

**GLOBAL PLANT CHARACTERISATION
AND DISTRIBUTION WITH EVOLUTION
AND CLIMATE**

JAMES NICHOLAS FURZE

Ph.D.

2014

GLOBAL PLANT CHARACTERISATION AND DISTRIBUTION WITH EVOLUTION AND CLIMATE

JAMES NICHOLAS FURZE

A thesis submitted in partial fulfilment of the requirements of
the University of the West of England, Bristol, UK
for the degree of Doctor of Philosophy

Engineering, Design and Mathematics / Geography and
Environmental Management
Faculty of Environment and Technology
University of the West of England, Bristol, UK

May 2014

This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

To my parents and my wife

ABSTRACT

Since Arrhenius published seminal work in 1921, research interest in the description of plant traits and grouped characteristics of plant species has grown, underpinning diversity in trophic levels. Geographic exploration and diversity studies prior to and after 1921 culminated in biological, chemical and computer-simulated approaches describing rudiments of growth patterns within dynamic conditions of Earth. This thesis has two parts:- classical theory and multidisciplinary fusion to give mathematical strength to characterising plant species in space and time.

Individual plant species occurrences are used to obtain a Species-Area Relationship. The use of both Boolean and logic-based mathematics is then integrated to describe classical methods and propose fuzzy logic control to predict species ordination. Having demonstrated a lack of significance between species and area for data modelled in this thesis a logic based approach is taken. Mamdani and T-S-K fuzzy system stability is verified by application to individual plant occurrences, validated by a multiple interfaced data portal.

Quantitative mathematical models are differentiated with a genetic programming approach, enabling visualisation of multi-objective dispersal of plant strategies, plant metabolism and life-forms within the water-energy dynamic of a fixed time-scale scenario. The distributions of plant characteristics are functionally enriched through the use of Gaussian process models. A generic framework of a Geographic Information System is used to visualise distributions and it is noted that such systems can be used to assist in design and implementation of policies.

The study has made use of field based data and the application of mathematic methods is shown to be appropriate and generative in the description of characteristics of plant species, with the aim of application of plant strategies, life-forms and photosynthetic types to a global framework. Novel application of fuzzy logic and related mathematic method to plant distribution and characteristics has been shown on a global scale. Quantification of the uncertainty gives novel insight through consequent trophic levels of biological systems, with great relevance to mathematic and geographic subject development. Informative value of Z matrices of plant distribution is increased substantiating sustainability and conservation policy value to ecosystems and human populations dependent upon them for their needs.

Key words: sustainability, conservation policy, Boolean and logic-based, fuzzy logic, genetic programming, multi-objective dispersal, strategies, metabolism, life-forms.

ACKNOWLEDGEMENTS

I would like to take this opportunity to express my sincere gratitude to my supervisors, Professor Quanmin Zhu, Associate Professor Jennifer Hill (both from UWE, UK) and Professor Feng Qiao (Faculty of Information and Control Engineering, Shenyang Jianzhu University, China) for their encouragement and guidance during the past three years of my studies. I also give my thanks to colleagues from the Engineering Design and Mathematics and Geography and Environmental Management departments and the staff from the Faculty of Environment and Technology of UWE with whom I share the excellent and comfortable research environment. I am also grateful for the encouragement of staff of the Faculty of Applied Sciences at UWE.

I would like to extend my gratitude to colleagues within the Intergovernmental Panel on Climate Change in Switzerland and the Global Biodiversity Information Facility who provided excellent communication regarding server updates enabling the supply of high quality data to be obtained. I also wish to thank colleagues from the many Institutes and Universities who took part in conference and research events, which I attended throughout my time at UWE.

I am grateful to the financial contributions from the Faculty of Environment and Technology at UWE UK and excellent award schemes at UWE, which I benefitted from.

Most importantly, I must underline that this thesis would not have been possible without the loving support of my family and my wife, I will be indebted to them forever. Thank-you to my parents-in-law and sister-in-law, too.

Finally, I would say “THANK YOU” to Feng Qiao who has acted both as a supervisor during the project and a close friend for many years. This research is the culmination of a life-long personal interest. All friends over my life have contributed in some way, and needless to say I am grateful to you all.

Bristol,

UK

TABLE OF CONTENTS

ABSTRACT	III
ACKNOWLEDGEMENTS	IV
TABLE OF CONTENTS	V
APPENDICES LIST	IX
ABBREVIATIONS	X
SYMBOLS	XII
1. INTRODUCTION	1
1.1.Motivations	1
1.2.Outline of the thesis	3
1.3.Contributions	5
1.4.Research questions	6
1.5.Published work	7
2. LITERATURE SURVEY	9
2.1. Introduction	9

2.2. Plant richness distribution and bio-geography	11
2.3. Qualitative plant life-form description	14
2.3.1. Plant life-history strategies and continuums	15
2.4. Computer generated plant strategy modelling	19
2.5. Quantitative pattern identification	23
2.5.1. Fuzzy logic	24
2.5.2. Genetic algorithms	26
2.5.3. Generative fusion of genetic-fuzzy systems	28
2.6. Knowledge guidance hybrid systems	30
2.7. Summary	33
 3. PRELIMINARY METHODOLOGIES	 34
3.1. Introduction	34
3.2. Data sources	35
3.2.1. Validation and reasoning for chosen biodiversity sources	35
3.2.2. Climatic and topographic data sources	36
3.2.3. High-resolution data for enhanced accuracy of modelling	38
3.3. Image processing	38
3.4. The Species-Area relationship	39
3.5. Variable partitioning	40
3.6. Groups to distinguish in plant characterisation	44
3.7. Fuzzy logic	49
3.7.1. Mamdani systems	50
3.7.2. Takagi-Sugeno systems	53
3.7.3. Expert intuition	55
3.8. Genetic algorithm optimisation of objective variables	55

3.8.1. Hybrid systems to explore utopian relations of plant characters	58
3.9. Summary	59
4. THE SPECIES-AREA RELATIONSHIP IN CONTEXT	61
4.1. Introduction	61
4.2. Species-Area relationship implementation	62
4.3. Information-rich modelling	66
4.4. Validation of algorithms and global mapping of plant strategies	73
4.5. Summary	82
5. USING FUZZY LOGIC CONTROL TO ORDINATE PLANT STRATEGIES	84
5.1. Introduction	84
5.2. Takagi-Sugeno-Kang modelling of plant strategies	85
5.3. Fine-scale resolution data enhances model formation	87
5.4. Implementation of fuzzy logic control to predict species occurrence	92
5.5. Summary	98
6. STOCHASTIC PROCESSES TO MODEL UNCERTAINTY	100
6.1. Introduction	100
6.2. Fuzzy-genetic programming of plant life-history strategies	103
6.2.1. Nodal structure and algorithmic detail	107
6.3. Fuzzy-genetic programming hybrid methodology	109
6.3.1. Multi-objective genetic algorithm dispersal of plant strategies	110

6.4. Multi-objective dispersal of plant strategies and metabolism	112
6.4.1. Linear and quadratic rules of multi objective dispersed elements	115
6.5. Summary	118
7. PLANT LIFE-FORM DISTRIBUTION AND INSTRUCTIVE GEOGRAPHIC INFORMATION SYSTEMS	120
7.1. Introduction	120
7.2. Using DEM and climatic data to substantiate T-S-K systems	122
7.2.1. ANFIS and efficiency of reduced model components	127
7.3. Multi-objective optimisation selection and dispersal of plant life-forms	132
7.3.1. Functional approximation within utopia of plant life-forms	136
7.4. ANFIS based algorithms control strategy of plant strategies, plant metabolism and plant life-forms	139
7.4.1. Functional approximation algorithm for plant communities	143
7.5. Geographic information systems - applying plant characterisation	144
7.5.1. Recommendations for policy formation using mathematical method	147
7.6. Summary	149
8. FUTURE RESEARCH WORK AND CONCLUSIONS	151
8.1. Thesis Summary	151
8.2. Answering the research questions	152
8.3. Future research directions	156
8.4. Conclusions	158

APPENDICES LIST

APPENDIX 1. Data and citations of collections for each example location used

APPENDIX 2. Algorithmic statements for Guyana, Cuba, Democratic republic of the Congo (Africa), Georgia (Southern Europe), Guinea (Africa) and Macedonia (Southern Europe)

APPENDIX 2.1. Example Matlab (Version 2010a ©) code using downloaded IPCC file

APPENDIX 2.1.1. Example Matlab (Version 2010a ©) code using USGS DEM data, GTOPO30

APPENDIX 2.1.2. Additional Matlab (Version 2010a ©) programming code to modify axis of DEM images

APPENDIX 2.2. Mamdani FL system surface area

APPENDIX 2.3. Matlab (Version 2010a ©) code for T-S-K (E2)

APPENDIX 2.4. 16 Rules for Macedonia, E7, Life-forms

APPENDIX 3. Climatic and DEM data for E1, E2, E3, E4, E5, E6, E7

APPENDIX 4. 2-Dimensional and 3-Dimensional surface of E2, T-S-K

APPENDIX 5. Matlab (Version 2010a ©) code for MOGA

APPENDIX 5.1. 19 Rules for Azerbaijan, E6 plant strategies ANFIS

APPENDIX 5.2. Cuba, E3 ANFIS (3 input variables) diagram

APPENDIX 5.3. Residual variance plot of Cuba MOGA

APPENDIX 6. MOGA Pareto plot for strategies

APPENDIX 6.1. MOGA Pareto plot for life-forms

APPENDIX 7. Matlab (Version 2010a ©) code for life-form MOGA process

APPENDIX 7.1. Utopia Line and Curve rules for life-forms

APPENDIX 7.2. Matlab (Version 2010a ©) code for Rastrigin's function

APPENDIX 7.3. Lesser number of characters Rastrigin's function

APPENDIX 7.4. Expanded probability of 18 life-forms

APPENDIX 8. Example Matlab (Version 2010a ©) code for GIS mapping

ABBREVIATIONS

Abbreviation	Description
ANFIS	Adaptive Neural Fuzzy Inference System
C3	3 Carbon photosynthetic pathway
C4	4 Carbon photosynthetic pathway
C	Competitive strategy
CA	Cellular Automation
CAM	Crassulacean Acid Metabolism
CC	Cloud Cover
CIA	Central Intelligence Agency
DEM	Digital Elevation Model
DZ	Diversity Zone
<i>E</i>	Environment
EC	Evolutionary Computation
EDA	Estimation of Distribution Algorithm
<i>eR</i>	Rule Based
F	Function
FAA	Functional Approximation Algorithm
FCM	Fuzzy C Means
FL	Fuzzy Logic
FLB	Fuzzy Logic Based
FLC	Fuzzy Logic Control
FRBS	Fuzzy Rule Base System
FSPM	Functional Structural Plant Model
GA	Genetic Algorithm
GARP	Genetic Algorithm for Rate set Prediction
GBIF	Global Biodiversity Information Facility
GFF	Ground Frost Frequency
GFRBS	Genetic Fuzzy Rule Base System
GIS	Geographic Information System
I	Moran's I
IFNIS	Intuitionistic Fuzzy Negative Ideal Solution
IFPIS	Intuitionistic Fuzzy Positive Ideal Solution
IFS	Intuitionistic Fuzzy Set
IPCC	Intergovernmental Panel on Climate Change
IRL	Iterative Rule Learning

K	K strategist
KB	Knowledge Base
Km	Kilometre
Kp	K proximity
L	Lindenmayer
Log	Logarithm
M	Metre
MaxTemp	Maximum Temperature
MCDM	Multiple Criteria Decision Making
MeanTemp	Mean Temperature
MinTemp	Minimum Temperature
MOGA	Multi-Objective Genetic Algorithm
MOO	Multiple-Objective Optimisation
MSPM	Mutable Single Plant Model
NIS	Negative Ideal Solution
P	Precipitation
PEP	Phospho Enol Pyruvate
PIS	Positive Ideal Solution
R	r strategist
R	Ruderal strategy
rK	Relative to K strategy
Rp	r proximity
RUBISCO	Ribulose-bis Carboxylase Oxygenase
S	Second
S	Stress-tolerant strategy
SAR	Species Area Relationship
TOPSIS	Technique for Order Preference by Similarity to Ideal Solution
T-S-K	Takagi-Sugeno-Kang
UNEP	United Nations Environment Programme
USGS	United States Geological Survey
VP	Vapour Pressure
WDF	Wet Day Frequency
W-E	Water-Energy

SYMBOLS

1 Geographic

Symbol	Definition
$^{\circ}$	Degrees
A	Area in the SAR equation
C	Environmental constant specific to geographic area
m	Number of species
$X_{i,k}$	The abundance of the k^{th} species in quadrat/sample i
$X_{j,k}$	The abundance of the k^{th} species in quadrat/sample j
S	Number of species in SAR equation
Z	Taxon specific constant

2 Distribution

Symbol	Definition
$\geq$	Greater than or equal to
$\leq$	Less than or equal to
∂	Differential
∞	Infinity
λT	Lambda, specifying occurrences in a specified time interval
m	Mean of the distribution
ϕ^{-1}	Inverse function of one dimensional Gaussian function
ϕ_G	Generator of the distribution
ϕ_r	Multivariate Gaussian distribution with correlation r
π	Pi, 3.14'
σ	Standard deviation
σ^2	Variance
a, b, c	Intervals a, b, c
C	Copula
$f(x)$	Function of x
H	Multivariate distribution function
I	Interval

$k!$	Factorial of constant k
Max	Maximum
Min	Minimum
Pr	Probability
T	Students t statistic
$u = (u_1, u_2, \dots, u_n)$	Universe equal to universal set values

3 Logic

Symbol	Definition
$\subseteq$	Subset
$\wedge$	And
ε	Error
$\forall$	For all
$\int_U$	Union of fuzzy singletons
$\cdot$	Multiplier
$\dots$	Undefined data / operation usually repeat iterations
$*$	Multiplication operator
μ	Grade of membership
$\prec$	Precedes
Σ	Summation
$A(x, y)$	Antecedent variables (x, y)
B	Consequent variable
D	Decision variable
Dom	Dominant
$i = 1, 2, \dots, n$	Crisp value
$j = 1, 2, \dots, m$	Crisp value
Lim	Limit
R	Relational Index
$\bar{x}$	Mean of x
X_i	Data set, universe of discourse of x_i
$x_i \in X_i$	Fuzzy system input element of X_i
$\bar{y}$	Mean of y
Y_j	Data set, universe of discourse of y_j
$y_j \in Y_j$	Fuzzy system output element of Y_j

Z

Utopian, combined objective space

CHAPTER 1

Introduction

1.1 Motivations

Plant characterisation is a multi-disciplinary subject, there being classical/historical and modern approaches to the main subject area. Areas of relevance are plant exploration, biogeographic and species distribution patterns, qualitative vegetation description, metabolic pathways, and plant life-history based strategies. Globally there are ambiguities in distribution of species and their functional groups, as many traits of the species and the conditions in which they exist exhibit uncertainties in structural and conditional relations in time and space.

Species exist in trophic levels and the main motivation of this study is to describe the parameters and external forces acting on species of the primary producing level. Using an engineering based technique, stable and robust control systems are crucial to the modelling

of any systems elements. Modelling plant species elements and distribution using engineering based techniques may reduce uncertainties and enable inferences clarifying the importance of plant species in trophic systems and provide structure more prone to analysis and prediction than previous methods. Although modelling species on a global level seems impossible as there are too many uncertainties and parameters, this study presents a simplified model of the global earth system with mathematical, biological and geographical foundation (Furze et al., 2013d). In considering the applications of this study the author would primarily like to underline that plants are one of the main kingdoms of life. They are primary producers and their importance therefore filters through every level of research and life existing on the planet by the same principle of trophic levels within ecosystems (Kreft et al., 2008; Wright, 1983). Measuring the impact of the global characterisation of plants, therefore, may be seen to be an asymptotic exercise. However, the following provide some of the benefits in research terms: expanding the knowledge base of plant science and furthering demographic studies of areas previously only studied from pure numeric estimation with a more systematic foundation. Modelling forwards (with experimental use of different scenarios of climatic data) enables an accurate prediction of plant distribution with climate change to further add to environmental models. Using fuzzy logic in this way reduces error, which may be present in methods previously used (Kreft and Jetz, 2007), and further specifies novel categories of plant distribution patterns.

In the design of control systems with structural uncertainties, variable parameters and dynamic relations the designer usually tries to minimise the perturbations with the use of a detailed mathematical model combined with solid statements of a historical and scientific nature in order to form certainties. One approach for this kind of non-linear control design is the use of differential methods in order to form closed loop systems. Control algorithms and generative mathematic method are key in the description of evolving and dynamic conditions in which species and their external influences are based. Such methods constantly monitor the relevant data patterns and change the numerical distributions by integrating an intuitive nature with a residual knowledge base.

Species distribution and elements of plant characterisation have been described using qualitative bio-geographic methods (MacArthur and Wilson, 1967) and quantitative scientific methods (Kraft et al., 2008; Massant et al., 2010). Attempts have been made to

present summaries of data within geographic information systems. However the use of sophisticated control theory has not been employed.

