	0	***	0	**
C-TDR	- -	**	*	0	*
C-RHWtQ	++	*	0	0	*
C-TMnCdM	++	0	0	***	*
M-AIMxQ	++	*	0	0	0
C-TSn	- -	0	**	0	*
C-PAn	++	**	******	0	0
S-OCtop	+	0	0	0	*
C-TCdQ	+	***	**	0	0
C-TIso	+	*	*	0	0
M-AISn	+	0	0	0	0
C-PWmQ	+	**	0	0	**
C-RHCdQ	+	0	0	0	0
S-OCsub	+	0	0	0	0
C-RHWmQ	+	**	0	0	*
S-TNtop	+	*	0	0	***
M-AIMn	+	0	0	0	0
C-RHDyQ	+	**	0	0	0
C-SDWtQ	-	0	0	0	*
C-TMxWmM	-	**	*	****	*
S-TNsub	+	0	0	0	*
M-AIAn	+	*	0	0	*
C-PDyM	+	0	0	0	0
C-PDyQ	+	*	0	0	*
C-PCdQ	+	*	0	0	0
S-pHsub	-	0	0	0	0
C-TWmQ	-	*	*	**	0
S-pHtop	-	0	0	0	0
M-AIWmQ	+	0	0	0	0

§ The symbol + means that the species was found in locations with higher values than average. The symbol –means the reverse. The greater the number of symbols, the higher the correlation (the more the species mean differs from the global mean). 0 indicates a very weak correlation. Each symbol is representive of a unit of 0.1 for the coefficient value (e.g. ++ = 0.20 – 0.29).

#The symbol * means the species was found occupying a narrower range of values than available. The greater the number of asterix, the narrower the range (the more the species variance differs from the global variance). 0 indicates a very low specialization. Each symbol is representative of a unit of 0.1 for the coefficient value (** = (+/-0.20 – 0.29)).

$ See appendix for table with all variables used in this model and all 7 factors retained to produce suitability map

Climate: C-TDR Mean Diurnal Range(°C), C-TIso Isothermality(°C), C-TSn Temperature Seasonality, C-TMxWmM Maximum Temperature of Warmest Month(°C), C-TMnCdM Minimum Temperature of Coldest Month(°C), C-TRn Temperature Range(°C), C-TWmQ Mean Temperature Warmest Quarter(°C), C-TCdQ Mean Temperature Coldest Quarter(°C), C-PAn Annual Precipitation(mm) , C-PWtM Precipitation of Wettest Month (mm), C-PDyM Precipitation of Driest Month(mm), C-PWtQ Precipitation of Wettest Quarter(mm), C-PDyQ Precipitation of Driest Quarter(mm), C-PWmQ Precipitation of Warmest Quarter(mm), C-PCdQ Precipitation of Coldest Quarter(mm), C-RHDyQ Relative Humidity of Driest Quarter, C-RHWmQ Relative Humidity Warmest Quarter (%), C-RHWtQ Relative Humidity Wettest quarter(%), C-SDWtQ Mean Sunshine Duration Wettest Quarter(%). ***Moisture Availability:*** M-AISn Moisture Availability Index Seasonality, M-AIWmQ Mean Moisture Availability Index of Warmest Quarter, M-AIMxQ Mean Moisture Availability Index of Maximum Quarter, M-AIAn Mean Annual Moisture Availability Index, M-AIMn Minimum Monthly Moisture Availability Index. ***Soil:*** S-OCsub Organic Carbon content subsoil(% by weight), S-OCtop Organic Carbon content topsoil(% by weight), S-pHsub pH subsoil, S-pHtop pH topsoil, S-TNsub Total nitrogen subsoil(% by weight), S-TNtop Total nitrogen topsoil(% by weight).

4.3.2.1.4 Australasia region "all variable" model

The regional model for Australasia gave a marginality score 1.49, a specialisation 5.35 and a tolerance of 0.187. This indicates that tamarind habitat differed dramatically from that of the mean of the study area and was highly restricted in term of niche breadth in this region. Tolerance was lower than that of the other regional models indicating a narrow niche width within this region.

The six factors retained to produce the suitability map explained 88% of the information (the total sum of the eigenvalues), 100% of the marginality, and 76% of the specialisation.

Table 20 (below) gives details of the specialisation explained and the marginality and specialisation coefficient values of variables for the first 5 of the 6 ecological factors retained. From these factor correlation coefficients (Table 20) it is possible to identify the variables on which tamarind showed high marginality and specialisation (the variables which are high correlated with marginality and specialisation factors). These are the variables that the model used to discriminate suitable from non-suitable habitat. The identified variables are listed below.

Temperature: Within this region tamarind showed high positive marginality in terms of mean annual temperature, minimum temperature coldest month (19.6 °C vs 14.9 °C) and mean temperature during coldest quarter (22.3 °C *vs* 16.1°C). Minimum temperature coldest quarter with a high value for the first and third specialisation factors indicated, that tamarind does not vary far from this mean in relation to this variable. Temperature driest quarter (23.6 °C *vs* 19 °C) was also found to have a high value on the marginality factor indicating a preference for a higher average for this variable than found in the study area. Similar reasoning on the other variable coefficients showed that tamarind favoured areas of lower temperature variation than average in the study area in relation to diurnal range (10.1 °C vs 11.7 °C), temperature seasonality (2135 vs 4562) and temperature range (17.6 °C vs 25 °C). For these variables tamarind showed high specialisation indicating that it could not tolerate areas very different from the average conditions.

Water Availability: Although precipitation in this region is higher than in Africa, tamarind in the Australasian region still showed favourability for higher than average annual precipitation (1523 mm vs 1240mm), higher than average precipitation during the wettest month (309mm vs 214mm) and higher than the average precipitation during the wettest quarter (757mm vs 651mm). This corresponded with a higher than average MAI for maximum quarter. Tamarind did however show preference for lower average precipitation during the driest month precipitation and driest quarter, corresponding with a lower than average MAI during these periods. Tamarind was found in areas with higher average water availability during the wet periods and lower average water availability during the dry periods compared to that of the study area. This preference for varying water availability during the different seasons corresponds with a higher average MAI seasonality and precipitation seasonality found in tamarind habitat in comparison with the study area.

Sunshine: Tamarind habitat also shows a preference for higher average conditions than that of the total Australasian region in regard to sunshine during the coldest periods and wettest periods. This is likely to be due to high cloud cover found in this region during these periods.

Soil: In this region tamarind also appears to discriminate suitable habitat based on soil structure, preferring soils with higher clay content and bulk density than the average for the study area. It is also found in areas with a higher average for the total nitrogen content in comparison with the whole Australasia region.

Table 20 Specialisation explained by the first 5[$] of the 6 ecological factors retained and correlation coefficient symbols for marginality and specialistion for the most important 28[$] variables, for the Australasian region "All variable" model

Environmental Variable	Factor 1 (24%)	Factor 2 (24%)	Factor 3 (15%)	Factor 4 (7%)	Factor 5 (6%)
	Marginality§	Spec.1#	Spec. 2#	Spec. 3#	Spec.4#
C-TCdQ	+++	****	**	****	-0.09
C-TMnCdM	+++	*	*	0	****
C-TRn	- - -	***	**	*****	*****
C-TSn	- -	0	*	**	**
C-TDyQ	++	0	0	0	0
C-Tan	++	**	******	**	0
C-PWtM	++	0	0	**	0
C-TDR	- -	**	*	**	0
M-AISn	++	0	*	0	0
C-Tlso	+	0	0	*	0
M-AIMxQ	+	0	0	0	0
C-RHWtQ	+	**	0	0	*
C-PWtQ	+	*	0	*	0
C-PDyM	-	0	0	0	0
S-BDtop	+	0	0	0	0
C-PAn	+	0	0	*	0
S-BDsub	+	0	0	0	0
C-PSn	+	0	0	0	0
S-ClyPsub	+	0	0	0	0
M-AIMnM	-	0	0	0	0
M-AIMnQ	-	0	*	0	0
M-AICdQ	-	0	0	0	0
C-PDyQ	-	0	0	0	0
C-SDWtQ	-	*	0	0	0
C-RHWmQ	+	**	0	**	0
S-TNsub	+	*	0	**	0
C-SDCdQ	+	*	0	0	*
S-TNtop	+	0	0	*	0

§ The symbol + means that the species was found in locations with higher values than average. The symbol –means the reverse.
The greater the number of symbols, the higher the correlation (the more the species mean differs from the global mean). 0 indicates a very weak correlation. Each symbol is representive of a unit of 0.1 for the coefficient value (e.g. ++ = 0.20 – 0.29).

#The symbol * means the species was found occupying a narrower range of values than available. The greater the number of asterix, the narrower the range (the more the species variance differs from the global variance). 0 indicates a very low specialization. Each symbol is representative of a unit of 0.1 for the coefficient value (** = (+-0.20 – 0.29)).

$ See appendix for table with all variables used in this model, all 6 factors retained to produce suitability map

Climate: C-TAn Mean Annual Temperature(°C), C-TDR Mean Diurnal Range(°C), C-Tlso Isothermality(°C), C-TSn Temperature Seasonality, C-TMnCdM Minimum Temperature of Coldest Month(°C), C-TRn Temperature Range(°C), C-TDyQ Mean Temperature Driest Quarter(°C), C-TCdQ Mean Temperature Coldest Quarter(°C), C-PAn Annual Precipitation(mm) , C-PWtM Precipitation of Wettest Month (mm),C-PDyM Precipitation of Driest Month(mm), C-PSn - Precipitation Seasonality(mm), C-PWtQ Precipitation of Wettest Quarter(mm), C-PDyQ Precipitation of Driest Quarter(mm), C-RHWmQ Relative Humidity Warmest Quarter (%), C-RHWtQ Relative Humidity Wettest quarter(%), C-SDCdQ Mean Sunshine Duration Coldest Quarter(%), C-SDWtQ Mean Sunshine Duration Wettest Quarter(%). ***Moisture Availability:*** M-AICdQ Mean Moisture Availability Index of Coldest Quarter, M-AISn Moisture Availability Index Seasonality, M-AIMxQ Mean Moisture Availability Index of Maximum Quarter, M-AIMn Minimum Monthly Moisture Availability Index, M-AIMnQ Mean Moisture Availability Index Minimum Quarter. ***Soil:*** S-BDsub Bulk Density subsoil(g/cm^{-3}), S-Bdtop Bulk Density topsoil(g/cm^{-3}), S-ClyPsub Percentage Clay subsoil(%), S-TNsub Total nitrogen subsoil(% by weight), S-TNtop Total nitrogen topsoil(% by weight).

4.3.2.1.5 America's region "All variable" model

The regional model for "Americas" gave a marginality score 1.442, a specialisation 4.005 and a tolerance 0.25 indicating that tamarind habitat differed dramatically from that of the mean of the study area and was highly restricted in terms of niche breadth in this region.

The six factors retained to produce the suitability map explained 87% of the information 100% of marginality and 74% of the specialisation.

Table 21 (below) gives details of the specialisation explained and the marginality and specialisation coefficient values of variables for the first 5 of the 6 ecological factors retained. From these factor correlation coefficients (Table 21) it is possible to identify the variables on which tamarind showed high marginality and specialisation (the variables which are highly correlated with marginality and specialisation factors). These are the variables that the model uses to discriminate suitable from non-suitable habitat. The identified variables are listed below.

Soil: Unlike the other regions marginality factor was most closely correlated with soil characteristics, both relating to soil nutrition (Total nitrogen, Organic carbon), soil chemistry (Effective ECEC, pH subsoil), and structure (Percentage silt). Although these variables showed high marginality, they showed low specialisation within the region. Hence although tamarind prefers these conditions it is not restricted to them.

Water availability: Tamarind showed a preference for habitat with higher averages than found in the study area for for both precipitation and moisture availability index seasonality. It was also found to show a preference for habiat with higher than average water availability during the wet periods and a lower than average water availability during the dry periods.

Temperature: Within this region Tamarind showed highly positive marginaliy for mean annual temperature, minimum temperature coldest month and temperature during the coldest quarter. This Indicates that tamarind show a preference for habitat with warmer temperature than that found in the cold period on average in this region. A high specialisation coefficient for temperature coldest quarter showed indiacted tamarind is realtivily restricted to these warmer than average conditions during the cold periods. Unlike other regions, in the Americas tamarind showed a preference for higher than average for the study area mean temperature warmest quarter (although show no preference for maximum temperature warmest month). Tamarind as with all other regions showed preferences to lower than average temperature variation and showed high specialisation for these variables.

Table 21 Specialisation explained by the first 5$ of the 6 ecological factors retained and correlation coefficient symbols for marginality and specialisation for the most important 30$ variables, for the America's region "All variable" model

Environmental Variables	Factor 1 (7%) Marginality§	Factor 2 (30%) Spec.1#	Factor 3 (14%) Spec. 2#	Factor 4 (10%) Spec. 3#	Factor 5 (8%) Spec.4#
S-ECECsub	+++	0	0	0	0
S-ECECtop	+++	0	0	0	0
C-PSn	++	0	0	0	0
S-SltPtop	++	0	0	0	0
S-TNsub	++	**	*	0	0
S-pHsubsoil	++	0	0	*	0
C-TWtQ	+	0	**	*	0
S-BDtop	-	0	0	0	0
S-OCtop	+	0	0	0	0
C-Tan	+	0	******	*	****
S-SltPsub	+	0	0	0	0
S-pHtop	+	0	0	*	0
S-OCsub	+	*	*	**	*
C-TWmQ	+	*	0	0	**
TMnCdM	+	*	0	****	*****
M-AISn	+	0	0	0	0
C-PWtM	+	0	0	0	*
C-TCdQ	+	**	*****	**	**
C-TDyQ	+	0	0	0	0
S-TNtop	+	******	0	0	*
M-AIMnQ	-	*	0	0	0
C-SDWmQ	-	0	0	0	0
C-PWmQ	-	0	0	0	0
S-SndPtop	-	0	0	0	*
C-SDDyQ	-	0	0	0	0
C-PWtQ	-	0	0	0	0
C-TDR	-	*	0	0	0
C-TRn	-	***	**	******	****
C-TMxWmM	+	0	0	***	*
M-AIMn	-	0	0	0	0
C-PDyQ	-	0	0	0	0

§ The symbol + means that the vulture was found in locations with higher values than average. The symbol –means the reverse. The greater the number of symbols, the higher the correlation (the more the species mean differs from the global mean). 0 indicates a very weak correlation. Each symbol is representive of a unit of 0.1 for the coefficient value (e.g. ++ = 0.20 – 0.29).

#The symbol * means the vulture was found occupying a narrower range of values than available. The greater the number of asterix, the narrower the range (the more the species variance differs from the global variance). 0 indicates a very low specialization. Each symbol is representative of a unit of 0.1 for the coefficient value (** = (+-0.20 – 0.29)). $ See appendix for table with all variables used in this model, all 6 factors retained to produce suitability map.

Climate: C-TAn Mean Annual Temperature(°C), C-TDR Mean Diurnal Range(°C), C-TMxWmM Maximum Temperature of Warmest Month(°C), C-TMnCdM Minimum Temperature of Coldest Month(°C), C-TRn Temperature Range(°C), C-TWtQ Mean Temperature Wettest Quarter(°C), C-TDyQ Mean Temperature Driest Quarter(°C), C-TWmQ Mean Temperature Warmest Quarter(°C), C-TCdQ Mean Temperature Coldest Quarter(°C), C-PWtM Precipitation of Wettest Month (mm),C-PSn - Precipitation Seasonality(mm), C-PWtQ Precipitation of Wettest Quarter(mm), C-PdyQ Precipitation of Driest Quarter(mm), C-PWmQ Precipitation of Warmest Quarter(mm), C-SDDyQ Mean Sunshine Duration Driest Quarter(%), C-SDWmQ Mean Sunshine Duration Warmest Quarter(%). ***Moisture Availability:*** M-AISn Moisture Availability Index Seasonality, M-AIMxQ Mean Moisture Availability Index of Maximum Quarter, M-AIMn Minimum Monthly Moisture Availability Index, M-AIMnQ Mean Moisture Availability Index Minimum Quarter. ***Soil:*** S-BDtop Bulk Density topsoil(g/cm^{-3}), S-SndPtop Percentage Sand topsoil(%), S-SltPsub Percentage Silt subsoil(%),S-SltPtop Percentage Silt topsoil(%),S-ECECtop Effective CEC topsoil (cmolc kg^{-1}), S-ECECsub Effective CEC subsoil(cmolc kg^{-1}), S-OCsub Organic Carbon content subsoil(% by weight), S-OCtop Organic Carbon content topsoil(% by weight), S-pHsub pH subsoil, S-pHtop pH topsoil, S-TNsub Total nitrogen subsoil(% by weight), S-TNtop Total nitrogen topsoil(% by weight).

4.3.2.2 Reduced variable models

The reduced variable models found the similar variables to be highly correlated with marginality and specialisation. Although not all variables were included, the relative rank of variables with the same ecological meaning remained the same. The variables which were included in both "all variable" models and "reduced variable" models had higher loading coefficients on the "reduced variable" ENFA. This indicated they were more closely fitted by the marginality and specialisation factors in the "reduced variable" ENFA.

4.3.2.2.1 Global "Reduced Variable" model

The global model for tamarind showed a marginality of 0.817, a specialization of 1.625 and a tolerance 0.615. This showed that tamarind habitat differs quite a lot from the mean conditions across the study area.

The five factors retained to produce the suitability map for the global model explained 83% of the information (the total sum of the eigenvalues), 100 % of marginality and 67% of the specialisation

Table 22 (below) gives details of the specialisation explained and the marginality and specialisation coefficients of variables for the 5 ecological factors retained.

From these correlation coefficients (Table 22) it is possible to identify the variables on which tamarind showed high marginalisation and specialisation (which variables were highly correlated with marginality and specialisation factors). These are the variables that the model used to discriminate suitable from non-suitable habitat. The identified variables are listed below.

Temperature: A high negative coefficient on the marginality factor showed tamarind habitat to have a lower averages than the study areas for temperature variation both in terms of daily range (diurnal range) and variation across the year (temperature seasonality). Based on the speciality factors tamarind distribution has a restricted range and showed narrow niche width in relation to minimum and maximum temperatures along with temperature variation. Tamarind shows a higher average minimum temperatures and lower average maximum temperature in comparison with the study area.

Water availability: The marginality factor showed an association of tamarind with higher averages than the study areas in terms of water availability both in relation to moisture availability and precipitation, during wettest periods (moisture availability index wettest quarter, precipitation wettest quarter) and variation in moisture availability (moisture availability index seasonality). It also showed a preference for habitat with a higher average than that found in the study area for relative humidity wettest quarter.

Soil: Tamarind habitat showed a positive correlation with marginality for total nitrogen subsoil.

Table 22 Specialisation explained by the 5 factors retained and correlation coefficient symbols for marginality and specialisation for the most important 14$^{\$}$ variables, for the Global "Reduced variable" model.

Environmental Variable	Factor 1 (20%) Marginality	Factor 2 (19%) Spec. 1	Factor 3 (12%) Spec. 2	Factor 4 (9%) Spec. 3	Factor 5 (7%) Spec. 4
C-TMnCdM	++++	*******	*******	*****	0
M-AIMxQ	+++	0	***	*	0
C-PWtQ	+++	0	0	0.01	0
C-TSn	- - -	*****	***	****	*****
C-TDR	- - -	*	***	***	*
C-RHWtQ	+++	0	*	0	0
M-AISn	++	0	0	0	*
S-TNsub	++	0	0	0	0
S-ECECsub	++	0	0	0	*
C-SDWtQ	-	*	0	*	0
C-TWtQ	+	*****	0	0	****
S-SndPsub	-	0	0	0	0
C-TMxWmM	-	0	***	*****	******
C-SDWmQ	-	0	0	0	**

§ The symbol + means that the vulture was found in locations with higher values than average. The symbol –means the reverse. The greater the number of symbols, the higher the correlation (the more the species mean differs from the global mean). 0 indicates a very weak correlation. Each symbol is representive of a unit of 0.1 for the coefficient value (e.g. ++ = 0.20 – 0.29).

#The symbol * means the vulture was found occupying a narrower range of values than available. The greater the number of asterix, the narrower the range (the more the species variance differs from the global variance). 0 indicates a very low specialization. Each symbol is representative of a unit of 0.1 for the coefficient value (** = (0.20 – 0.29)).

$ See appendix for table with correlation coefficients for all variables used in this model

Climate: C-TDR Mean Diurnal Range(°C), C-TSn Temperature Seasonality, C-TMxWmM Maximum Temperature of Warmest Month(°C), C-TMnCdM Minimum Temperature of Coldest Month(°C), C-TWtQ Mean Temperature Wettest Quarter(°C), C-PWtQ Precipitation of Wettest Quarter(mm), C-RHWtQ Relative Humidity Wettest quarter(%), C-SDWmQ Mean Sunshine Duration Warmest Quarter(%), C-SDWtQ Mean Sunshine Duration Wettest Quarter(%). ***Moisture Availability:*** M-AISn Moisture Availability Index Seasonality, M-AIMxQ Mean Moisture Availability Index of Maximum Quarter. ***Soil:*** S-SndPsub Percentage Sand subsoil(%),S-ECECsub Effective CEC subsoil(cmolc kg^{-1}), S-TNsub Total nitrogen subsoil(% by weight).

4.3.2.2.2 Africa region "Reduced Variable" model

The regional model for Africa gives a marginality value of 1.147, a specialisation of 1.832 and tolerance 0.546. This showed that tamarind habitat differs drastically from the mean conditions in the African region.

The five factors retained to produce the suitability map for the global model explained 85% of the information (the total sum of the eigenvalues) 100% of marginality and 69% of the specialisation.

Table 23 (below) gives details of the specialisation explained and the marginality and specialisation coefficient values of variables for the 5 ecological factors retained. From these factor correlation coefficients (
Table 23), it is possible to identify the variables on which tamarind showed high marginalisation and specialisation (which variables were highly correlated with marginality and specialisation factors). These are the variables that the model uses to discriminate suitable from non-suitable habitat. The identified variables are listed below.

Temperature: A high negative coefficient value on the marginality factor showed tamarind habitat to have a lower average than the study areas for temperature variation in relation to diurnal range and temperature seasonality. Tamarind habitat was also shown to have higher average temperatures during the coldest periods (Minimum temperature coldest month). Based on the specialisation factors tamarind distribution appears to be restricted and shows narrow niche breadth in relation to minimum temperatures along with variation in temperature.

Water Availability: High marginality coefficients for water availability during the wettest periods (mean moisture availability index maximum quarter) and total yearly rainfall (annual precipitation), indicated tamarind prefers wetter conditions than those found on average in the study area both during the wet season and throughout the year.

Soil: A high marginality coefficient for organic carbon and total nitrogen indicates that tamarind shows a preference for soils likely to have high soil nutrient content.

Table 23 Specialisation explained by the 5 factors retained and correlation coefficient symbols for marginality and specialisation for the most important 19$^{\$}$ variables, for the Africa region "Reduced variable" model.

Environmental Variable	Factor 1 (21%) Marginality	Factor 2 (27%) Spec.1	Factor 3 (8%) Spec. 2	Factor 4 (7%) Spec. 3	Factor (5%) Spec. 4
C-TDR	- - -	*	*	*	0
C-RHWtQ	+++	*	**	***	0
C-TMnCdM	+++	*****	**	***	*
M-AIMxQ	+++	0	0	***	****
C-TSn	- - -	****	**	**	0
C-PAn	+++	0	0	0	****
S-OCtop	++	0	*	**	0
C-RHCdQ	++	0	0	*	*
S-TNtop	++	*	**	0	0
C-RHDyQ	++	*	***	0	0
C-SDWtQ	- -	0	***	0	0
S-pHsub	-	0	*	**	**
S-pHtop	-	0	0	**	*****
C-SDCdQ	-	0	0	***	**
S-SndPsub	-	0	***	0	0
C-SDdyQ	-	0	*	**	0
C-SDWmQ	-	0	0	**	**
C-TWtQ	0	******	**	*	*

§ The symbol + means that the vulture was found in locations with higher values than average. The symbol –means the reverse.
The greater the number of symbols, the higher the correlation (the more the species mean differs from the global mean). 0 indicates a very weak correlation. Each symbol is representive of a unit of 0.1 for the coefficient value (e.g. ++ = 0.20 – 0.29).

#The symbol * means the vulture was found occupying a narrower range of values than available. The greater the number of asterix, the narrower the range (the more the species variance differs from the global variance). 0 indicates a very low specialization. Each symbol is representative of a unit of 0.1 for the coefficient value (** = (0.20 – 0.29)).

$ See appendix for table with correlation coefficients for all variables used in this model

Climate: C-TDR Mean Diurnal Range(°C), C-TSn Temperature Seasonality, C-TMnCdM Minimum Temperature of Coldest Month(°C), C-TWtQ Mean Temperature Wettest Quarter(°C), C-PAn Annual Precipitation(mm), C-RHCdQ Relative Humidity of Coldest Quarter(%), C-RHDyQ Relative Humidity of Driest Quarter, C-RHWtQ Relative Humidity Wettest quarter(%), C-SDCdQ Mean Sunshine Duration Coldest Quarter(%), C-SDDyQ Mean Sunshine Duration Driest Quarter(%), C-SDWmQ Mean Sunshine Duration Warmest Quarter(%), C-SDWtQ Mean Sunshine Duration Wettest Quarter(%).
Moisture Availability: M-AIMxQ Mean Moisture Availability Index of Maximum Quarter.
Soil: S-SndPsub Percentage Sand subsoil(%), S-OCtop Organic Carbon content topsoil(% by weight), S-pHsub pH subsoil, S-pHtop pH topsoil, S-TNtop Total nitrogen topsoil(% by weight).

4.3.2.2.3 Australasia region "Reduced variable" model

The regional model for Australasia gave a marginality score of 0.819, a specialisation of 2.366 and a tolerance of 0.423. This indicated that tamarind habitat differed from that of the mean, although was not particularly restricted in the study area.

Four factors were retained to produce the suitability map. These factors explained 88% of the information, 100% of the marginality and 76% of the specialisation.

Table 24 (below) gives details of the specialisation explained and the marginality and specialisation coefficient values of variables for the 4 ecological factors retained. From these factor correlation coefficients (Table 24), it is possible to identify the variables on which tamarind showed high marginalisation and specialisation (the variables which are highly correlated with marginality and specialisation factors). These are the variables that the model uses to discriminate suitable from non-suitable habitat. The identified variables are listed below.

Temperature: Tamarind showed a preference for higher temperatures than those of the mean of the study area particularly during the cold periods (Minimum temperature coldest month and Annual temperature)

Water availability: Tamarind shows a preference for a higher average water availability during the wettest part of the year (Moisture availability maximum quarter, Relative humidity wettest quarter) than found in the whole study area. It also shows a preference for higher average conditions for annual precipitation. During the driest period of the year it found in regions with a lower average rainfall than foind in the whole study area.

Table 24 Specialisation explained by the 4 factors retained and correlation coefficient symbols for marginality and specialisation for the most important 10[$] variables, for the Australasian region "Reduced variable" model.

Environmental Variable	Factor 1 (24%) Marginality	Factor 2 (32%) Spec.1	Factor 3 (12%) Spec. 2	Factor 4 (8%) Spec. 3
C-TMnCdM	+++++	**	*****	*
C-TAn	++++	******	***	**
M-AIMxQ	+++	*	***	*
C-RHWtQ	+++	**	*	***
C-PAn	++	0	0	0
S-BDsub	++	0	**	****
S-ClyPsub	++	0	***	*
C-PDyQ	- -	0	0	*
S-OCsub	+	0	**	****
C-TWtQ	0	******	***	*

§ The symbol + means that the vulture was found in locations with higher values than average. The symbol –means the reverse.
The greater the number of symbols, the higher the correlation (the more the species mean differs from the global mean). 0 indicates a very weak correlation. Each symbol is representive of a unit of 0.1 for the coefficient value (e.g. ++ ≈ 0.20 – 0.29).

#The symbol * means the vulture was found occupying a narrower range of values than available. The greater the number of asterix, the narrower the range (the more the species variance differs from the global variance). 0 indicates a very low specialization. Each symbol is representative of a unit of 0.1 for the coefficient value (** ≈ (0.20 – 0.29)).

[$] See appendix for table with correlation coefficients for all variables used in this model

Climate: C-TAn Mean Annual Temperature(°C), C-TMnCdM Minimum Temperature of Coldest Month(°C), C-TWtQ Mean Temperature Wettest Quarter(°C), C-PAn Annual Precipitation(mm), C-PDyQ Precipitation of Driest Quarter(mm), C-RHWtQ Relative Humidity Wettest quarter(%).
Moisture Availability: M-AIMxQ Mean Moisture Availability Index of Maximum Quarter
Soil: S-BDsub Bulk Density subsoil(g/cm^{-3}), S-ClyPsub Percentage Clay subsoil(%),S-OCsub Organic Carbon content subsoil(% by weight).

4.3.2.2.4 Americas region "Reduced variable" model

The regional model for America gave a marginality score of 1.147, a specialisation of 1.832 and a tolerance of 0.546. This indicated that tamarind habitat differed from that of the mean, although it was not particularly restricted in the study area.

Four factors were retained to produce the suitability map. These factors explained 81% of the information, 100% of marginality and 63% of the specialisation.

Table 25 gives details of the specialisation explained and the marginality and specialisation coefficient values of variables for the 4 ecological factors retained. From these factor correlation coefficients (Table 25), it is possible to identify the variables on which tamarind showed high marginalisation and specialisation (the variables which are highly correlated with marginality and specialisation factors). These are the variables that the model used to discriminate suitable from non-suitable habitat. The identified variables are listed below.

Soil: The Marginality factor was most closely correlated with soil characteristics, both relating to chemical characteristics (pH subsoil) and structure (Percentage silt). Although these variables showed high marginality, they were not very specialised within the region.

Water availability: The Marginality factor is also highly positively correlation with precipitation for the wettest and warmest quarters, indicating a preference for a higher average for these variables than found in the study area.

Temperature: High marginality coefficients for mean temperature coldest quarter, maximum temperature warmest month, indicates that tamarind shows a preference for conditions of higher temperature than those found in the study area. The specialisation factors show tamarind niche width to be restricted by minimum temperatures (Minimum temperature coldest month) and temperature variation (temperature range).

Table 25 Specialisation explained by the 4 factors retained and correlation coefficient symbols for marginality and specialisation for the most important 12[$] variables, for the America's region "Reduced variable" model.

Environmental Variable	Factor 1 (5%) Marginality	Factor 2 (30%) Spec.1	Factor 3 (18%) Spec. 2	Factor 4 (9%) Spec. 3
S-pHsub	++++	0	0	0
S-SltPsub	++++	0	0	0
C-TMnCdM	+++	******	******	******
C-PWmQ	++	0	0	0
C-PWtQ	++	0	0	0
C-TRn	- -	******	******	******
C-TMxWmM	++	***	***	**
C-PCdQ	- -	0	0	0
S-BDsub	-	0	0	0
M-AIWmQ	-	0	0	0
S-ClyPsub	-	0	0	0
S-SndPsub	-	0	0	0
C-RHWtQ	+	0	0	**
C-RHcdQ	+	0	0	**

§ The symbol + means that the vulture was found in locations with higher values than average. The symbol –means the reverse. The greater the number of symbols, the higher the correlation (the more the species mean differs from the global mean). 0 indicates a very weak correlation. Each symbol is representive of a unit of 0.1 for the coefficient value (e.g. ++ = 0.20 – 0.29).

#The symbol * means the vulture was found occupying a narrower range of values than available. The greater the number of asterix, the narrower the range (the more the species variance differs from the global variance). 0 indicates a very low specialization. Each symbol is representative of a unit of 0.1 for the coefficient value (** = (0.20 – 0.29)).

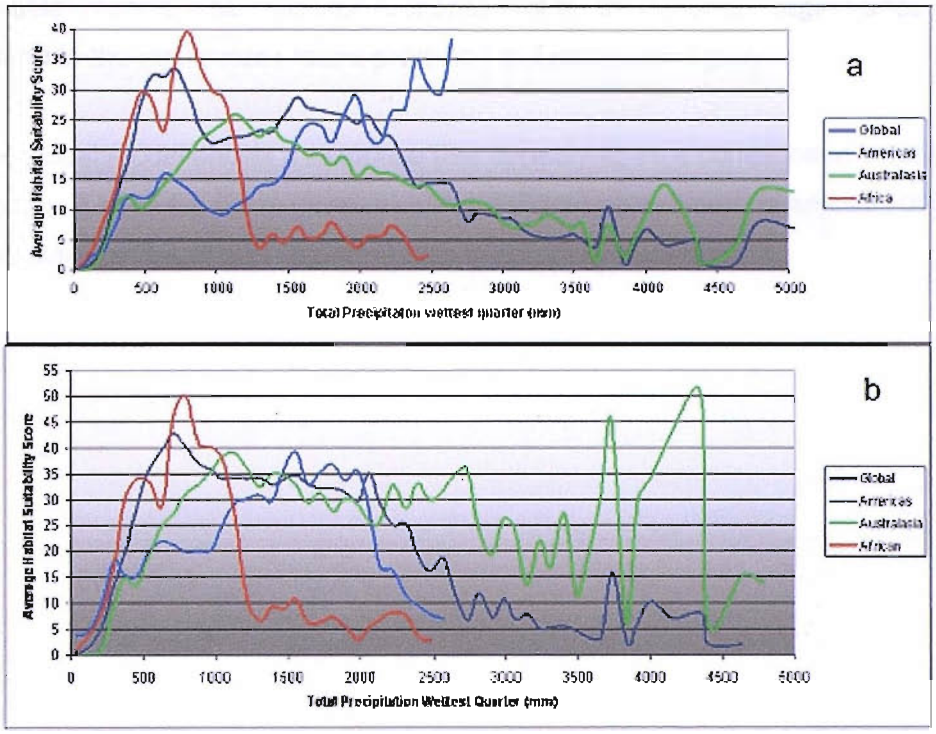
$ See appendix for table with correlation coeffeicentsfor all variables used in this model.

Climate: C-TMxWmM Maximum Temperature of Warmest Month(°C), C-TMnCdM Minimum Temperature of Coldest Month(°C), C-TRn Temperature Range(°C), C-PWtQ Precipitation of Wettest Quarter(mm), C-PWmQ Precipitation of Warmest Quarter(mm), C-PCdQ Precipitation of Coldest Quarter(mm), C-RHCdQ Relative Humidity of Coldest Quarter(%), C-RHWtQ Relative Humidity Wettest quarter(%).
Moisture Availability: M-AIWmQ Mean Moisture Availability Index of Warmest Quarter.
Soil: S-BDsub Bulk Density subsoil(g/cm^{-3}), S-ClyPsub Percentage Clay subsoil(%), S-SndPsub Percentage Sand subsoil(%),S-SltPsub Percentage Silt subsoil(%), S-pHsub pH subsoil.

4.3.2.3 Variable response plots

By plotting predicted suitability scores from the suitability maps (see section 4.3.4) against environmental variable value from the environmental datasets, it is possible to derive a **predicted univariate response plots** for each region. This provides an indication about the predicted relationship between the environmental variable and suitability of the species. Predicted response plots were created by classifying global environmental variables into 40 or 60 equal interval classes and plotting this against the average suitability score for each class. Graphs are plotted based on the mean value of the environmental variable for all cells in each class and average suitability score within that class. It should be noted that it is the general shape of the curve that is of importantance. Low number of grid cells in classes at the end of the range can lead to high variation causing fluctuation in the tails of the curves, these variations should not be considered. Response plots for “all variable” models and “reduced variable” models showed similar response patterns for most variables.

Total precipitation coldest quarter



Figures 25 a and b Response plot for total precipitation wettest quarter for (a) all variable model and (b) reduced variable model.

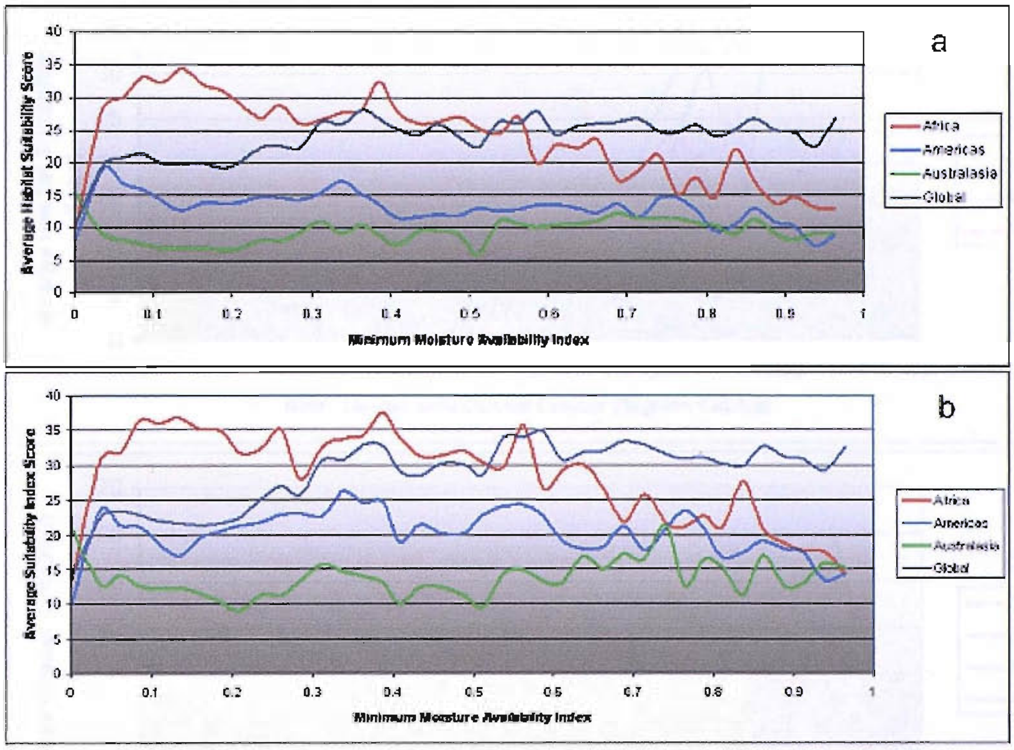
Figures 25 a and b show high variation in habitat suitability response between regions in relation to total precipitation wettest quarter. In Africa tamarind shows a very restricted range of suitable conditions, increasing to its optimum of 750mm before decreasing rapidly to a very low suitability at approximately 1250mm.

In the Australia region tamarind shows a much less restricted range, with a gradual increase to an optimum of approximately 1200 mm (close to the edge of the limit for tamarind found in Africa) before gradually declining, reaching low suitability beyond 3000mm.

For the Americas the response plot shows predicted suitability increasing up to around 2000 mm for both "all variable" and "reduced variable" model, however beyond 2000 mm the response curve for the "all variable" model (Figure 26a), continues to show a rise in suitability, where as the reduced variable model response surve shows a decline in suitability. This is most likely due to a difference in interactions caused by the additional variable included in the “all variable” models. The predicted response plot for the Americas region does not extend to the high values found predicted in Australasia region.

The global model shows a response that appears to be a combination of all the regions. It shows a sharp increase in suitability to an optimum of approx 700mm plataeuing until just over 2000mm before gradually declining.

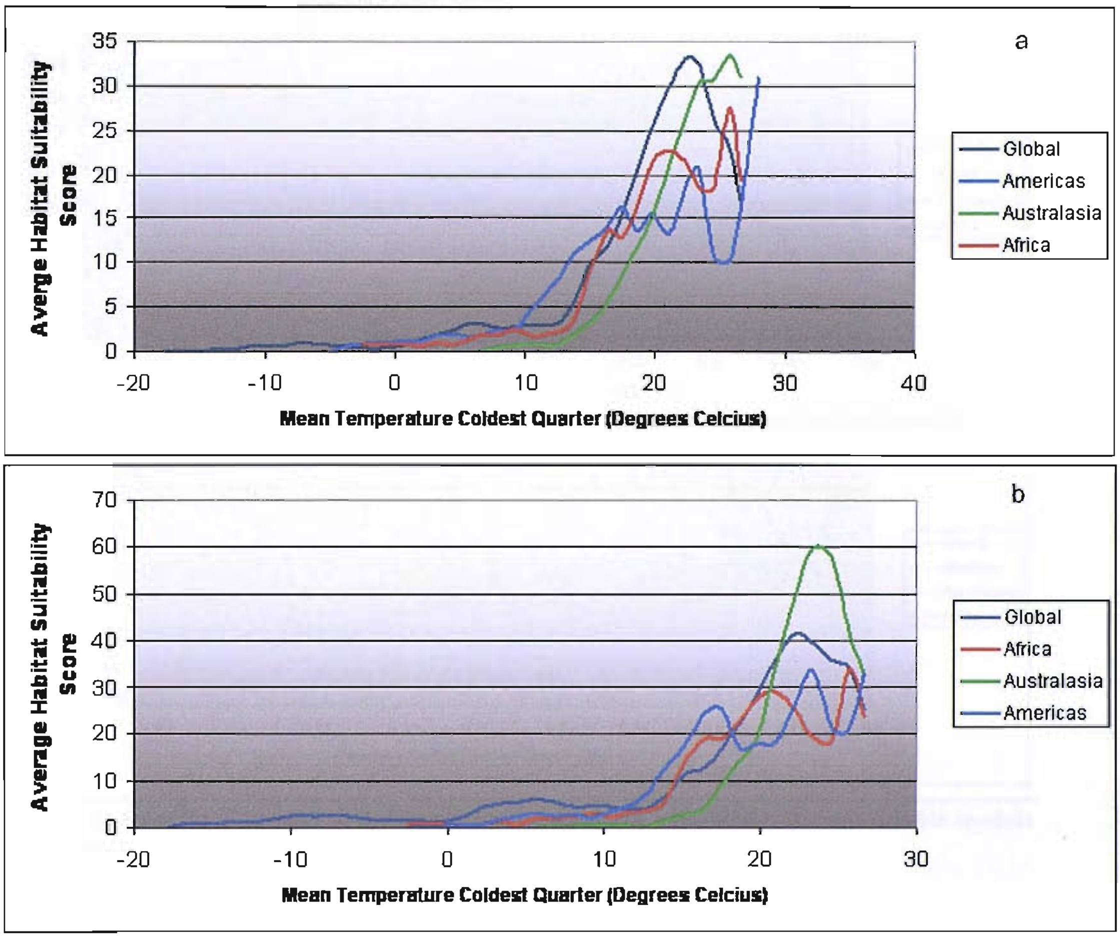
Minimum moisture availability index



Figures 26 a and b; Response plot for minimum moisture availability index (a) all variable model and (b) reduced variable model

Figures 26 a and b show the predicted response plot of tamarind to **minimum moisture availability index** for all regions, for all variable and reduced variable models. Both in the Americas and Africa region tamarind suitability gradually declines as minimum moisture availability index increases, though this decline seems to be more pronounced in Africa. In the Australasia region tamarind suitability shows little variation with the increase in minimum moisture availability index. The Global model shows an increase in suitability with increasing minimum moisture availability before plauteauing at around 0.3.

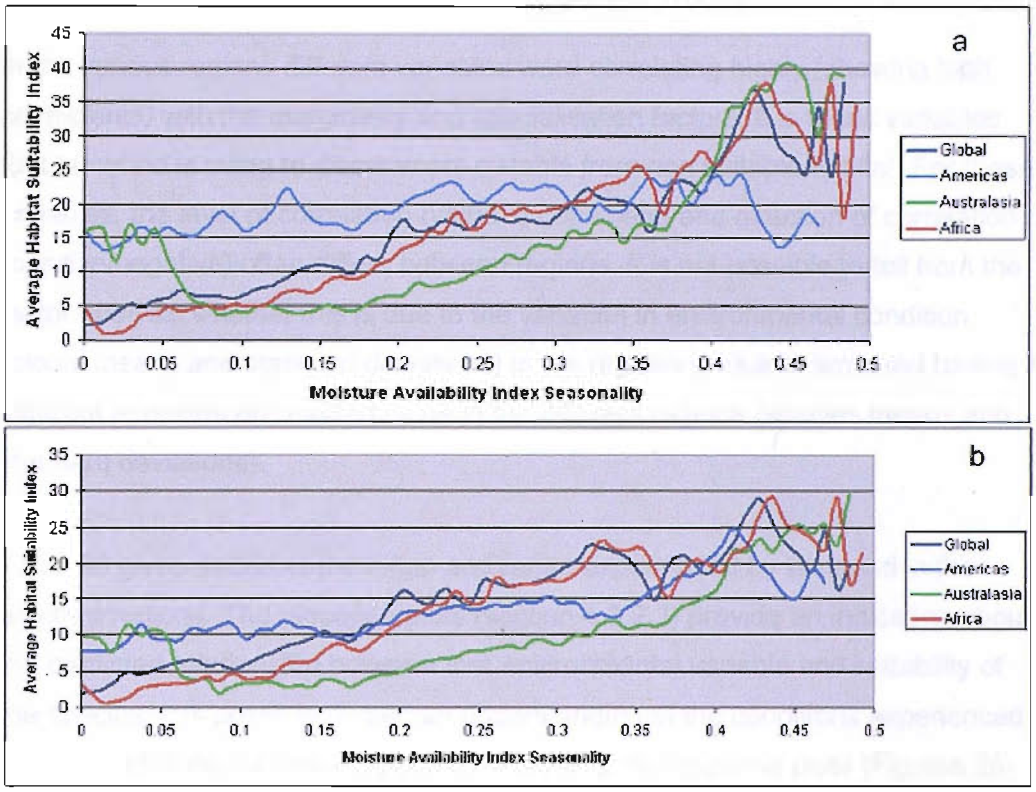
Mean temperature coldest quarter



Figures 27 a and b; Response plot for mean temperature coldest quarter (a) all variable model (b) reduced variable model

Figures 27 a and b show the predicted response of tamarind suitability to mean temperature coldest quarter for all regions. Tamarind showed a similar response for all regions, only showing suitability above 10 °C and declining or plateauing above approx 20 °C.

Moisture availability index seasonality



Figures 28 a and b; Response plot for moisture availability seasonality (a) all variable models (b) reduced variable models

Figures 28 a and b show tamarind predicted response to *moisture availability seasonality*. A similar response is observed for all regions, and between regional and the global models. For all regions and at the global extent, tamarind suitability is increasing with moisture availability index seasonality. Although in the Americas the curve has a lower gradient indicating a less pronounced change.

4.3.3 Characterisation of the regional niche

In the various regions different variables were correlating highly (showing high coefficients) with the marginality and specialisation factors. It is these variables that tamarind is using to discriminate suitable from non suitable habitat. For these variables, the level of correlation (value of coffecient) and direction of correlation (positive/negative) often differs between regions. It is not possible to tell from the factor analysis whether this is due to the variation in environmental condition (global means and standard deviations) in the regions or due to tamarind having a different experienced/realised niche in the different regions (species means and standard deviations).

Table 26 gives details of the mean and range experienced by tamarind in the different regions. The response plots (section 4.3.2.3) provide an indication about the predicted relationship between that environmental variable and suitability of the species. It is possible to gain an understanding of the conditions experienced by tamarind in the different regions by analysing the response plots (Figures 25-Figures 28), mean and range experienced (Table 26) by tamarind in the different regions and correlation coffecients from the factor analysis (Table 18 - Table 25).

Table 26 Comparison of the regional environmental variable mean and ranges on for tamarind occurrence points

	Mean			Range		
Environmental Variable	Africa	Australasia	Americas	Africa	Australasia	America
C-TAn	24.4	25.1	24.6	18.2-28.6	20.5 - 27.9	15.8 - 28
C-TDR	11.9	10.1	11.2	73 - 96	58 - 143	65 - 169
C-TIso	6.6	6.0	6.7	42 - 86	4.2 - 8.9	48 - 88
C-TSn	1683.0	2135.2	1755.7	310 - 4660	269 - 4717	363 - 4627
C-TMxWmM	33.4	33.6	33.0	25.9 - 40.2	27.7 - 41.2	271 - 399
C-TMnCdM	15.2	16.0	16.3	4.6 - 22.4	7.4 - 24.1	4.2 - 21.1
C-TRn	18.2	17.6	16.8	10.7 - 30.5	7.5 - 29.8	9.6 - 25.9
C-TWtQ	24.6	25.9	25.9	14.2-28.2	21.6 - 31.1	17.6 - 28.9
C-TDyQ	23.1	23.6	23.7	14.8-28.7	15.5 - 29	13.6 - 28.8
C-TWmQ	26.5	27.7	26.6	19.8 - 32.5	23.3 - 32.6	18 - 30.9
C-TCdQ	22.2	22.3	22.4	14.2 - 26.5	14.7 - 27	13.5 - 26.8
C-PAn	1058	1523	1532	32 - 2689	346 - 3583	337 - 3817
C-PWtM	212	309	276	9 - 487	56 - 1373	86 - 753
C-PDyM	7	25	27	0-53	0 - 170	0 - 190
C-PSn	83	80	74	35 - 148	14 - 151	10 - 111
C-PWtQ	537	757	714	15 - 1260	155 - 2902	198 - 1971
C-PDyQ	31	92	97	0 - 220	0 - 548	0 - 599
C-PCdQ	181	183	178	0 - 1260	2 - 1829	11 - 765
C-PWmQ	223	376	406	3 - 788	69 - 1161	62 - 1240
C-RHCdQ	66.54	63.64	75.55	19.63-84.8	30.4 - 86	53.8 - 90.3
C-RHDyQ	56.93	59.38	72.56	17.7-78.8	30.4 - 84	44.5 - 89.4
C-RHWmQ	60.51	64.33	73.95	30.8-83.6	31.8 - 84.3	60 .1 - 90.7
C-RHWtQ	75.57	76.00	78.53	61.4-84.6	36.2 - 86	64.5 - 90.7
C-SDCdQ	65.22	69.76	54.61	26.8-87.7	38.3- 91.4	18.3 - 73.3
C-SDDyQ	72.46	73.29	57.28	49.3-88.6	46.2 - 93.6	18.9 - 76.5
C-SDWmQ	64.58	63.47	51.22	48.5-85.4	37.2 - 75.5	21.8 - 71.6
C-SDWtQ	53.20	46.96	47.72	26.8-80.7	24.9 - 75.5	21.8 - 71.5
M_AIAn	0.667	0.722	0.836	0.03 - 0.997	0.276 - 1	0.384 - 1
M-AISn	58.917	49.285	31.531	0.8 - 122	0 - 100	0 - 107.5
M-AICdQ	0.509	0.530	0.803	0-1	0.01 - 1	0.036 - 1
M-AIWmQ	0.663	0.751	0.823	0.001-1	0.276 - 1	0.236 - 1
M-AIMn	0.128	0.249	0.444	0 - 0.97143	0 - 1	0.404 - 1
M-AIMx	0.981	0.996	1.000	0.1 - 1	0.7 - 1	1 - 1
M-AIMnQ	0.197	0.304	0.517	0 - 0.991	0 - 1	0.012 - 1
M-AIMxQ	0.971	0.991	1.000	0.06 - 1	0.502 - 1	0.990 - 1
S-BDsub	1.44	1.48	1.32	1.16 - 1.76	1.25 - 1.76	0.8 - 1.58
S-BDtop	1.39	1.44	1.27	1.1 - 1.65	1.22 - 1.65	0.76 - 1.54
S-Clysub	29.20	33.95	32.04	3.75 - 68	7 - 68	5 - 68
S-Clytop	24.69	25.10	28.86	5 - 60	6 - 56	4 - 57
S-ECECsub	15.57	16.72	21.26	2 - 49	2 - 49	4 - 59
S-ECECtop	14.82	15.36	20.50	3 - 50	3 - 50	5 - 55
S-GrPsub	10.31	8.54	8.09	3 - 41	3 - 27	3 - 23
S-GrPtop	9.95	8.32	7.97	1 - 27	1 - 35	1 - 35
S-OCsub	0.416	0.392	0.661	0.2 - 1.12	0.2 - 0.69	0.26 - 1.9
S-OCtop	0.993	0.864	1.672	0.35 - 2.72	0.4 - 1.74	0.41 - 7
S-pHsub	6.31	6.10	6.51	4.6 - 8.3	4.7 - 8.5	4.8 - 8.9
S-pHtop	6.32	6.03	6.31	4.7 - 8.3	4.6 - 8.125	4.4 - 9
S-SltPsub	21.45	22.98	26.89	4 - 45	Apr-38	6 - 53
S-SltPtop	25.39	26.91	32.31	5 - 47	May-43	10 - 55
S-SndPsub	41.74	39.50	35.61	3 - 89	4.25 - 89	4.5 - 86
S-SndPtop	49.97	48.12	38.97	16 - 90	16 - 90	7 - 87
S-TNsub	0.0502	0.0478	0.0715	0.02 - 0.11	0.03 - 0.08	0.03 - 0.18
S-TNtop	0.0927	0.0885	0.1425	0.03 - 0.28	0.04 - 0.18	0.05 - 0.34

In the **Americas and Australasia** regions tamarind shows high a preference for higher annual **temperatures** and average temperatures during the driest period than found in these regions. In the **Americas** it showed a preference for higher temperature during the warmest periods. In the **African** region tamarind showed a preference for lower maximum temperatures during the warmest periods. While the different regions show variation in terms of marginality for these varaiables Table 26 and Figures 27 indicate that tamarind experienced very similar realised niche across all regions in terms of temperature. It is therefore due to the varying temperature conditions (differing global averages) in the regions that cause tamarind to respond differently between regional models in terms of marginality and specialisation.

In all regions tamarind is discriminating to select higher **water availability during the wet periods** than found on average in the study areas. In **Africa** tamarind showed some specialisation in terms of water availability, particularly during the wettest period this may be due to the drier conditions found in Africa. Although tamarind shows a preference for higher than average water availability in Africa, it is still experiencing a lower average precipitation (annual precipitation (1058mm), precipitation wettest month (212mm) and precipitation wettest quarter (536mm)) than at the global extent (1333, 257,650mm respectively) in Australasia (1523,309,757mm respectively) and the Americas (1532, 276, 713mm respectively). This lower precipitation corresponds to lower mean annual moisture availability index experienced by tamarind in the African region (0.67) in comparison to globally (0.73) in the Australasian region (0.73) and Americas region (0.83). The response plot for precipitation wettest quarter shows a much more restricted range for tamarind in Africa in comparison to the other regions.

Tamarind shows a preference for lower than average water availability during the dry periods in **Australasia and the Americas** where precipitation is relativily high and evenly distributed. Although discriminating for drier than average areas during the dry periods in these regions it is still expereriancing considerable higher water availability during these periods than in Africa.
It is only in **Africa** where tamarind favours higher than average precipitation and moisture availability during the dry season. However the response plot for minimum moisture availability shows that in relation to the other regions, high

water availability during the driest periods have a more negative effect on suitability. Indicating that although tamarind does not favours wet conditions during the dry periods of the year; average conditions across Africa become too dry during this period.