The motivation for the research work for this thesis is to characterise primary producing elements (plants) of trophic systems thereby revealing keys to sustainability and stability of ecological networks within a dynamic system; to define characters of plants under the principal descriptors of plant life history strategy (Grime et al., 1995), life-form (Raunkier, 1934) and metabolic pathway. The author was further motivated to produce a modelling framework to determine the numerical distribution of the above within time and space, in order to understand and better respond to increasing human, biotic and abiotic pressures imposed on plant species.

It is the author's belief that uniting engineering based mathematic technique with bio-geographic methods reduces uncertainty and provides a predictive ability, which may facilitate policy formation at a national and international level. Additionally, the author recognises that interdisciplinary unification of mathematic, scientific and qualitative methods provides a synergistic function, which generates evolution of knowledge bases related to the component disciplines.

1.2 Outline of the thesis

This thesis contains eight chapters. Chapter 1 introduces the thesis. The main body of the thesis can be classified into two parts. The first part includes Chapters 2-4, which concentrate on theoretical studies and classical approaches. The second part includes Chapters 5-7, which each have particular focus dealing with the application of fuzzy logic control to describe species patterns, stochastic methods defining life history based strategies / life form / metabolism and geographic information system structural capability in order to answer research questions and to thereby inform policy formation. Chapter 8 concludes the thesis. The thesis outline is itemized as follows:

Chapter 1. The introduction and outline of the thesis.

Chapter 2. The literature survey covers classical methods of plant characterisation, scientific categorisation and background of control theory to deal with the diverse nature of plant elements offering examples of different approaches.

Chapter 3. The essential knowledge and principal methodology related to the research work of this thesis is set out, and some clarification and conceptual definition is given.

Chapter 4. The classical species area relationship is applied in context and description of plant distributions in global locations are given. The deficiency of this approach is shown in statistical terms, substantiating the use of subsequent mathematic technique.

Chapter 5. Fuzzy logic control algorithms are constructed to provide information rich modelling of plant species variation, relevant variables are minimised and the stability of the algorithms is verified and validated using large scale experimental data.

Chapter 6. The adaptive fuzzy logic control brought about by Chapter 4 and implemented in Chapter 5 is further developed through stochastic processes, which may be used to model uncertainty within the utopia plane of primary producers.

Chapter 7. The application of fuzzy logic and hybrid genetic programming to construct geographic information systems is detailed and research questions of life history strategies, life-forms and metabolism are answered. Further, the process of geographic information system design is used to inform ecological sustainability and policy formation.

Chapter 8. Research questions are revisited, future research work is proposed and conclusions are drawn.

1.3 Contributions

The main contributions of this thesis are summarised as follows:

- Subject to a comprehensive literature survey in the research area, concepts and definitions of the related subjects are revisited and resolved with critical justifications and revisions to improve the descriptions and examples from a novel perspective.
- Classical, qualitative methodology is elucidated in multiple locations and is subjected to statistical testing to underline the necessity for quantitative analysis, original use of Mamdani and Takagi-Sugeno-Kang rule bases are explained and optimisation methods detailed in order to cater for multiple categories of plant elements on a global scale.
- Novel adaptive neural fuzzy logic control systems are designed for named locations and the rule bases are minimised to increase efficiency. Following numerical simulation and categorisation using validated global climatic data, the fuzzy logic schemes are tested in alternative locations in order to illustrate the robustness.
- Control systems are simulated to combine detailed description of high-resolution biogeographic data with genetic methods enabling original predictive function within the hyper-plane of primary producers. ANFIS and hybrid GA are suggested to inform geographic information systems.
- The structural basis of Geographic Information Systems (GIS) is used to query dynamic global modelling of plant characteristics, including life-forms and metabolism. Original inferential relations to ecological network stability and recommendations to national policy formation securing sustainability are made.

1.4 Research Questions

Research questions addressed in this thesis were:

- **RQ. 1.** Is it possible to predict the global occurrence of plant species and if so, what is the most efficient form of prediction? (Addressed in Chapters 4 and 5).
- **RQ. 2.** How can the trade off between accuracy and interpretability of predictive models be balanced? (Addressed in Chapter 5).
- **RQ. 3.** What sub-groups of plant characters should we differentiate to give adequate informative value to different subject areas in terms of plant characterisation? (Addressed in Chapter 6).
- **RQ. 4.** How are functional groups of plants distributed a) mathematically, b) geographically? (Addressed in Chapter 6 and 7).
- **RQ. 5.** Can we make predictions of climatic conditions using the occurrence of plant species and if so, how? (Addressed in Chapter 7).
- **RQ. 6.** Given that we can develop geographic information systems of plant characters with a refined mathematical basis, what further conclusions for conservation and sustainability policy formation may be made? (Addressed in Chapter 7).
- **RQ. 7.** Is there a required structure of groups of plant species in ecosystems to enable stability within a dynamic climatic system, if so, can we use a mathematical function to show this? (Addressed in Chapter 7).

The above research questions are made to give thorough coverage to plant characterisation, with evolutionary and climatic relevance, according to the authors motivations for this study, answers are given throughout as indicated. Summarisation of research questions is made in Chapter 8.

1.5 Published work

The papers published during the period of my studies at UWE are listed below:

- 1) J. Furze, Q. Zhu and F. Qiao, 'Collaborative proposal development towards sustainable communities', *Global Education Magazine*, Human Rights Day, Issue No. 6, pp. 28-32, 10 Dec. 2013.
- 2) J. N. Furze, Q. Zhu, F. Qiao and J. Hill, 'Functional enrichment of utopian distribution of plant life-forms', *American Journal of Plant Science*, Vol. 4, No. 12A, pp. 37-48, 2013.
- 3) J. N. Furze, Q. Zhu, F. Qiao and J. Hill, 'Utopian exploration of global patterns of plant metabolism', *Proceedings of the Fifth International Conference on Modelling, Identification & Control (ICMIC 2013)*, Cairo, Egypt, pp. 47-52, Aug 31st-Sept. 3rd, 2013.
- 4) J. Furze, Q. Zhu and J. Hill, 'The use of plant characterisation modelling studies to substantiate national conservation and sustainability policies', *Global Education Magazine*, International Day of Democracy, Issue No. 5, pp. 68-75, 15 Sept. 2013.
- 5) J. N. Furze, Q. Zhu, F. Qiao and J. Hill 'Linking and implementation of fuzzy logic control to ordinate plant strategies', *International Journal of Modelling, Identification and Control*, Vol. 19, No. 4, pp. 333-342, 2013.
- 6) J. N. Furze, Q. Zhu, F. Qiao and J. Hill, 'Mathematic methods to quantify and characterise the primary elements of trophic systems', *International Journal of Computer Applications in Technology*, Vol. 47, No. 4, pp. 314-325, 2013.
- 7) J. N. Furze, Q. Zhu, F. Qiao and J. Hill, 'Implementing stochastic distribution within the utopia plane of primary producers using a hybrid genetic algorithm', *International Journal of Computer Applications in Technology*, Vol. 47, No. 1, pp.68-77, 2013.

- 8) J. N. Furze, Q. M. Zhu, F. Qiao and J. Hill, 'Facilitating description of fuzzy control algorithms to ordinate plant species by linking online models', *Proceedings of the Fourth International Conference on Modelling, Identification & Control (ICMIC 2012)*, Wuhan, China, pp. 933-938, June 24-26, 2012.
- 9) J. Furze, J. Hill, Q. M. Zhu and F. Qiao, 'Algorithms for the characterisation of plant strategy patterns on a global scale', *American Journal of Geographic Information Systems*, Vol. 1, No. 3, pp. 72-99, 2012.
- 10) J. Furze, Q. M. Zhu, F. Qiao and J. Hill, 'Species area relations and information rich modelling of plant species variation', *Proceedings of the 17th International Conference on Automation and Computing (ICAC)*, pp. 63-68, Sept. 10, 2011.

CHAPTER 2

Literature survey

2.1 Introduction

The nature of plant characterisation is multidisciplinary, consisting of different subject areas that are generative in producing new areas of study. Although it would be difficult to provide complete reference of all areas that have contributed to progress within the subject, the author attempts to provide a balanced view of the areas that are of relevance to ongoing research and progress within the subject. This chapter is important because it presents a clear picture of the frontier for research work and justifies the necessity for potential concept/technique development in the thesis.

Plant characterisation originated in geographic exploration of plants. Alexander von Humboldt (1769-1859) published on botany and geography, founding biogeography itself (Wallace, 1878). Humboldt was one of the first to describe the increase in species richness towards the equator (Humboldt, 1806), the 'latitudinal gradient' as it later became known (Humboldt, 1808; Rosenzweig, 1995). Humboldt (1845-1858) was the first to identify the

Chocó region and Andean forests of Columbia, South America as one of the mega centres of plant diversity:

‘Die dem Äquator nahe Gebirgsgegend ... von Neugranada [today: Columbia] ... ist der Teil der Oberfläche unseres Planeten, wo im engsten Raum die Mannigfaltigkeit der Natureindrücke [today: biodiversity] ihr Maximum erreicht’ [Humboldt (1845), p. 12].

(English translation by Otté (1860, p. 10): ‘The countries bordering on the equator [meant is the present-day country of Colombia] possess another advantage ... This portion of the surface of the globe affords in the smallest space the greatest possible variety of impressions from the contemplation of nature [today: biodiversity]’ (Barthlott et al, 2005).

Humboldt (1808) hypothesized explanations for this diversity including complex topography and the variety of suitable climatic conditions in the Chocó region. He made the statement that plant richness declines at higher latitudes due to the fact that many species are frost intolerant and may not survive in the comparatively cooler temperatures of temperate zone winters, substantiating the water-energy dynamic (Wright, 1983). Wright continued that plant productivity is limited primarily by energy from the sun and water availability. He added, however, that the solar energy that transfers through each trophic level is what constrains richness as opposed to the total energy within a geographic area (Hawkins, 2003; Jetz et al., 2009; Wright, 1983).

Addressing these issues, this chapter is organised as follows: species richness distributions and bio-geographical approaches are outlined in section 2.2. Life-form categorisation and plant life-history strategies are detailed in section 2.3. The use of computer generated modelling of plant strategies is covered in section 2.4. Quantitative methods of pattern identification are explored in section 2.5, with an elaboration of fuzzy logic, genetic algorithms and the fusion of these two methods included in subsections. Knowledge guidance underpinning intuition in evolving systems of plant strategy estimation is laid out in section 2.6 and the literature survey is summarised in section 2.8.

2.2 Plant richness distribution and bio-geography

Barthlott (1996) used the term geo-diversity to describe the reasoning behind plant distribution patterns, geo-diversity being the variety of earth materials and processes that make up and shape a specific area of the Earth itself. Estimations of species numbers and their fundamental descriptions were greatly enhanced by the work of Russian geneticist Nikolay Vavilov (1926), his major work being on the ‘Centres of Cultivated Plant Origins’. Vavilov covered India, China, Indo-Malaya, Central Asia, the Mediterranean, the Near East, Ethiopia, South Mexico/Central America, and South America. The work of Vavilov provided the basis for a great Russian tradition, with the quantitative nature of his genetic approach facilitating later thinking on diversity mapping (Barthlott et al. 1996; Barthlott et al. 2005; Malyshev, 1975; Wulff, 1935).

There are two main approaches to producing maps of species richness; these are taxon and inventory based. The more accurate of the two approaches is taxon based, being maps of a species or higher taxon compiled from detailed sources of distributional information. The information may be gained from gridded maps, locality data from natural history collections or from expert drawn ‘polygon’ maps (Barthlott et al., 2007). The most detailed taxon based maps may extend to include field data of a particular species or group of species and in some cases this may include every single individual within the studied area. In inventory based mapping summary data of the floras of different countries or areas (geographic units) have been used in analyses (Mutke and Barthlott, 2005).

The methods of Barthlott and his group have become key in the estimation of numbers of species and their mapping in different areas. Recently, Barthlott has seminally published on numbers of plant species on a global scale (Barthlott et al. 2005). This work showed great progression from the times of Arrhenius and his expression of the power model of the species-area relationship (Arrhenius, 1921).

Barthlott proposed that data gaps in the SAR are filled by interpolation using additional information on climate, vegetation or geo-diversity (Barthlott et al., 2005; Barthlott et al., 2007; Kreft and Jetz, 2007; Sommer et al., 2010). The standard area used for species

number estimation, is 10000km². Surveys are typically carried out in measured areas (usually square) called quadrats. Quadrat size varies with the type of vegetation. The size of the measured area doubles until no further new species are found. The most effective size following this method in Barthlott et al. papers above was 10000km² when measuring vegetation on a global scale. Numbers of species at this resolution were measured in thousands. The beauty of Barthlott's approach was that it enabled description of areas on a global scale in terms of diversity. In a series of papers, progressing to the present day, members of the group of the Institute for Biodiversity of Plants, at the University of Bonn in Nees, Germany, where Barthlott is located, described 'diversity zones' (DZ). Indeed, Barthlott et al. (2005) proposed a leading description of over twenty different locations, covering description of the zones in which over 3000 species/10000km² were found. The most prominent climatic, geologic and floristic features of the 20 diversity zones (DZ) were given. Areas within DZ 8-10 were detailed, providing an invaluable reference, which is made use of in this study. Increasingly advanced methods are used to monitor plant diversity and its interaction with various patterns that are present at a global scale. The top 5 centres of diversity were described as those that have greater than 5000 species per 10000km² (DZ 10). Examples were Costa Rica-Chocó, Tropical Eastern Andes, Atlantic Brazil, Northern Borneo, and New Guinea (Barthlott et al., 2005; Mutke et al., 2010).

The work of Barthlott, and many others, can be said to be foremost in the development of geographic information systems (GIS). A GIS is any system that captures and summarises information about any geographic area (Trauth, 2006). Work on species distribution mapping therefore describes the remit of GIS. Today GIS systems are increasingly complex and may store and utilise vast amounts of information of many qualitative and quantitative types for species. One of the foremost ecological systems stored and updated electronically by multiple interfaces (users) is the Global Biodiversity Information Facility (GBIF).

Global biodiversity mapping is of great importance in the setting of conservation priorities. The assessment of species richness and endemism has had great impact on the setting of conservation objectives (Soosairaj et al., 2007). Endemism is a value of the extent to which a species originates from a specific area. Endemism is often strongly positively correlated with species richness, although in island situations there are instances of there being no relationship. The likely explanation for this is that the isolated nature of islands puts more genetic pressure on species to develop, which may have resulted in high levels of species

suited only to the conditions of the island (Kier et al., 2009). Whatever the nature of the relationship between endemism and species richness, it was seen as a truism that studies of endemism have great value in assessing relationships between and within species groups (Kreft et al., 2006) and, further, the effect of land use changes on species distribution (Barthlott et al. 2005). In an analysis of plant distribution and floristic knowledge it has been found that land-use change ('the human footprint') is very rarely positively associated with species richness (Kier et al. 2005). The analysis of species richness on islands was particularly useful in identifying environmental determinants, equilibrium theory and the high level of significance of the water-energy dynamic on biodiversity patterns per se (Kreft et al. 2008; Mutke and Barthlott, 2005). This is a currently developing field.

Areas with extremely high levels of diversity in the equatorial regions (particularly in South America) were also of great use in assessing the nature of the factors that drive diversity due to the unique combinations of water and energy related factors. Such locations contained species within complex ecological (e.g. neophytic) relationships and extraordinary high levels of growth (Bass et al., 2010; Gentry and Dodson, 1987; Kreft et al., 2004; Kreft and Jetz, 2007; Silvera et al., 2009). It is difficult to measure the great contribution that studies in these areas have had, both on the people of the areas and the protection status of species within the areas. Such areas have become increasingly vulnerable to human development due to the high levels of species richness and variation in all the major kingdom groups there and also due to the mineral (e.g. fossil fuels, oil) reserves found there (Bass et al., 2010).

There was much evidence to substantiate a mid-elevation rise in diversity found in neotropical areas. This 'burst' of diversity was found within 'cloud forests' in tropical montane regions, characteristically over the midpoint of increasing incline (Gentry and Dodson, 1987; Küper et al., 2004). This phenomenon further contributed to the significance of the water-energy dynamic on plant diversity and adaptive radiation evolutionary patterns (Silvera et al., 2009).

2.3 Qualitative plant life-form description

It was a remarkable coincidence that at roughly the same time as Wulff (1935) was making great progress in the mapping of species on global scale, a Danish Botanist called Raunkier (1934) was advancing the theory of describing plant form over different geographic patterns.

It is incredible that such an early theory, which appears to remain for the most part in its original state, is still of great relevance today in descriptions of plant communities. Gradients are often analysed and shown to have significantly different proportions of Raunkier life-forms. The latitudinal gradient has recently been restated in terms of plant life-form differing across Burkina Faso (Schmidt et al. 2005), a study in which a combination of field and herbarium data were used for increased accuracy. In the drier areas of northern Burkina Faso the number of therophytes was seen to be highest; towards the more humid southerly areas the number of hemicryptohytes and phanerophytes increased. The categorisation shows its flexibility and robust quality for inferring climatic data when it was again used across the Himalayas in Eastern Nepal (Bhattarai and Vetaas, 2003). Variation was shown along a subtropical gradient, highlighting the importance of water and energy in the distribution and plant species richness of such areas. This thesis makes great use of displaying results in 3 dimensions in order that the above effect can be clearly visualised, quantified and differentiated. Dimensions of x , y and z planes are proposed to stress the water-energy dynamic effect.