In the Americas region tamarind is experiencing much higher water availability during dry periods, than the other regions. Though the amount of precipitation experienced by tamarind is not higher than in Australasian region; the higher relative humidity during these periods, combined with a finer soil structure and more cloud cover (less sunshine), results in less evapotranspiration.Therefore MAI minimum quarter does not fall below 0.4. Although the Australasian region receives as much precipitation as in the Americas, a more coarse soil structure, more sunshine with lower humidity results in tamarind experiencing a lower moisture availability particularly during the drier periods; (although still considerably higher than in the Africa).

Tamarind shows a preference for potentailly high nutrient soil conditions in all regions; although does not appear to be restricted to these conditions. In the Americas region soil variables show the highest marginality coefficients. In this region the global averages for soil nutritional variables are lower than the Australasia or African regions, yet the species inhabits regions of higher soil nutrition conditions in comparison to other regions.

In all regions tamarind was highly correlated with specialisation in terms of temperature variation, and minimum temperatures (Table 18 - Table 25), indicating that it was highly restricted in terms of its distribution in relation to these variables. In all regions tamarind show a preference for higher annual temperatures during the cold periods of the year. A similar response was seen for all regions with suitability response to minimum temperature coldest month.

4.3.4 Suitability maps

Two sets of suitability maps are presented as outputs; those from the "all variable" models and those from the "reduced variable" models. Maps are classified based on a suitability index from 0 – 100. Cells with a suitability index of 0 -10 were grouped with the no data (habitat identified as unsuitable by the model or not included in analysis as no data was available for one or more of the predictor environmental variables (i.e water bodies)), as the score was so low it is highly unlikely the species could exist in these conditions.

4.3.4.1 "All Variable" suitability maps

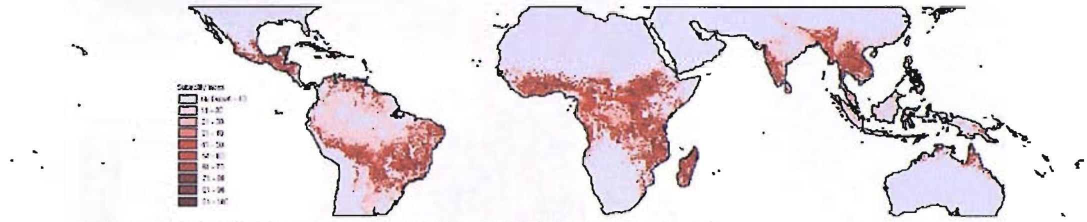


Figure 29 Habitat suitability map for global "all variable" model

The global habitat suitability map for the "all variable model" (Figure 29) shows a high suitability scores to be predicted across the tropics, predominantly in the wet-dry tropical zones. In Africa tamarind is predicted as suitable in the wet tropical zones of the Congo basin; however the suitability score here is lower. Suitable habitat is also identified in the humid highland conditions of Sudan.

In the Americas tamarind is predicted as suitable in the wet-dry tropical climates of South America. It is not identified as suitable in wet tropical zones of the Amazon basin. In much of the wet-dry tropics of Central America (north of Costa Rica), it is identified as highly suitable and the predicted distribution extends into the more humid conditions of central Mexico. The northern range of the predicted suitability in Central America appears to be limited by the northern desert regions of Mexico. The map shows a high suitability score predicted for the wet-dry tropical climate found in the Caribbean region, and the northern coast of Venezuela.

In Asia high suitability scores are predicted in the wet –dry tropical areas of south western regions and north eastern regions of the Indian Subcontinent and South East Asia. However the wet tropics of the East Indies show low suitability scores. The wet-dry tropics of Northern Australia were also predicted with high suitability scores.

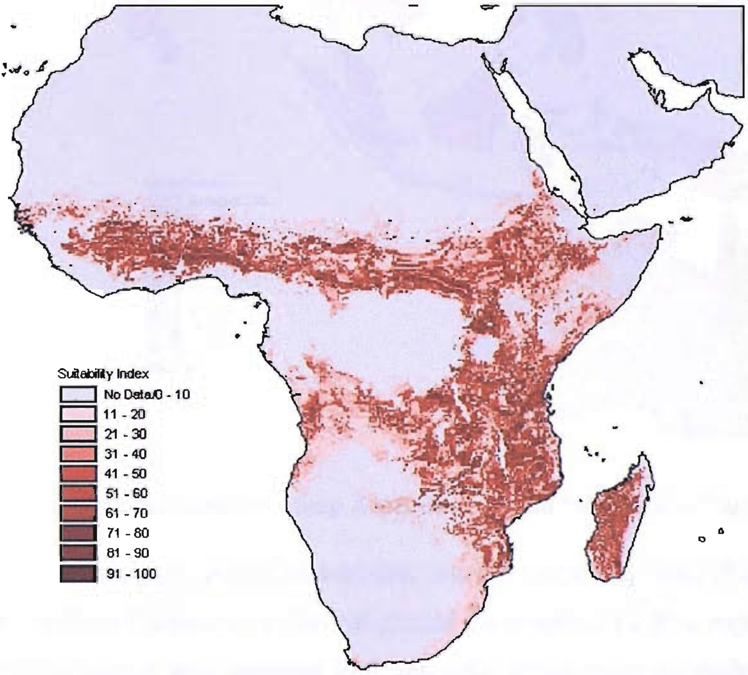


Figure 30 Habitat suitability map for Africa region "all variable" model

The Africa Region "all variable" model suitability map (Figure 30) shows a different distribution pattern of areas predicted as suitable, to that predicted by the global model for this region. As with the prediction from the global model predominant areas identified with suitability score above 10 are found in the wet-dry tropical regions. However unlike the global model suitability map, tamarind is predicted unsuitable in the wet tropics of the Congo basin and the southern coastal regions of West Africa.

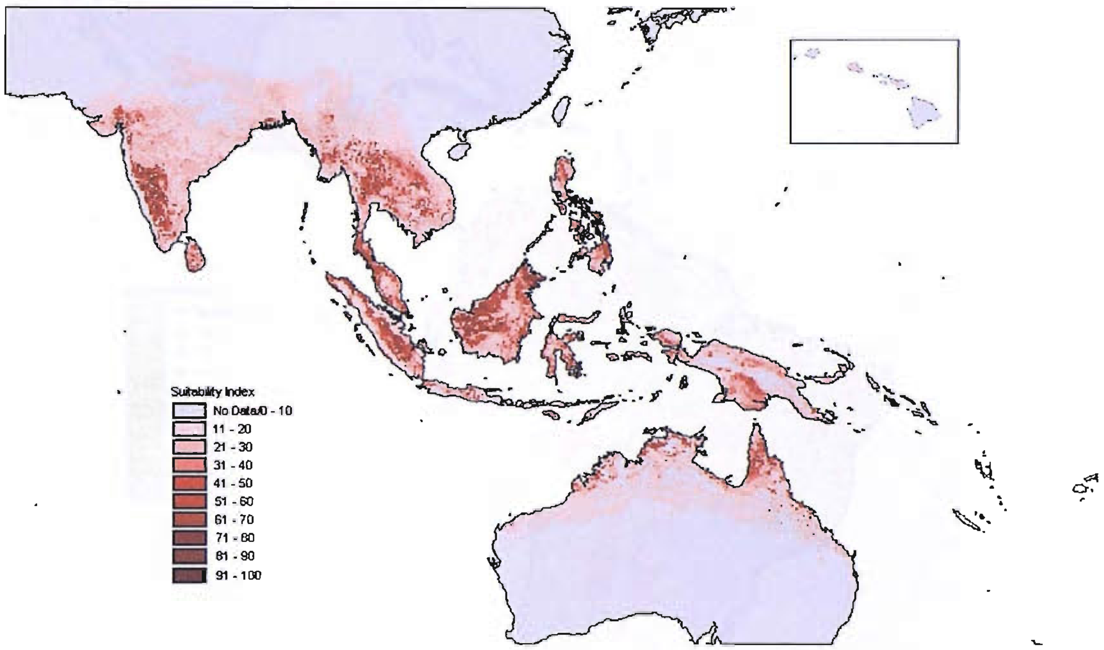


Figure 31 Habitat suitability map Australasia region "all variable" model

The Australasian region all variable habitat suitability map (Figure 31) shows a very different distribution to that global distribution for this region (Figure 29). In both the global and regional map suitable areas were predicted in the wet-dry tropics of South East Asia. However in wet tropical areas of the East Indies few areas are predicted as suitable on the global map (Figure 29), while large areas are predicted with high suitability scores by the regional map (Figure 31).

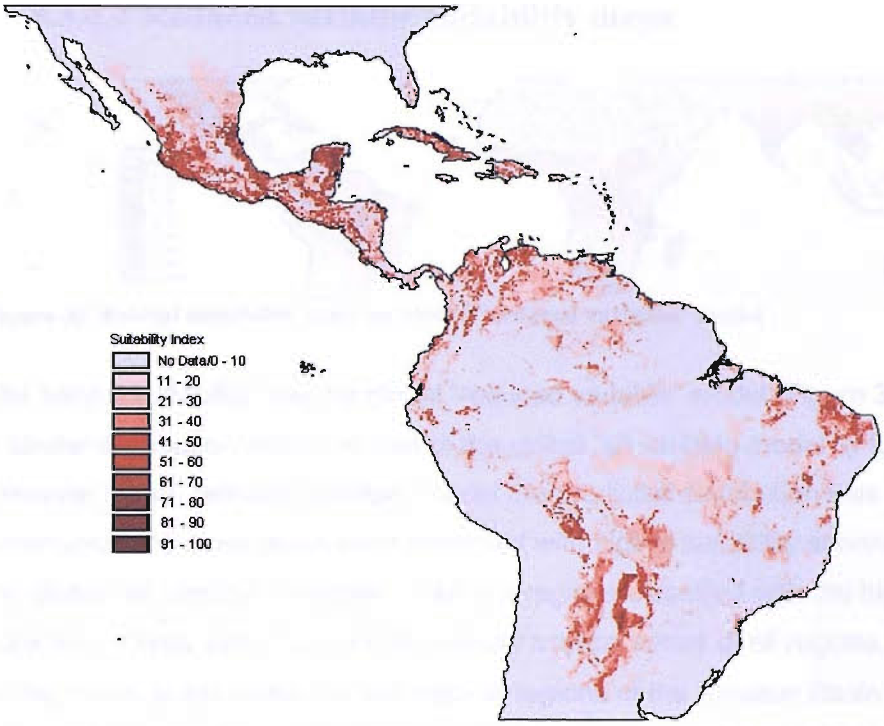


Figure 32 Habitat suitability map for Americas region "all variable" model

The habitat suitability map for the Americas regions (Figure 32) shows a much more restricted suitability distribution to that of the global model (Figure 29), as well as a very different distribution pattern. As with the global suitability map, regions of high suitability are found in Central America, although areas of suitability are patchier. The pattern in South America is very different to that predicted by the global suitability map, being much more restricted and patchy. Areas identified as suitable (not found on the global suitability map), include areas stretching further south beyond the extent of the wet-dry tropical region to the humid and dry winter zones of Argentina. There are also regions of low suitability within the Amazonian basin, along the river catchments, predicted as unsuitable by the global model.

4.3.4.2 Reduced variable suitability maps

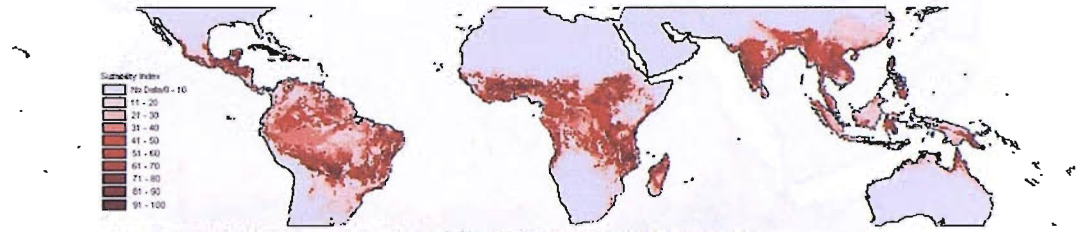


Figure 33 Habitat suitability map for global "reduced variable" model

The habitat suitability map for global "reduced variable" model (Figure 33) showed a similar distribution pattern to that of the global "all variable model" (Figure 29). However in the "reduced variable" model the predicted distribution was more continuous and more areas were predicted with higher suitability scores. As with the global "all variable" suitability map, the regions classified with the highest suitability scores were found in the wet-dry tropical zones of all regions. However in this model areas within the wet tropical regions of the Amazon Basin were also predicted as suitable. A larger area in the Indian subcontent was also predicted with a higher suitability score than in the "all variable" model. As with the "all variable" model few areas of high suitability scores were seen within the wet tropical region of the East Indies.

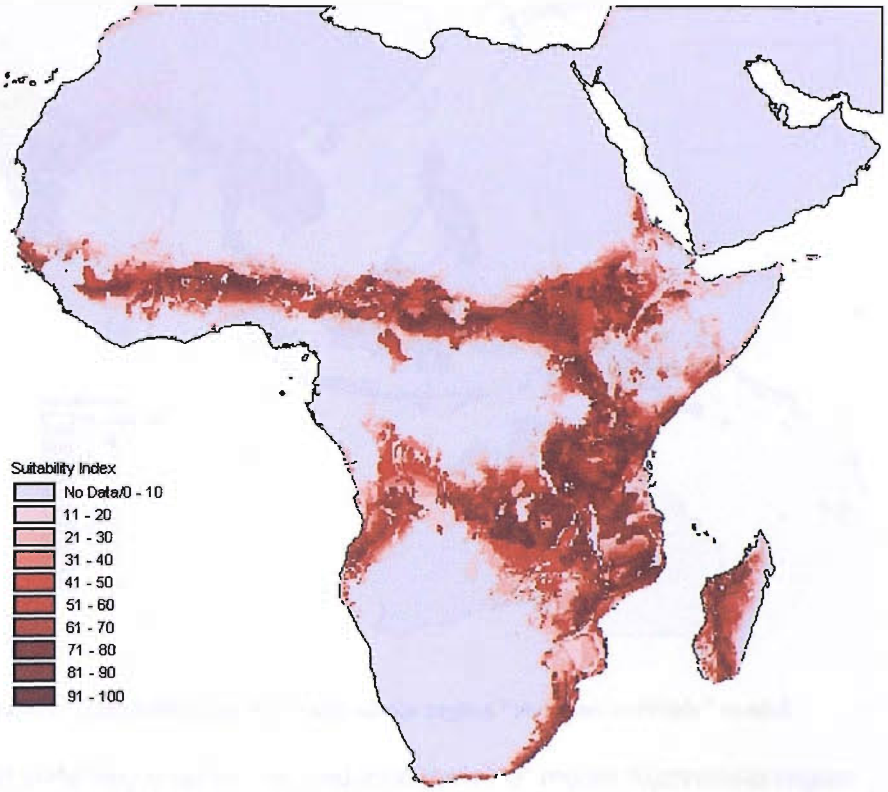


Figure 34 Habitat suitability map for Africa region”reduced variable” model

The Africa region “reduced variable” model suitability map (Figure 34) showed a different distribution pattern to the “reduced variable” global distribution (Figure 33) for the Africa region. The “reduced variable” Africa region map showed a very similar distribution pattern to the “all variable” Africa region map (Figure 30). The “reduced variable” map was showed a more continuous patterns and had a smoother gradient from cells with a high suitability score to cells with a low suitability score. However some differences did occur, the “reduced variable” model (Figure 34) predicted a much higher suitability score along the coast of South Africa. It also showed areas in the north of Africa, along the Mediterranean coast to be suitable, which were predicted as unsuitable by the all variables model. Both of these are found at the edge of tamarind geographic range. This could indicate that the reduced variable model may be predicting more favourably at the edge of tamarinds range.

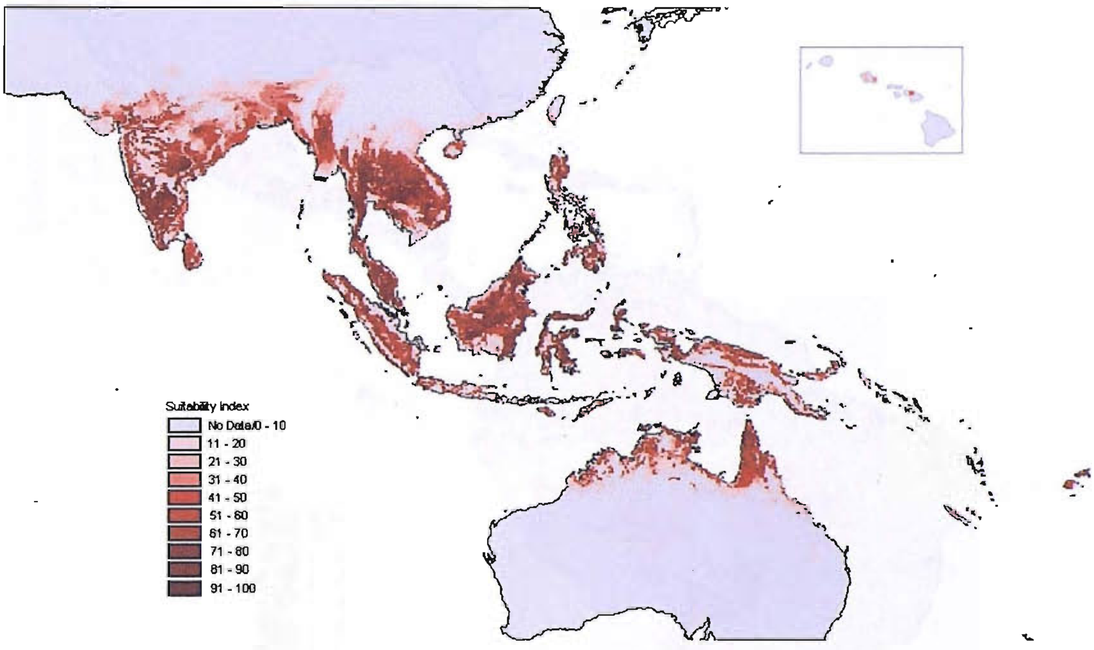


Figure 35 Habitat suitability map for Australasia region "reduced variable" model

The habitat suitability map for the "reduced variable" model Australasia region (Figure 35) showed a very different distribution pattern to that of the global model "reduced variable" map (Figure 33) for this region. It shows a similar pattern to the "all variable" model Australasia region suitability map (Figure 31), however shows a less restricted distribution pattern of suitable areas.

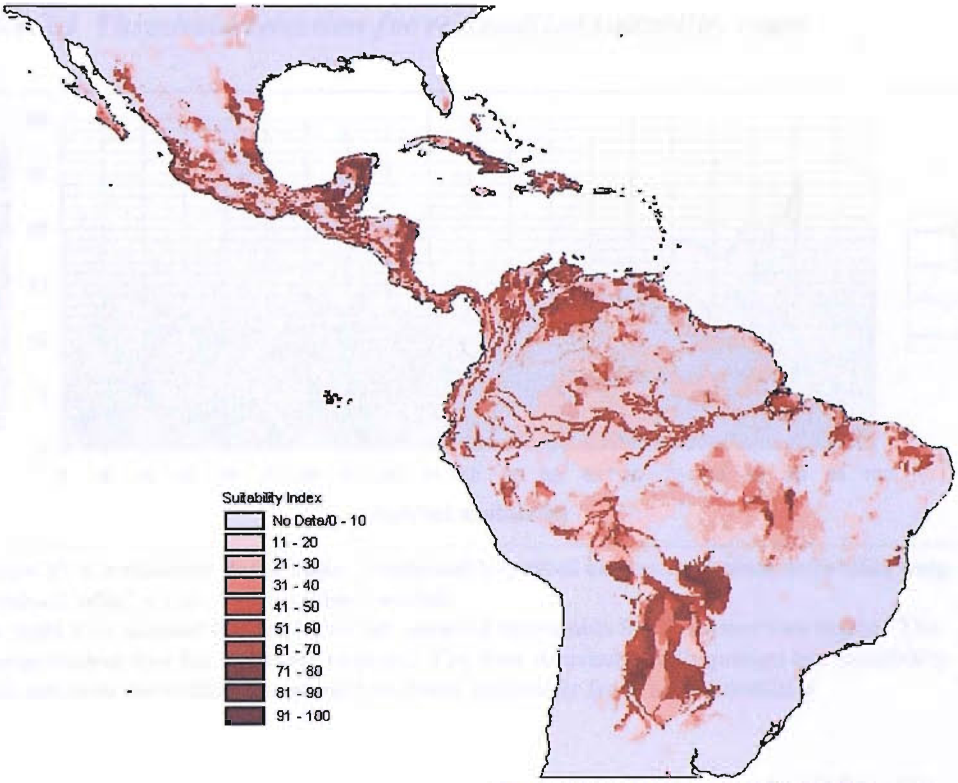


Figure 36 Habitat suitability map for Americas region "reduced variable" model

The habitat suitability map for the "reduced variable" model for the Americas regions (Figure 36) shows a very different distribution pattern to that of the "reduced variable" global model suitability map (Figure 33), for this region. It shows a similar distribution pattern for suitability as the "all variable" Americas regional model suitability map (Figure 32), however shows a less restricted distribution pattern of suitable areas.

4.3.4.2.1 Threshold selection for reclassified suitability maps

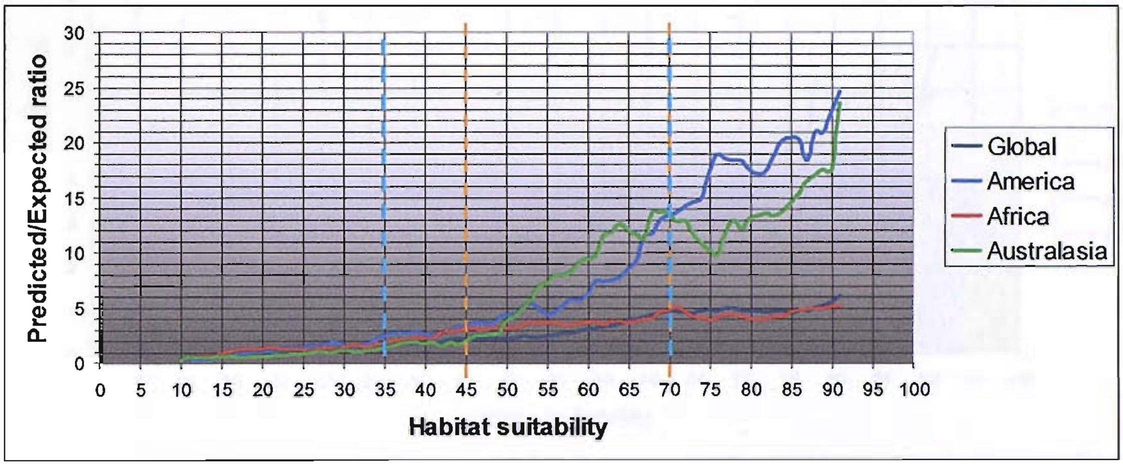


Figure 37 Continuous Boyce Index Predicted/Expected curves for habiat suitability map threshold selection for "all variable" models.
The light blue dashed line indicates the selected thresholds for the Americas region. The orange dashed line for the other regions. The first threshold distinguishes low suitability from medium suitability, the second medium suitability from high suitability

In this study a threshold of a mean F_i of above 2 was selected to distinguish unsuitable habitat from suitable habitat (see section 4.2.8.). For this reason the first threshold for "all variable" models (Figure 37) was selected at a habitat suitability index value of 35 for the Americas region model and 45 for all other models. This threshold distinguished low suitability habitat for medium suitability habitat. The second threshold for the "all variable" models was selected at a habitat suitability index values of 70, as this appeared as a natural break for all region. This threshold separated medium suitability from high suitability habitat.

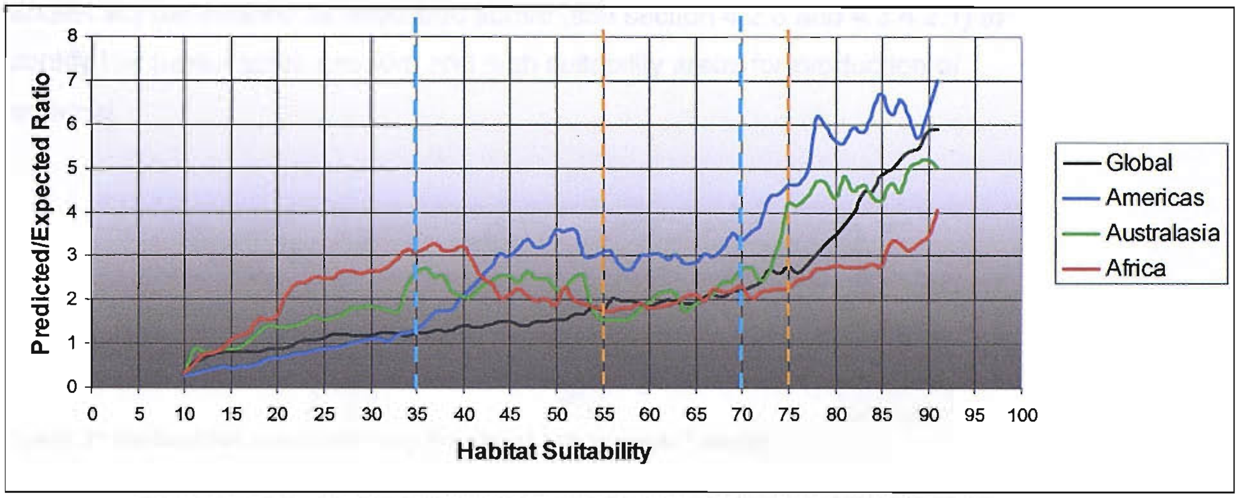


Figure 38 Continuous Boyce Index Predicted/Expected curves for habitat suitability map threshold selection for "reduced variable" models.
The light blue dashed line indicates the selected thresholds for the Americas region. The orange dashed line for the other regions. The first threshold distinguishes low suitability from medium suitability, the second medium suitability from high suitability

The first threshold for the "reduced variable" models (Figure 38) was selected at a habitat suitability index value of 40 for the America model and 55 for all other models. This threshold distinguished low suitability habitat for medium suitability habitat. The second threshold for all reduced "variable models" was selected at a habitat suitability of 70 for the Americas and 75 for all other regions and the global model. This threshold separated medium suitability from high suitability habitat.

4.3.4.2.2 Reclassified suitability maps

Models are reclassified as described above (see section 4.2.8 and 4.3.4.2.1) to identify low (unsuitable), medium and high suitability areas for production of tamarind.