The classification of Raunkier can be described as qualitative. A categorisation more dependent on robust numerical partitions was formed by Ellenberg (1991). Ellenberg distinguished 7 major scales on which to define the relationship between the vegetation and the surrounding climate. This was originally undertaken across mid-European areas. The categories were as follows: temperature, continentality, light, moisture, reaction (soil acidity), nitrogen, salt. The main basis of the values is a species' realized ecological niche. However, mainly due to the geographic climatic variance there was a difficulty in transferability across different areas. In a recent paper it was found that Ellenberg values may be used to monitor environmental change, though transfer from one area to another required a strict algorithmic approach. The factor of continentality must be left out in such

cases, as it was difficult to transfer due to climatic patterns in effect across large bodies of land such as Europe and not across smaller islands such as the United Kingdom (Hill et al., 2000). The method of Hill et al. (2000) was, in summary, a two way weighted averaging followed by a local regression of values.

Using the method of Hill et al. (2000) it was possible to repredict values from a British sample. However, discrepancies arise between the values obtained from the study of Hill as compared to Ellenberg. These were due to different ecological requirements of the species measured, differences in microhabitats of the species, sampling error (due to human bias for homogenous sample areas), including species that were not closely associated and too large a sample area being used resultantly misleading values were obtained. Indeed, there were great problems with linking indicator values to direct measurements, particularly where a range of habitats or a large number of species were being measured (Hill et al., 2000). Difficulties were due to a shift in species behaviour across environmental gradients. A different scale may be used or refined definition to the same scale. Competition may be seen to vary in different environments. Additionally, species may be selected differently across different habitat types (Godefroid and Dana, 2007). Massant et al. (2009) combined GIS with an extrapolation of plant strategy values on a scale larger than 50m², made use of graph analysis to determine the spatial aggregation of strategies, logistic regression and elegantly used Ellenberg values in discussion of the differences between vegetation structure and strengthened the argument for plant strategies (Massant et al., 2009). The latter study also made use of conventional Boolean statistics in order to differentiate between the biotypes of plant strategies. Plant strategies are further discussed in the next section.

2.3.1 Plant life-history strategies and continuums

Plant strategies were based on plants' life history descriptions (Grime, 1979). It was stated 'A plant strategy may be defined as a grouping of similar or analogous genetic characteristics which recurs widely among species or populations, such that they show similarities in ecology' [Grime et al. (1995), p. 15]. The main categories are: Competitors (C), Stress-tolerating (S) and Ruderal (R) species.

The environment may infer the strategy and vice versa. Furthermore, in a description of the usefulness of plant strategy it was stated “Understanding the distribution of plant species across environmental gradients requires bringing theories together regarding the construction of plants, as well as their interactions with the environment, and the assembly of communities” [Craine (2005), p. 1041]. The possible flaws and merits of the plant strategy theories of Grime and Tilman, were given full discussion by Craine (2005; 2007). Craine recognised that both authors have developed exemplary theories regarding natural selection, the strategies of plants and the functioning of ecosystems. The question of nutrient availability and resource allocation (light) was key in Craine’s critique of Grime’s strategies. Craine accepted that high nutrient availability enables the dominance of C strategy, and low nutrient availability enabled the dominance of S strategists. Craine considered that Grime’s theory does not cater for low nutrient supply at constant rates, rather he recognised that a spectrum of the strategies exists. The traits of species of each strategy allowed more than one strategy to be adopted at highly variable nutrient resource rates. Considering S strategy adaptability to low levels of light, Craine stated that amongst certain groups of species S strategies alternatively be stated as C strategists due to the fact that species more tolerant to low levels of light were, in fact, better competitors in relative terms. Craine (2007) limited his arguments of Grime’s strategies and did not consider the mixture of strategies that clearly occur, especially given varying levels of water and energy factors. This thesis concentrates on the above matter in Chapter 5 answering RQ 1 and 2. Craine (2007) criticised Tilman’s strategy theory as having placing too much importance on modelled strategies and did not recognise that Tilman provides sufficient evidence in support of his models. Craine (2007) did recognise that future paradigms of strategies must combine elements of both Grime and Tilman, which is what we saw in the development of Massant et al. (2009) and this thesis progresses this development yet further in order to predict the numbers of species in strategy based environments.

One of the drawbacks of strategies in the classical form is the difficulty/complexity with which plants are ordinated into each of the different types, making the categorisation sometimes time consuming and impractical when applied to large numbers of species. For example, Grime et al. (1995) listed 20 characteristics of the three strategies that have proved useful in classifying plants: main groups included types of morphology (including life-forms), elements of specific life history, physiological descriptions of growth rate (including photosynthetic mechanisms) and miscellaneous elements (such as litter description, palatability to other unspecialized herbivores and DNA content, e.g. numbers

of chromosomes, numbers of sets of chromosomes). Some or all of these factors may vary in different environments, requiring that the criteria for plant strategy characterisation are clearly stated in the specific studies where it is applied (Grime et al., 1995; Hodgson et al., 1999; Kraft et al., 2008; Massant et al., 2009; Mustard et al., 2003).

In accordance with the original definition of Grime et al. (1995) it was logical to attempt or to predict specific definition within the patterns of plant species as genetic characteristics are distributed in this way (Hunt and Colasanti, 2007). This was made use of by Massant et al. (2009), where spatial patterns are considered on a meso scale (larger than 50 m x 50 m). Clustering of the strategies was found which was then explained using available environmental factors. In an aggregated pattern, the strategies showed non-random and unequal distribution. Using multi-variate statistical methods, 'clouds' of data were shown, which indicated the formation of definite 'biotopes' (areas in which certain strategies predominate). Biotopes were seen to form both in accordance with habitat and disturbance (e.g. high competitive values indicated competitive (C) biotopes found under pine; stress tolerant plants (S biotopes) were found under mixed oak-beech and pure beech stands of 100 to 150 years old; and ruderal plants (R biotopes) were found nearest roads). Detrended component analysis should be employed with caution when used with environmental data (Grime et al., 1995) due to the 'arching effect' of data trends on the resulting analysis. Massant et al. (2009) also relied on several methods of analysis (variance values of C, R and S, with defined measurement positions of samples, logistic regression giving probability of C, R and S, calculation of weighted averages, rank order of Ellenberg values) before making their conclusions. The work provided a useful linear interpolation of plant strategy, converting into 3 dimensions: C = (10,0,0); S = (0,10,0); R = (0,0,10); in intermediates S-R, C-S-R, C/C-S-R, etc. the C-S-R values sum to 10.

On consideration of Grime's strategies in *Quercus cerris* L. var. *cerris* woodland in Samsun, northern Turkey, Kiliç et al. (2010) stated that the categorisation was considered appropriate as habitat diversity and environmental factors were present in different combinations, equating to various functional modifications in plants. The study found that plant species conformed to definite strategies and it further confirmed that plant traits are subject to key factors of competition and disturbance in their selection. However, using the classical predictor values described by Hodgson et al. (1999) was not appropriate for *Ruscus aculeatus* due to the stem being the photosynthetic organ as opposed to the leaf (which was used in the calculation for other species). Species which store large amounts of

water in their leaves may also be mis-classified due to the classification requiring calculation of specific leaf area and leaf dry matter. The problem also exists for species adapted to live in very shaded and high salt environments. This is also in accordance with the conclusions of Hodgson et al. (1999) and it was recommended that future application of strategies should be by different calculation using more robust methods (Kilinc et al., 2010).

Allowing for progression of the theory as identified by Craine (2005; 2007), and adjusting methods or combining analysis as in Massant (2009), Kilinc et al. (2010) showed that further application of plant strategies is a powerful approach in describing vegetation at a range of scales. The idea of plant strategies is a very elegant approach that may be seen as logical, compared with another approach which categorised strategies of plants and other species, the rK strategy theory of MacArthur and Wilson (1967).

The basis of the rK selection theory was that evolutionary pressures work in two directions, allowing two ends to the selection continuum (MacArthur and Wilson, 1967). In unstable or unpredictable environments the ability to reproduce quickly and with many offspring (seed and or asexual propagules in the case of plants) is important – such species were termed r-strategists. Though they have a high probability of reproduction, r-strategists were poor competitors, as there is little advantage to competing with others due to the changing environment. Typical traits of r-selection were short life span, early onset of breeding, high fertility, early onset of maturity, short generation time and poor maternal quality. In terms of Grime's strategy, r-strategists may be described as being ruderal (R) species. In diverse, stable environments the ability to compete was more important as there were available but limited resources, hence K-strategists existed in higher proportions. Populations with more K-strategists were described as being closer to their carrying capacity (or in a state of climax). Organisms with K-selected traits were typically large and had lengthy life spans or extended generation time. In terms of Grime's strategy, K-strategists may be described as being competitive (C) species. It was a logical progression to recognize that there is an rK continuum with some species having elements of r and K selection. Although rK selection theory was originally developed in the context of biogeographical patterns of species distribution on islands (MacArthur and Wilson, 1967, Simberloff, 1974), it is possible to make use of the theory in many ecological models and population studies. Knowledge and recognition of the rK continuum is often causal, it being fundamental in the distribution of evolutionary patterns. Various authors have made

use of Grime's theory in simulation modelling, although the application of similar models in the context of rK is one that was useful in current (and future) distribution prediction (Barreto, 2008), and is developed in this thesis. Simulations of strategies are discussed in the following section.

2.4 Computer generated plant strategy modelling

Mustard et al. (2003) recognised that 3 main strategies exist for plants, as proposed and supported by Grime et al. (1995), and they acknowledged that environmental factors correlate strongly with the main axes of strategy variation (resources and disturbance). However Mustard et al. (2003) continued in their analysis and concluded that the nature of predictability for evolutionary outcome was as yet undefined. Increased definition may help to assert the primary drivers of plant strategy variation. Mustard et al. (2003) studied the evolution of model plant populations in environments, the object of their study was to determine whether the previously described plant strategies would evolve, and to ascertain whether the pattern was consistent with previously described theoretical and field evidence. After simulating nitrogen availability and disturbance frequency alone, they observed the emergence of 3 distinct strategies.

The virtual plant population model of Mustard et al. (2003) was set up in the following way: A computer generated spatial array constituted the virtual environment in which a mutable model of single plant growth (MSPM) was the second part of the model. In the virtual environment the 'plants' were iterated through constant plant growth with constant variables. In a parameter-rich, complex model plant traits developed, being confined to competition for space, light and nitrogen. All other potentially limiting factors such as herbivory, 'biomechanical' constraints and water and temperature variation were excluded from the model. Thus, the simulated plants that evolved differed only in their life history and ability to capture resources. Simple patterns were imposed upon the plants by not allowing them to combine sexually but to produce seeds asexually. Each trait mutated independently, with no genetic linkage between traits assumed. In the above model there were 29 mutable parameters/plastic responses. These can be seen as those which were measured to distinguish plant strategy, compared to the 20 that were provided as measurement criteria by Grime et al. (1995). Mustard et al. (2003) set the probability of

mutation such that $1/29$ of the traits mutated for every new seed produced. Structurally, the model plants consisted of a stem, leaves and nine roots all determined by specified mathematical relationships. The model simulated natural selection for plant strategy and resulted in improved reproductive performance in the simulated environment. The model developed was therefore biased towards *r* or ruderal selection due to the exclusion of the other limiting factors listed above. However, Mustard et al. (2003) were able to suggest that plant strategies emerge in changing conditions of resources and disturbance. They provided a highly valid modelling tool that could be made use of in the future to indicate different features of populations over time.

The use of computer modelling as a tool to monitor plant growth was further developed by Hunt and Colasanti (2007). They set the aim of identifying the ‘centre of organization’ of plants in response to environmental stimuli. To ascertain whether plants can be self-assembling, Hunt and Colasanti proposed the use of a cellular automation (CA) model of plant growth, comprised of identical ‘modules’ making up branching structures of the whole plant. The modular activity was driven by morphological, physiological and reproductive rule-sets, which were derived from comparative plant ecology. Experimentation with the virtual plants showed a wide range of whole plant, population and community level responses. Impressively, it is concluded that a C-S-R version of the CA model simulated the most essential properties of natural vegetation and its inter-relationship with environment. Hunt and Colasanti (2007) recognised and defined the 7 combinations of plant strategies in terms of three groups of factors: morphological, physiological and reproductional. CA is currently being used to investigate trophic levels beyond the plant environment interface.

A later author, Barreto (2008) carried out work in which the *rK* continuum was proposed to accommodate the seven life history strategies of the C-S-R model of plant life history based strategies. Barreto ‘mapped’ primary and secondary strategies of the C-S-R model in the linear *rK* continuum, using an algorithmic approach. The paper made use of the calculation of the entropy measure of importance, which calculated the amount of organization within data (discussed in a later section) and the concept of geometric euclidian proximity to calculate relative proximities. Growth parameters required to set growth vectors were established by reference (Brzeziecki and Kienast, 1994) to calculate the proximity (*p*) to archetypes at either end of the continuum (i.e. *r_p* and *K_p*). For a given point in the scale $r_p + K_p = 1$, the higher a species *K_p*, the closer it is to the paradigm of the

K strategist. Thus, for the r (ruderal) archetype $K_p=0$, and for the K (competitive) archetype $K_p=1$ (Barreto, 2008).

Bornhofen and Lattaud (2008) made use of strict definition of Grime (1995) in prescribing the evolution of C-S-R strategies in virtual plant communities using artificial life concepts. They developed the work carried out by Mustard et al. (2003), using a mutable model of single plant growth. Bornhofen and Lattaud (2008) recognized that the model is restricted in its parameters and therefore gave only a highly simplified morphology, unable to evolve.

To progress, Bornhofen and Lattaud (2008) made use of an alternative approach, moving towards the use of 'L-systems' (programmed visual systems with defined morphology) in evolutionary plant modelling. L-systems (Lindenmayer systems, Lindenmayer (1968)) were developed greatly by Damer et al. (1998) in the form of 'The Nerve Garden', a computer application that allowed users to grow and interact with virtual plant communities created in online worlds. Ebner (2006) showed the development of competitive aspects of virtual plants in response to light and noted that the use of interacting algorithms enabled co-evolution in virtual plant communities. In virtual communities, functional structural plant models (FSPM) were a developing form of mutable single plant model (MSPM). FSPM combined morphology with physiological processes based on artificial life concepts, and produced 3-D architecture combined with resource assimilation, flow and allocation. An artificial genome was described as the mutable code which has morphological aspects and physiological processes. Similar to a real life scenario, the code was acted on by evolution and favoured individuals adapted to the selection process (Bornhofen and Lattaud, 2008). Using an 'L-system alphabet of the used plant model' and a defined environment in which rules are set out for development of plant phenotype, genotype and life cycles, experiments were carried out simulating evolutionary adaptations in environments with heterogeneous levels of disturbance and mineral stress. Propagation dynamics of the resulting population of virtual plants was measured, as were physiological and morphological adaptations. Plant strategies were visualized using multi-variate analysis of the data.

The results of Bornhofen and Lattaud (2008) were consistent with those of Grime (1995) in that ruderal plants showed low maturity (i.e. in response to frequent disturbance early seed production took place). Low seed biomass allows production of many seeds. As

ruderals were evolved with high growth rates, their selection favoured high resource use, which accelerated their life cycle. The collection of traits of ruderals also matched the life history strategies of r-selected plants (Barreto, 2008; Pianka, 1970). Stress tolerators evolved the longest life span, with slow growth and reproduction due to few soil resources.

As well as distinguishing the emergence of the same relationships of plants as in the Grime (1995) C-S-R triangle, Bornhofen and Lattaud (2008) were also able to describe the morphological forms of the L-system plants and recognized the formation of three distinct types. The evolved morphology of competitors included a high stem without branches in order to reach light rapidly in crowded environments, and root systems poorly developed due to the presence of abundant mineral nutrients. Ruderals showed the most simple, condensed morphologies, with poorly developed roots, producing seed as a priority. Stress tolerators displayed well developed root systems and variable shoot systems. The descriptions of FSPM by Bornhofen and Lattaud (2008) were both quantitative and qualitative, and showed their great potential in evolutionary studies at both the population and larger scales by integration into adaptive algorithms (Prusinkiewicz and Lindenmayer, 1990).

L-systems continue to develop in their complexity and capacity to model biological systems. For example, Bornhofen et al. (2011) increased the iterative power of their original model by producing one that has levels for the individual plant, community and wider environment. In brief, the model operated in the following way: Individual genotypes with L-system, physiological and life-history parameters gave plant state at a defined time. This iterated through the level of the individual and community, and interacted with soil nutrients and light in the wider environment. As with the Bornhofen and Lattaud (2008) study, the gradients of disturbance and stress intensity were distributed over a 5x5 square grid and strategies evolved in patches. Conditions imposed upon the 'plants' were more strictly defined than in Bornhofen and Lattaud (2008) and hence Bornhofen et al. (2011) showed more explicit examples of physiology and architecture. The work gave examples of the calculations used to determine shoot and root compartments in the virtual plants and described the relationship by which carbon and nutrients are assimilated and cycle back to the environment. Such equations may be applied in future L-systems to model the effects of plant growth in variable environments, for example with temperature influencing fixed growth constants and water influencing nutrient concentration.