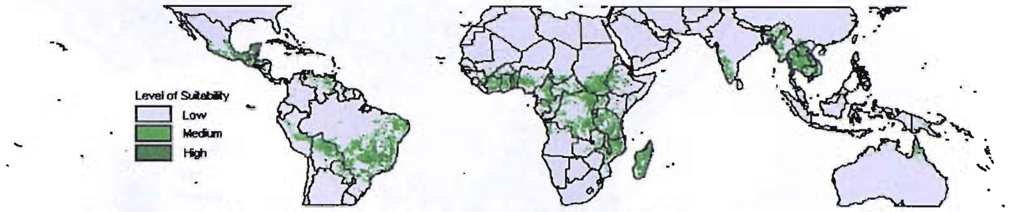


Figure 39 Reclassified suitability map for global "all variable" model

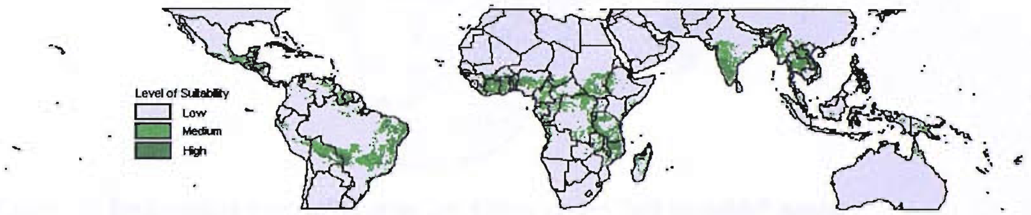


Figure 40 Reclassified suitability map for global "reduced variable" model

The reclassified suitability maps for the global "all variable" model (Figure 39) and the "reduced variable" model (Figure 40), show very similar distribution pattern for medium and high suitability areas. In both cases medium and high suitability habitats are limited predominantly to the wet-dry tropics. However there are some differences, the "all variable" map shows a greater number of areas of medium and high suitabilities areas in, Africa (in the wet tropics of the Congo basin and within the semi arid regions of Kenya and Tanzania), Latin America (in the humid regions of Paraguay and the humid highland regions of Peru) and in south East Asia (Cambodia). However the reduced variable model predicts more areas as suitable in India and the East Indies. The "all variable" model shows high suitability regions interspersed with medium suitability, however the "reduced variable" model suitability map shows more clearly defined areas of medium and high suitability.

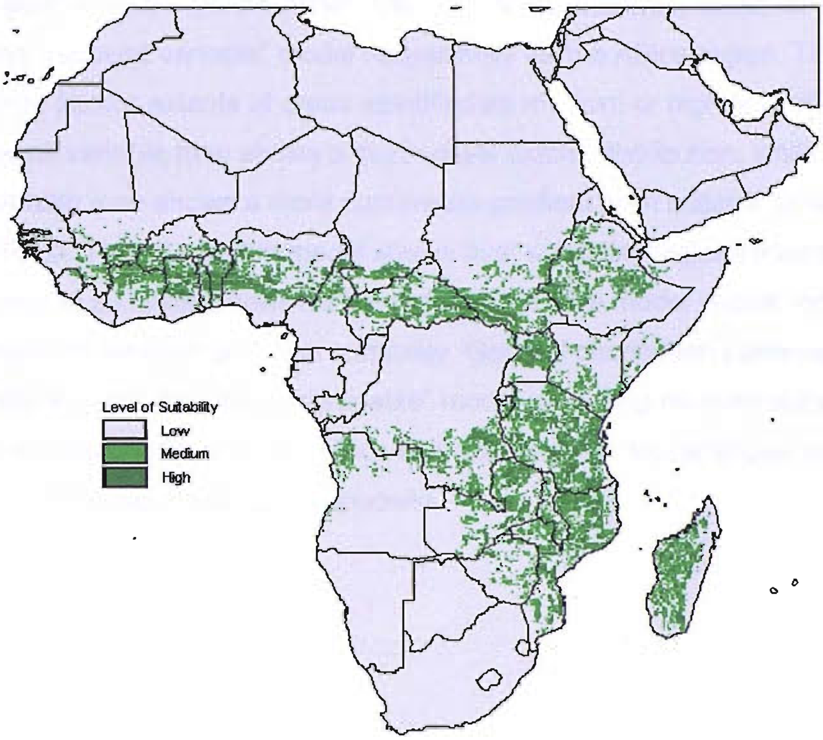


Figure 41 Reclassified Suitability map for Africa region "all variable" model

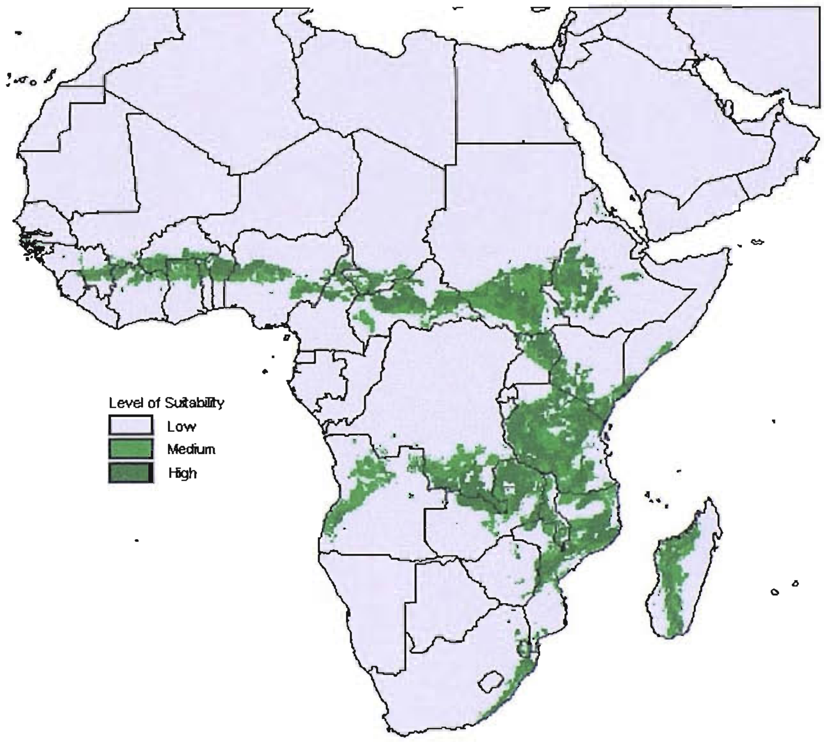


Figure 42 Reclassified suitability map for Africa region "reduced variable" model

Figure 41 and Figure 42 show the reclassified suitability maps for “all variable” and “reduced variable” model respectively for the Africa region. The Africa region show similar extents of areas identified as medium or high suitability. However the all variable map shows a much more patchy distribution, while the reduced variable map shows a more continuous gradient form suitable to unsuitable habitat. The all variable model shows high suitability regions interspersed with medium suitability, however the reduced variable model shows more defined region of medium and high suitability. General distribution patterns were slightly different with the “reduced variable” model predicting medium suitability and high suitability habitat in South Africa and not Southern Mozambique and the “all variable” map showing the opposite.

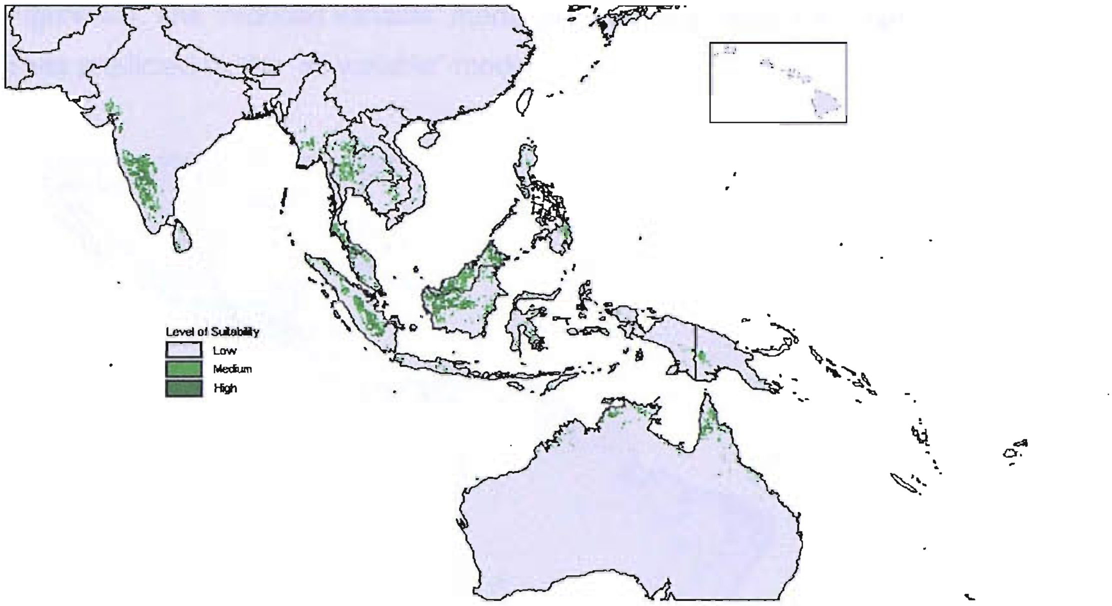


Figure 43 Reclassified suitability map for Australasia region "all variable" model

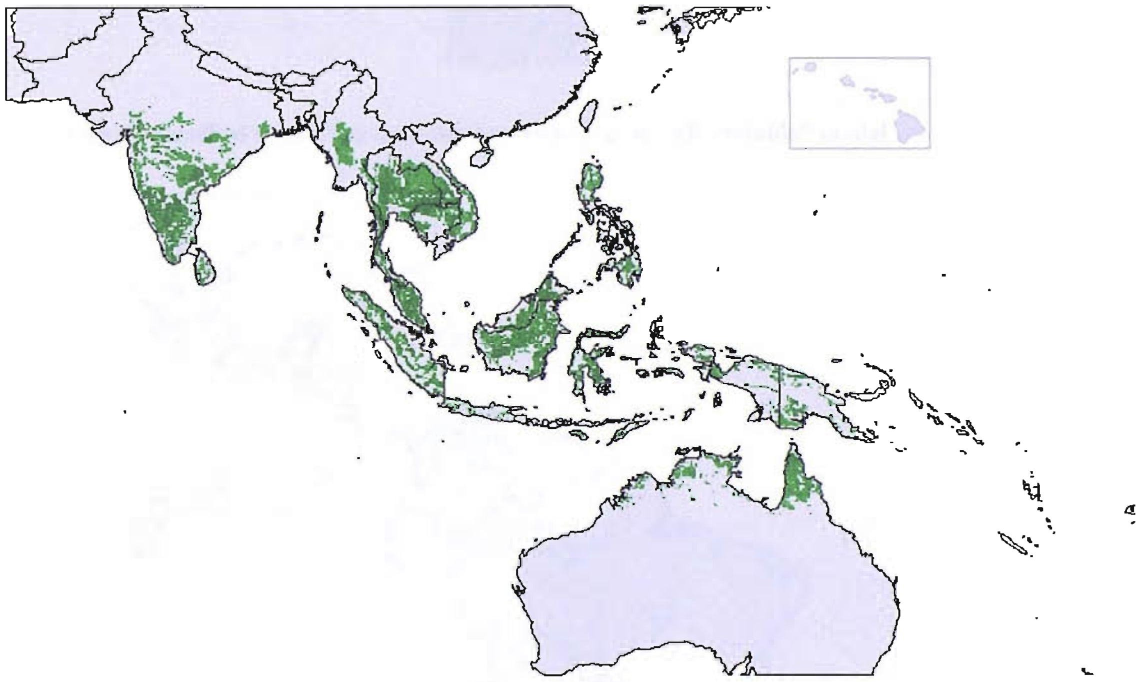


Figure 44 Reclassified suitability map for Australasia Region "reduced variable model"

The reclassified suitability map for "all variable" model Australasia region (Figure 43) shows a more restricted distribution than of the "reduced variable" model

(Figure 44). The "reduced variable" model predicts as suitable the majority of the areas predicted by the "all variable" model

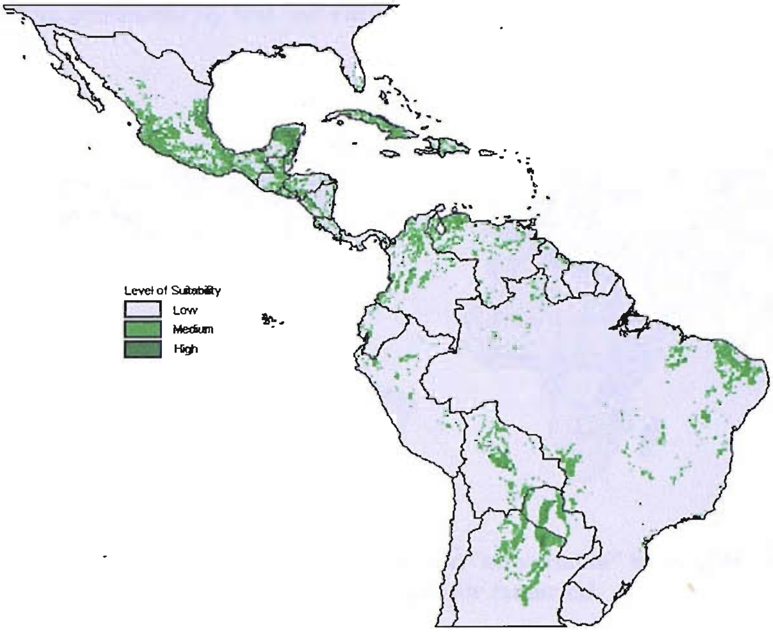


Figure 45 Reclassified suitability map for Americas region "all variable" model

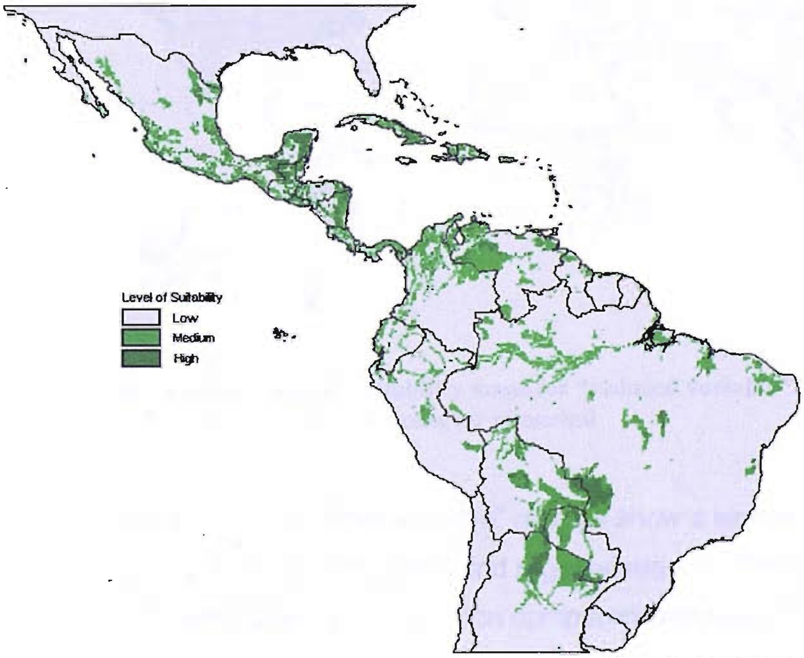


Figure 46 Reclassified suitability map for the Americas region "reduced variable" model

The reclassified suitability map for “all variable” model Americas region (Figure 45) shows a more restricted distribution than of the “reduced variable” model (Figure 46). The “reduced variable” model predicts as suitable the majority of the areas predicted by the “all variable” model

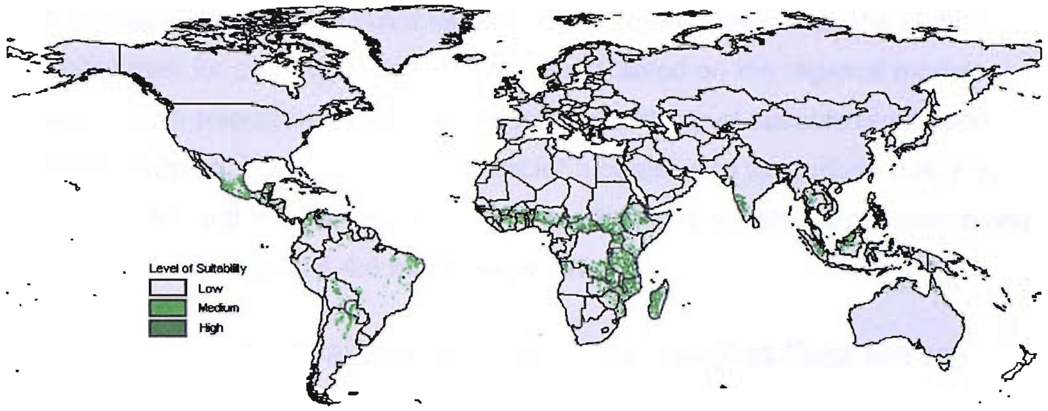


Figure 47 Reclassified regional models for "all variable" model, identifying global distribution of suitable production areas for tamarind

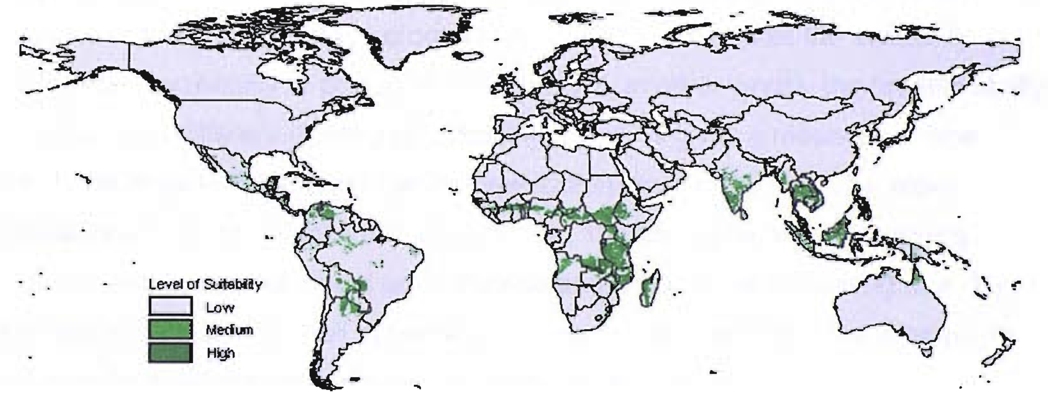


Figure 48 Reclassified regional suitability maps for “reduced variable” identifying global distribution of suitable production areas for tamarind

The “all variable” and “reduced variable” models show a similar pattern of predicted distribution for both global and regional models. The global model predicted different pattern of distribution compared with regional models. In most cases the regional models also predicted more restricted distributions (the exception being Australasia). The “reduced variable” models predicted a more continuous distribution, while the "all variable" maps showed a more fragmented distribution. This is due to the algorithm used (Hirzel1 and Arlettaz, 2003b). In most cases the "all variable" maps showed a more restricted distribution, than the "reduced variable" map predicting fewer areas as suitable.

Figure 47 shows the "all variable" regional maps combined and Figure 48 the "reduced variable" regional suitability maps combined. This shows the global suitability map for production areas of tamarind based on the regional models. Figure 47 was selected as the "best" map for identifying of suitable production areas of tamarind. The regional "all variable" models used to produce this map were selected as the "best" models for the prediction of suitable production areas of tamarind (see section 4.4.4 for full explanation).

4.3.5 Cross validation and validation against field survey data

Two measures of model quality were used to validate the model. The Continuous Boyce cross validation Index and a cross validated Contrast Validation Index (CVI). The Continuous Boyce cross validation Index measures the suitability index's proportionality to probability of presence, in other words, the models ability to distinguish different classes of suitability. The CVI gives a measure of how much the model differs from chance expectation or randomness. The cross validation partitions all occurrence points into 10 geographical non-overlying subsets and builds the model on 9 of these and tests on the remianing one. This process is repeated until all subsets have acted as the test set. This attempts to account for spatial nonstationarity within the validation process.

Table 27 Cross validation CVI and Continuous Boyce Index for all models

		Global	Africa	Australasia	Americas
Contrast Validation Index (CVI)	All variable	0.387	0.401	0.462	0.397
	Reduced variable	0.33	0.284	0.367	0.413
Continous Boyce Index	All variable	0.903	0.599	0.667	0.734
	Reduced variable	0.811	0.244	0.398	0.681

CVI - proportion of validation points predicted >=50 minus proportion of total study area predicted >=50.

Table 27 gives the results of the cross validation analysis for all models. The global models received the highest values for the Continuous Boyce Index,

indicating that these models are good at distinguishing different classes of suitability. However the regional models showed a higher CVI score than the global models (other than the Africa region "reduced variable" model). Indicating that although the global model are good at distinguishing suitable from unsuitable habitat they do not deviate much from what would be expected by chance.
The all variable model received the highest validation scores both in terms of continous boyce validation index and CVI score.

Due to the unknown reliability and accuracy of the data records used in the study, a further validation was conducted in addition to the cross validation. The Global and Australasia models were built using a subset of the presence points which excluded the field survey points collected in India. The habitat suitability index scores predicted by these models were validated against the excluded field survey data.

Table 28 Validation of Global and Australasia models predicting habitat suitability index against excluded field survey points

Validation measure	Geometric Mean – Reduced variable models		Harmonic mean – All variable models	
	Global	Australasia	Global	Australasia
Contrast Validation Index (CVI HS 50)	0.56	0.29	0.05	0.28

CVI - proportion of validation points predicted >=50 minus proportion of total study area predicted >=50.

Table 28 shows how the Global and Australasia models perform when validated with excluded field survey data. For the "reduced variable" models the global model performed better than the Australiasia regional model when validated against the field survey data. For the "all variable" models the Australasia regional model preformed better than the Global model when validated against the field survey data.

The global "reduced variable" model preformed best out of all models tested showing the highest CVI score as The Australasia "reduced variable" model preformed next best, followed by the Australasia regional "all variable" model and then the Global "all variable model".

It should be noted that although the validation against the field data, validates against occurrence records that were collected in the field using a GPS in 2003 and therefore are highly reliable and accurate. The validation is only carried out on a small proportion of the map in Western India. The cross validation uses 10 geographical partitioned subsets to carry out validation and therefore validates across the whole study area. This needs to be considering the results of the validation against the field data.

4.4 Discussion

4.4.1 Identifying the variables discriminating suitable from unsuitable habitat.

From the factor table analysis it was possible to identify which variables the model used to discriminate suitable tamarind habitat from unsuitable habitat. Results were similar for both "all variable" and "reduced variable" models. Tamarind responded to gradients for a number of variables at the global extent and at all regional extents. These were measures of minimum temperatures, temperature variation and water availability during the wettest periods. Although these variables are not the most important gradients in all regions, it is these variables that appear to be dictating the overall distribution of tamarind within the extents and grain size used in this study.

In all regions tamarind showed a preference for lower than the average for the study area for *temperature variation* variables. Tamarind also showed high specialisation, indicating that tamarind had a narrow niche width in relation to these variables. Although the extent was reduced to the 35 °N and 35 °S, the study area still included a broad range of temperatures due to the extremes of desert regions and cold mountainous regions. Occurrence records from the species dataset (Table 26) showed tamarind to exist in temperatures as low as 4 °C and a mean annual range of 16 – 28 °C. This was relatively restricted in comparison to the variation found across the large extents in the study areas used in this project.

In all regions and at the global extent, tamarind showed a preference for higher than average *minimum temperatures*. The response plot for mean temperature coldest quarter (Figures 27) shows tamarind is not suitable in regions below 10 °C, which indicates a lack of tolerance to cold temperatures. This concurs with information from the literature, Troup (1921);FAO (1988);Chundawat (1990); Mahoney (1990); Vogt (1995); who all state that tamarind is sensitive to frost. According to NFTA (1993) and Morton (1987) it is neither tolerant to persistent cold nor to brief frost and does not do well in cold temperatures. However older trees are said to be more resistant to extreme temperatures than younger trees (Coronel, 1991; Gunasena and Hughes, 2000) (see section 2.1.6.1.2.1 for further details).

The models indicated that tamarind shows a preference for a bimodal climate in terms of *water availability*. It was selecting areas with higher than average water availability (precipitation and moisture availability) during the wet periods of the year and areas of lower than average water availability during the dry periods. This would indicate that tamarind is well adapted to the monsoonal, wet-dry tropical climate in which it occurs across much of its distribution range.

Tamarind's preference for drier than average conditions during the dry periods in the wetter regions of Australasia and Americas concurs with information from the literature. Many authors have noted that precipitation during the flowering period (flowering and fruiting of tamarind take place in the dry season (Gunasena and Pushpakumara, 2006)) will effect fruit yield. If heavy precipitation occurs during flowering, the tree will not bear fruit (El-Siddig et al., 2006). An extended dry season is said to be essential for fruit development, (Allen and Allen 1981, (Von Maydell, 1986) (Morton, 1987), and lower precipitation is required during the dry periods to allow flower initiation. The Africa and Americas region show a decline in predicted suitability with increasing minimum moisture availability index (MAI) (Figures 26a,b), which is consistent with the above. However the Australasia region shows no change in suitability with increasing minimum MAI.

Tamarind showed a preference for high water availability areas in all regions during the wet season (wettest period of the year), and in two regions, higher average annual precipitation. Tamarind is often referred to as a drought resistant species (Troup, 1921);(Coronel, 1991)) due to its deep and extensive rooting system (Williams, 2006c). Williams noted that tamarind can withstand up to six months without precipitation and that it has been seen to grow in very dry areas, but with supplementary irrigation. It has however been noted that premonsoonal droughts have been known to effect growth. Tamarind showed a preference for a dry interval during the year to allow fruit initiation and is tolerant to extended periods without rainfall. However it appears to require reliable precipitation for some period during the year.

All models showed that tamarind showed a preference for high nutrient soils in all regions although it did not appear to be restricted to these conditions. Literature sources have noted that tamarind can grow in a wide range of soil (Sozolnoki, 1985) and it has been suggested to have no particular soil requirements (Galang, 1955; Chatturvedi, 1985). The variables on which tamarind correlated best with the marginality factor for Africa, Australia, were related to water availability and temperature. However in the America's marginality was most highly correlated with soil variables (see Table 21 and Table 25), primarily with chemical properties but also with soil structure. The species is showing a preference for higher average values for soil variables than in the other regions. However the regional global average in this region is lower than found in other regions, indicating tamarind is being highly selective in this region for these conditions.

4.4.2 Variation in global and regional plant environment responses

Plotting of the tamarind presence points in component space (Figure 23) and ENFA space (Figure 24) indicated that there may be differentiation in environmental conditions experienced by tamarind in the different regions. Results from the modelling of the data, for both the "all variable" models and "reduced variable" models, appears to confirm this. The predicted environmental response plots showed different responses to environmental conditions on a number of variables (Figures 25a,b and Figures 26a,b), both between the global and regional dataset and between each of the regional datasets.

Tamarind showed a marked difference in the predicted suitability response to "precipitation wettest quarter" (Figures 25a,b) between regions. In Africa tamarind showed a much more restricted range in terms of precipitation. Tamarind also showed a differing response in terms of minimum moisture availability (Figures 26a,b) between regions (As discussed above). These differences in modelled environmental response to a number of the variables, is the likely cause of the different predicted distributions described by the global model for each region, and the distribution predicted by each of the regional models

The likely explanation for this is spatial nonstationarity. When modelling species that exist in large and highly variable areas, heterogeneity in the predictor variables will exist and species may respond to habitats in different ways because of different ecological status (Osborne and Suarez-Seoane, 2002; Estrada-Pena et al., 2006). Interregional differences in modelled relationships can arise through models not being fully specified, because habitat availability differs spatially, and because wide-spread species show variation in ecological characteristics across their range (Osborne et al., 2007). Therefore the observed geographical patterns and measured species environment responses tend to be spatially variable (Jetz et al., 2005), a concept termed spatial non-stationary. Causes of nonstationarity include the variation in local conditions (Jetz et al., 2005), variation in community structure due to the influence of inter-specific competition (Peterson and Holt, 2003) or niche evolution leading to geographical variation in the niche (Peterson and Holt, 2003).

The characteristically dry climate in the Africa region may mean tamarind is typically found in drier climates in this region. However Africa shows a similar range in terms of precipitation to that of the Americas. Tamarind however is not found or predicted by the regional model in wetter regions of Africa. As discussed previously at the extent and scale of which this project is conducted, factors such as community structure and competition are likely to have little effect on the shape of the response curve (Peterson and Holt, 2003; Guisan and Thuiller, 2005; Soberon and Townsend Peterson, 2005). Geographical genetic variation in the niche is a possible explanation for the variation in the conditions experienced and shape of response curves between regions. It may be the case that the tamarind population has adapted to different conditions in terms of water availability in the different regions.

4.4.3 Global and regional suitability maps

In this study tamarind was modelled over its global extent to ensure the capture of the full climatic range of the species. It may be assumed that the tamarind global population has no genetic difference in terms of the niche due to the effect of niche conservation (Huntley et al., 1989; Holt and Gaines, 1992; Beerling et al., 1995; Holt, 1996; Peterson et al., 1999; Peterson and Vieglas, 2001; Prinzing et al., 2001; Peterson and Holt, 2003; Peterson, 2003a; Thuiller et al., 2005; Estrada-Pena et al., 2006) (see section 2.2.3.5). Apparent geographical variation within the niche may be due to phenotypic plasticity or other causes of spatial nonstationarity (Schluter, 2000; Peterson and Holt, 2003). This suggests that carrying out reciprocal planting of a tamarind plant from one location in the predicted global distribution to another; the population would survive in the new location. In this case it must be assumed that the global model may give the most accurate prediction.

However Peterson and Holt (2003) noted a high level of human introduction increases the likelihood of evolutionary effects in the form of geographic variation in niche characteristics, leading to geographical subpopulations and therefore spatial nonstationarity. Miller and Knoft (2006) suggested that changes in the niche between wild and cultivated populations could reflect artificial selection during domestication. Tamarind has been referred to as a “specialised domesticate” (Davis and Bye, 1982; Williams and Haq, 2003), These species are considered to be protected by local people and used in replacement planting and rely on man for dispersal and are co-evolved in disturbed local environments.

Therefore differences in the responses of tamarind are less likely to be affected by factors such as changes in community structure and competition and are more likely to be due to the influence of man. This includes the facilitation of dispersal and reproduction and the selection of plants for adaptation to particular habitats and uses. Although tamarinds origin is uncertain, it is thought to have originated and have its indigenous range in Africa (El-Siddig et al., 2006) (see Section 2.1.1). There is evidence that it was introduced to South and Southeast Asia from Africa by Arabian seafarers in the first millennium BC (NAS, 1979; Williams, 2006a) and is now naturalised in many areas (Coronel, 1991). From Asia it was thought to have been introduced to the Americas (Ecoport, 2001).

It may be that the tamarind niche influenced by human introduction and selection, has evolved/adapted to the new wetter environments found in Asia and the Americas . This leads to differentiated populations (on the basis niche characteristics) existing across the full range of the species (Pulliam, 2000; Schluter, 2000; Peterson and Holt, 2003; Estrada-Pena et al., 2006). These populations can be obscured when niche based models are applied to the entire area in which the species is distributed. If this is assumed then regional models may give the most accurate predicted distribution.

4.4.4 Selecting a model for use in predicting suitable regions for the production of tamarind

In order to select a map to identify suitable regions for tamarind production from the global or regional models, it is important to identify which model provides the most accurate prediction. Two measures of model quality have been used to validate the model. The Continuous Boyce cross validation Index and a cross validated Contrast Validation Index (CVI). The Continuous Boyce cross validation Index measures the suitability index's proportionality to probability of presence, in other words, the model's ability to distinguish different classes of suitability. The CVI gives a measure of how much the model differs from chance expectation or randomness. The cross validation partitions all occurrence points into 10 geographical non-overlying subsets and then builds the model on 9 of these and tests with the remaining 1. This process is repeated until all subsets have acted as the test set. By validating the model with geographical subsets it attempts to account for spatial nonstationarity within the validation process.

The global model received the highest values for the Continuous Boyce Index indicating that these models showed a straight line increase in Predicted/Expected ratio as Habitat suitability index increased. However the regional models showed a higher CVI than the global models (other than the Africa region "reduced variable" model), indicating that although the global model are good at distinguishing suitable from unsuitable habitat they do not deviate much from what would be expected by chance.

Owing to spatial nonstationarity, global models built over a large area may have weak local predictive power because of differences in the habitats available or selected (Osborne and Suarez-Seoane, 2002; Foody, 2004). Global models sacrifice local fit for generality and the predictive maps and response curves they produce are averages that mask the underlying environmental data (Osborne et al., 2007). The danger is that the averaged relationship may not exist in nature (Osborne et al., 2007). Osborne and Suarez-Seoane (2002) and Estrada-Pena et al. (2006) both found large scale models improved when data was geographically partitioned before analysis. In this study regional models showed variation in predicted environmental responses. This is likely to be due to spatial nonstationarity. Genetic variation in the niche is assumed, as a possible explanation for the variation in the conditions experienced and shape of response curves between regional and global models.

Owing to the unknown reliability and accuracy of the herbarium data records used in the study, a further validation was conducted against the field survey data collected in India. The field data was excluded from the occurrence dataset and the global and Australasia models rebuilt; the field survey data was then used to validate these models. A Contrast validation index was calculated for the Global and Australasia region models.

As with the cross validation CVI, when validated against the field survey data the Australasia "all variable" regional model performed better than the "all variable "global model. However when the "reduced variable" models were validated against the field survey data, the global model showed a higher CVI than the Australasia model. Although the Global "reduced variable" model may be predicting well in India it may not be in other areas, causing it to receive a low score in the cross validation CVI. It may also be the case that the removal of field survey occurrence points from the regional model has a greater effect on the models prediction power, as a greater proportion of the total points were removed.

Other than the global "reduced variable" model, when validated against the field survey data the models preformed poorly in comparison to when validated with the cross validation CVI. Indicating that although overall the regional models are receiving high CVI values they are not predicting well in all locations. However this may be due in part to the reduction of occurrence points used to train the models.

Osborne et al. (2007) argue that local methods are complementary to global methods, revealing habitat associations and data properties which global methods average out and miss. In this study regional models showed variation in predicted environmental responses. Global models were unable to encompass regional variation. Tamarind is known to have undergone human selection and introduction over a long period of time and genetic variation has been seen in other characteristics of the species. Wide phenotypic variation in tamarind (i.e. sweet tamarind) has been attributed to geographic isolation and gene mutation (Evenson and Gollin, 2003). Assuming no other causes of spatial nonstationarity are affecting the model, genetic variation occurring in the species niche leading to geographical subpopulations is the likely cause. However without reciprocal planting or genetic analysis this is impossible to prove.

Although regional models were shown not to have as good an ability to distinguish classes of suitability as global models, they are better at predicting distributions which differ from random, therefore having greater predicting power. This is likely to be due to their ability to incorporate the regional variation incurred by non-stationary assumed to be due to genetic variation across the niche. The "all variable" models received the higher validation scores than "reduced variable model both in terms of continous Boyce validation index and CVI score. The "all variable regional model out performed the global models both in the cross validation and validation against the field data. Therefore the "all variable" regional models were selected as "best" for predicting potential distribution of tamarind.

Global models apply generalised response curves which are unable to account for regional variation. While regional models improve the prediction they may not predicting well locally in all locations. Both global and regional models in this study were built on a large scale. This allows us to map the tamarind's full environmental niche. It also provides information on regional niche variation which may exist across the distribution. Using this information, potential distribution maps have been produced at the regional and global extent, which give an indication of regions of potential production. However caution should be taken when using these models to predict potential production areas at a local scale. Osborne et al. (2007) suggest that modellers interpolating data for practical actions (i.e. conservation) should consider local modelling methods. In order to predict suitable locations for production at a practical scale for use by extension workers or farmers, models built at a local scale may be more appropriate, incorporating the complex factors which may influence tamarind distribution at this scale.

4.4.5 Predicting regions of high yield

Many of variables which tamarind used to discriminate suitable from non-suitable habitat relate to survival (i.e. minimum temperatures, precipitation wettest periods). Some of the variables with smaller coefficients appear to be discriminating areas relevant for good flowering and fruiting conditions. These include nitrogen, organic carbon, high temperatures and low precipitation during dry periods (Hughes, in press). This would indicate that the model is based on populations which are naturally regenerating. This suggests that although the model is built purely on presence data, area's predicted as high suitability, are likely to concur with high yields.

4.4.6 Limitations

The data used in this study was collected from an extensive search of herbaria from across the world and is thought to encompass a true sample of the entire ecological space of the species. Caution must be noted in regard to using density based distance algorithms when using herbarium data or other data which has not been systematically sampled. Some of the individual points may be representative of more densely populated points in ecological space which have been poorly sampled. To carry out a systematic sampling at a global extent as used in this study would be near impossible, and so the database created for this project is a viable alternative.

In this study rather than modelling the species across the entire global datasets we restricted the study areas to 35 °S and 35 °N. Therefore to only encompassing areas where a tamarind a tropical tree species (Williams, 2006a) is likely to occur; however even at this extent the model may have been "blinded" by large scale regional differences in temperature and precipitation (Hirzel pers comm.). It is these large scale variables which appeared to be having greatest influence on the model. By removing areas where we know tamarind is highly unlikely to exist and reducing the study area extent further; it may be possible to identify variables which are important to tamarind distribution at smaller scales. The selection of areas for removal could be conducted using expert knowledge or based on the suitability maps of previous models such as those produced in the study.

5 Chapter 5 Final Discussion and Conclusions

5.1 Conclusions

As awareness grows regarding concerns of global climate change and rising populations, the benefits provided by underutilised fruit tree species such as tamarind in combating increasing malnutrition, hunger and poverty in a changing world become more apparent. The importance of learning more about underutilised species niche requirements, their environmental adaptation and ecogeographic distribution has been identified. Such information allows the prediction of the potential distribution of underutilised species such as tamarind and the identification of suitable areas for production both under current and future climate scenarios.

There is little data available on the eco-physiology or yield records of tamarind and so traditional crop distribution modelling techniques could not be applied. In this study a species occurrence dataset was created from herbarium records for the full extent of tamarind's distribution. Such datasets often carry error and inherent spatial and temporal bias. However procedures were carried out in this study to identify and remove erroneous data and account for bias. This has allowed the creation of an extensive dataset which is thought to be representative of the full geographical and ecological distribution of tamarind and suitable for modelling the potential distribution of the species.

The creation of this dataset has allowed the use of the statistical modelling tool ENFA to model the full response of tamarind to the abiotic environment and quantify its full niche. This niche has then been mapped into geographic space in order to identify suitable areas for production. Tamarind was modelled at a global and at 3 regional extents. Models were run with all environmental variables and a reduced set of variables. The species appeared to experience different condition in different regions for a number of variables. For certain variables the predicted environmental response varied between regional models and between the regional and global models. This resulted in different predicted patterns in distribution between global and regional models. Owing to spatial nonstationarity, global models built over a large area may have weak local predictive power because of differences in the habitats available or selected (Osborne and Suarez-Seoane, 2002; Foody, 2004). Global models sacrifice local fit for generality and the predictive maps and response curves they produce are averages that mask the underlying environmental data (Osborne et al., 2007). The danger is that the averaged relationship may not exist in nature (Osborne et al., 2007). Osborne and Suarez-Seoane (2002) and Estrada-Pena et al. (2006) both found large scale models improved when data was geographically partitioned before analysis. Genetic variation in the niche is assumed, as a possible explanation for the variation in the conditions experienced and shape of response curves between regional and global models.

Although regional models were shown not to have as good an ability to distinguish classes of suitability as global models, they are better at predicting distributions which differ from random, therefore having greater predicting power. This is likely to be due to their ability to incorporate the regional variation incurred by non-stationary assumed in part to be due to genetic variation across the niche. The all variable model received the higher validation scores than "reduced variable model both in terms of continous boyce validation index and CVI score. The "all variable regional model out preformed the global models both in the cross validation and validation against the field data. Therefore the "all variable" regional models are as selected as "best" for predicting potential distribution of tamarind.

Global models apply generalised response curves which are unable to account for this variation. While regional model improve the prediction they may not predicting well locally in all locations. Both global and regional model in this study were built on a large scale. This allows us to map the tamarind's full environmental niche in ecological space. It also provides information on regional niche variation which may exist across the distribution. Using this information, potential distribution maps have been produced at the regional and global extent, which give an indication of regions of potential production. However caution should be taken when using these models to predict potential production areas at a local scale.

Osborne et al. (2007) suggest that modellers interpolating data for practical actions (i.e. conservation) should consider local modelling methods. In order to predict suitable locations for production at a practical scale for use by extension workers or farmers, models built at a local scale may be more appropriate, incorpoarting the complex factors which may influence tamarind distribution at this scale.

The lack of data on the eco-physiological response of underutilised species to the environment means few of these crops have been modelled. The exceptions include global modelling of the potential yield of bambara ground nut (Azam-Ali et al., 2001) However development of such "process" crop models requires extensive time and financial investment in greenhouse and field experiments, in order to measure physiological responses. There is increasing availability of plant species location data in the form of passport information from herbarium records and germplasm collections; this combined with the methods used in this study enables the modelling of the potential distribution of a large number of underutilised crops with relative ease. Such methods will allow the quantitative modelling of the potential production areas for crop species for which limited or no eco-physiological, empirical growth and yield information exists.

It should be noted that choice of the species grown should reflect the farmer's priorities. However this method provides a tool which allows the farmer to be able to see what options are available at a particular location.

5.2 Applications of model

There exists a considerable land area where tamarind production could be expanded in its native range, but due to low priority allocation, many countries have not identified areas that could be used for expansion (Nyadoi, 2004). Potential production areas for tamarind depend on the demand for tamarind products. The initial spread of plantations is likely to occur around the current production centres where technology, skill and marketing channels are already in place. Such locations could be identified and mapped on to suitability maps to identify distance to or amount of suitable habitat around such areas.

Cultivation is likely to spread to resource-poor areas and wastelands where other crops cannot grow, because such land usage is receiving increased attention (El-Siddig, 2006). Suitability maps could be combined with maps of marginal land, to identify were both suitability of tamarind and resource poor land overlap.

With increasing demand for processed tamarind products there is a need for large plantations established for processing and export. Such plantations would most likely be situated close to the cities, where the facilities for packing and transport are close at hand (El-Siddig, 2006). Suitability maps could be combined with maps of urban areas or population density and areas of production could be selected from high suitability areas close to the cities.

5.3 Selection of global and regional model for predicting suitable location for production of underutilised species in climate change scenarios

With effects of global climate change, it is becoming increasingly important to identify the plants response to the environment and how this influences a species distribution patterns or in the case of crops where it has the potential to be cultivated.

Global temperatures are expected to rise between 1.4 and 5.8 °C during this century depending on the level of green houses gases we emit (IPCC, 2001). Impacts can be expected on many atmospheric systems leading to fundamental changes in weather patterns and extreme events. This will result in a greater incidence and magnitude of hurricanes, floods, and droughts (IPCC, 2001; Jarvis et al., 2006). A number of underutilised crops show potential for both mitigation and adaptation to climate change (Bowe, 2007). It is important to be able to identify where it is possible to grow these crops. This can also be applied to identifying where germplasm for the species may exist, in order to conserve genetic resources and identify planting material that may be adapted to specific environments.

This study identifies where tamarind can be cultivated under present conditions. The models developed in this study can be applied to future climate scenarios in order to identify distribution and yield under future climate senarios. Peterson et al (1999) and Peterson and Holt (2003) noted that use of ecological niche models to predict future distribution based on global climate change, depends fundamentally on the identification on whether niches are relatively stable over the geographical extent of the species range and over time. Although these authors were mainly referring to wild populations this is also highly relevant in terms of crops, particularly for underutilised species such as tamarind for which populations show high variability and few known cultivars exist.

In this study we modelled the entire species distribution range, as well as regional subsets. The model showed species to respond differently on certain variables between global and regional extents, resulting in very different predicted distribution patterns. Which of these (the global or regional models) is more accurate depends to a great extent on the population ecology of the species. If a species fundamental niche is consistent across its entire population (phylogenetic conservation in niche characteristics), it will be more accurate to model the global distribution, incorporating the full response curve of the species and predicting this into future climate. A number of species distribution models have been based on this assumption. Such models were developed over a broad extent in order to capture the full species range before projecting the species distribution on finer scale grids or into new environments (i.e. future climate scenarios) (Pearson et al., 2002; Pearson et al., 2004; Araujo et al., 2005; McPherson et al., 2006). However if the niche varies within a regional extent it will be more accurate to model (delineate the response curve at a regional basis) and predict into future climate at a more regional scale. As Peterson and Holt (2003) noted a high level of human introduction increases the likelihood of evolutionary effects in the form of geographic variation in niche characteristics, leading to geographical subpopulations. Variation has been reported for tolerance to drought, wind, poor soils, water logging, high and low pH and grazing for tamarind (Gunasena and Hughes, 2000). For underutilised crops such as tamarind a more regional approach may therefore be considered the more prudent approach. However decisions should be made on the basis of each individual species.

Farmers select plants with morphological characteristics and properties which meet their needs; they develop landrace, cultivars or varieties. The models used in this study provide information on the environmental response at a species scale, and subpopulations at a broad scale. However they provide little information on whether phylogenetic niche variation occurs at the level of landraces, cultivars or varieties.

5.4 Future Work

The focus should now be on investigating niche requirements and niche variation on land races, varieties or cultivars of underutilised species identified as suitable to meet (through farmer participation and research) nutritional, medicinal and market requirements under both current and future climate scenarios (Bowe, 2007).

Future research should look at using herbarium distribution data, and morphological and chemical characteristics from germplasm collection, with empirical species distribution models to investigate the ecological niche of semi domesticated populations of underutilised species. This should investigate the following

- Is there a link between niche conservation, niche evolution and morphological and chemical characteristics, both in terms of physiological adaptations to the environment in terms quality determinates
- How much plasticity (both in terms of phenotypic plasticity and genetic variation is found across the population)?
- Effect of domestication on the niche, the more domesticated a crop, the wider its ecological amplitude?

Such studies would give us a better idea of the effects of climate change on the distribution and morphological characteristics of populations both in relation to conservation and utilisation of genetic resources and potential production areas for crops.

6 References

ACUC (2004) ACUC Newsletter. **1**

Aggarwal PK (1995) Uncertainties in crop, soil and weather inputs used in growth models: implications for simulated outputs and application. Agricultural Systems **48:** 361 - 384

Aggarwal PK, Kalra N (1994) Analysing the limitations set by climatic factors, genotype, and water and nitrogen availability on the productivity of wheat II. Climatic potential yields and management strategies. Field Crop Research **38**

Alexandratos N (1995) Agriculture: Towards 2010:An FAO Study. John Wiley and Sons, Chichester, United Kingdom

Allen KE, Bardley RD, Monk RR, Knyazhnitskiy OV, Parker NC, Schmidly DJ, Baker RJ (2001) Determining the Accuracy of Museum Voucher Specimen Data. Occasional Papers, Museum of Texas Tech University **210**

Allen ON, Allen EK (1981) The Leguminosae: A Source Book of Characteristics, Uses and Nodulation. University of Wisconsin Press, Madison, WS

Anon (2005) Meeting the Millennium Development Goals with Agricultural Biodiversity. *In.* International Plant Genetic Resources Institute, Global Facilitation Unit for Underutilized Species, MS Swaminathan Research Foundation

Anon (2005) Underutilized' species, Rich potential is being wasted. People and Biodiversity **Issue paper, GTZ**

Anthony K, Haq N, Cilliers B (1995) Genetic Resources and Utilization of underutilisd crops in Southern and Eastern Africa,

Araujo BM, Thuiller W, Williams PH, Reginster I (2005) Downscaling European species atlas distributions to a finer resolution: implications for conservation planning. Global Ecology and Biogeography **14:** 17-30

Austin MP (2002) Case Studies of the Use of Environmental Gradients in Vegetation and Fauna Modeling: theory and Practice in Australia and New Zealand. *In* JM Scott, PJ Heglund, ML Morrison, JB Haufler, MG Rapheal, WA Wall, FB Samson, eds,

Austin MP (2002) Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. Ecological Modelling **157:** 101-118

Austin MP, Belbin L, Meyers JA, Doherty MD, Luoto M (2006) Evaluation od statistical models used for predicting plant species distributions: Role of artificial data and theory. Ecological Modelling

Austin MP, Meyers JA (1996) Current approaches to modelling the environmental niche of eucalyptus: implication for management of forest biodiversity. Forest Ecology and Management **85:** 95-106

Azam-Ali S, Bonkoungou E, Bowe C, deKock C, Godara A, Williams JT (2006) Ber. Southampton Centre for Underutilised Crops, Southampton UK

Azam-Ali SN, Sesay A, Karikari SK, Massawe FJ, Aguilar-Manjarrez J, Bannayan M, Hampson KJ (2001) Assessing the potential of an underutilsied crop - a case study using Bambara groundnut. Expl. Agric. **37:** 433-472

Azurdia C (2004) Compilador. Frutales Sapotaceos Centro Americanos Proyecto Diversidad, Convervacion, Cooperacion y Uso Sostenible de los Recursos Ceneticos de Frutales Tropicales Nativos de America Tropical. Parte1. Sapotaceas. IPGRI. *In*,

Azurdia C (2006) Tres Especies de Zapote en America Tropical. Southampton Centre for Underutilised Crops , Universidad de Southampton, Southampton, UK.

Baguette M (2004) The classical metapopulation theory and the real natural world: a critical appraisal. Basic and Applied Ecology **5:** 213-224

Baker RJ, Phillips CJ, Bradley RD, Burns JM, Cooke D, Edson GF, Haragan DR, Jones C, Monk RR, Montford JT, Schmidly DJ, Parker NC (1998) Bioinformatics, Museums and Society: Integrating Biological Data for Knowledge-Based Decisions. Occasional Papers Museum Texas Tech Univ. **187:** 1-4

Barry S, Elith J (2006) When things go wrong: error and uncertainty in habitat models. Journal of Applied Ecology **43:** 413-423

Barry SC, Welsh AH (2002) Generalized additive modelling and zero inflated count data. Ecological Modelling **157:** 179-188

Batjes NH (1995) A Global dataset of Soil pH Properties. *In* Technical Paper 27. ISRIC, Wageningen

Batjes NH (1996) A Global Data Set of Derived Soil Properties by FAO-UNESCO Soil Unit. Draft, ISRIC Wageningen

Batjes NH (1997) A world dataset derived soil properties by the FAO-UNESCO soil unit for global modelling. Soil use and Management **13:** 9-16

Batjes NH (2002) Soil Parameter Estimates for the Soil Types of the world for use in Global and Regional Modelling (version 2.1) ISRIC Report 2002/02c. *In* IFPRIaISRaI Centre, ed,

Batjes NH, Bridges EM (1994) Potential emissions of radiatively active gases from soil to atmosphere with special reference to methane: development of a global database (WISE). Journal of Geophysical Research **99(D8):** 16479-16489

Batjes NH, Bridges EM, Nachtergaele FO (1995) World Inventory of Soil Emission Potentials: development of a global soil data base of process-controlling factors. *In* S Peng, KT Ingram, HU Neue, LH Ziska, eds, Climate change and Rice. Springer-Verlag, Heidelberg, pp 102-115

Beaumont LJ, Hughes L, Poulsen M (2005) Predicting species distribution: use of climatic parameters in BIOCLIM and its impact on predictions of species' current and future distributions. Ecological Modelling **186:** 250-269

Beerling DJ, Huntley B, Bailey JP (1995) Climate and distribution of *Fallopia japonica*: use of an introduced species to test the predictive capacity of response surface. Journal of Vegetation Science **6:** 269-282

Binzenhofer B, Schroder B, Strauss B, Biedermann R, Settele J (2005) Habitat models and habitat connectivity analysis for butterflies and burnet moths – The example of *Zygaena carniolica* and *Coenonympha arcania*. Biological Conservation **126:** 247-259

Bisby FA (2000) The quiet resolution, biodiversity infomatics and the Internet. Science **289:** 2309-2312

Bolliger J, Kienast F, Bugmann H (2000) Comparing models for tree distributions: concept, structures and behavior. Ecological Modelling **134:** 89-102

Booth TH (1995) From Third International Conference/Workshop on Integrating GIS and Environmental Modelling, Santa Fe, New Mexico, 21-25 January 1995, Proceedings published on WWW and CDROM by National Center for Geographic Information and Analysis, Santa Barbara.

Booth TH (1999) Matching Germplasm to Geography: Environmental Analysis for Plant Introduction. *In* SL Greene, L Guarino, eds, LInking Genetic Resources and Geography: Emerging Strategies to Conserving and Using Crop Biodiversity, CSSA Special Publication Number 27,

Bouman BAM, van Kculen H, van Laar HH, Rabbinge R (1996) The 'School of de Wit' Crop Growth Simulation Models:A Pedigree and Historical Overview. Agricultural Systems **52:** 171-198

Bowe C (2007) Potential Answers to the Adaptation too and Mitigation of Climate Change through the Adoption of Underutilised Crops. Tropical Agriculture Association Newsletter **27** 9-13

Box EO (1996) Plant functional types and climate at the global scale. Journal of Vegetation Science **7:** 309-320

Boyce MS, Vernier PR, Nielsen SE, Schmiegelow FKA (2002) Evaluating resource selection functions. Ecological Modelling **157:** 281-300

Bragazza L, Gerdol R (1996) Response surfaces of plant species along water-table depth and pH gradients in a poor mire on the southern Alps (Italy). Annales Botanici Frennici **33:** 11-20

Brandis T (1921) Indian Trees 13. *Tamarindus* Linn.. Constable & Co. Ltd., London

Brenan JPM (1967) Leguminosae, Subfamily Caesalpinioidae. In: Flora of Tropical East Africa: 151-153. Agents for Overseas Governments and Administrations London

Bryan TL, Mataxas A (2007) Predicting suitable habitat for deep-water gorgonian corals on the Atlantic and Pacific Continental Margins of North America. Marine Ecology Progress Series **330:** 113-126

Bui EN, Moran CJ (2001) Disaggregation of polygons of surficial geology and soil maps using spatial modelling and legacy data 2. Geoderma **103:** 79–94

Bui EN, Moran CJ (2003) A strategy to fill gaps in soil survey over large spatial extents: an example from the Murray– Darling basin of Australia. Geoderma **111:** 21–44

Burrough PA, McDonnell RA (1998) Principles of Geographical infomation Systems. Oxford University Press, Oxford UK

Busby JR (1991) BIOCLIM - A Bioclimatic Analysis and Prediction system. *In* CR Margules, MP Austin, eds, Nature Conservation: Cost Effective Biological Surveys and Data Analysis,

Bydekerke L, Van Ranst E, Vanmechelen L, Groenemans R (1998) Land suitability assessment for chermoya in southern Ecuador using expert knowledge and GIS. Agriculture, Ecosystems and the Environment **69:** 89-98

Cain ML, Damman H, Muir A (1998) Seed Dispersal and the Holocene migration of woodland herbs. Ecological Monographs **68:** 325-347

Campbell BM (1987) The use of wild fruit in Zimbabwe. Economic Botany: 375-385

Carpenter G, Gillison AN, Winter J (1993) DOMAIN - a flexible modelling procedure for mapping potential distribution of plants and animals. Biodiversity and conservation **2:** 667-680

Cassinello J, Acevedo P, Hortal J (2006) Prospects for the population expansion of the exotic aoudad (*Ammotragus lervia*; Bovidae) in the Iberian Peninsula: clues from habitat suitability modelling. Diversity and Distributions **12:** 666-678

Chanda K (2001) Evaluation of Tamarind (*Tamarindus indica* L.) Genotypes for Productivity and Grafting. University of Agricultural Sciences, Dharwad

Chapman AD (1999) Quality Control and Validation of Point-Sourced Environmental Resource Data. *In* Kim Lowell, Annick Jaton, eds, Spatial accuracy assessment: Land Information Uncertainty in Natural Resources, pp 409-418

Chapman AD (2004) Internal Report No. 5 to Environmental Data Quality. *In* Internal Report to CIRA,

Chapman AD, Munoz ME, Koch I (2005) Environmental Information: Placing Biodiversity Phenomena in a Ecological and Environmental Context. Biodiversity Infomatics **2:** 24-41

Chatturvedi AN (1985) Firewoood farming on degraded lands of the Gangetic plain. U.P. *Forest Bulletin*, Lucklow, India, Government of India Press 1 **50:** 286

Chefaoui RM, Hortal J, Lobo JM (2005) Potential distribution modelling, niche characterisation and conservation status assessment in using GIS; a case study of Iberian *Copris* species. Biological conservation **122:** 327-338

Chen SC, Da Fonseca LB (1980) Corn yield model for Ribeirao Preto, Sao Paulo State, Brazil. Agricultural Meteorology **22:** 341-349

Chundawat DJ (1990) Arid Fruit Culture Oxford IBH Publishers, New Delhi

Clark JS (1998) Why Trees Migrate So Fast: Confronting Theory with Dispersal Biology and the Paleorecord. The American Naturalist **152:** 204-224

Coates-Palgrave K (1998) Trees of Southern Africa 10. Tamarindus indica L.. C. S. Struik Publisher, Cape Town

Collins W, Hawtin G (1998) Conserving and using crop plant biodiversity in agroecosystems. In W.W.Collins snf C. Qualset (eds.), Biodiversity in agroecosystems. CRC Press, Washington, D.C.