2.5 Quantitative pattern identification

When considering large numbers of species in virtual or real-life situations there are various methods that can be used to show clustering or grouping patterns. The recent work discussed on L-systems by Bornhofen et al. (2011), for example, made use of multi-variate principle component analysis. However, it is possible to use alternative, sometimes simpler, methods. In reality, use of principle component analysis and detrended correspondance analysis may be difficult due to the nature of data being analysed and or due to the ‘arching effect’ of continuous variables. In such situations, as in Barreto (2008), it was more pertinent to make use of basic clustering methods. One such method, nearest neighbour, takes advantage of the pattern of data in their own (geometric) space. Using this method the coefficient of squared euclidean distance may be obtained. The coefficient is based on the properties of a right-angled triangle, the square on the hypotenuse being equal to the sum of the squares on the opposite two sides. Thus, if two or more species occur in two or more samples or areas 1 and 2, n , the similarity or ‘distance’ (D) between the species in geometric space is defined as:

$$D_{i,j} = \sqrt{\sum_{k=1}^m (X_{i,k} - X_{j,k})^2} \quad (2.5.1)$$

$D_{i,j}$ = squared Euclidean distance between quadrats/samples i, j .

m = number of species

$X_{i,k}$ = the abundance of the k^{th} species in quadrat/sample i

$X_{j,k}$ = the abundance of the k^{th} species in quadrat/sample j

Euclidean distance is often made use of in more complex ordination calculations such as polar ordination. Geometric calculations were part of the calculation for relative proximity (Barreto, 2008).

In GIS systems basic clustering methods such as Moran’s I are used to show association/dispersion of species with each other. This can be seen to great effect in Kreft and Jetz (2007) where the clustering method was used to show the importance of various climatic variables versus the species richness on a global scale.

In the novel approaches of the current study, euclidian proximity may be inferred with measures of the amount of entropy (organization). Barreto (2008) made use of the calculation of entropy between various plant growth related parameters as part of his algorithm COMPTO in the calculation of relative proximity values as detailed in sections 2.3.1 (p. 18) and 2.4 (p. 20). Entropy was first established for use in thermodynamics, but in information theory entropy is used as a measure of uncertainty associated with random variables. Shannon entropy is a measure of the ‘surprise’ or probability of an event occurring. Thus, entropy has found a place in modelling distribution of biological factors in ecology (Elith et al., 2011; Phillips et al., 2006). Elith et al. (2011) clearly explained how to calculate entropy using mathematical software. Entropy was calculated using the species presence data and a combined covariate value obtained from environmental data of the region being modelled. This is very useful as previous description such as that of Shannon (1948), have great application in machine learning, although are only accessible to ecologists through complex statistical software code. Full description was made of the process of the calculation in an informed paper and accompanying appendix (Elith et al., 2011). Key components were defined (e.g. covariates, features and the landscape) and the process of model fitting (feature selection, constraints and regularization) was elucidated. Case studies were described for a range of both plant and animal species, the models were shown and interpretation given. The major use of Elith et al. (2011) is that the process of mechanistic species niches is eloquently described using quantitative means. It is speculated that such description may make the calculations accessible using alternative software platforms (for example, R statistical software or Matlab). This is discussed in Chapter 3 of the current study. The possible use of the calculation of entropy, as in Barreto et al. (2008) for the approximation of relative proximity matrices on which rK continuums may be based, is considered.

2.5.1 Fuzzy logic

The concept of fuzzy logic is most easily considered with respect to fuzzy sets. Fuzzy sets were conceived by Lotfi Zadeh, a member of the University of California, Berkely. Professor Zadeh (1965) proposed that it is possible to form answers and operate perfectly well without concise numerical input, as humans operate under these conditions. It follows, therefore, that feedback systems controlling (defining) large operations can be

programmed to accept ‘noisy’, or inaccurate input, which would render them a great deal more effective and easier to implement than more complex systems. In formal definition of fuzzy sets, Professor Zadeh (1965) stated: “Let X be a space of points (objects), with a generic element of X denoted by x . Thus $X = \{x\}$. Then, a fuzzy set (class) A in X is characterized by a membership (characteristic) function $f_A(x)$ which associates with each point in X a real number in the interval $[0, 1]$, with the value of $f_A(x)$ at x representing the ‘grade of membership’ of x in A .”

Fuzzy logic can be seen both as a method to control systems or to classify them. The beauty of fuzzy logic is that it concentrates on what a system should do rather than trying to model how it works. The approach therefore may be seen to be completely devoid of semantic definition. The concept of the membership as applied to fuzzy and classical (crisp) sets of data is a curve defining how each point in the input space relates (is mapped) to a membership value of the set. This may also be termed the degree of membership between 0 and 1. The values 0 to 1 are used to describe the degree of certainty that the point belongs to the set. Membership function is denoted as μ . The values within fuzzy sets may be any value between 0 and 1, whereas the membership function of classical sets is more discretely defined as 0 or 1.

Fuzzy sets are often defined using existing mathematical models. Such a mathematical model is Gaussian or ‘normal’ distribution. The use of this and other weighted distributions was discussed in terms of environmental indices (Silvert et al., 2000; Sivanandam et al., 2007). Gaussian distribution was used to great effect when using fuzzy logic in biological data (Taheriyoun et al., 2010). The paper by Taheriyoun et al. (2010) is very useful to the current author. Although Taheriyoun et al. considered reservoir water quality evaluation, they do so in such a way that 3 different classes of water body are implicated, using 4 different indicators. A parallel with vegetation distribution is drawn in the context of the current study. Competitive, stress tolerant and ruderal types of plant strategy could be considered as classes, with the seven possible combinations of these and their environments being the membership functions.

At this stage of the review it is pertinent to state that it would be possible to make use of the classification latterly given using the Silvert (2000) definition of multiple membership function:

$$\mu/(1-\mu) = \{[\mu_1/(1-\mu_1)]A[\mu_2/(1-\mu_2)]B[\mu_3/1-\mu_3]C...\}1/(A+B+C+...) \quad (2.5.1.1)$$

The inception of (2.5.1.1) is rooted in normal distribution, hence, recent developments in fuzzy logic included examples that facilitated characterisation of plant species distribution patterns (Broekhoven et al., 2007; Herrera, 2005; Nasibov and Peker, 2011; Taheriyoun, 2010; Wang and Yang, 2010; Wendt et al., 2010). Broekhoven et al. (2007) made use of linguistic terms in fuzzy classification in ecology. Herrera provided a useful review of genetic-fuzzy systems. Nasibov and Peker (2011) detailed the process by which increasing the frequency of observation resulted in an expansive differentiation of membership function. Wang and Yang (2010) provided an example of the fusion of objective clustering and genetic algorithms to elaborate and investigate data patterns. Wendt et al. (2010) detailed how knowledge guidance may be used to estimate and optimise parameters in environmental modelling. The methods of these authors demonstrate the power of mathematic methods to expand environmental and species information within dynamic condition. As such, they are seminal to the author of this thesis and enable formation of the novel application fuzzy logic to more than 300 000 plant species. The use of fuzzy logic and related methods enable variables used in modelling to be concisely expressed in terms of x and y and further expand the informative value of combined objective planes. This point is further discussed in the context of genetic algorithms in the following section.

2.5.2 Genetic algorithms

Further methods (commonly in the field of evolutionary computation (EC)) used to prescribe patterns of species modelling include the genetic algorithm (GA) for rule set prediction (GARP). GARP creates ecological niche models for species, using defined points. Probability of the species occurrence can then be mapped elsewhere depending on the agreement or rule sets. GARP is based on genetic algorithms. These were described as being searching functions, which generated the best fit to criteria using the genetic parameters of evolution and natural selection- inheritance, mutation, selection and crossover events (Goldberg, 1989). They were robust, stochastic evolutionary computational algorithms (Su et al., 2009). It is possible to calculate the genetic algorithm of a defined pattern using statistical software. Cao and Wu (1999) made a concise description of the use of genetic algorithm using the statistical software, Matlab. GA are

also made use of in the teaching of various computer languages. GA are adaptive algorithms for finding the best (global) solution to optimization problems. The stages of GA are as follows: start with a population of randomly generated ‘chromosomes’ (not actual chromosomes but sequences of defined length)- initialisation. The collection of ‘chromosomes’ (sequences) then evolve through a form of natural selection. Each iteration of selection is known as a generation. The chromosomes are rated for their ‘adaption for solutions’ or potential to solve the problem. On the basis of the evaluation a new population of sequences is formed using a process of selection. At this point genetic processes analogous to crossover and mutation take place. After further selection, given the solution is found, an output is given. An evaluation or fitness function must be devised for each problem to be solved. Specific solutions are devised, being a sequence or chromosome; the evaluation function returns a single numeric value proportional to the adaptation of the solution represented by the chromosome or sequence (Cordón et al., 2004). Populations may go through continual cycles of the GA depending on complexity of the population itself and the changing conditions in which it exists.

Many systems and models have been improved using the intuitive nature of genetic algorithms. Examples include the successful estimation of parameters included in a biochemically-based model of photosynthesis (Su et al., 2009), climatic parameter estimation in terms of quantitative precipitation forecasting (Lee et al., 2009), and the wider use of a knowledge guided genetic algorithm for input parameter optimization in environmental modelling (Wendt et al., 2010). The latter study used GAs to great effect in the elucidation of unpredictable events such as forest fire occurrence. GAs were seen to produce solutions that are as good or better than those found by non-linear curve fitting and they also allowed complex problems within populations to be taken into account, formulated and solved (Su et al., 2009). GAs hold great potential in future modelling applications. Indeed, Lee et al. (2009), studied heavy rainfall in Korea in 2005, quantified related variables such as the reduction rate of potential energy for cloud structure and filtered numerical stability. They devised a fitness function for the quantitative precipitation forecast covering the Korean Peninsula and concluded that optimisation of computer-generated data as well as that of physical parameters improved the overall GA model performance.

2.5.3 Generative fusion of genetic fuzzy systems

Herrera (2005) clearly described linguistic rules that govern fuzzy sets; the ‘IF-THEN’ rules are related by the concepts of fuzzy implication and the compositional rule of inference. A fuzzy rule based system (FRBS) is made up of a knowledge base (KB) including expert information making the linguistic rules, a fuzzification interface to transfer crisp data into fuzzy sets and an inference system to use the fuzzy rules with the KB in reasoning. Importantly there is also a de-fuzzification interface that translates the fuzzy rule action into a real action. Genetic learning systems and an algorithmic approach is key in genetic fuzzy systems. Further, there are ‘Michigan’, iterative rule learning (IRL) and cooperative-competitive approaches that may facilitate the system.

Two main problems occur when generating a knowledge base in genetic fuzzy rule base systems (GFRBSs). Firstly, in taking specification of the ranges (universes of discourse) of the data and resolving the number of labels for each linguistic variable, which fuzzy memberships are associated to each label. Secondly, it is important to define the number of rules and their composition (i.e. which specific label is associated with each linguistic variable). Methods of resolving the problems incorporate genetic learning of the rule base, genetic tuning and genetic learning of the knowledge base and the database. For complex systems there are alternative methods as documented by Cordon et al. (2004). Complex methods documented included fuzzy sliding mode control, neural network, neuro-genetic fuzzy systems, parallel genetic algorithms, co-operative co-evolutionary algorithms and Lamarckian co-adaptation.

An example application of a FRBS was given in Broekhoven et al. (2007), where a genetic learning process was made use of to increase the crisp application of linguistic terms used in a fuzzy ordered classification. Habitat suitability models were defined for 86 species and the information made use of by application of genetic algorithm to present an FRBS. Definition of linguistic values was given along with the variables considered in habitats. Linguistic values were also shown graphically, providing a useful model for alternative applications. Qualitative terms applied are absent, low, moderate and high (Broekhoven et al., 2007). Linguistic terms may be optimized with use of entropy, allowing the precise statement of a genetic algorithm. The genetic algorithm may be stated to give precise description of the membership functions particularly effectively with use of optimization intervals for the membership function parameters. The use of binary coded as well as real-

coded genetic algorithm may be applied. The representation of the membership function parameters by a binary vector using Gray encoding (Goldberg, 1989) restricts the values the parameters can take to a limited set of values which are defined by the upper and lower bound of the optimization interval and the length of the binary string. Accuracy of allocation of the data points within the 'input space' can be further increased by use of weighted averages. The functionality of this approach is shown to great effect, in concise statements of ecological network organisation which enabled greater accuracy in analysis of ecological relations (Broekhoven et al., 2007; Taheriyoun, 2010). The latter unveils the process by which we may extrapolate functions of species and variables within the current limits of our knowledge of the processes and also gives us motivation for novel investigations of further processes. The novelty of this approach is made use of in this study in terms of plant characterisation, further discussed in Chapters 3, 4, 5, 6 and 7.

Fusion of techniques to form 'hybrid' methods is seen to be highly productive. One such example was recently described forming an iterative fuzzy identification method, which hybridized a modified objective clustering method with genetic algorithm (Wang and Yang, 2010). The method involved combining the fuzzy c-means (FCM) algorithm. A robust, compact fuzzy partition was obtained within the input space, obtained by iteration. Resulting parameters were estimated using Kalman filter based algorithm (Angelov and Filev, 2004). Combining the methods in this way maximizes the robustness and accuracy of the partition. The method was shown to be superior to other methods using an electronic simulation example (Wang and Yang, 2010). The method shown with application to a continual fuzzy identification used by Angelov and Filev (2004) may have important consequences in the present case when used together with more recent methods such as those discussed above by Broekhoven et al. (2007). Parts of the essential methods are developed in Chapters 3, 5 and 6. Speculatively, the methods will enable concise statements to be made which will facilitate future monitoring and greater accuracy of plant richness, rudimentary characteristics (plant strategies, life-forms and photosynthetic type), clarifying the relationship between antecedent data input and consequent distributions.

2.6 Knowledge guidance hybrid systems

The development of rule based (*eR*) systems may use a summary of the information, in terms of the potential of a new data sample (such as accumulated spatial proximity information), to trigger the new rule base. Greater generality of the structural changes to the data was therefore catered for (Angelov and Filev, 2004). In terms of plant strategy estimation, for example, we may state the membership functions in terms of the seven environments (Barreto, 2008) and obtain the data for the placement of species within each by considering optimization limits (given that the initial number of species is known (Broekhoven, 2007)). Using this method we may obtain stochastic matrices, effectively a *Kp* or *rK* continuum.

Angelov and Filev (2004) gave detailed description of an approach, which may be used when a combination of monitored (supervised) and unmonitored data were used. The work shows how a rule base and parameters of it evolved with the introduction of new data. Initial rule sets and linguistic parameters may be defined as in the Takagi-Sugeno-Kang reasoning method. For Gaussian-like antecedent fuzzy sets an important expression was given thus defining the distribution:

$$\mu_{i,j} = e^{-a[x_j - x_{i,j}]^2} : i = \{1, R\}, j = \{1, n\} \quad (2.6.1)$$

$a = 4/r^2$ and r is a positive constant, which defines the spread of the antecedent and the zone of influence of the i^{th} model (radius of the neighbourhood of a data point); too large a value of r leads to averaging, too small a value to over fitting (causing distortion of the data pattern). The methods of Angelov and Filev (2004) demonstrated an approach applicable to complex chaotic patterns.

An alternative approach to dealing with complex data sets was mentioned earlier in the section on computer generated modelling of plant strategies. The technique for order preference by similarity to ideal solution (TOPSIS) was used effectively in fuzzy systems models (Hung and Chen, 2009). TOPSIS was particularly efficient when used in combination with an entropy based weighting within ‘intuitionistic’ fuzzy sets (IFS). In brief, in TOPSIS the most preferred alternative has the shortest distance (or least) from the

positive ideal solution (PIS), and the farthest distance (or greatest) from the negative ideal solution (NIS). TOPSIS is simple, rational and easily understood. It was efficiently computed and has a great ability to measure relative performance in a simple mathematical form (Hung and Chen, 2009). In essence, one may consider TOPSIS as a multiple criteria decision making (MCDM) process by clustering data. Its use with entropy calculation may thus be seen as complementary.

Combining entropy calculation with TOPSIS requires a six step process using the method of Hung and Chen (2009). The steps are detailed using theory and an illustrative example with data in the paper. The steps were: construction of an intuitionistic fuzzy matrix; determination of the criteria weights using the entropy-based method; construction of the weighted intuitionistic fuzzy decision matrix, determination of the intuitionistic fuzzy positive ideal solution (IFPIS, A^+) and intuitionistic fuzzy negative ideal solution (IFNIS, A^-); calculation of the distance measures of each alternative A_i from IFPIS and IFNIS; calculation of the relative closeness coefficient (CC) of each alternative and ranking the preference order. The illustrated example of Hung and Chen (2009) considered five companies with respect to three criteria. The example could easily be extended to accommodate seven environments with respect to three criteria (plant strategies) in the current study.

An example of the rK continuum was provided by Barreto (2008), as described in a previous section. In the approach Barreto (2008) made use of his algorithm COMPTO. He combined euclidian proximity with entropy calculations within growth parameters of the species he was considering by comparison to previously calculated values (Brzeziecki and Kienast, 1994). Barreto (2008) achieved reconciliation of the rK and C-S-R models by generating the following values within the rK continuum (0-1): K_p 0.000001-0.142857 equal to the r3 archetype, ruderal strategy; 0.142858-0.285714 equal to the r2 archetype, stress tolerant ruderal strategy; 0.285715-0.428571 equal to the r1 archetype, competitive ruderal strategy; 0.428572-0.571428 equal to the rK archetype, competitive strategy; 0.571429-0.714285 equal to the K1 archetype, competitive stress tolerant ruderal strategy; 0.714286-0.857142 equal to the K2 archetype, competitive stress tolerant strategy; 0.857143-0.999999 equal to the K3 archetype, stress tolerant strategy. Barreto's study successfully applied a linear numerical basis to the discrete categories of Grime's strategies. This novel approach allowed unification of Grime's strategies with Raunkier's life forms. Such a unified categorisation provides substantiation to how strategies, life-

forms and photosynthetic type are quantified, the novel application is used at a global scale in the current study.

In the current study, as species numbers are known, Broekhoven et al. (2007) provided a good simplistic template to follow. However, the author of the present study postulates that characterisation of the plant strategies with the example rK continuum given by Barreto et al. (2008) may also be used (using values as the fuzzy centres of data for the seven environments) and then calibrated for each of the environments under consideration with an iterative genetic algorithm approach. An alternative knowledge base may be created, by summing species numbers equally allocated across seven possible environments (Bornhofen et al., 2011; Bornhofen and Lattaud, 2008; Grime et al., 1995). This approach is given further discussion in Chapter 3.

Climatic data may provide probability of species occurrence (Elith et al., 2011). The distributions of each knowledge base may be compared in order to increase the accuracy of the solutions by successive iteration (Wendt et al., 2010). Climate data (using data and calculated co-variates from Intergovernmental Panel for Climate Change (IPCC), advised water/energy factors (Kreft and Jetz, 2007)) will provide the current study with factual entropy based optimization values. Applying these values to the equally distributed estimated knowledge base with a genetic algorithmic approach, until the number of individuals by iteration equals the number in the data samples provided, should produce accurate model allocation (Cordón et al., 2004). Insertion of data in the genetic algorithm approach takes place during mutation of the sequences (chromosomes) representing the individuals. Knowledge guidance may infer a greater probability of mutation on the population (Wendt et al., 2010). Furthermore, algorithms can be written to summarise the distribution of species into different strategies across the different environments. Additional weighting can be placed with use of Euclidian distances, increasing the knowledge-guided retrieval of the data, by nearest neighbour, as described in section 2.5, or by heterogenous Euclidean overlap measuring the similarity between individual events (characterisations) (Wendt et al., 2010).

2.7 Summary

In this critical review of global plant characterisation and its distribution over evolution and with climate change, the author has attempted to present an overview of progress within the subject ranging from work of great historical significance (Humboldt, 1806), through computer simulated situations and mathematical model formation to current methods of modelling, identification and control previously untested in the specific case of global plant characterisation but with great application potential.

Approaches to characterizing current plant distribution may use several methods. There is a great choice of future potential methodology, and these are given further consideration in Chapters 6-8, Chapter 6 answering RQ. 3 and 4, Chapter 7 resolving RQ. 4, 5, 6 and 7.

In this thesis adaptive neural fuzzy inference systems (ANFIS) are developed to concisely state the ordination of plant characteristics on a global scale. Mathematic method is advanced by hybridization of genetic fuzzy systems with use of expert based intuition, and explorations into the (Z) hyper-plane of plants are made. The use of mathematics is applied to global models in order to answer research questions related to plants, climatic prediction and other component disciplines such as plant science, biochemistry and sustainability of ecosystems, together with the communities dependent upon them.

CHAPTER 3

Preliminary methodologies

3.1 Introduction

This chapter discusses preliminary methodologies in order to add definition and give clarity to the characterisation of plant species. Detail of the data sources used in this thesis is also given.

The chapter is structured as follows: 3.2 covers the method by which data have been sourced including validation and reasoning for the chosen biodiversity source. The types and reasoning for differentiated climatic scenario and topographic data are given and the use of high resolution data is justified in this section; section 3.3 outlines image processing carried out and its potential application and use in this study. The species area relationship is covered in section 3.4; section 3.5 details the process of variable partitioning. Section 3.6 outlines the background of groups of characters under which plant distribution and characterisation is carried out. Section 3.7 introduces the use of fuzzy logic systems, with coverage of both Mamdani and the Takagi-Sugeno-Kang approach. A brief explanation of the weight of expert intuition in mathematic reasoning for plant characterisation is also

outlined in this part. Section 3.8 presents genetic algorithm optimization and hybrid genetic programming is detailed. Finally the chapter is summarized in 3.9.

3.2 Data sources

3.2.1 Validation and reasoning for chosen biodiversity sources

Initially, literature was consulted documenting the areas of greatest plant species numbers. There are areas in which more than 3000 species per 10000km² (Diversity Zone (DZ) 8-10) have been recorded (Barthlott et al., 2005). The representative areas were recorded and actual numbers of plant species were sourced from the Global Biodiversity Information Facility (GBIF). DZ8-10 areas are made use of in this study due to it being suggested that these areas have greatest numbers of species by Barthlott and co-workers. High numbers of species' individual occurrences must be used as they show the greatest impact through trophic levels. DZ8-10 areas are most at risk to land use change, inordinate climatic change and imbalances within their ecosystems. The author has taken a similar approach to Myers (2000), prioritising the areas of greatest richness for study in order to protect, conserve and or ensure sustainability in the areas. Effort is made to establish a relative (gradual) difference between these areas in Chapter 7.

The GBIF is an international organisation that focuses on making scientific data on biodiversity freely available and it exists as a database hosted online with multiple interfaces (users and contributors). This source was chosen as it has been statistically validated as being the largest available source of biological presence data, being representative of the kingdoms it holds (Yesson et al., 2007). Key objectives of the GBIF are open sharing of data in order to advance scientific studies and relevant policies for the betterment of conservation and sustainability. Individual occurrences of species are recorded by field workers of multiple institutions and uploaded to the website with precise location details, in some cases accurate to within 2 metres on site. After searching for each DZ8-10 area by name, the files containing the names of all plant species recorded as present were stored and dated. These files can be said to be a representative picture of biodiversity on the date at which the files were stored (Ponder et al, 2001) in terms of individual occurrence data. Acknowledgement and reference is given to each contributing institution in Appendix 1.

Occurrence overview

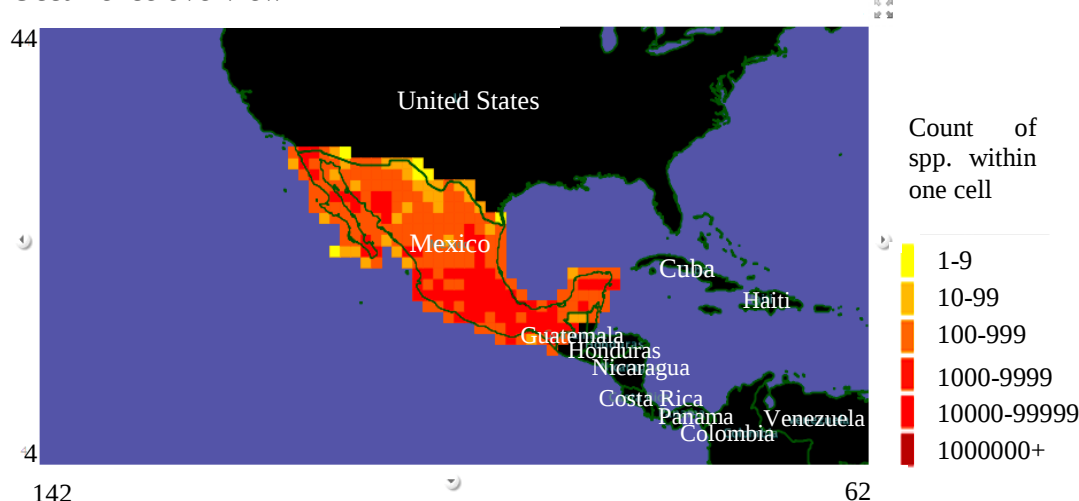


Figure 3.2.1 Global Biodiversity Information Facility Map of Mexico

Fig. 3.2.1 is shown as the first step by which data can be obtained from the GBIF. In **Fig. 3.2.1**, the occurrence overview of the country of Mexico is shown within 142-62 degrees West, 4-44 degrees North. Reference is made to the numbers of all species recorded in the location. After obtaining a map of each area to act as an initial indication of its coordinates, species recorded in the location are queried and subsequently the kingdom plantae is chosen in order to obtain the list of plant species recorded. The lists of individual species occurrences for 7 locations (chosen at random) of DZ8-10 used in this thesis are shown with reference to their collecting institutions in Appendix 1. The areas were Mexico, Guyana, Cuba, Democratic Republic of the Congo, Georgia, Guinea and Macedonia. Data for Azerbaijan was also used, in Chapter 6.

3.2.2 Climatic and topographic data sources

Climatic data used in this thesis are of two categories- water related and energy related. This choice is made following recent work inferring the geographical importance of water and energy in the patterning of species distributions (Hawkins et al., 2003; Sommer et al., 2010). The data are validated by their use-age through the Intergovernmental Panel on Climate Change (IPCC). Data were originally assessed for accuracy and presented by New et al. (1999) and are available in both spreadsheet and graphical format at defined resolution through the IPCC.

The choice of variables to be considered is listed in **Tab. 3.2.2** by category.

Table 3.2.2 Water-Energy Dynamic variables (New et al., 1999)

Variable	Category
Cloud cover (CC)	Water-Energy
Ground frost frequency (GFF)	Water-Energy
Maximum Temperature (MaxTemp)	Energy
Mean Temperature (MeanTemp)	Energy
Minimum Temperature (MinTemp)	Energy
Precipitation (P)	Water
Vapour pressure (VP)	Water
Wet day frequency (WDF)	Water

Data were collected at regional meteorological stations, collated by the Climatic Research Unit of the University of East Anglia, UK and disseminated under different time scale scenarios. Quarterly data between 1961-90 of TS2.1 30 year monthly mean observation climatology are made use of in this study; as use of these data in this context has been validated (Evans et al., 2005), Kreft et al. (2007) have also made use of specific water and energy related variables in the identification of rudiments of plant characteristics and processes. Throughout the process of modelling, detailed in chapters 4-7, the above variables are minimised to enable efficient statements of the water-energy dynamic, due to possible co-variability or redundancy in the variables.

It has been well documented that variation in climatic variables corresponds to variation in altitude, other elements of topography (aspect, slope), and with regard to species distribution (Mitchell and Jones, 2005; Bhattarai and Vetaas, 2003). Measurements of altitude are made use of, using first ‘range data’ from the central intelligence agency (CIA) with unspecified resolution and, subsequently, enhanced resolution digital elevation model (DEM) data from the United States Geological Survey (USGS) in order to illustrate greater accuracy in the frameworks presented. These data are given further mention in Subsection 3.2.3.

3.2.3 High-resolution data for enhanced accuracy of modelling

The work for and conclusions drawn from this thesis are based on the results of modelling frameworks. It has been well documented that low resolution data may be a useful tool for the construction of linguistic models. However, using higher resolution data enables enhanced accuracy of a modelling (algorithmic) framework on which analytical models can be based (Cleveland and Meystel, 1990; Shackelford and Davis, 2003).

As detailed in subsection 3.2.2, ranges of digital elevation data are made use of in early stages of this work (Chapter 4). However, greater accuracy (higher resolution) is required for more analytical conclusions, associated with detailed algorithmic frameworks, which can be used to chart the (uncertain) distribution of groups of plants and their characters (Su et al., 2009; Zadeh, 1973). High resolution DEM data used in this study comprise global topography at 30 second / 1 km resolution (GTOPO30). These data are the result of a collaborative effort, and are under constant development, the requirements for which are summarised by Gesch et al. (1996). GTOPO30 data were accessed from the USGS online pages (<http://www.usgs.gov/> accessed January, 2012).

The resolution of climatic data used in initial work (Chapter 4) is taken from 30-arc minute (50km) grid cells. Later work (Chapters 5-7) uses 10-arc minute (18.5km) grid cells (New et al., 1999). The resolutions of the examples given are quoted here. These values may not be used unilaterally as the distance equating to 1 arc minute varies with latitudinal and longitudinal coordinates due to the shape of the Earth (Haswell, 1920). For enhanced accuracy the coordinates of each location are also given.

3.3 Image processing

Rudimentary water and energy climatic and topographic imagery constitute the initial knowledge base used for the antecedent modelling framework used in this thesis. Images were first extracted from the sources detailed in section 3.2.2, after which they were uploaded into the software platform Matlab (Version R2010a ©). Global areas (countries of DZ8-10) explored in Chapter 4 were obtained, 'alpha' and 'cdata' (Matlab Coded) data files were generated upon uploading. Images may be plotted from these files.

Alternatively, the data for climatic variables may be obtained in spreadsheet format from the IPCC website and downloaded in .csv format. After converting to .xls format the data may be imported into Matlab and saved as a Matlab file. Subsequently the location of the .xls file was used to open the data following Appendix 2.1 entered into the Matlab Command Prompt. The example of using ‘Mean Precipitation January 1961-90.xls’ is shown in Appendix 2.1. The code method was used to display climatic data of which plots were constructed and quantified in order to calculate algorithmic frameworks.

Following code given in Appendix 2.1, and subsequent generation of data files, images were displayed as figures for algorithmic analysis as detailed in 3.5 and 3.7. Alternatively, further information may be extracted from the graphical figures using the ‘red’, ‘green’, ‘blue’ balance within the image. Useful application of this information is to construct algorithms for characters of species distribution patterns as detailed in section 3.6.

DEM data were extracted using Matlab after uploading the compressed files from the source mentioned in section 3.2.3. The 1 km resolution data were outlayed on a global scale, there being 33 tiles which cover the Earth. After identifying which tile and area was of interest using a graphical user interface within Matlab, brief example code was applied in order to display a figure of the chosen region (Appendix 2. 1. 1). This is an efficient method of processing data from multiple contributors of the USGS and has great informative value for topographic systems (Trauth, [ch. 1], 2006).

Additional programming in Matlab modified the axis of the chosen figure, adding more labels onto latitude and longitude, enabling more accurate reading of the coordinates of the chosen locations, for example (Appendix 2.1.2).

3.4 The Species-Area relationship

Initial calculations carried out in this thesis on the numbers of plant species in each of the identified areas is by the species area relationship (Arrhenius, 1921).

$$S = cA^z \quad (3.4.1)$$

Where S refers to the number of species; c is an environmental constant specific to the area; A refers to the area and z is a taxon specific constant. To estimate what the values of c and z are from our data we take logarithms of both sides of the equation:

$$\log S = z \log A + \log c \quad (3.4.2)$$

Equation (3.4.2) is that of a straight line; the constant z is its slope, the constant $\log c$ is its intercept, $\log A$ is the independent variable and $\log S$ is the dependent variable.

A positive result in the above indicates that species do indeed increase with area, whilst a negative result would indicate that species decrease with area. The significance of the resultant weighted least squares regression shown in (3.4.2) is tested using a t-test significance at 5% probability indicating a perfect fit. That is the constants and independent variable explain the number of species present. The species area relationship is applied in the real context of this thesis in Chapter 4.

3.5 Variable partitioning

Variables are described by their probability of occurring. Distributions are defined by continuous and discrete classes. It should be noted that in probability theory the spread of normal (Gaussian) variation is defined by:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (3.5.1)$$

Where μ is the mean of the distribution, σ is the standard deviation σ^2 is the variance of the distribution and e is the base of the natural logarithm (2.71828). If for example, a particular plant's distribution is defined continuously in the interval a, b , the following probability density function integral applies:

$$\Pr[a \leq X \leq b] = \int_a^b f(x) dx \quad (3.5.2)$$

Continuous variables may be plotted over a bell shaped curve or parabola. This category of distribution is made use of in this thesis in patterning of biotic and abiotic variables used

for the modelling framework of this study. The derivative of Gaussian or normal deviation is the central sample of the variation, which is proportional to the standard deviation or greatest incline of the population range of variables under consideration. Gaussian patterns are central to the inception of fuzzy techniques (Zadeh, 1965), which are given further mention in section 3.7 and applied in Chapter 5.

Discrete variables are a further category of distribution, each vector in a discrete class is independent. Discontinuous variables were used to classify defined classes of variation. Commonly used discrete distributions include the Poisson distribution, binomial and Pareto distributions, which are implemented in later chapters as they show stochastic organisation qualities.

The Poisson distribution function is defined in the following form:

$$f(k; \lambda) = \Pr(X = k) = \frac{\lambda^k e^{-\lambda}}{k!} \quad (3.5.3)$$

Where e is the base of the natural logarithm, $k!$ is the factorial of constant k , λ is λT when the number of events observed in a specified time interval is 1.

This thesis concentrates on the spread of distribution, which may be transgressed through generations of the populations considered, hence binomial, Pareto and other n dimensional distributions are most easily discussed with relation to the family of estimation of distribution algorithms (EDAs) known as copulas.

Copula theory separates a joint probability distribution function into the component univariate margins of the distribution and a copula which represents the domain of the random variables. There are copulas that represent all classes of distribution in probability theory, there being Elliptical (Gaussian) copulas and Archimedean (representing classes of discontinuous variation). Essentially, the copula distribution may be used to generate the distribution of a random variable and identify dependencies between the variables used in modelling frameworks via differentiation of the terms contained in their expressions.