Conn BJ (1996) HISPID3. Herbarium Information Standards Protocols for Interchange of Data. Version 3. *In*, Sydney: Royal Botanic Gardens

Conn BJ (2000) HISPID4. Herbarium Standards and Protocals for Interchange of Data. Version 4 - Internet version only. *In*, Sydney: Royal Botanic Gardens

Constable JVH, Friend AL (2000) Suitability of process-based tree growth models for addressing tree response to climate change. Environmental Pollution **110:** 47-59

Coronel RE (1991) *Tamarindus indica* L. In Plant Resources of SouthEast Asia Wageningen, Pudoc No. 2 Edible fruits and nuts (eds) Verheij, E. W. M. and Coronel, R. E. PROSEA Foundation, Bogor, Indonesia 298-301

Coudun C, Gegout JC, Piedallu C, Rameau JC (2006) Soil Nutritional factors improve models of plant species distribution: an illustration with *Acer campestre* (L.) in France. Journal of Biogeography **33:** 2006

Dale IR, Greenway PJ (1961) Kenya Trees and Shrubs: 654. Buchanan's Estate Ltd

Dalziel JM (1937) The Useful Plants of West Tropical Africa *In*. Crown Agents for Overseas Governments and Administrations: 612 p.

Davis T, Bye RA (1982) Ethnobotany and progressive domestication of Jaltomata (*Solanaceae*) in Mexico and Central America Economic Botany **36:** 225-241

De'Ath G (2002) Multivariate regression tree: a new technique for model species-environment relationships. Ecology **83:** 1105-1117

de Koning GHJ, Jansen MJW, Boons-Prins ER, van Diepen CA, Penning de Vries FWT (1993) Crop growth simulation and statistical validation for regional yield forecasting across the European Community.

Desai PN (2005) Challanges of agro-biotechnologies, intellectual property rights and globalisation: role of academic insitutions in achieveing the millennium development goals. World Review of Science, Technology, and Sustainable Development **2:** 60-71

Dingkuhn M, Penning de Vries FWT, Meizan KM (1993) Improvement of rice plant type concepts: systems research enables interaction of physiology and breeding. *In* FWT Penning de Vries, P Teng, K Metselaar, eds, Systems approaches for sustainable agricultural development. Kluwer Academic Publishers, Dordrecht, The Netherlands

Dobos E, Micheli E, Baumgardner MF, Biehl L, Helt T (2000) Use of combined digital elevation model and satellite radiometric data for regional soil mapping. Geoderma **97:** 367– 391

Dobos E, Montanarella L, Negre T, Micheli E (2001) A regional scale soil mapping approach using integrated AVHRR and DEM data. International Journal of Applied Earth Observation and Geoinformation **3:** 30–41

Drummond ST, Sudduth KA, Joshi A (2000) Predictive ability of neural networks for site-specific yield estimation. The Second International

Geospatial Information in Agriculture and Foresty Conference, Lake Buena Vista, Florida, 10–12 January.

Duke JA, Terrell EE (1974) Crop diversification matrix: Introduction. Taxon, 23 (5-6):. 759-799

Ecoport (2001) *Tamarindus indica* L. *In*, Vol 2007, p http://ecoport.org/

El-Siddig K (2006) Agronomy and Production. *In* JT Williams, RW Smith, N Haq, Z Dunsiger, eds, Tamarind *Tamarindus indica* L., Vol Southampton. Southampton Centre for Underutilised Crops

El-Siddig K, Ebert G, Luedders P (1999) Tamarind (*Tamarindus indica* L.): A review on a multipurpose tree with promising future in the Sudan. Journal of Applied Botany 73, 5-6

El-Siddig K, Gunasena HMP, Prasad BA, Pushpakumara DKNG, Ramana KVR, Vijayanand P, Williams JT (2006) Tamarind, *Tamarindus indica* L. Southampton Centre for Underutilised Crops, Southampton, UK

Elith J, H. Graham C, P. Anderson R, Dudik M, Ferrier S, Guisan A, J. Hijmans R, Huettmann F, R. Leathwick J, Lehmann A, Li J, G. Lohmann L, A. Loiselle B, Manion G, Moritz C, Nakamura M, Nakazawa Y, McC. M. Overton J, Townsend Peterson A, J. Phillips S, Richardson K, Scachetti-Pereira R, E. Schapire R, Soberon J, Williams S, S. Wisz M, E. Zimmermann N (2006) Novel methods improve prediction of species' distributions from occurrence data. Ecography **29:** 129-151

Engels JMM, Arora RK, Guarino L (1995) A introduction to plant germ plasm exploration and collecting: planning, methods and procedures, follow up. *In* L Guarino, V Ramanatha Rao, R Reid, eds, Collecting plant genetic diversity, Technical guidelines. CAB International, Wallingford, United Kingdom, pp 31-63

ESRI (2000) Arc View 3.2a.

ESRI (2000) ArcView 3.2 - World data.

Estrada-Pena A, Venzal JM, Sanchez Acedo C (2006) The tick *Ixodes ricinus*: distribution and climate preferences in the western Palaerctic. Medical and Veterinary Entomology **20:** 189-197

Evenson RE, Gollin D (2003) Assessing the Impact of the Green Revolution, 1960 to 2000. Science **300:** 758-762

FAO (1971-1981) The FAO-Unesco Soil Map of the World. Legend and 9 volumes. FAO Unesco, Paris.

FAO (1974) Soil Map of the World 1:5000000. Volume I. Legend, UNESCO Paris

FAO (1977) Crop water requirements. *In* R Food and Agriculture Organisation of the United Nations, ed, Irrigation and Drainage paper 56,

FAO (1979) Yeild response to water, Vol 33. FAO, Rome

FAO (1984a) Land Evaluation for Development. , Food and Agriculture Organization of the United Nations Rome.

FAO (1984b) Guidelines: Land Evaluation for Rain-fed Agriculture. FAO Soils Bulletin 52. Food and Agriculture Organization of the United Nations, Rome.

FAO (1985) Guidelines: Land Evaluation for Irrigated Agriculture. FAO Soils Bulletin 55.Food and Agriculture Organization of the United Nations, Rome. *In*,

FAO (1988) Fruit Bearing Trees Technical notes FAO-SIDA Forestry Paper 165-167

FAO (1995) Digital Soil Map of the World and Derived Soil Properties,

FAO (1996) The Digitized Soil Map of the World Including Derived Soil Properties. CD-ROM, FAO Rome

FAO (1996a) Global plan of Action for Conservation and Sustainable Utilisation of Plant Genetic Resources and the Leipzig Declaration, adopted by the International Technical Conference in Plant Genetic Resources. Leipzig, germany 17 - 23 June 1996

FAO (1998) Crop evapotranspiration - Guidelines for computing crop water requirements. FAO Irrigation and drainage paper 56

FAO (2002) Maximum avialable soil moisture. *In*. GEO NETWORK

FAO (2002) TERRSTAT Global land resources GIS models and databases and land degradation assessment, Beta version 1.0. *In* FAO Land and Water Digital Media Series 20., Food and Agricultural Organisation of the United Nations Rome, Land and Water Development Division

FAO (2003) FAO Digital Soil Map of the World (ver.3.6),

FAO (2006) Towards the Summit commitments, The way ahead: strengthening efforts for eradicating hunger. *In* The State of Food Insecurity in the World; Eradicating world hunger-taking stock ten years after the World Food Summit,

FAO (2006) Undernourished around the world;Counting the hungry: trends in the developing world and countries in transition. *In* The State of Food Insecurity in the World; Eradicating world hunger-taking stock ten years after the World Food Summit,

FAO/IIASA (2000) Global Agro-Ecological Zones (Global -AEZ). Food and Agricultural Organisation of the United Nations, Rome

Ferguson ME, Jarvis A, Stalker TH, Williams D, Guarino L, Valls JFM, Pittman RN, Simpson CE, Bramel PJ (2005) Biogeography of wild Arachis (Leguminosae): distribution and environmental characterisation. Biodiversity and conservation **14:** 1777–1798

Feria TP, Conaiglio T, Espinosa D, Ivan J (2007) Niche breath and niche position predict global and regional range size across *Bursera* species (BURSERACEAE). *In* Botany and Plant Biology 2007 - Joint Congress, Chicago, Illinois

Ferrier S (2002) Extended statistical approaches to modelling spatial pattern in biodiversity in northeast New South Wales. II. Community level modelling. Biodiversity and Conservation **11:** 2309-2338

Fischer G, Van Velthuizen HT, Nacjtergaele FO (2000) Global Agro-ecological Zones Assessment: Methodology and results, IR-00-064,. *In* FaAOotUNaIIfASA (IIASA), ed,

Fischer G, Van Velthuizen HT, Prieler S (2001) Assessment of potentail productivity of tree species in China, Mongolia and Former Soviet Union:

Methodology and Results. IR-01-015. *In* L International Institute for Applied Systems Analysis (IIASA), ed,

Foody GM (2004) Spatial nonstationarity and scaledependency in the relationship between species richness and environmental determinants for the sub-Saharan endemic avifauna. Global Ecology and Biogeography **13:** 315-320

Fresco LO (2003) “Which Road Do We Take?”Harnessing Genetic Resources and Making Use of Life Sciences, a New Contract for Sustainable Agriculture, Food and Agriculture Organization of the United Nations (FAO)

Galang FG (1955) Fruits and Nuts Grown in the Phillippines Rizal AIA Printing Press, Malabon

Gallego D, Canovas F, Esteve MA, Galian J, F. S (2004) Descriptive biogeography of Tomicus (Coleoptera : Scolytidae) species in Spain. Journal of . Biogeography **31:** 2011-2024

Gaubert P, Papes M, Peterson TA (2006) Natural history colletcions and the conservation of poorly known taxa: Ecological niche modeling in central African rainforest genets (*Genetta* spp.). Biological Conservation **130:** 106-117

Gevrey M, Worner SP (2006) Predictions of Global Distribution of Insect Pest Species in Relation to Climate by Using an Ecological Informatics Method. Journal Of Economic Entomology **3:** 2006

GNS- National Geospatial-inteligence agency (2005) U.S. National Imagery and Mapping Agency's (NIMA) database of foreign geographic feature names - Country Files.

Gonzalez P, Castro E, Rocha OJ (2000) Ecogeographic survey of the distribution of Lima beans (Phaseolus lunatus) in the Central Valley of Costa Rica (poster). *In* International Conference on Science and Technology for Managing Plant Genetic Diversity in the 21st Century. IPGRI, Kuala Lumpur, Malaysia

Graham CH, Ferrier S, Huettman F, Moritz C, Peterson TA (2004a) New developments in museam-based infomatics and applications in biodiversity analysis. Trends in Ecology and Evolution **19:** 497-503

Graham CH, Ron SR, Santos JC, Schneider CJ, Moritz C (2004) Integrating Phylogentics and Environmental Niche Models to Explore Speciation Mechanisms in Dendrobatid Frogs. Evolution **8:** 1781-1793

Grinnel J (1924) Geography and evolution. Ecology **5:** 225-229

Grinnel J (1928) Presence and absence of animals. Univ. Calif. Chron **30:** 429-450

Guisan A, Theurillat J-P (2000) Equilibrium modeling of Alpine plant distribution: How far can we go? Phytocoenologia **30:** 353-384

Guisan A, Thuiller W (2005) Predicting species distribution: offering more than simple habitat models. Ecology Letters **8:** 993-1009

Guisan A, Zimmermann NE (2000) Predictive habitat distribution models in ecology. Ecological Modelling **135:** 147-186

Gunasena HMP, Pushpakumara DKNG (2006) Chapter 8. Genetic Resources *In* Tamarind, *Tamarindus indica* L. Southampton Centre for Underutilised Crops, Southampton

Gunasena HPM, Hughes A (2000) *Tamarindus indica* L. International Centre for Underutilised Crops, University of Southampton

Gunasena HPM, Pushpakumara DKNG (2006) Chapter 7. Genetic improvement. *In* JT Williams, RW Smith, N Haq, Z Dunsiger, eds, Tamarind, Tamarindus indica L. Southampton Centre for Underutilised Crops, Southampton

Hackett C, Vanclay JK (1998) Mobilizing expert knowledge of tree growth with PLANTGRO and INFER systems. Ecological Modelling **106:** 233-246

Hall JB, O'Brien EM (2002) Ecology. *In* JB Hall, EM O'Brien, FL Sinclair, eds, Sclerocarya birrea, a monograph. School of Agricultural and Forest Sciences Publication Number 19, Univeristy of Wales Bangor, p 157

Haq N (1989) Crop Plants: Potential for Food and Industry. *In* Wickens, N Haq, eds, New Crops for Food and Industry. Chapman and Hall

Haq N (1995) The Development of underutilised crops and their role in sustainable production. *In* P de Groot, N Haq, eds, Promotion of Traditional and Underutilised Crops. Commonwealth Science Council, Valletta,Malta 1992

Haq N (2004) Women reintroducing neglected crops. LEISA Magazine **20:** 28 - 29

Haq N (2007) Underutilised Fruits: A Resource for Sustainable Livelihoods. *In* Breeding of negelected and Underutilised crops

Hausser JE (1995) Saugetiere der Schweitz. Atlas de Suisse. Mammiferi della Svizzera.

Heglund PJ (2002) Foundations of Species-Environment Realtions. *In* JM Scott, PJ Heglund, ML Morrison, JB Haufler, MG Rapheal, WA Wall, FB Samson, eds, Predicting Species Occurrences: Issues of Accuracy and Scale. Island Press, Washington

Hijmans1 RJ, Cameron SE, Parra JL, Jones PG, Jarvis A (2005) Very High Resolution Interpolated Climate Surfaces for Global Land Areas. Int. J. Climatol. **25:** 1965–1978

Hijmans R, Guarino L, Jarvis A, O'Brien R (2007) DIVA-GIS.

Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A (2004) The WorldClim interpolated global terrestrial climate surfaces. Version 1.3. *In* Aa http://biogeo.berkeley.edu/, ed,

Hijmans RJ, Garrett KA, Huaman Z, Zhang DP, Schreuder M, Bonierbale M (2000) Assessing the Geographic Representativeness of Genebank Collections: the Case of Bolivian Wild Potatoes. Conservation Biology **14:** 1755-1765

Hijmans RJ, Schreuder M, De la Cruz J, Guarino L (1999) Using GIS to check coordintes of genebank accessions. Genetic Resources and Crop Evolution **46:** 291-296

Hirzell AH, Arlettaz R (2003a) Modeling Habitat Suitability for Complex Species Distributions by Environmental-Distance Geometric Mean. Environmental Management Vol. **Vol. 32:** pp. 614-623

Hirzell AH, Arlettaz R (2003b) Environmental-envelope based habitat-suitability models. *In* BFJ Manly, ed, Ist Conference on Resource Selection by Animals. Omnipress, Laramie, USA

Hirzel AH, Arlettaz R (2003a) Modeling Habitat Suitability for Complex Species Distributions by Environmental-Distance Geometric Mean. Environmental Management Vol. **Vol. 32:** pp. 614-623

Hirzel AH, Hausser J, Chessel D, Perrin N (2002) Ecological Niche Factor Analysis: How to compute Habitat-Suitability Maps without Absence Data. Ecology **83:** 2027-2036

Hirzel AH, Hausser J, Perrin N (2006) Biomapper 3.2. Lab of Conservation biology. Department of Ecology and Evolution, University of Lausanne, Switzerland. URL: http:\\www.unil.ch/biomapper.

Hirzel AH, Helfer V, Metral F (2001) Assessing habitat-suitability models with a virtual species. Ecological Modelling **145:** 111-121

Hirzel AH, Le Lay G, Helfer V, Randin C, Guisan A (2006) Evaluating the ability of habitat suitability models to predict species presences. Ecological MOdelling **199:** 142-152

Hirzel AH, Posse B, Oggier P-A, Crettenand Y, Glenz C, Arlettaz R (2004) Ecological requirements of reintroduced species and implications for release policy: the case of the bearded vulture. Applied Ecology **41:** 1103 - 1116

Holt RD (1996) Demographic constraints in evolution: towards unifying the evolutionary theories of senescence and niche conservation. Evolutionary Ecology **10:** 1-11

Holt RD, Gaines MS (1992) Analysis of adaptation in heterogeneous landscapes: implications for the evolution of fundamental niches. Evolutionary Ecology **6:** 433-447

Horie T, Yajima M, Nakamura M (1992) Yield Forcasting. Agricultural Systems **40:** 211-236

Hort A (1916) Theophrastus; Enquiry into Plants. Heinemann, London

Houlder D, Hutchinson M, Nix H, McMahon J (2000) ANUCLIM user's guide. CRES

Hughes A (in press) Tamarind. *In* J Janick, ed, Encyclopaedia of Fruits and Nuts,

Hunt LA (1993) Designing improved plant types : a breeder's view point. *In* FWT Penning de Vries, P Teng, K Metselaar, eds, Systems approches for sustainable agricultural development. Kluwer Academic Publishers, Dordrecht, The Netherlands

Huntley B, Bartlein PJ, Prentice IC (1989) Climate control of the distribution and abundance of beech (*Fagus* L.) in Europe and North America. Journal of Biogeography **16:** 551-560

Hutchinson GE (1957) Concluding remarks. Cold Spring Harbor Symposium on Quantitative Biology **22:** 415-427

Hutchinson MF (1995) Interpolating mean rainfall using thin plate smoothing splines. International Journal of Geographical InformationSystems **9:** 385–403

Hutchinson MF (1996) A locally adaptive approach to interpolation of digital elevation models. *In* NCGIA, ed, Proceedings of the Third International Conference Integrating GIS and Environmental Modelling. University of California, Santa Barbara, National Center for Geographic Infomation and Analysis, Santa Fe, New Mexico

Hutchinson MF (2003a) Topographic and Climate Database for Africa. Version 1.0. *In*. Canberra: Centre for Resource and Environmental Studies, Australia National University

Hutchinson MF (2004) Anusplin Version 4.3. *In*. Centre for Resource and Environmental Studies. The Australian National University: Canberra, Australia

Hutchinson MF, Bischof RJ (1983) A new method for estimating the spatial distribution of mean seasonal and annual rainfall applied to the Hunter Valley, New South Wales. *In* Austral. Met. Mag, pp 179-184

ICUC (2006) The International Centre for Underutilised Crops, Championing underutilised plant species for food nutrition and sustainable development.

ILDIS (International Legume database and Information Service) (2007) Tamarindus indica. Tamarind **Report generated by LegumeWeb from the ILDIS World Database of Legumes, version 10.01 (accessed march 2007)**

IMWI (2005) IMWI Climate Atlas. http://www.iwmi.cgiar.org/WAtlas/atlas.htm **2005**

InWEnt/GFU (2003) Proceedings of the International Workshop on Underutilized Plant Species. 6-8 May 2003, Leipzig Germany. *In* S Gündel, I Höschle-Zeledon, B Krause, K Probst, eds,

IPCC (2001) Technical Summary, Climate change 2001: Impacts, Adaptation and Vulnerability. *In* IPCC Third Assessment Report – Climate Change 2001,

Irvine FR (1961) Woody Plants of Ghana: 868. Oxford University Press, London

James FC, Johnston RF, Wamer NO, Niemi GJ, Boecklen WJ (1984) The Grinnellian Niche of the Wood Thrush. The American Naturalist **124:** 17-47

Jarvis A, Ferguson ME, Williams DE, Guarino L, Jones PG, Stalker TH, Valls JFM, Pittman RN, Simpson CE, Bramel PJ (2003) Biogeography of Wild Arachis: Assessing Conservation Status and Setting Future Priorities. Crop Science **43:** 1110-1108

Jarvis A, Fisher M, Jones P, Cook S, Guarino L (2006) Agriculture, Risk and Climate Change, CGIAR (http://www.cgiar.org/pdf/2006_Jarvis%20and%20others-Ag_Risk_ClimateChange_2006_FINAL.pdf),

Jarvis A, Guarino L, Williams D, Williams K, Vargas I, Hyman G (2002) Spatial analysis of wild peanut distributoins and implications for plant genetic resources conservation. Plant Genetic Resources Newsletter: 28-34

Jeffree EP, Jeffree CE (1994) Temperature and Biogeographical Distributions of Species. Functional Ecology **8:** 640-650

Jetz W, Rahbek C, Lichstein JW (2005) Local and global approaches to spatial data analysis. Global Ecology and Biogeography **14:** 97-98

Johnston JF, Chapin SF (2003) Non-equilibrium sucession dynamics indicate continued northern migration of lodgepole pine. Global Change Biology **9:** 1401-1409

Jones PG, Gladkov A (1999) FloraMap: A Computer Tool for the Distribution of Plants and Other Organisms in the Wild. CIAT, Cali, Colombia

Jones PG, Thornton PK (2003) The potential impacts of climate change on maize production in Africa and Latin America in 2055. Global Environmental Change **13:** 51-59.

Kandji ST, Verchot V, Mankensen J, Boye A, van Noordwijk M, Tomich T, Ong C, Albrecht A, C. P (2006) Opportunities for linking climate change adaptation and mitigation through agroforestry systems. *In* DP Garrity, A Okono, M Grayson, S Parrott, eds, World Agroforestry into the Future. World Agroforestry Centre - ICRAF, Nairobi, Kenya

Kiambi DK, Ford-Lloyd BV, Maxted N, Newbury HJ, Guarino L, Jarvis A (2000) Predicting the distribution of five wild *Oryza* spp in East and Southern Africa (Poster). *In* International Conference on Science and Technology for Managing Plant Genetic Diversity in the 21st Century. IPGRI, Kuala Lumpur, Malaysia

Knox JW, Mathews RB, Wassmann R (2000) Using a crop/soil simulation model and GIS techniques to assess methane emissions from rice fields in Asia. III: Databases. Nutrient Cycling in Agroecosystems **58:** 179-199

Knyazhnitskiy OV, Monk RR, Parker NC, Baker RJ (2000) Assignment of Global Information System Co-oridinates to Classical Museum Localities for Relational Dabase analysis. Occasional Papers, Museum of Texas Tech University **199**

Kropff MJ, Haverkort AJ, Aggarwal PK, Kooman PL (1995) Using system approaches to design and evaluate ideotypes for specific enviornments. *In* JK Bouma, A Kuyvenhoven, BAM Bouman, JC Luyten, HG Zandstra, eds, Ecoregional approaches for sustainable land use and food production. Kluwer Academic Publishers, Dordrecht, The Netherlands

Kuhnlein HV (2001) Progress report: Documenting traditional food systems of indigenous peoples in Asia. FAO, Rome

Kwesiga F, Mwanza S (1997) Underexploited wild genetic resources: the case of indigenous fruit trees of eastern Zambia. *In*,

Kyddl J, Dorward A, Morrison J, Georg C (2004) Agricultural Development and Pro-poor Economic Growth in Sub-Saharan Africa: Potential and Policy. Oxford Development Studies **32:** 37-57

Kydd J, Dorward A, Morrison J, Georg C (2004) Agricultural Development and Pro-poor Economic Growth in Sub-Saharan Africa: Potential and Policy. Oxford Development Studies **32:** 37-57

Landau S, Mitchell RAC, Barnett V, Colls JJ, Craigon J, Payne RW (2000) A parsimonious, multiple-regression model of wheat yield response to environment. Agricultural and Forest Meteorology **101:** 151–166

Leathwickl JR (1998) Are New Zealand's Nothofagua species in equilibrium with their environment? Journal of Vegetation Science **9:** 719-732

Leathwick JR (1992) Forest pattern, climate and valcanism in central North Island, New Zealand. Journal of Vegetation Science **3:** 603-616

Leathwick JR (1998) Are New Zealand's Nothofagua species in equilibrium with their environment? Journal of Vegetation Science **9:** 719-732

Lehmann A, Overton JM, Leathwick JR (2002) GRASP: generalized regression analysis and spatial prediction. Ecological modelling **157:** 189-207

Leonard J (1957) Genera des Cynometreae et des Amherstieae africaines (Leguminosae-Caesalpinioideae). *Memoire Acad. Royal Belgique* **30:** 1-314

Levin SA (1992) The problem of pattern and scale in ecology. Ecology **73:** 1943-1967

Lewis G, Schrire B, Mackinder B, Lock M (2005) Legumes of the World, Royal Botanic Gardens, Kew.

Lillesø JPB, Dhakal LP, Shrestha TB, Nayaju RP, Shrestha R, Kjaer ED (2001) Tree Planting Zones in Nepal - an ecological approach based on vegetation types. DFSC Case Study No.1. TISC Technical Paper No. 103. Danida Forest Centre, Humlebæk. Tree Improvement and Silviculture Component, Kathmandu.