Essential notation for definition of copulas (Nelson, 2006): interval $\mathbf{I}$ as $[0,1]$, an n dimensional copula is a function C from $\mathbf{I}$ to $\mathbf{I}^n$ with the following property: for every

$u = (u_1, u_2, \dots, u_n)$ in $\mathbf{I}^n$, $C(u) = 0$ if at least one vector of u is 0, if all vectors of u are 1 except u_k then $C(u) = u_k$.

Gaussian copula is therefore defined:

$$C_r(u_1, u_2, \dots, u_n) = \phi_r(\phi^{-1}(u_1), \phi^{-1}(u_2), \dots, \phi^{-1}(u_n)) \quad (3.5.4)$$

Where ϕ_r is the multivariate Gaussian distribution with correlation r , ϕ^{-1} is the inverse function of the standard one dimensional Gaussian distribution function (3.5.2).

Gumbel copula (Gumbel, 1960) is:

$$\phi_G(u) = (-\log u)^{\phi_G} \quad (3.5.5)$$

Where ϕ_G is the generator of the distribution, related to the Laplace transform which may be used in differential / inferential equations.

Sklar's theorem (1959) states that multivariate distribution functions can be expressed as a copula function evaluated as a 2 dimensional distribution function. Uni-variate distribution functions (e.g. $F_1, \dots, F_n$) can be linked to a multivariate distribution function, H as shown:

$$H(x_1, \dots, x_n) = C(F_1(x_1), \dots, F_n(x_n)) \quad (3.5.6)$$

Elements of the water-energy dynamic, with bi-variate dependency structures (Schölzel and Friedrichs, 2008) are pertinently described using copula theory (Gumbel type, (3.5.4)) as they may show stochastic distribution in a Pareto Type I / poisson domain as may be seen in Chapter 6.

It remains an important point that the most effective manner in which copula distribution may be estimated is by an intuitive nature related to its maximum likelihood. Often a stepwise process may be carried out in copula theory estimation, and involves partitioning of the variables under consideration and derivation of a rule structure (Gaspar et al., 2007; Schölzel and Friedrichs, 2008).

A criticism of conventional / Boolean mathematics is that the methods used often lead to a distorted view of the data. However, operation of logic-based mathematics requires operation of set theory, which processes vectorised data using the central premise of the distribution pattern shown within the data (Zadeh, 1965). Popular spreads of variation include the normal or Gaussian pattern (continuous), dependent on a balance of probability as in (3.5.2), or the use of binomial (discrete) functions, which are essentially based on a factorial nature (e.g. (3.5.3)) with the natural logarithm of the sample.

It is usual to consider sections of Gaussian distribution as the distribution covers the complete range from $0 - \infty$ (covering the maximum range of variables considered). 5 sections are considered here. The variables discussed in subsection 3.2.2 are first allocated notation, being $A_1, \dots, A_n$. Values were converted to percentage value and equally partitioned to enable 5 concise linguistic partitions, after the method shown by Alcalá et al. (2007). The partitions are shown in **Tab. 3.5**.

Table 3.5 Separation of variables into quintile partitions

Percentage Range	Linguistic Category	Notation
0-20	Low	1
20-40	Low-Medium	2
40-60	Medium	3
60-80	Medium-High	4
80-100	High	5

The partitions shown in **Tab. 3.5** were used to quantify the constituent parts of inference A , to create the union which, in combination with variables membership weight, express the consequent B (Zadeh, 1965). The framework for the partitioned variables (3.5.7) and Laplace transform union (3.5.8) is as follows:

$$IF A_1 \dots n_{(n)} - A_1 \dots n_{(n)} THEN B_1 \dots n_{(n, \dots, n)} \quad (3.5.7)$$

Here, the linguistic connectors IF and THEN are used to unify inferential variables (numerically defined from partitioned data) to give the consequent range (numerically defined from the knowledge base in subsection 3.2.1).

$$A = \int_U \mu_A(y) / y \quad (3.5.8)$$

Where $\int_U$ stands for the union of fuzzy singletons, μ is the grade of membership in A (or y). Candidate groups distributed in characterisation of plant species are given in the following subsection.

3.6 Groups to distinguish in plant species characterisation

Plant species are differentiated into the following groups to enable quantitative characterisation on a global scale: 1) life-history strategies (Grime et al., 1995); 2) primary metabolism (photosynthetic) categories (Keeley and Rundel, 2003; Wang et al., 2012) and 3) life-form (Raunkier, 1934).

Life-history strategies

Groups of plant life-history strategies (Grime et al., 1995) are as follows: Competitive species (C) are fast growing, often aggressive species, with rapid nutrient absorption and rapid root and leaf growth. They develop a consolidated growth form with vigorous lateral spread above and below ground, thriving in high nutrient soils. Stress-tolerating species (S) are slow growing, capturing and retaining scarce resources in a continuously hostile environment. Their leaves are long-lived and often heavily defended against predation. Ruderal species (R) have a potentially high growth rate within the seedling phase, and display early onset of the reproductive phase. The allocation of resources to flowers and seeds is suited neither to development of extensive root and shoot systems needed for dominance of habitats, nor to highly stressed environments dependent on conservative patterns of resource use. Species combine the above described strategies (e.g., C-R, S-R, S-C and C-S-R), integrating different growth forms to suit the environment.

In a study of the evolution of model plant populations in computer simulated environments, where nitrogen availability and disturbance frequency alone were used as variables, the evolution of expected plant strategies and patterns proved consistent with described theoretical and field evidence (Grime, 1979; Mustard et al., 2003). The illustration of plant strategies in computer simulated systems supports the existence of patterns on all scales, which may be modelled in real space and time. Barreto (2008)

linearly spaced plant species using simulation techniques to obtain the rK ('Kp') continuum, with seven definite partitions is shown in **Fig. 3.6.1**.

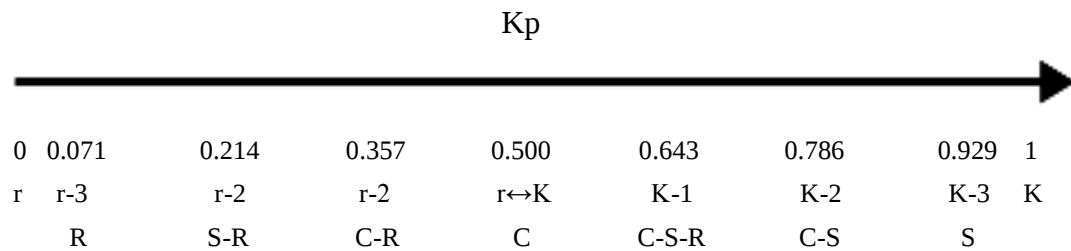


Figure 3.6.1 Separation of the rK and C-S-R theories in 2 dimensions (Baretto, 2008)

Fig. 3.6.1 of Baretto's (2008) 2 dimensional rK and C-S-R reconciliation shows low to high Kp value: plants of ruderal (R) strategy were isolated in places of high disturbance and productivity; stress tolerant-ruderal plants (S-R) were seen in lightly disturbed habitats with low productivity; competitive-ruderal plants (C-R) were present in habitats where disturbance brought moderated competition by a relatively low level of stress; competitive plants (C or r = K) were found in environments with low disturbance and high productivity; competitive-stress tolerant-ruderal plants (C-S-R) were found in environments where there was a moderate intensity of stress and disturbance; competitive-stress tolerating plants (C-S) were found in environments where a moderate intensity of stress and a situation of relative non-disturbance existed; stress tolerating plants (S) were found in environments where there was low productivity and low disturbance.

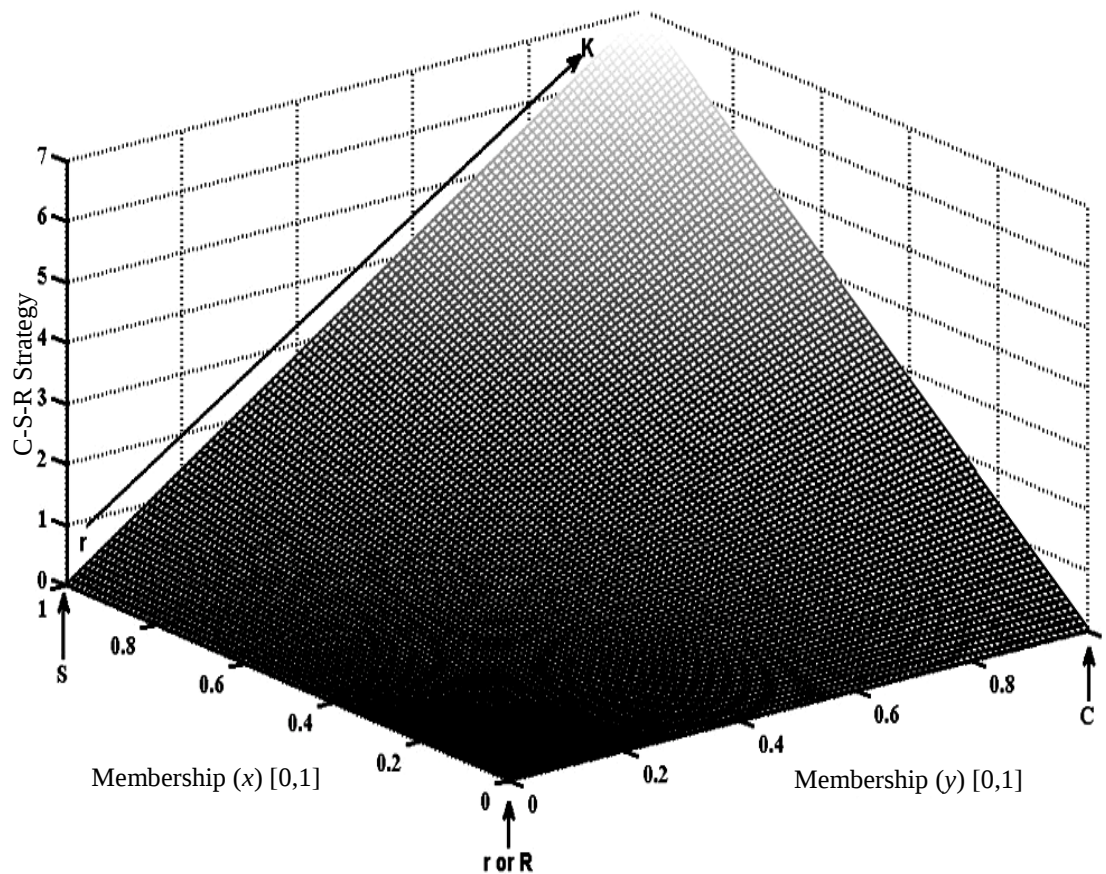


Figure 3.6.2 C-S-R and rK Strategies in 3 dimensions

An algorithm was combined with the technique for order preference and similarity to ideal situation (Barreto, 2008; Hung and Chen, 2009), which effectively reconciled both the rK and C-S-R theories (Grime et al., 1995; Raunkier, 1934). Once again, computer simulated systems were efficiently used to illustrate plant strategy with a more complex suite of input variables (Barreto, 2008; Bornhofen et al., 2011).

Ordination of strategies is applied in Chapter 5. This was carried out to determine the result of modelling of strategy characters in accordance with the water-energy dynamic. Numerical dispersal of strategies is applied in Chapter 6.

Primary metabolism

Photosynthesis is the primary metabolic process by which plants grow. There are three main groups of photosynthesis in plants, C₃, C₄ and Crassulacean Acid Metabolism (CAM). C₃ plants have 3 carbon compounds in the first step of photosynthesis, C₄ have 4

carbon compounds in the first step of the process and CAM plants store carbon in the form of an acid before photosynthesis.

C3 plants' stomata are open during the day, allowing gaseous and water exchange via photorespiration (breakdown of sugars formed in photosynthesis, releasing CO₂ and H₂O). Ribulose bis-Carboxylase Oxygenase (RUBISCO) is the enzyme involved in uptake of CO₂ in C3 photosynthesis. Photosynthesis takes place throughout the leaves of the plants. C3 plants represent the largest group of plant species, the process being highly efficient under cool and moist conditions (Niu et al., 2005).

C4 plants' stomata are open during the day. Phosphoenolpyruvate (PEP) carboxylase is the enzyme involved in uptake of CO₂ with RUBISCO processing CO₂ in photosynthesis. Photosynthesis takes place in specialized Kranz cells, compartmentalized inner layers of the leaf. C4 plants photosynthesize faster than C3 plants under circumstances of high energy (e.g. light, temperature) and have much lower rates of photorespiration as the enzyme RUBISCO is more saturated with CO₂ for photosynthesis due to PEP activity. CO₂ uptake is more efficient in C4 plants, coupled with highly efficient water use in photosynthesis due to spatial separation. Stomata close, the latter enables less loss of water from the plants and hence greater efficiency under warm and drier conditions. They are mainly summer annual species, occurring in over 19 plant families (Keeley and Rundel, 2003; Salisbury and Ross, 1992; Wang et al., 2012).

CAM plants keep their stomata closed during the day, and during both day and night in periods when water must be conserved (known as CAM-idle). During CAM-idling photosynthesis and photorespiration couple, the oxygen given off in photosynthesis is used during respiration and CO₂ given off in respiration is used in photosynthesis. CAM-idling leads to a build up of toxic compounds over very dry periods. When moisture is available, the stomata reopen and CAM occurs as before. Opening of stomata at night enables more efficient use of water as temperatures and wind speeds are lower than during the day. CAM is an adaptation to very hot, dry conditions. Most cacti and succulent plants use this metabolism. It is also found in orchids and epiphytic bromeliads (Lüttge, 2003). Intermediates occur between C3-CAM and C4-CAM. C3 metabolism evolved primarily, with C4 photosynthesis serving as an adaptation to warmer temperatures and CAM enabling plants to cover arid zones and extreme environments.

Algorithmic dispersal of the three photosynthetic metabolism types is applied in chapter 6, from which spatial separation of the metabolic types is inferred between and within individual environment types.

Life-forms

In this thesis life-forms (Raunkier, 1934) are divided as follows:

i) Phanerophytes are plants with ‘growing buds’ on above ground parts of the plant and are of three types: a) evergreen phanerophytes with bud scales; b) evergreen phanerophytes without bud scales; c) deciduous phanerophytes with bud scales. Phanerophytes are further divided according to height: Mega- (>30m); Meso- (8-30m); Micro- (2-8m); Nano- (<2m).

ii) Chamaephytes are plants with growing buds above ground, always below 2m. There are woody and herbaceous types. Chamaephytes are broken into: a) Suffrutescent- (after the main growth period upper shoots die, only lower parts of the plant remain in ‘unfavourable’ period); b) Passive- (in unfavourable conditions upper shoots become procumbent, protecting from environmental stresses); c) Active- (shoots only produced along the ground and remain so); d) Cushion- (similar to passive type but shoots are so closely packed together they form a ‘cushion’).

iii) Hemicryptophytes are plants of which all above ground parts die back in unfavourable conditions and buds are borne at ground level. Hemicryptophytes are further divided into: a) Proto- (leaves are well developed up the stem of the plant, partially developed leaves protect growing buds); b) Partial-rosette- (developed leaves form a rosette at the base of the plant, the following year a long aerial shoot may grow); Rosette (leaves restricted to a basal rosette, long exclusively flower bearing aerial shoot forms).

iv) Cryptophytes are plants with buds or shoot tips surviving below ground or water. Cryptophytes are divided into: a) Geophyte- (underground organs such as bulbs, rhizomes, tubers, shoots emerge in growing season); b) Helophyte (growing buds are in soil or mud under water producing shoots above water); c) Hydrophyte- (buds lie under water, unfavourable period spent completely below water).

v) Therophytes are annual plants, which survive the unfavourable period as seeds, completing their life cycle in the summer months.

The continual dispersion of the five life form classifications is applied in functional numeric terms in Chapter 7.

The following section provides background to fuzzy logic control and details the methods of forming instructive Fuzzy Logic Rule Base System (FRBS) for ordination of plant characteristics.

3.7 Fuzzy Logic

Fuzzy logic is underpinned by set theory and, as such, is used to mathematically describe the quantification of complex design problems and form control strategies on which one bases a rule structure. The ordination of individuals (x) within sets may be stated as generic elements (X) on which we may place constitutive elements according to their membership value (μ) in terms of an interval between 0-1. The basic assumption is that independent variables share Gaussian patterns of distribution, a premise which allows a rule base of component variables to be formed enabling individual vectors to be given a membership value of the set. In essence, therefore, there are two main parts to the resultant instructive algorithms that may be developed: a constitutive inferential statement and a consequential statement. The goal of fuzzy logic is to provide a founding for approximate reasoning whilst using imprecise arguments, the training of which may establish a higher degree of certainty in uncertain systems. In this it is an ideally placed technique by which plant species are ordinated under extremely variable dynamics. There are two basic types of fuzzy logic control (FLC); Mamdani and Takagi Sugeno Kang (T-S-K) systems. The basic difference is that Mamdani FLC may be defined with fuzzy inference and fuzzy consequence, whereas T-S-K FLC uses defined statements of consequence. Mamdani and T-S-K fuzzy systems are detailed below.

3.7.1 Mamdani systems

Mamdani systems (Mamdani, 1974) are the most common type of fuzzy logic systems. They consist of four basic parts as shown in **Fig. 3.7.1**. In this initial form a static mapping function traces inputs to outputs.

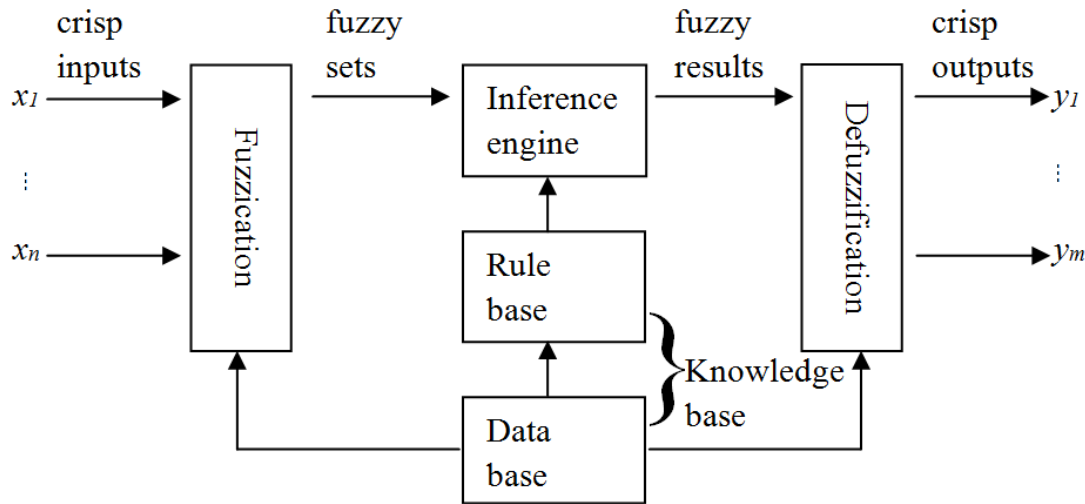


Figure 3.7.1 Fuzzy logic systems

Fuzzy system inputs are $x_i \in X_i$ (where $i = 1, 2, \dots, n$) and outputs are $y_j \in Y_j$ (where $j = 1, 2, \dots, m$). Inputs and outputs are both crisp, real numbers as provided and validated. The process of fuzzification converts inputs into membership values, establishing the fuzzy sets. The rule base is constructed and operates on the membership values after which the defuzzification process results in crisp outputs from the system. As indicated in **Fig. 3.7.1** fuzzy values may flow back through the system in a loop, thereby updating the database and providing more accurate allocation of the membership values for the system.

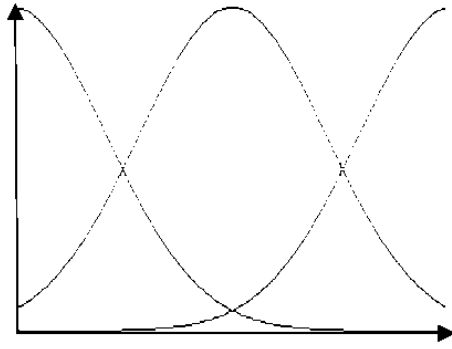
Definitions for fuzzy systems are provided by Zadeh (1965) and are summarised below.

D 3.7.1. Universe of discourse: All information relevant to inferential or consequential variables, e.g. data sets X_i and Y_j are the “universes of discourse” of elemental sets x_i and y_j respectively.

D 3.7.2. Linguistic variables: A variable expressed as a word or sentence, used in description of fuzzy inputs and outputs.

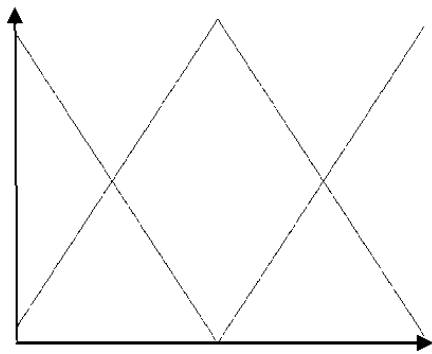
D 3.7.3. Linguistic rules: A simplification of an algorithmic statement which ordinales inputs of a fuzzy system to outputs of the same. These rules make use of conditional statements formed with linguistic connectors, e.g. ‘IF’ *inference*, ‘AND’,... ‘THEN’ *consequence*.

D 3.7.4. Membership function: expresses the degree of truth or value of a variable within the interval 0,1. The type of membership function employed within a fuzzy system is determined by the spread of distribution of the characters being considered within the system. Most commonly used membership functions are Gaussian, Triangular and Trapezoidal types. The spreads of distribution and key equations for these are given below.



$$f(x) = \exp\left(\frac{0.5(x-c)}{\sigma^2}\right) \quad (3.7.1.1)$$

Figure 3.7.1.2 a) Gaussian Membership Function



$$f(x; a, b, c) = \max\left(\min\left(\frac{x-a}{b-a}, \frac{c-x}{c-b}\right), 0\right) \quad (3.7.1.2)$$

Figure 3.7.1.2b) Triangular Membership Function

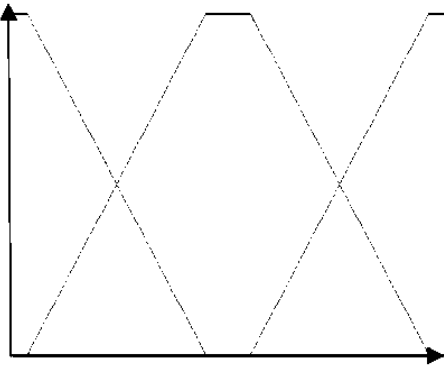


Figure 3.7.1.2c) Trapezoidal Membership Function

Figure 3.7.1.2 Generic membership function structures

In (3.7.1.1) c is the mean and σ^2 the variance; b) Parameters a , c (feet) and b (peaks) are defined within (3.7.1.2). Gaussian and Triangular functions of **Fig. 3.7.1.2** expressed in these equations are accepted alternate views of considering continual variation.

The structure of the trapezoidal membership function is shown here as an example although the Gaussian and Triangular functions are used in later stages of this thesis.

Membership functions are non-negative and differ from probability density functions in that the area under the membership function curve may be less than unity, as more than one membership function may be used in any one system, depending on the number of rules which require differentiation in the rule base. Gaussian functions are often used in biological and or continuously variable systems (Broekhoven et al., 2007).

D 3.7.5 Fuzzy sets: the integers of a universe of discourse expressed through their associated membership values. Set values are denoted as x_i and y_j . However in later parts of this thesis linguistic terms $A_{(1,...,n)}$ and $B_{(1,...,n)}$ are used to summarise x and y variables where more than one variable is used within the antecedent and consequential statements respectively. Fuzzy singletons are class variables, elements of fuzzy sets.

3.7.2 Takagi-Sugeno systems

T-S-K fuzzy systems (Takagi and Sugeno, 1985) are more easily applied to multiple input and multiple (ranged) output. The general form of the i th rule as applied to T-S-K systems is as follows:

$$\text{If } x_1 \text{ is } A_1 \wedge x_2 \text{ is } A_2 \wedge \dots \wedge x_n \text{ is } A_n \text{ Then } y_i = f_i(x_1, \dots, x_n) \quad (3.7.2.1)$$

Here we see a constructive breakdown of the antecedent term A which may be dispersed through multiple sets of linguistic variables such as those present in climatic systems. These are detailed as:

$$X = \{x_1, \dots, x_n\} \quad (3.7.2.2)$$

Where $x_1, \dots, x_n$ are set values of generic set X . In this thesis, the following definitions using multiple elements in T-S-K systems are used:

$$\bar{x} = [x_1, x_2, \dots, x_n] \quad (3.7.2.3)$$

$$\bar{y} = [y_1, y_2, \dots, y_n] \quad (3.7.2.4)$$

$$f_A(x; y) = X_\mu; Y_\mu \quad (3.7.2.5)$$

In (3.7.2.3) the mean of set x is represented by normal set elements $x_1, \dots, x_n$. Accordingly, in (3.7.2.4) the mean of set y is represented by normal set elements $y_1, \dots, y_n$. In (3.7.2.5) f_A refers to function A in x and X_μ is the grade of membership of X and may also be used to express function A in y and Y_μ is the grade of membership of Y . Both X and Y are used in the antecedent linguistic A matrix in order to form:

$$\bar{y} = [\bar{x} * R] \quad (3.7.2.6)$$

Where R is the relational index matrix as a function of the combined arguments of $\bar{x}$ and $\bar{y}$. In this it may also be stated to be the result of the Laplace transform union as in (3.5.8), or succinctly defined as ‘*IF A... THEN B*’ (Zadeh, 1965).

T-S-K fuzzy logic allows greater accuracy in modelling as rule consequents are functions of crisp inputs whereas Mamdani fuzzy logic use fuzzy inputs as well as consequents. For this reason, accuracy is a key strength of T-S-K systems as opposed to interpretability for Mamdani systems.

Fuzzy logic has been applied in quantification, classification and to affect control of many artificial and man made systems. It has also found application in ecological systems (Cui et al., 2012; Su et al., 2009; Taheriyoun et al., 2010). However this thesis documents the novel use of fuzzy logic and consequent systems in order to characterise over 300 000 plants (primary species). The original application of these techniques, which was useful in description of non-linear systems (Herrera, 2005), is of principle use in precise analytical statements of ecological relations and parameters.

Adaptive neural fuzzy inference systems (ANFIS) are developed which effectively map inputs to outputs as in the system diagram of **Fig. 3.7.1**. These are discussed in greater detail where they are implemented in Chapter 5.

Use of fuzzy logic is what underpins the development of instructive algorithms, which are used in combination with expert intuition to form geographic information systems (GIS) shown in Chapter 7, where research questions of plant distribution are also answered. The following subsection details expert intuition.

3.7.3 Expert intuition

Expert intuition may be used on fuzzy classification in order to make value judgements of the sets deployed in both antecedent and consequent expressions. These value judgements have two basic components.

Firstly, in choosing variables to use in modelling, variables should be from those that have a broad inference on the patterns of dispersion of plant species. These choices are made from the knowledge base built up by the researcher and that which is dispersed across literature.

Secondly, the variables to be considered in the modelling framework may be minimised to maximise efficiency of the algorithms deployed. Minimisation is carried out given that the general trend of the variable does not change the distribution of the species as predicted by the control structures given in the original algorithmic statement (i.e. the pattern of distribution type is not altered from one form to another. Error, therefore, is less than 1 for the variable to be minimised). The end result of the use of intuition is ultimately a product of the mathematic inference shown by the variables in question (Nasibov and Peker, 2011).

Certainty in the algorithmic statement is strengthened, although as researchers we must be aware of the expanding relations of the research work we carry out and at all times consider a balanced (objective) view of our expert intuition (Furze et al., 2013a; Nasibov and Peker, 2011; Zhao, 2012). In order to consider the most accurate, efficient and interpretable system of balancing variables, optimisation methods are carried out. Some of these are detailed in the following section and expanded in Chapters 6 and 7.

3.8 Genetic algorithm optimisation of objective variables

The choice of the objectives should be made by knowledge guided expert intuition (subsection 3.7.3), followed by their successive minimisation in order to make the most efficient choice for optimisation. Optimisation of objective variables is the process of improving the variables so that they are optimally expressed. In this thesis characteristics of plant species (plant strategies, life-forms and photosynthetic metabolism) are dispersed

optimally through different conditions and scenarios via mathematic quantification based on T-S-K fuzzy logic. These are described in Chapters 6 and 7.

There are many techniques by which optimisation may be carried out (e.g. Particle Swarm Optimisation, Simulated Annealing and Genetic Algorithm) and subsequently many categories within these techniques. This thesis focuses on the use of optimisation within constrained parameters. The variables expressed as objectives in climatic systems may indeed be described as being stochastic, that is they operate within functional domains, hence the chosen area of technique is by genetic algorithm (GA).

GAs are algorithms based on natural genetics, providing robust search capabilities in complex (objective) space. The design of a genetic algorithm is such that elements of the character being optimised are represented by a string of chromosomes. After random selection of the chromosomes, they then run through a series of iterations of evaluation, selection and recombination, followed by re-evaluation. Given that the best solution to the specified objective parameters has been found, the best global solution in the chromosome population is found. The algorithm continues with other chromosomes until all the best solutions are found. The process of these algorithms is outlined in **Fig. 3.8**.

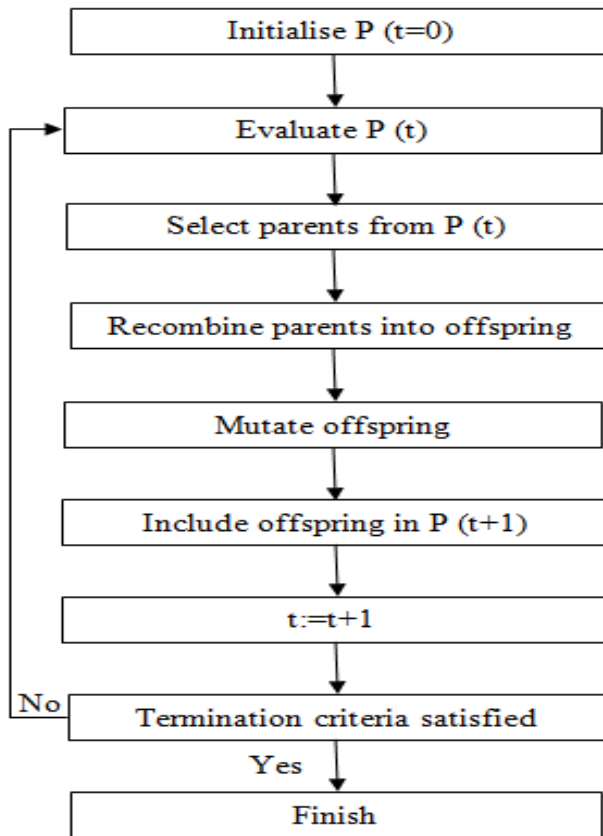


Figure 3.8 Genetic algorithm basic flow (Cordon et al., 2004)

In **Fig. 3.8**, P represents the population of chromosomes, t represents the population stage (0 being initialisation), parameters are chosen against which to evaluate the (strings) chromosomes of the population after which the best individuals are selected via a fitness function (set) and used as parents to produce offspring. The offspring undergo mutation and are included in the next iterative generation of the population. Should the termination criteria (eg. number of generations; time limit; change greater than the weighted average change in vectors of chromosomes; no change being achieved within a defined time limit) be met, then the chromosomes of the population disperse over objective space to form ‘non-inferior’ or Pareto optimal solutions.

The search capabilities of genetic algorithms enables their use in the tuning of membership functions, or indeed in representation of them, which have both been explored in Pittsburgh (Smith, 1980) and Michigan (Holland and Reitman, 1978) approaches and more recently fuzzy methods have been combined with a basic genetic algorithm structure to produce efficient sets of fuzzy rules (Alcala et al., 2007). The combination of T-S-K systems with genetic algorithms is an intuitive step in categorisation and control studies

and one that has found application in studies of continual and discrete variables (Alcala et al., 2007).

Genetic algorithms are described here in order to clarify the technique expanded upon in subsection 3.8.1. As has been described in subsection 3.2.2 there is more than one objective in question in characterisation of plant species. A novel modified (fuzzy hybrid) multi-objective genetic algorithm method is described in the following subsection.

3.8.1 Hybrid systems to explore utopian relations of plant characters

T-S-K fuzzy systems are accurate systems by which we may derive fuzzy rule bases suited for the analytical power of genetic computational methods. The variables to be taken into consideration include more than one objective set in the case of the water-energy dynamic (Hawkins et al., 2003). As such, the vector of objectives is defined as follows:

$$F(x)=[F_1(x),F_2(x),...F_m(x)] \quad (3.8.1.1)$$

Here, the multi-objective optimisation function of x (in the case of elements of water and energy) is equal to the functions of x (set value(s)) and y (set value(s)), with the assumption made that the most efficient choice of objectives has been carried out.

The process is simply expressed in the following order: ANFIS of objective variables to form efficient sets of rules for construction of the fuzzy rule base leading to accurate identification of consequent environments for strategy, life-form or photosynthetic identification (expanded in Chapters 5, 6 and 7); dispersal of elements of characters to be optimised via intuitive use of genetic computation (using a random selection process expanded in Chapter 6). This methodology is carried out after agreement of the distribution type of variables to be considered (Schölzel and Friedrichs, 2008).

Given that the genetic computation method identifies (3.8.1.1) via the dispersal of objective elements in a strength Pareto across the defined objective space, it may be stated that:

$$Z = \{x \in R^n\} \quad (3.8.1.2)$$

Where Z is equal to the objective (utopian) space over which x is an element of R^n , in agreement with (3.7.2.6).

Plotting a least squares regression through the dispersed elements may produce a utopia line and hence calculation of the objective values at any point within the objective space may be made. Furthermore, complex (polynomial) estimations of the precise distribution pattern of dispersed elements enable further methods to be developed for the estimation of the presence of the dispersed elements in similar dynamic objective conditions. In this regard the modified objective genetic algorithm mentioned here and expanded on in Chapter 6 enables a precise stochastic structure, which may be extrapolated onto a geographic information system in order to answer research questions posed in Chapter 7.