Lindenmayer DB, Nix HA, McMahon JP, Hutchinson MF, Tanton MT (1991) The Conservation of leadbeaters Possum, *Gymmobelideus leadbeateri* (McCoy): A Case Study of the Use of Bioclimatic Modelling. Journal of Biogeography **18:** 371-383

Loisellel BA, Howell CA, Graham CH, Goerck JM, Brooks T, Smith KG, Williams PH (2003) Avoiding the Pitfalls of Using Species Distribution Models in Conservation Planning. Conservation Biology **17:** 1591 - 1600

Louw JH, Scholes MC (2006) Site index functions using site descriptors for Pinus patula plantations in South Africa. Forest Ecology and Management **225:** 94-103

MacArthur RH (1960) On the Relative Abundance of Species. American Naturalist **94**

Mahoneyl D (1990) Trees of Somalia *Tamarindus indica* Oxfam Press Oxford

Manel S, Dias JM, Buckton ST, Ormerod SJ (1999a) Alternative methods for predicting species distribution: an illustration with Himalayan river birds. Journal of Applied Ecology **36:** 737-747

Margules CR, Redhead TD (1995) BioRap. Guideline for using the BioRap Methodology and Tools. *In*. CISRO, Canberra, p 70

Matson PA, Parton WJ, Power AG, Swift MJ (1997) Agricultural Intensification and Ecosystem Properties. Science **277:** 504

Matthews RB, Kropff MJ, Bachelet D, van Laar HH (1995) Modelling the impact of climate change on rice production in Asia. CAB International, Wallingford, UK

McBratney AB, Mendonca Santos ML, Minasny B (2003) On Digital soil mapping. Geoderma **117:** 3-52

McGarigal K, Cushman S, Stafford S (2000) Multivariate Statistics for Wildlife and Ecology Research. Springer - Verlag New York, Inc

McNeely JA, Schutyser F (2003) Let a thousand flowers bloom: How can underutilised plant species contribute to biodiversity objectives. *In* InWEnt/GFU, ed, Proceedings of the International Workshop on Underutilized Plant Species. 6-8 May 2003, Leipzig, Germany

McPherson JM, Jetz W, Rogers DJ (2006) Using coarse-grained occurrence data to predict species distributions at finer spatial resolutions-possibilities and limitations. Ecological Modelling **192:** 499-522

Microsoft (2003) Microsoft Office Excel 2003.

Microsoft (2003) Visual Studio.net. *In,*

Midgley GF, Hannah L, Millar D, Thuiller W, Booth A (2003) Developing regional and species-level assessments of climate change impacts on biodiversity in the Cape Floristic Region. Biological Conservation **112:** 27-97

Miller AJ, Knouft HJ (2006) GIS-based characterization of the geographic distributions of wild and cultivated populations of the Mesoamerican fruit tree, Spondias purpurea (Anacardiaceae). Journal of American Botany **93:** 1757-1767

Miller AJ, Knouft HJ (2006) GIS-based characterization of the geographic distributions of wild and cultivated populations of the Mesoamerican fruit tree, *Spondias purpurea* (Anacardiaceae). Journal of American Botany **93:** 1757 - 1767

Morrison ML, Hall LS (2002) Standard Terminology: Toward a Common Language to Advance Ecological Understanding and Application. *In* JM Scott, PJ Heglund, ML Morrison, JB Haufler, MG Rapheal, WA Wall, FB Samson, eds, Predicting Species Occurrences: Issues of Accuracy and Scale. Island Press, Washington

Morton J (1987) Fruits of Warm Climates (Miami FL). 115-121

Nachtergaele FO (1976) Studies of some Saline and Sodic Soils in Sudan. *In* Technical Report No. 24. Soil. Survey Admnistration. FAO/UNDP Project, Wad Medani, p 91

Nachtergaele FO (1999) From the Soil Map of the World to the Digital Global Soil and Terrain Database, 1960–2002. *In* ME Summer, ed, Handbook of Soil Science. CRC Press, Boca Raton, pp H5-17

NAS (1979) Tropical Legumes: Resources for the Future, Washington DC: 117 - 121

New M, Hulme M, Jones P (1999) Representing Twentieth-Century Space-Time Climate Variability. Part1 Development of a 1961-90 Mean Monthly Terrestrail Climatology. Journal of Climate **12:** 829-856

New M, Lister D, Hulme M, Makin I (2000) A high-resolution data set of surface climate over global land areas. Climate Research **21:** 1-25

New M, Lister D, Hulme M, Makin I (2002) A high-resolution data set of surface climate over global land areas. Clim Res **21**

Nix HA, McMahon JP, Mackenzie D (1977) Potential areas of production and the future of pigeon pea and other grain legumes in Australia. *In* In: The potential for pigeon pea in Australia. Proceedings of Pigeon Pea (*Cajanus cajan* (L.) Millsp.) Field Day (eds Wallis, E.S. & Whiteman, P.C.). University of Queensland, Queensland, Australia, pp. 5/1–5/12.,

Okafor JC, Lamb A (1992) Fruit trees: diversity and conservation strategies in Leakey and Newton A. C.: Tropical Trees: the potential for domestication and rebuilding of forest resources. *In* ITE symposium no. 29, ETCF symposium no. 1, Heriot Watt University Edinburgh

Olivier F, Wotherspoon SJ (2005) Modelling habitat selection using presence-only data: Case study of a colonial hollow nesting bird, the snow petrel. Ecological Modelling

Ong CK, Leakey RRB (1999) Why tree crop interactions in agroforestry appear at odds with tree-grass interactions in tropical savannah. Agroforestry Systems **45:** 109-129

Ortega M, Vackavik T (2007) Modeling Potential Distribution of Norway Maple (Acer platanoides) in Massachusetts, USA. *In* Clark University workshop, USA,

Osborne4 PE, Suarez-Seoane S (2002f) Should data be partioned spatially before building large-scale distribution models. Ecological Modelling **157:** 249-259

Osborne PE, Foody GM, Suarez-Seoane S (2007) Non - Stationarity and local approaches to modelling the distributions of wildlife. Biodiversity Research **13:** 313-323

Osborne PE, Suarez-Seoane S (2002) Should data be partioned spatially before building large-scale distribution models. Ecological Modelling **157:** 249-259

Padulosi S, Eyzaquirre PB, Hodgkin T (1999) Challenges and Strategies in Promoting Conservation and Use of Neglected and Underutilised Crop Species. *In* J Janick, ed, Perspectives on new crops and new uses, ASHA Press, Alexandria, VA

Padulosi S, Giuliani A, Noun J (2003) Underutilzed species: what are they? *In* International Workshop on Underutilized species, 6-8 May 2003, Leipzeg, Germany

Padulosi S, Hodgkin T, Williams JT, Haq N (2003) Underutilized crops: trends, challenges and opportunities in the 21st Century. CABI-IPGRI.

Padulosi S, Hoeschle-Zeledon I (2004) Underutilized plant species: What are they? ILEIA **20.1**

Park SJ, Hwang CS, Vlek PLG (2005) Comparison of adaptive techniques to predict crop yield response under varying soil and land management conditions. Agricultural Systems **85:** 59-81

Pasinelli G, Naef-Daenzer B, Schmid H, Keller V, Holzang O, Graf R, Zbinden N (2001) An avifaunal zonation of Switzerland and its realtion to environmental conditions. Global Ecology and Biogeography **10:** 261-274

Pearce JL, Boyce MS (2006) Modelling distribution and abundance with presence-only data. Journal of Applied Ecology **43:** 405-412

Pearson RG, Dawson TP, Berry PM, Harrison PA (2002) SPECIES:A Spatial Evaluation of Climate Impact on Envelope of Species. Ecological Modelling **154:** 289-300

Pearson RG, Dawson TP, Liu C (2004) Modelling species distribution in Britain: a hierarchical intergration of climate and land-cover data. Ecography **27:** 285-298

Penning de Vries FWT, Jansen DM, ten Berge HMF, Bakema A (1989) Simulation of Ecophysiological Processes of Growth in Several Annual Crops, Simulation Monographs, Pudoc/Internation Rice Research Institute, Wageningen,

Penning de Vries FWT, van Keulen H, Rabbinge R (1995) Natural resources and limits to food production in 2040. *In* JK Bouma, A Kuyvenhoven, BAM Bouman, JC Luyten, HG Zandstra, eds, Eco-regional approaches for sustainable land use and food production. Kluwer Academic Publishers, Dordrecht, The Netherlands

Perez P, Ardlie N, Kuneepong P, Dietrich C, Merritt WS (2002) CATCHCROP: modelling crop yield and water demand for integrated catchment assessment in Northern Thailand. Environmental Modelling & Software **17:** 251-259

Peterson TA (2003a) Predicting the geography of species' invasions via ecological niche modeling. The Quaterly Review of Biology **78:** 413-433

Peterson TA, Egbert SL, Sanchez-Cordero V, Price KP (2000) Geographic analysis of conservation priority: endemic birds and mammals in Veracruz, Mexico. Biological Conservation **93:** 85-94

Peterson TA, Holt RD (2003) Niche differentiation in Mexican birds: using point occurances to detect ecological innovation. Ecology Letters **6:** 774-782

Peterson TA, Holt RD (2003) Niche differentiation in Mexican birds: using point occurrences to detect ecological innovation. Ecology Letters **6:** 774-782

Peterson TA, Soberon J, Sanchez Acedo C (1999) Conservatism of thre ecological niches in evolutionary time. Science **285:** 1265-1267

Peterson TA, Soberon J, Sanchez Acedo C (1999) Conservatism of three ecological niches in evolutionary time. Science **285:** 1265-1267

Peterson TA, Vieglas A (2001) Predicting species invasions using ecological niche modeling: New Approaches from bioinfomatics attack a pressing problem. Bioscience **51:** 363-371

Pettigrew CJ, Watson L (1977) On the classification of Caesalpinioideae. Taxon **26:** 57-64

Phillips SJ, Anderson RP, Schapire RE (2006) Maximum entropy modeling of species geographic distributions. Ecological Modelling **190:** 231-259

Phillips SJ, Anderson RP, Schapire RE (2006) Maximum entropy modelling of species geographic distributions. Ecological Modelling **190:** 231-259

Pinto P, Gegout Jc (2005) Assessing the nutritional and climatic response of temperate forest tree species in the Vosges Mountians. Annals of Forest Science **62:** 1-10

Polhill D (1988) Flora of Tropical East Africa; Index of Collecting Localities. Royal Botanical Gardens, Kew

Pope GV, Pope DG (1998) Flora Zambesiaca, Collecting localities in the Flora Zambesiaca area.

Porte A, Bartelink HH (2002) Modelling mixed forest growth: a review of models for forest management. Ecological Modelling **150:** 141-188

Pramanik K (2003) Properties and use of jatropha curcus oil and diesel fuels blens in compression ignition engine. Renewable energy **28**

Prentice IC, Cramer W, Harrison SP, Leemans R, Monserud RA, Solomon AM (1992) A global biome model based on plant physiology and dominance, soil properties and climate. Journal of Biogeography **19:** 117-134

Prinzing A, Durka D, Klotz S, Brandl R (2001) The niche of higher plants: evidence of phylogenetic conservation. Proc. R. Soc. Lond. B **268:** 2383-2389

Prinzing A, Durka D, Koltz S, Barandl R (2002) Geographic variability of ecological niches of plant species: are competition and stress relevant? Ecography **25:** 721-729

Pulliam HR (1988) Source, Sink and Population regulation. American Naturalist **132:** 652-661

Pulliam HR (2000) On the realtionship between niche and distribution. Ecology Letters **3:** 349-361

Punjani BL, Kumar V (2002) Folk medicinal plants used for skin disorders in tribal pockets of Sabarkantha district, Gujarat. Journal of natural remedies **2**

Purseglove JW (1987) Tropical Crops, Dicotyledons: 204-206. Longman Science and Technology

Rabbinge R (1986) The bridge function of crop ecology. Netherlands Journal of Agricultural Science **3:** 239-251

Randin CF, Dirnbock T, Dullinger S, Zimmermann NE, Zappa M, Guisan A (2006) Are niche-based species distribution models transferable in species. Journal of Biogeography

Reddy YTN (1997) Orchard effeciency analysis in mango. Final report of Indian Council of Agricultural Research, Ad hoc scheme, 14 pp.

Reese GC, Wilson KR, Hoeting JA, Flather CH (2005) Factors affecting species distribution predictions: a simulation modelling experiment. Ecological Application **15:** 554-564

Relwani LL (1993) Tamarind, a multipurpose tree for diverse environments. The Bharathia Agricultural and Industrial Foundation Journal **13**

Reutter BA, Helfer V, Hirzel AH, Vogel P (2003) Modelling habitat-suitability using museum collections: an example with three sympatric *Apodemus* species from the Alps. Journal of Biogeography **30:** 581-590

Rich TCG (1997) Is *ad hoc* good enough? Trans. Sufolk Nat. Hist. Soc. **33:** 14-21

Ridley HN (1922) The Flora of the Malay Peninsula. 47. *Tamarindus* Linn. Vol. 1 Polypetalae L. : 636 Reeve & Co. Ltd. , London

Rimbau V, Cerdan C, Vila R, Iglesias A (1999) Antiinflamatory activity of some extracts from plants used in tradional medicine of North African Countries. Phytotherapy Research **13:** 128 - 132

Rondolo (2002) ICUC-UTFANET and its role in Asia. *In* N Haq, A Hughes, eds, Fruits for the Future in Asia. International Centre for Underutilised Crops

Rotter R, Dreiser (1994) Extrapolation of maize fertiliser trial results by using crop growth simulation: results for Murang'a District, Kenya. *In* LO Fresco, L Stroosnijder, JK Bouma, v Keulen, eds, The future of the land: mobilising and intergrating knowledge for land use options. John Wiley & Son Ltd., West Sussex, UK, pp 249-260

Sanchez AP, Buresh JR, Leakey RB (1997) Trees, soil and food security. 949-961

Scheldemanl X (2002) Distribution and Potential pf Cherimoya (*Annona cherimola* Mill.) and Highland Papayas (*Vasconcellea* spp.) in Ecuador. Universiteit Gent

Schluter D (2000) Ecological Character Displacement in Adaptive Radiation.

Schultz A, Wieland R, Lutze G (2000) Neural networks in agroecological modelling - stylish application or helpful tool? Computers and Electronics in Agriculture **29:** 73-97

Schulze RE, Kunz RP (1995) Potentail Shifts in Optimum Growth Areas of Selected Commercial Tree Species and Subtropical Crops in Southern Africa Due to Global Warming. Journal of Biogeography **22:** 679-688

Seligman NG (1990) The crop model record: promice or poor show? *In* R Rabbinge, H Goudriaan, H van Kculen, FWT Penning de Vries, HH van Laar, eds, Theoretical Production Ecology: reflection and prospects. PUDOC, Wageningen, The Netherlands, pp 294 - 264

Simmonds NW (1984) Evolution of Crop Plants. Longmans London

Soberon J (1999) Linking biodiversity information sources. Trends in Ecology and Evolution **14**

Soberon J, Townsend Peterson A (2005) Interpretation of models of fundamantal ecological niches and species distributional areas. Biodiversity Infomatics **2:** 1-10

Sokal RR, Rohlf FJ (1981) Biometry: the principle and prastices of statistics in biological research. W.H. Freeman, New York, New York, USA

Sombroek WG (1984) Towards a Global Soil Resources Inventory at Scale 1:1 Million. Discssion paper.

Sozolnoki TW (1985) Food and Fruit Trees of Gambia. Hamburg, Federal Republic of Germany: 132 Tiftung Walderhatung in Afrika

Stape JL, Ryan MG, Binkley D (2004) Testing the utility of the 3-PG model for growth of *Eucalyptus grandis* urophylla with natural and manipulated supplies of water and nutrients. Forest Ecology and Management **193:** 219–234

Stein BR, Wieczorek J (2004) Mammels of the world: MaNIS as an example of data integration in a distributed network environment. Biodiversity Informatics **1:** 14-22

Stephens W, Middleton T (2002) Why has the uptake of decision support systems been so poor. *In* R Mathews, W Stephens, eds, Crop-soil Simulation Models. CAB International, pp 129 - 147

Stockwell D, Peters D (1999) The GARP modelling system: problems and solutions to automated spatial prediction. International Journal of Geographical Information science **13:** 143-158

Stockwell D, Peterson TA (2001) Controlling Bais in biodiversity Data. *In* MJ Scott, PJ Heglund, ML Morrison, JB Haufler, MG Rapheal, WA Wall, FB Samson, eds, Predicting Species Occurances: Issues of Accuracy and Scale. Island Press, Wahington Covelo London

Suarez AD, Tsutusi ND (2004) The value of museum collections for research and society. BIOSCIENCE **54:** 66-74

Svenning J-C, Skov F (2004) Limited filling of the potential ranges in European tree species. Ecology Letters **7:** 565-573

Thuiller W, Araujo BM, Lavorel S (2003a) Generalized models vs. classification tree analysis: Predicting spatial distributions of plant species at different scales. 14

Thuiller W, Brotons L, Araujo BM, Lavorel S (2004) Effects of restricting environemntal range data to project current and future distributions. Ecography **27:** 165-172

Thuiller W, Richardson MD, Pysek P, Midgley GF, Hughes GO, Rouget M (2005) Niche-based modelling as a tool for predicting the risk of alien plant invasions at a global scale. Global Change Biology **11:** 2234-2250

Thuiller W, Vayreda J, Pino J, Sabate S, Lavorel S, Gracia C (2003) Large-scale environmental correlates of forest tree distribution in Catalonia (NE Spain). Global Ecology and Biogeography **12:** 313-325

Tilman D (1998) The greening of the green revolution. Nature **396:** 211-212

Titeux N (2006) Modelling species distribution when habitat occupancy departs from suitability; Application to birds in a landscape context. Universite catholique de Louvain

Tole L (2006) Choosing reserve sites probabilistically: A Colombian Amazon Case Study. Ecological modelling **194:** 344-356

Tole L (2006) Choosing reserve sutes probabilistically: A Colombian Amazon Case Study. Ecoloigical modelling **194:** 344-356

Troup RS (1921) The Silviculture of Indian Trees, Vol.11. Leguminosae (Caesalpinieae) to Verbenacae.7. *Tamarindus indica* L. Oxford, Clarendon Press

Udvardy MDF (1975) A classification of the biogeographical provinces of the world. IUCN Occasional Paper no. 18. Morges, Switzerland: IUCN.

UNCTAD Biofuels Initiative team (2006) The UNCTAD Biofuels Initiative (United Nations Conference on Trade and Development) (http://r0.unctad.org/ghg/projects/Biofuels%20flyer.pdf),

UNEP/ISSS/ISRIC/FAO (1995a) Global and National Soils and Terrain Digital Databases (SOTER). *In* Procedures Manual. World Soil Resources Report No. 74. FAO, Rome

Ungar ED (1990) Management of agropastoral systems in a semiarid region., Wageningen, The Netherlands

University of Colorado (2003) MaPSTeDI. A Guide to Georeferencing.

University of Kansas Biodiversity Reserach Centre (2002) Lifemapper.

USGS (1981) Geographic Names Information System.

Van Horn B (2002) Approaches to habitat modelling: the tension between pattern and process and between specificity and generality. *In* JM Scott, PJ Heglund, ML Morrison, JB Haufler, MG Rapheal, WA Wall, S F.B., eds, Predicting Species Occurrences: Issues of Accuracy and Scale. Island Press, Covelo, CA, pp 63-72

van Keulen H, Penning de Vries FWT (1993) Farming under uncertainty: terminology and techniques. *In* CSSo America, ed, International Crop Science1, 677 S.Segoe Road, Madison USA, pp 139-144

Vanclay JK (1994) Modelling Forest Growth and Yield Applications to Mixed Tropical Forests. CAB International, Wallingford UK

Vandermeer (1972) Niche Theory. Annual Review of Ecology and Systematics **3:** 107-132

Verchot LV, Mackensen J, Kandji S, van Noordwijk M, Tomich T, Ong C, Albrecht A, Bantilan C, Anupama KV, Palm C (2005) Opportunities for linking adaptation and mitigation in agroforestry systems. *In* C Robledo, M Kanninen, L Pedroni, eds, Tropical forests and adaptation to climate change: In search of synergies., Bogor, Indonesia. Center for International Forestry Research (CIFOR). P. 103-121.

Verheij EWM, Coronel RE (1991) Plant Resources of SouthEast Asia No. 2.,. Edible fruits and nuts (Wageningen)

Vetaas OR (2002) Realized and potential climate niches: a comparison of four Rododendron tree species. Journal of Biogeography **29:** 545-554

Vogt K (1995) A field workers guide to identification propagation and uses of common trees and shrubs of dry land Sudan 52p. SOS Sahel International, London

von Braun J, Swaminathan MS, Rosegrant MW (2004) IFPRI Annual Report Essay Agriculture 2003 - 2004, Food Security, Nutrition and the Millennium Development Goals. *In*. IRPRI

Von Maydell HJ (1986) Trees and shrubs of the Sahel, their characteristics and uses. Rossdorf TZ-Verlagsgesellschaft

Walker PA, Cocks KD (1991) HABITAT- a procedure for modeling a disjoint environment envelope for plant or animal species. Global Ecology and Biogeography **1**

Walker PA, Cocks KD (1991) HABITAT- a procedure for modelling a disjoint environment envelope for plant or animal species. Global Ecology and Biogeography **1**

WCMC (2001) Udvardy's Ecoregions. *In*,

Wieczorek J (2001) MaNIS Georeferencing Guidelines.

Wieczorek J, Qinghua G, Hijmans RJ (2004) The point-radius method for georeferencing locality descriptions and calculating associated uncertainty. International Journal of Geographical Information science **18:** 745-767

Wiens JA (1989) Spatial Scaling in Ecology. Functional Ecology **3:** 385-397

Wiens JA, Garham CH (2005) Niche Consevatism: Intergrating Evolution, Ecology and Conservation Biology. Annual Review of Ecology and Systematics **36:** 519-539

Wiley EO, McNyset KM, Peterson TA, Robins CR, Stewart AM (2003) Niche Modeling Perspective on Geographic Range Predictions in the Marine EnvironmentUsing a Michine-learning Algorithm. Oceangraphy **16:** 120-127

Williams JT (2006a) Chapter 1. Introduction, Taxanomy, Description and Distribution. *In* JT Williams, RW Smith, N Haq, Z Dunsiger, eds, Tamarind, *Tamarindus indica* L. Southampton Centre for Underutilised Crops, Southampton

Williams JT (2006b) Chapter 3. Uses and Products *In* JT Williams, RW Smith, N Haq, Z Dunsiger, eds, Tamarind, *Tamarindus indica L.* Southampton Centre for Underutilised Crops, Southampton

Williams JT (2006c) Chapter 4 Ecological Requirements. *In* JT Williams, RW Smith, N Haq, Z Dunsiger, eds, Tamarind *Tamarindus indica* L. Southampton Centre for Underutilised Crops, Southampton

Williams JT, Haq N (2002) An assessment of current activities and proposal for an enhanced cooperation." *In* Internatioal Centre for Underutilised Crops , 46,

Williams JT, Haq N (2002) An assessment of current activities and proposals for enhanced cooperation. International Centre for underutilised crops 46

Williams JT, Haq N (2003) Underutilised Crops for Food Security and Poverty Alleviation.

Williams PH, Margules CR, Hilbert DW (2002) Data requirements and data sources for biodiversity priority area and selection. Journal of Bioscience **27:** 327-338

Wolf J (1993) Effects of climate change on wheat production potentail in the European Community. European Journal of Agronomy **2:** 281-292

Yaacob O, Subhadrabhandu (1995) The production of economic fruits in SouthEast Asia. Oxford University Press

Yadav IS, Srinivasa Rao NK, Reddy BMC, Rawal RD, Srinivasan VR, Sujatha NT, Bhattacharya C, Nageswara Rao PP, Ramesh KS, Elango S (2002) Acreage and production estimation of mango orchards using Indian Remote Sensing (IRS) satellite data. Scientia Horticulturae **93:** 105-123

Zerger A (2004) Gridsampler. CSIRO Sustainable Ecosystems

7 Appendix

7.1 Description of program used to create environmental quarterly dataset "ProcessTemp"

The input for the dataset is a monthly data set for 2 eco-geographical variables Var1 and Var2. The program produces 12 moving quarterly averages or totals (depending on the variable) for a year, i.e. January, February, March through to December, January, February for both monthly dataset (Var1 and Var2). It then calculates the average or total value for Var 2 for the minimum and maximum quarterly periods of Var1 and the total or average value for Var 1 for the minimum and maximum quarterly periods of Var2.

For example the monthly datasets (January-December) for Rainfall (Var1) and Temperature (Var2) are entered into the 'Processtemp' program and 12 moving totals and 12 moving quarterly averages were calculated for rainfall and temperature respectively. The program then identifies the wettest quarter (maximum rainfall) and driest quarter (minimum rainfall) and the hottest (maximum temperature) and coldest quarter (minimum temperature). The program then calculates the average temperature that occurs in the wettest quarter and the average temperature for the driest quarter and the total rainfall for the hottest quarter and the total rainfall for the coldest quarter.

7.2 Description of Water Balance dataset program "WATBAL"

As the start (January) water balance value is not known, the WATBAL programme assumes that the water balance (W)/soil water storage (WST) value for January is 0mm. The program then calculates all monthly water balance for all months for a year (year 0) based on this value. It will then repeat the whole process once (year 1) and compare Januarys WST value with the value for the first run through (year

0). If all values for January WST (year 1) are equal to WST (year 0) the program will finish, if not it will continue to run until all values for January WST are the same as the January WST value from the previous year run through. This lets the soil moisture value stabilize so that when the same sequence of precipitation and evaporation is applied again, the model has a realistic soil moisture store to begin with. The programs continues to run producing a year of monthly water balances until all January WST values are equal for two consecutive years.

7.3 *Ecological Niche Factor Analysis (ENFA) factor tables*

Specialisation explained by the first 7 ecological factors and coefficients values for all variables for the Global "All Variables", ENFA

Environmental Variable	Factor 1 (18%) Marginality	Factor 2 (14%) Spec.1	Factor 3 (14%) Spec. 2	Factor 4 (7%) Spec. 3	Factor 5 (6%) Spec.4	Factor 6 (4%) Spec,.5	Factor 7 (4%) Spec.6
C-TMnCdM	0.28	-0.45	0.26	-0.41	0.04	-0.47	-0.47
C-TRn	-0.27	-0.78	0.63	-0.50	0.02	-0.62	-0.50
C-TCdQ	0.26	-0.08	0.30	0.03	0.28	0.03	-0.12
C-PWtM	0.24	0.00	-0.05	0.08	0.11	0.02	0.13
M-AIMxQ	0.22	-0.01	-0.05	-0.01	0.17	-0.08	-0.08
C-PWtQ	0.22	-0.01	-0.11	0.24	0.08	-0.15	-0.21
C-TSn	-0.22	-0.04	0.05	0.12	-0.06	0.15	0.00
C-TDR	-0.21	0.10	-0.20	0.01	-0.10	0.04	0.01
C-RHWtQ	0.21	-0.01	-0.03	-0.01	0.00	-0.04	0.05
M-AISn	0.20	0.00	-0.06	-0.05	0.05	0.07	0.08
S-OCtop	0.19	0.00	-0.02	0.01	-0.07	0.00	-0.07
C-TAn	0.18	-0.06	-0.25	0.24	-0.68	0.23	0.27
S-TNsub	0.18	0.00	-0.08	0.01	-0.07	-0.09	0.09
S-OCsub	0.17	-0.01	-0.01	-0.01	0.13	0.01	0.18
C-TIso	0.17	-0.08	0.18	-0.01	0.08	-0.02	-0.03
S-TNtop	0.17	0.01	0.16	0.01	0.02	0.15	-0.14
C-PAn	0.16	0.01	0.23	-0.43	-0.16	0.17	0.13
C-RHWmQ	0.16	0.01	0.10	-0.04	-0.04	-0.05	-0.08
C-PWmQ	0.15	-0.01	-0.11	0.06	0.03	0.01	-0.03
C-RHCdQ	0.14	0.00	0.03	-0.05	-0.02	-0.05	-0.05
C-RHDyQ	0.14	0.00	-0.14	0.05	0.11	0.12	0.09
S-ECECsub	0.13	0.00	0.01	0.01	-0.01	0.02	-0.02
S-ECECtop	0.12	0.00	-0.01	-0.01	-0.06	0.00	-0.01
C-PSn	0.12	0.01	0.12	-0.20	-0.20	0.02	0.02
C-TDyQ	0.11	0.00	-0.14	-0.05	-0.10	-0.05	-0.02
C-SDWtQ	-0.11	-0.01	-0.04	-0.01	0.03	-0.07	0.04
S-SltPtop	0.10	0.00	0.03	-0.03	0.02	-0.06	-0.04
C-TWtQ	0.09	0.03	0.29	-0.01	0.09	-0.17	0.04
S-SndPsub	-0.08	0.00	0.01	0.02	-0.02	0.00	0.03
S-SndPtop	-0.08	0.00	0.00	-0.07	0.09	-0.03	-0.08
C-TMxWmM	-0.08	0.38	-0.13	0.33	0.28	0.35	0.34
S-BDtop	-0.08	-0.01	0.01	0.00	-0.02	0.00	0.02
SDWmQ	-0.07	0.00	0.02	0.00	-0.04	0.00	-0.05
M-AIAn	0.07	0.04	-0.04	0.05	-0.17	-0.10	0.02
M-AIMnQ	-0.06	-0.02	-0.02	-0.09	0.07	0.10	0.08
C-SDDyQ	-0.06	0.01	-0.03	0.02	0.09	0.06	0.09
S-ClyPtop	0.06	-0.01	0.01	-0.04	0.10	0.00	-0.03
S-ClyPsub	0.05	0.01	0.01	-0.01	-0.03	-0.02	-0.07
M-AIMn	0.05	0.00	-0.01	0.00	0.05	0.01	-0.01
C-PCdQ	0.05	0.01	0.01	0.01	-0.03	0.00	0.02
M-AIWmQ	0.05	0.00	0.01	-0.03	-0.01	0.01	0.00
S-SltPsub	0.04	0.00	-0.03	0.01	0.03	0.00	-0.05
C-SDCdQ	-0.04	0.00	0.01	-0.03	-0.06	-0.04	-0.10
S-GrPsub	-0.04	0.00	0.01	-0.01	0.02	0.01	0.01
C-TWmQ	-0.03	0.08	-0.15	-0.28	0.31	-0.10	-0.23
C-PDyQ	0.02	-0.01	0.02	-0.03	0.00	0.01	0.09
M-AICdQ	0.02	-0.01	0.01	0.00	0.04	0.02	-0.04
S-pHtop	-0.01	0.01	0.01	0.02	0.03	-0.04	-0.14
S-GrPtop	-0.01	0.00	-0.01	0.01	-0.02	0.00	-0.02
C-PDyM	0.01	0.01	0.01	-0.01	-0.05	-0.01	-0.07
S-pHsoil	-0.01	-0.02	-0.01	-0.02	-0.01	0.03	0.11
S-BDsub	0.00	0.00	-0.01	0.01	0.00	-0.02	0.02

Specialisation explained by the first 7 ecological factors and coefficients values for all variables for the Africa "all variable" regional ENFA

Environmental Variable	Factor 1 (34%)	Factor 2 (17%)	Factor 3 (10%)	Factor 4 (8%)	Factor 5 (4%)	Factor 6	Factir 7
	Marginality	Spec.1	Spec. 2	Spec. 3	Spec.4	Spec.5	Spec.6
C-TRn	-0.25	-0.31	0.23	0.71	-0.49	0.77	0.51
C-PWtM	0.24	-0.14	-0.16	-0.02	-0.08	0.05	0.04
C-PWtQ	0.23	0.02	-0.37	-0.06	0.24	0.03	0.16
C-TDR	-0.23	0.21	-0.18	-0.09	0.15	-0.13	-0.15
C-RHWtQ	0.22	-0.12	0.04	0.02	0.14	0.02	-0.01
C-TMnCdM	0.22	0.07	-0.05	0.36	-0.14	0.42	0.19
M-AlMxQ	0.22	-0.17	0.03	-0.05	-0.04	0.03	-0.18
C-TSn	-0.22	-0.09	0.20	0.09	0.12	0.03	-0.06
C-PAn	0.21	0.24	0.61	0.08	-0.07	-0.08	-0.19
S-OCtop	0.19	0.03	0.01	-0.01	0.12	0.00	-0.19
C-TCdQ	0.19	-0.31	0.25	-0.02	0.05	0.09	-0.24
C-TIso	0.18	-0.19	0.19	0.08	-0.06	0.12	0.15
M-AlSn	0.18	-0.04	-0.04	0.03	0.09	0.00	0.23
C-PWmQ	0.17	-0.21	-0.06	0.00	-0.21	-0.03	-0.09
C-RHCdQ	0.17	-0.06	0.04	-0.01	-0.07	0.02	-0.01
S-OCsub	0.16	-0.04	-0.02	-0.01	-0.07	-0.01	0.29
C-RHWmQ	0.15	0.21	0.00	-0.03	0.10	0.00	-0.04
S-TNtop	0.15	0.13	0.00	-0.01	-0.32	0.03	0.05
M-AlMn	0.15	-0.02	-0.02	0.00	-0.04	0.00	0.01
C-RHDyQ	0.15	-0.21	-0.04	-0.01	-0.03	-0.03	0.00
C-SDWtQ	-0.15	-0.01	-0.02	0.01	0.12	-0.01	0.07
C-TMxWmM	-0.14	-0.23	0.12	-0.42	0.16	-0.33	-0.18
S-TNsub	0.14	-0.04	0.00	0.02	0.19	-0.02	-0.09
M-AlAn	0.13	0.14	-0.03	0.00	0.13	-0.06	0.20
C-PDyM	0.13	0.02	0.03	-0.01	0.08	0.00	-0.03
C-PDyQ	0.13	-0.11	0.04	0.00	-0.16	-0.01	0.00
C-PCdQ	0.12	0.11	0.02	0.02	0.07	-0.02	0.04
S-pHsub	-0.11	0.01	0.04	0.04	0.09	-0.01	0.12
C-TWmQ	-0.10	-0.11	-0.13	-0.22	-0.05	-0.01	-0.01
S-pHtop	-0.10	-0.03	-0.02	-0.05	-0.04	0.02	-0.19
M-AlWmQ	0.10	0.01	0.03	0.02	0.05	0.03	0.01
C-SDcdQ	-0.09	-0.02	0.10	0.01	-0.04	0.03	-0.05
C-TAn	0.08	0.34	-0.25	0.29	-0.31	-0.20	0.13
S-ClyPtop	0.08	-0.03	0.01	-0.02	0.03	-0.02	0.17
S-SndPsub	-0.08	0.04	-0.01	-0.02	0.09	0.01	-0.12
M-AlcdQ	0.08	-0.01	0.01	0.00	-0.02	0.03	-0.05
C-SDDyQ	-0.08	0.00	-0.07	0.00	0.00	-0.03	0.04
S-BDtop	-0.07	0.03	0.00	0.01	0.03	0.00	0.02
S-ClyPsub	0.07	0.00	0.03	0.00	0.04	0.02	-0.08
C-SDWmQ	-0.07	0.04	0.00	-0.01	0.00	0.01	-0.05
S-GrPsub	-0.05	-0.01	0.03	0.00	-0.01	0.00	0.03
S-SndPtop	-0.04	-0.05	0.07	0.01	0.01	-0.01	0.27
C-TWtQ	0.04	0.37	0.09	-0.07	0.32	0.12	0.11
C-TDyQ	-0.03	-0.24	0.11	-0.01	0.11	0.03	-0.04
S-ECECsub	0.03	0.00	-0.02	0.00	-0.03	0.01	0.01
S-ECECtop	0.02	0.02	0.04	0.02	0.01	-0.01	-0.01
S-SltPsub	-0.02	-0.05	-0.04	0.00	-0.14	-0.02	0.05
S-SltPtop	0.02	0.03	0.04	-0.01	0.16	0.02	0.07
M-AlMnQ	0.02	-0.07	-0.02	0.03	-0.04	0.01	0.05
C-PSn	0.01	0.06	0.29	0.02	-0.09	-0.04	-0.02
S-BDsub	0.01	0.00	-0.01	0.00	0.00	-0.01	0.00
S-GrPtop	0.00	-0.01	-0.03	0.00	0.00	0.00	-0.06

Specialisation explained by the first 6 ecological factors and coefficients values for all variables for the Australasian "all variables" model

Environmental Variable	Factor 1 (24%) Marginality	Factor 2 (24%) Spec.1	Factor 3 (15%) Spec. 2	Factor 4 (7%) Spec. 3	Factor 5 (6%) Spec.4	Factor 6 (4%) Spec.5
C-TCdQ	0.32	0.45	-0.18	-0.48	-0.09	0.40
C-TMnCdM	0.32	-0.16	-0.10	-0.06	0.45	-0.57
C-TRn	-0.30	0.35	0.20	-0.50	0.52	-0.37
C-TSn	-0.28	0.02	0.12	-0.28	-0.22	0.02
C-TDyQ	0.23	-0.04	0.03	0.01	-0.05	-0.04
C-TAn	0.23	-0.21	0.65	0.20	-0.09	-0.12
C-PWtM	0.20	0.07	0.07	0.22	-0.07	-0.06
C-TDR	-0.20	-0.25	-0.17	0.27	0.04	-0.11
M-AISn	0.20	0.08	0.11	-0.08	-0.04	0.20
C-TIso	0.19	0.06	0.07	-0.19	-0.03	-0.06
M-AIMxQ	0.18	-0.04	-0.04	0.04	0.01	-0.06
C-RHWtQ	0.18	0.28	0.05	0.05	-0.11	-0.09
C-PWtQ	0.17	-0.13	-0.05	-0.14	0.03	-0.07
C-PDyM	-0.16	0.03	0.00	0.00	0.00	0.06
S-BDtop	0.14	0.03	0.02	0.00	-0.02	0.00
C-PAn	0.14	0.08	-0.05	-0.11	0.02	0.21
S-BDsub	0.14	-0.03	-0.02	-0.03	0.02	0.02
C-PSn	0.13	0.05	-0.03	-0.07	0.05	0.00
S-ClyPsub	0.13	0.06	0.00	-0.01	-0.01	0.00
M-AIMnM	-0.12	0.05	0.00	-0.07	0.01	0.13
M-AIMnQ	-0.12	0.07	0.17	-0.08	-0.01	0.10
M-AICdQ	-0.11	-0.04	0.01	-0.04	-0.01	0.06
C-PDyQ	-0.11	-0.01	0.02	0.04	-0.02	-0.05
C-SDWtQ	-0.11	0.12	-0.01	0.03	-0.05	-0.03
C-RWnQ	0.11	-0.25	-0.05	-0.21	0.03	0.06
S-TNsub	0.10	0.14	0.06	0.20	0.09	0.04
C-SDCdQ	0.10	-0.17	-0.05	0.03	0.10	0.00
S-TNsub	0.10	-0.05	0.00	-0.10	-0.04	-0.01
C-SDDyQ	0.09	-0.02	0.00	0.05	0.00	-0.05
S-OCsub	0.08	-0.07	-0.05	-0.12	-0.05	-0.06
C-SDWmQ	0.08	-0.01	0.08	-0.12	-0.04	0.07
C-RHDyQ	0.06	-0.08	-0.07	0.08	0.04	-0.06
S-pHtop	-0.06	0.02	-0.02	-0.08	-0.10	-0.11
S-pHsub	-0.05	0.01	0.03	0.06	0.12	0.11
C-TMxWmM	-0.05	-0.06	0.06	0.09	-0.47	0.27
C-PWmQ	0.05	-0.06	0.03	0.04	-0.01	-0.07
S_OCtop	0.05	0.03	0.02	0.03	0.02	0.02
C-RHCdQ	0.04	-0.01	0.04	0.04	0.02	0.04
S-ClyPtop	0.04	-0.13	-0.06	-0.05	0.06	0.14
S-GrPtop	-0.04	-0.02	0.01	0.01	-0.01	0.00
S-Sndsub	-0.04	-0.04	-0.02	-0.06	-0.05	-0.02
S-GrPsub	-0.04	0.01	0.01	-0.03	0.01	0.00
C-TWtQ	0.04	0.43	-0.01	0.05	-0.11	0.00
M-AIWmQ	0.03	0.02	-0.03	-0.05	0.00	0.03
S-ECECtop	-0.02	0.02	0.01	0.08	0.01	0.02
C-PCdQ	-0.01	-0.01	-0.04	-0.01	0.04	0.01
S-PSltsub	-0.01	-0.01	0.02	0.05	0.02	-0.02
M-AIAn	0.01	0.01	-0.02	0.10	-0.02	-0.20
C-TWmQ	0.01	-0.24	-0.58	0.05	0.36	0.00
S-ECECsub	0.01	0.00	-0.01	-0.04	-0.05	-0.06
S-SndPtop	0.00	-0.03	-0.05	-0.01	0.08	0.15
S-Slttop	0.00	-0.03	-0.04	-0.06	0.02	0.07

Specialisation explained by the first 6 ecological factors and coefficients values for all variables for the America's "all variables" regional ENFA

	Factor 1 (7%)	Factor 2 (30%)	Factor 3 (14%)	Factor 4 (10%)	Factor 5 (8%)	Factor 6 (5%)
Environmental Variables	Marginality	Spec.1	Spec. 2	Spec. 3	Spec.4	Spec.5
S-ECECsub	0.32	0.03	-0.01	0.02	-0.05	0.01
S-ECECtop	0.30	-0.04	0.02	-0.06	0.09	-0.08
C-PSn	0.27	0.01	-0.03	-0.02	-0.03	0.03
S-SltPtop	0.24	-0.02	0.06	0.02	0.02	0.01
S-TNsub	0.22	-0.22	-0.13	-0.06	0.04	0.05
S-pHsubsoil	0.20	-0.09	0.01	-0.12	-0.04	0.01
C-TWtQ	0.19	0.00	0.28	-0.10	0.04	0.11
S-BDtop	-0.18	-0.06	-0.01	-0.02	0.02	-0.03
S-OCtop	0.18	-0.30	-0.01	0.01	-0.07	-0.06
C-Tan	0.17	-0.06	-0.64	-0.13	-0.47	0.03
S-SltPsub	0.17	0.00	-0.02	0.03	-0.06	0.00
S-pHtop	0.17	0.08	-0.01	0.15	0.03	0.02
S-OCsub	0.17	0.14	0.11	0.21	-0.17	0.00
C-TWmQ	0.16	0.14	-0.08	0.08	0.27	0.13
TMnCdM	0.16	-0.11	0.09	0.44	-0.53	0.50
M-AISn	0.16	0.07	-0.08	-0.08	0.04	-0.05
C-PWtM	0.15	-0.05	0.02	-0.03	0.11	-0.09
C-TCdQ	0.15	-0.20	0.58	0.28	0.22	-0.51
C-TDyQ	0.15	0.01	0.03	0.04	0.04	0.05
S-TNtop	0.13	0.68	-0.05	-0.05	0.19	0.06
M-AIMnQ	-0.13	0.17	-0.04	-0.01	0.09	0.06
C-SDWmQ	-0.12	-0.04	-0.05	0.01	-0.02	-0.07
C-PWmQ	0.12	-0.02	-0.01	0.01	0.02	-0.02
S-SndPtop	-0.12	0.03	0.01	0.00	-0.10	-0.02
C-SDDyQ	-0.12	0.09	0.01	-0.04	0.00	0.02
C-PWtQ	0.12	0.01	-0.02	0.03	-0.09	-0.01
C-TDR	-0.11	0.10	-0.08	-0.07	-0.05	0.19
C-TRn	-0.11	-0.31	0.29	0.66	-0.40	0.19
C-TMxWmM	0.11	0.05	0.01	-0.30	0.18	-0.27
M-AIMn	-0.11	-0.01	-0.05	-0.06	-0.02	-0.05
C-PDyQ	-0.10	-0.06	-0.03	-0.02	-0.05	-0.05
C-RHDyQ	0.09	0.06	0.02	0.08	-0.05	0.08
C-RHWmQ	0.09	0.04	-0.07	0.00	-0.01	-0.01
C-PCdQ	-0.09	0.00	0.02	0.01	-0.04	0.01
C-PDyM	-0.08	0.01	0.02	0.00	0.01	0.00
S-BDsub	-0.08	0.00	0.00	0.05	-0.05	0.03
M-AIWmQ	-0.08	0.03	0.00	-0.01	-0.02	0.01
S-ClyPsub	-0.08	0.06	0.02	-0.02	0.09	-0.02
S-SndPsub	-0.06	-0.06	0.01	0.04	0.01	0.00
C-RHWtQ	0.06	0.07	0.05	0.00	-0.02	0.12
M-AIAn	-0.05	-0.09	-0.01	-0.04	-0.03	-0.07
C-RHcdQ	0.05	-0.15	0.01	-0.08	0.02	-0.15
C-TSn	-0.03	-0.19	0.04	0.16	-0.01	-0.36
S-GrPsub	0.03	0.01	0.02	0.00	-0.03	0.01
M-AICdQ	-0.03	-0.01	0.00	0.03	0.04	0.02
C-SDWtQ	0.02	0.06	0.03	0.03	0.01	0.09
C-TIso	0.02	-0.12	0.03	0.09	0.10	-0.22
S-ClyPtop	-0.02	-0.05	0.03	0.07	-0.12	0.03
S-GrPtop	0.01	-0.02	0.01	0.02	0.05	-0.02
C-SDCdQ	-0.01	-0.12	0.01	0.02	0.01	0.00
C-Pan	0.00	0.09	0.00	0.00	-0.03	0.16

Specialisation explained by the first 5 ecological factors and coefficients values for all 20 variables for the Global "reduced variable" model

Environmental Variable	Factor 1 (20%) Marginality	Factor 2 (19%) Spec.1	Factor 3 (12%) Spec. 2	Factor 4 (9%) Spec. 3	Factor 5 (7%) Spec. 4
C-TMnCdM	0.42	-0.61	-0.72	-0.52	-0.05
M-AIMxQ	0.33	-0.05	0.33	0.10	-0.04
C-PWtQ	0.33	-0.04	-0.05	0.01	-0.03
C-TSn	-0.33	-0.52	-0.37	-0.45	-0.51
C-TDR	-0.32	-0.13	-0.31	-0.37	-0.16
C-RHWtQ	0.31	-0.02	0.13	-0.06	0.03
M-AISn	0.29	0.00	-0.01	0.02	-0.14
S-TNsub	0.27	0.04	-0.03	-0.05	-0.08
S-ECECsub	0.20	-0.03	-0.06	0.04	-0.13
C-SDWtQ	-0.16	-0.10	0.02	0.14	-0.03
C-TWtQ	0.13	0.55	0.05	0.03	-0.39
S-SndPsub	-0.12	0.02	-0.02	-0.04	-0.01
C-TMxWmM	-0.11	0.04	0.34	0.58	0.60
C-SDWmQ	-0.10	0.07	-0.01	-0.07	-0.28
S-ClyPsub	0.08	-0.01	-0.01	0.00	0.08
S-SltPsub	0.07	-0.03	0.03	0.12	-0.06
C-SDCdQ	-0.06	-0.09	-0.09	-0.03	0.20
S-GrPsub	-0.05	0.02	0.05	0.00	0.12
S-GrPtop	-0.02	0.00	-0.03	0.06	-0.12

Specialisation explained by the first 5 ecological factors and coefficients values for all 27 variables for the Africa regional model

Environmental Variable	Factor 1 (21%) Marginality	Factor 2 (27%) Spec.1	Factor 3 (8%) Spec. 2	Factor 4 (7%) Spec. 3	Factor (5%) Spec. 4
C-TDR	-0.33	-0.16	-0.17	0.13	-0.06
C-RHWtQ	0.32	-0.16	0.27	-0.32	0.00
C-TMnCdM	0.32	-0.51	-0.23	0.31	-0.10
M-AIMxQ	0.31	-0.06	0.05	0.34	0.40
C-TSn	-0.31	-0.42	-0.26	0.28	0.03
C-Pan	0.30	-0.07	-0.01	0.07	-0.41
S-OCtop	0.27	0.01	-0.15	-0.24	0.07
C-RHCdQ	0.24	0.03	0.05	0.19	0.18
S-TNtop	0.22	0.14	-0.20	0.04	0.02
C-RHdyQ	0.21	-0.10	-0.38	-0.05	-0.08
C-SDWtQ	-0.21	-0.07	0.30	-0.08	-0.07
S-pHsub	-0.16	-0.06	-0.12	-0.28	-0.27
S-pHtop	-0.14	0.06	0.02	0.29	0.51
C-SDCdQ	-0.13	-0.07	0.05	0.32	0.20
S-SndPsub	-0.11	0.05	-0.39	-0.01	0.15
C-SDdyQ	-0.11	0.01	-0.18	-0.25	-0.12
C-SDWmQ	-0.10	-0.03	-0.01	-0.25	0.26
S-BDtop	-0.11	-0.01	0.07	-0.15	-0.19
S-ClyPsub	0.10	-0.04	0.00	0.10	0.18
C-SDWmQ	-0.10	-0.03	-0.01	-0.25	0.26
S-GrPSub	-0.07	-0.02	-0.01	0.01	0.02
C-TwtQ	0.06	0.67	0.22	-0.18	0.17
S-ECECsub	0.04	-0.01	-0.11	0.02	0.07
S-ECECtop	0.03	0.00	-0.08	-0.06	-0.16
S-SltPsub	-0.03	0.02	0.44	0.09	-0.03
M-AIMnQ	0.02	0.01	0.07	-0.01	-0.02
S-BDsub	0.01	0.03	-0.07	0.07	0.03

Specialisation explained by the first 4 ecological factors and coefficients values for the most important 10 variables for the Australasian regional model

Environmental Variable	Factor 1 (24%) Marginality	Factor 2 (32%) Spec.1	Factor 3 (12%) Spec. 2	Factor 4 (8%) Spec. 3
C-TMnCdM	0.58	0.20	-0.52	0.12
C-TAn	0.42	-0.63	0.32	-0.22
M-AIMxQ	0.33	0.10	0.31	0.16
C-RHWtQ	0.33	0.23	0.17	0.32
C-PAn	0.25	-0.03	0.08	-0.08
S-BDsub	0.25	0.07	0.21	-0.49
S-ClyPsub	0.23	-0.05	-0.32	0.11
C-PDyQ	-0.21	0.02	0.06	-0.15
S-OCsub	0.15	0.07	0.29	-0.45
C-TWtQ	0.06	0.69	-0.36	0.15
S-GrPsub	-0.07	0.04	0.18	-0.06
S-SndPsub	-0.07	0.00	-0.08	0.05
C-TWtQ	0.06	0.69	-0.36	0.15
M-AIWmQ	0.06	-0.09	-0.03	0.15
S-SltPsub	-0.02	-0.06	0.10	-0.01
S-ECECsub	0.01	0.01	-0.20	0.51

Specialisation explained by the first 4 ecological factors and coefficients values for all 19 variables for the America's regional model

Environmental Variable	Factor 1 (5%) Marginality	Factor 2 (30%) Spec.1	Factor 3 (18%) Spec. 2	Factor 4 (9%) Spec. 3
S-pHsub	0.47	0.02	-0.01	-0.04
S-SltPsub	0.40	0.02	0.00	0.03
C-TMnCdM	0.37	0.65	-0.68	-0.61
C-PWmQ	0.27	0.02	0.00	-0.02
C-PWtQ	0.27	0.00	-0.01	0.02
C-TRn	-0.26	0.66	-0.67	-0.65
C-TMxWmM	0.26	-0.33	0.31	0.28
C-PcdQ	-0.21	0.00	0.01	-0.02
S-BDsub	-0.19	0.01	0.00	-0.02
M-AIWmQ	-0.19	-0.03	-0.01	0.06
S-ClyPsub	-0.18	0.00	0.00	0.01
S-SndPsub	-0.14	-0.01	0.00	-0.09
C-RHWtQ	0.13	-0.02	0.00	0.20
C-RHcdQ	0.12	0.01	-0.01	-0.23
C-TSn	-0.07	0.02	0.01	0.11
S-GrPsub	0.07	-0.01	0.00	-0.03
C-SDWtQ	0.05	0.04	0.01	-0.03
C-SDCdQ	-0.01	0.00	0.00	0.07
S_GrPtop	0.01	0.02	0.00	-0.06

Climate: C-TAn Mean Annual Temperature(ºC), C-TDR Mean Diurnal Range(ºC), C-TIso Isothermality(ºC), C-TSn Temperature Seasonality, C-TMxWmM Maximum Temperature of Warmest Month(ºC), C-TMnCdM Minimum Temperature of Coldest Month(ºC), C-TRn Temperature Range(ºC), C-TWtQ Mean Temperature Wettest Quarter(ºC), C-TDyQ Mean Temperature Driest Quarter(ºC), C-TWmQ Mean Temperature Warmest Quarter(ºC), C-TCdQ Mean Temperature Coldest Quarter(ºC), C-PAn Annual Precipitation(mm) , C-PWtM Precipitation of Wettest Month (mm), C-PDyM Precipitation of Driest Month(mm), C-PSn - Precipitation Seasonality(mm), C-PWtQ Precipitation of Wettest Quarter(mm), C-PdyQ Precipitation of Driest Quarter(mm), C-PWmQ Precipitation of Warmest Quarter(mm), C-PCdQ Precipitation of Coldest Quarter(mm), C-RHCdQ Relative Humidity of Coldest Quarter(%), C-RHDyQ Relative Humidity of Driest Quarter, C-RHWmQ Relative Humidity Warmest Quarter (%), C-RHWtQ Relative Humidity Wettest quarter(%), C-SDCdQ Mean Sunshine Duration Coldest Quarter(%), C-SDDyQ Mean Sunshine Duration Driest Quarter(%), C-SDWmQ Mean Sunshine Duration Warmest Quarter(%), C-SDWtQ Mean Sunshine Duration Wettest Quarter(%). ***Moisture Availability:*** M-AICdQ Mean Moisture Availability Index of Coldest Quarter, M-AISn Moisture Availability Index Seasonality, M-AIWmQ Mean Moisture Availability Index of Warmest Quarter,M-AIMxM Maximum Monthly Moisture Availability Index, M-AIMxQ Mean Moisture Availability Index of Maximum Quarter, M-AIAn Mean Annual Moisture Availability Index, M-AIMn Minimum Monthly Moisture Availability Index, M-AIMnQ Mean Moisture Availability Index Minimum Quarter. ***Soil:*** S-BDsub Bulk Density subsoil(g/cm^{-3}), S-Bdtop Bulk Density topsoil(g/cm^{-3}), S-ClyPsub Percentage Clay subsoil(%), S-ClyPtop Percentage Clay topsoil(%), S-SndPsub Percentage Sand subsoil(%),S-SndPtop Percentage Sand topsoil(%), S-SltPsub Percentage Silt subsoil(%),S-SltPtop Percentage Silt topsoil(%), S-GrPtop Percentage Grave topsoil(%), S-GrPsub Percentage Gravel subsoil(%), S-ECECtop Effective CEC topsoil (cmolc kg^{-1}), S-ECECsub Effective CEC subsoil(cmolc kg^{-1}), S-OCsub Organic Carbon content subsoil(% by weight), S-OCtop Organic Carbon content topsoil(% by weight), S-pHsub pH subsoil, S-pHtop pH topsoil, S-TNsub Total nitrogen subsoil(% by weight), S-TNtop Total nitrogen topsoil(% by weight).