3.9 Summary

In this chapter essential knowledge for sourcing of data frameworks within global plant characterisation is shown and preliminary methodologies are given to provide foundation for the work developed in the rest of the thesis. Each section does not provide a complete reference of the methods used in the thesis, methods are applied throughout and given further discussion. The main contributions of this chapter to the thesis are as follows:

- 1) Details of validated biodiversity sources and climatic modelling frameworks, along with reasoning and a brief discussion on the value of high-resolution data.
- 2) The basic outline of image processing is detailed, with examples of code, which may be employed for high resolution mapping. This may be is used later in construction of geographic information systems.
- 3) The species-area relationship is detailed, with background of the calculation. This section provides justification for further methodologies to be developed. Further detail is shown in Chapter 4.

4) Variable partitioning carried out in the thesis is outlined with mention given to distribution types and copular distribution, which provides justification for the use of stochastic methods shown in later sections, elaborated on in Chapters 4, 5, 6 and 7.

5) The groups that are characterised for all plant species are covered, being life-history strategies, life-forms and photosynthetic metabolism. An example is given of how life history strategies may be extrapolated to 2 and 3 dimensions. Strategy based environments are covered in global locations in Chapter 5.

6) Fuzzy logic systems are explained with reference to both Mamdani and T-S-K systems, which may be used for interpretable and accurate statements respectively. Some definitions were also provided. The algorithmic framework applied in Chapter 5 of the thesis was also described and a brief section on intuition included.

7) Genetic algorithms of objective variables are given simplistic description and hybrid genetic-fuzzy systems are detailed with the same principles. These are expanded on in chapter 6 (strategies and metabolism) and Chapter 7 (life-forms).

CHAPTER 4

The Species-Area Relationship in context

4.1 Introduction

As discussed in section 3.4 the species-area relationship (SAR) is the classical methodology by which one may quantify the numbers of species distributed in variable geographic locations. This is based on a non-specified relationship between the environment of a location, the area of the locations and the rate of increase of species presence. Though the SAR was formalized around 200 years ago, it is still used in essentially the same form.

The aims of this chapter are to construct a global plant species area relationship and to show how its use is becoming redundant on larger scales for estimation of species numbers due to measured species numbers being recorded in non standardized units. Many papers on the SAR do not control for sampling error, using data from tree species in Ghana illustrated that when sampling error was standardized a weak relationship emerges (Hill et al., 1994). Furthermore, this chapter introduces the use of algorithmic techniques in order

to demonstrate a more accurate information based plant species relation, allowing the examination of trends such as plant strategies, with genetically grouped origins at macro levels. As a result of the above, the author will identify the relevance and development of engineering techniques in biogeography, computer science and related fields.

The remainder of this chapter is organized as follows: in section 4.2 the basic methodology for the SAR is shown, detail of the numeric basis of plant species occurrences is illustrated and the statistical method utilized for analysis is elaborated. Further, the results of species presence against locations of increasing area are shown along with linear SAR and quadratic plots. In section 4.3 a fuzzy logic-based (FLB) approach is proposed and examples of climatic data are shown at low resolution. The framework of the FLB approach is shown with respect to 7 life-history strategy environments. In Section 4.4 the algorithm is validated and applied to field data of specific locations. Finally, the chapter is summarized in section 4.5.

4.2 Species-Area Relationship implementation

Calculation of species numbers were made by Barthlott et al. (2005). Twenty diversity zones (DZ) were described by these authors and standardized to the number of species/10000km². DZ 8-10 contained more than 3000 species/10000km² and these areas are investigated in the current study. Species recorded in terms of individual occurrences were sourced from the Global Biodiversity Information Facility (<http://www.gbif.org/> , accessed: December 2010, as validated by Yesson et al. (2007) in each of the DZ 8-10 locations (Barthlott *et al.*, 2005). Individual species occurrences against locations in latitudinal order were plotted using a histogram form. Species-area relations were indicated by plotting species presence against area, following the classic (non-standardized) form of the species area relationship of (3.4.1).

Least squares regression is applied (3.4.2) where exponent z is the gradient of the line (slope m) and the intercept of the line is the logarithm of c . Species-area relations were plotted and are shown below. The plant species occurrence numbers are listed in Appendix 1. **Fig. 4.2.1** shows the number of recorded species presences at the selected locations of DZ 8-10 in longitudinal order.

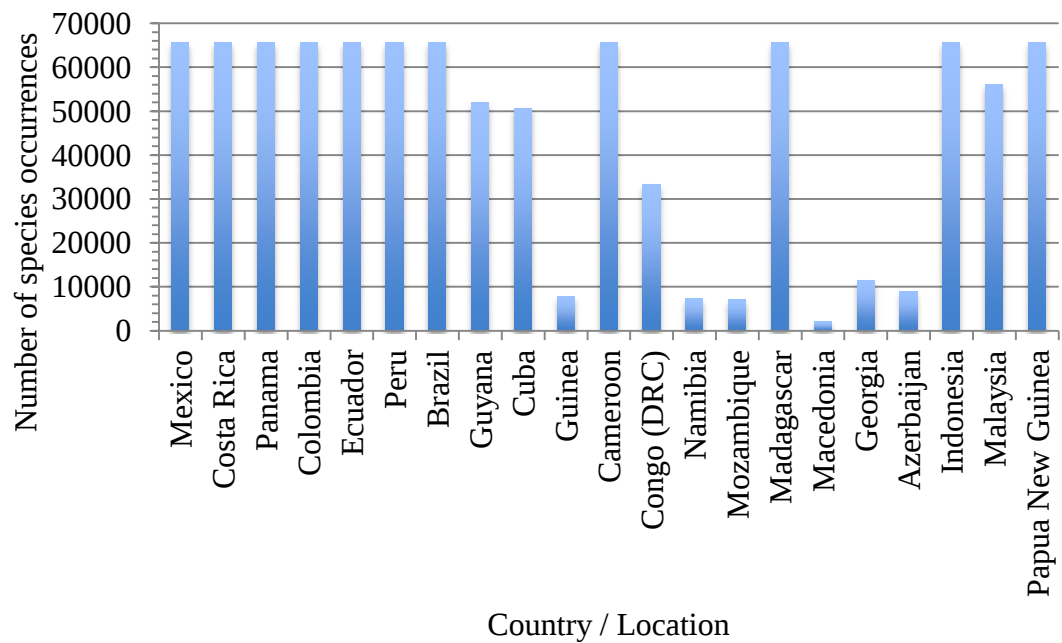


Figure 4.2.1. Species presence versus location for the diversity zones 8-10 of Barthlott et al. (2005)

Fig. 4.2.2 shows the number of species occurrences plotted against the area of each location in km².

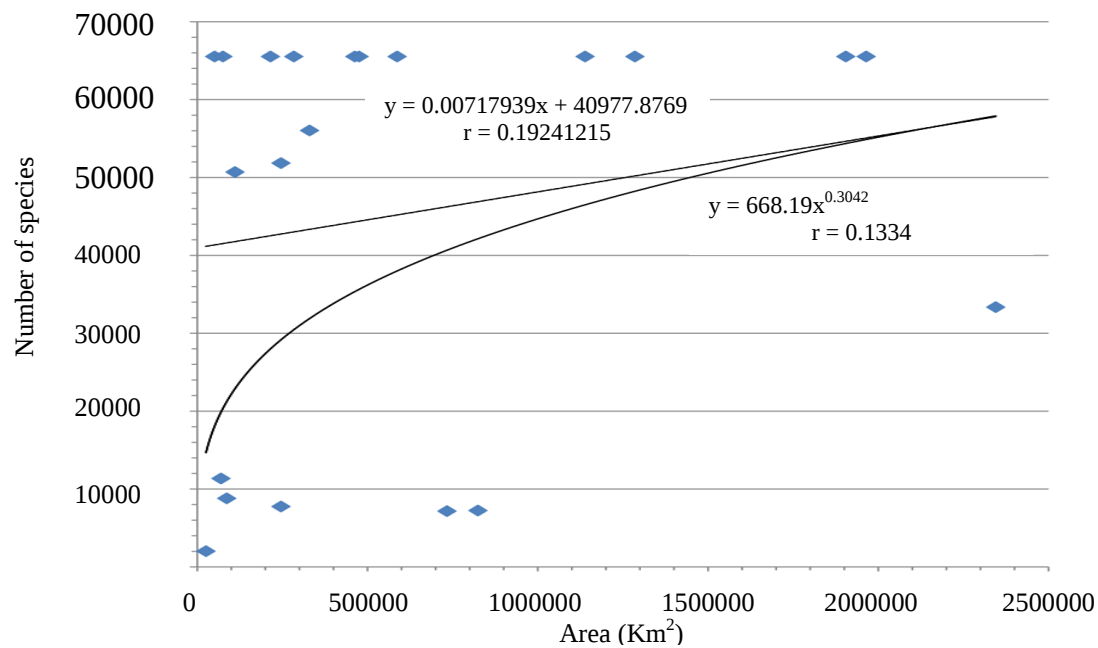


Figure 4.2.2 Number of individual occurrences versus area

We form the null hypothesis that there is no relationship of species with area.

In accordance with equation (3.4.2) the gradient of the straight line obtained is:

$$(m(z)) = 0.00717939 \quad (4.2.1)$$

Testing the significance of the relationship of species with area is simple. We derive the following 2 tailed t-test:

$$t = \frac{r\sqrt{N-2}}{1-r^2} \quad (4.2.2)$$

where r is the regression correlation, N is the number and 2 is the degrees of freedom.

$$t = \frac{0.19241215\sqrt{21-2}}{1-0.037022435467623} = 0.870949800130199 \quad (4.2.3)$$

The gradient of the linear curve shows a positive relationship between area and species numbers, but the correlation of 0.19 is insignificant at $p=0.05$.

In order to write the SAR we identify:

$$b = \log c = 40977.8769 \quad (4.2.4)$$

where b is the the intercept of the line.

The calculation for the species area relationship can therefore be written as

$$S = \log 40977.8769 A^{0.0071793} \quad (4.2.5)$$

The conclusion is made that the variation in species numbers cannot be explained by the increase in area, even when incorporating estimations of changes in environmental conditions (interpreted logarithmically from the intercept) and the rate of increase due to the species present (interpreted from the gradient of the regression line).

The author appreciates that such a conclusion compared to much published literature is rather unusual. However, the points on the graph each represent multiple users and data built up, in the highest cases, by in excess of 65000 individuals. Although the results indicated that a species-area relationship could be used as an indicator of ecological processes they do not specify the ecological processes themselves in an informative model. Various authors have made an effort to explain the nature of ecological processes in different terms, such as MacArthur and Wilson (1967) and Simberloff (1974), who invoked island biogeography theory (summarizing that species extinction rate decreases due to large areas supporting greater population totals which are less susceptible to random extinction). The habitat diversity hypothesis, covered in relation to equatorial areas of high diversity by Zimmerman and Biergaard (1986), illustrated how larger areas have greater diversity of habitats, each with their own sets of species (Jetz et al., 2009; Kier et al., 2009). Succession development was described by Houle (1990) and showed how larger islands possess more stages of community succession development, inferring that the SAR is shaped by competitive interaction of life-cycle history strategies of species within stages of succession.

Progressing from these early theories, Kreft and Jetz (2007) clearly show that there are multiple factors involved in global patterns of plant diversity and Kraft and Ackerly (2008)

explain the importance of developing perspectives of functional traits related to species niches in community assemblies in complex ecosystems. Although previous work enables us to hypothesise about ecological processes and community assemblies, criticism of these studies is that they make use of isolated cases, distorted statistical views, or largely unknown ‘black-box’ inferences describing the processes they cover. Lomolino (2001) summarised biogeographical factors that must be taken into consideration for the SAR to provide meaningful value for ecologists and biogeographers alike. He detailed that the relationship often requires multifactorial causal explanations as a range of processes are involved, which gives further justification for the research of this thesis.

Weaknesses may be dealt with by differentiation of the species-area relationship into different categories of modelling elements in a more information-rich modelling of plant species. This novel application is undertaken in this thesis with the context of all plant species on the planet using the water-energy hypothesis (Hawkins et al., 2003) as justification for differentiation of plants into groups distributed in a non-linear fashion. This is further discussed in the following section.

4.3 Information-rich modelling

In this subsection elaboration of the formation of a fuzzy-logic model to cater for variables of the water-energy dynamic is shown, which must be taken into consideration in order to predict the non-linear distribution of plant species (Hawkins et al., 2003). **Fig. 4.3.1** lays out the basic principle of the fuzzy logic model.

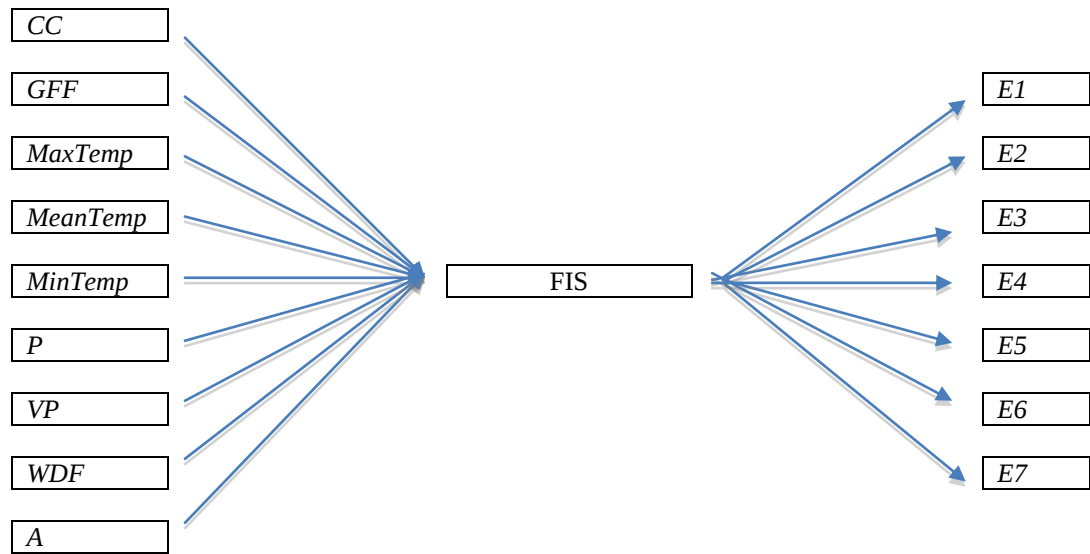


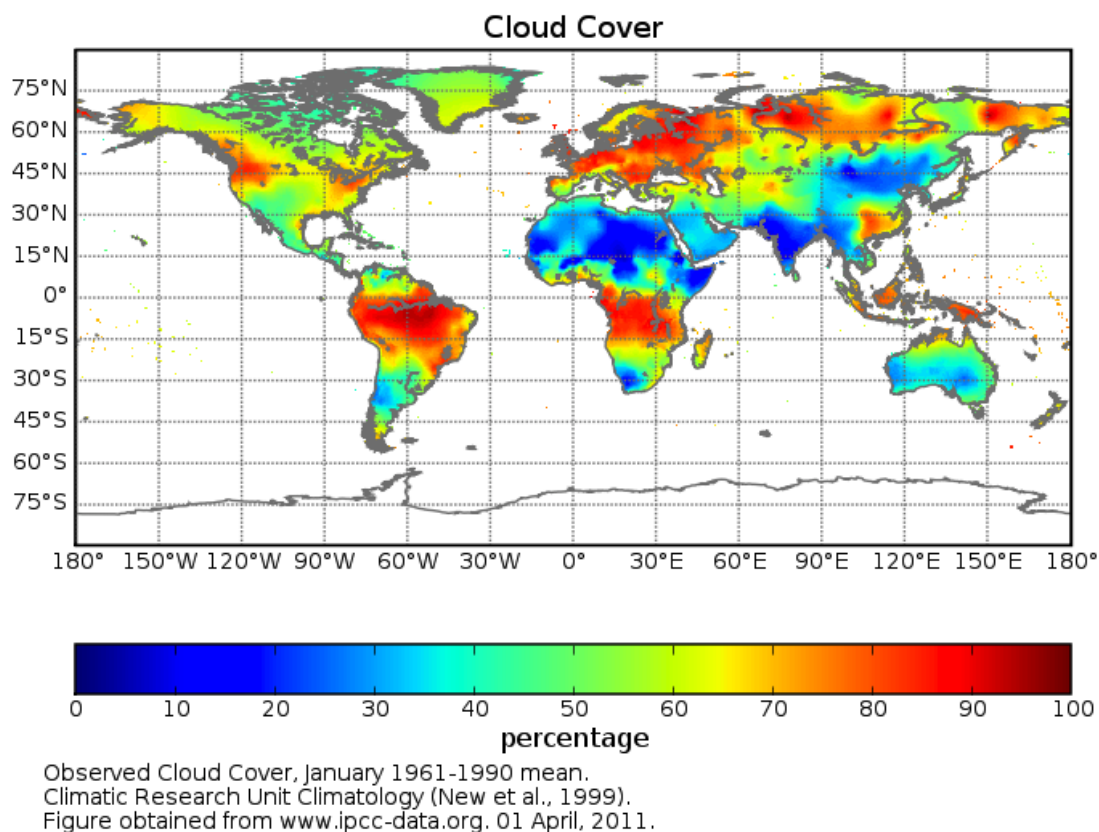
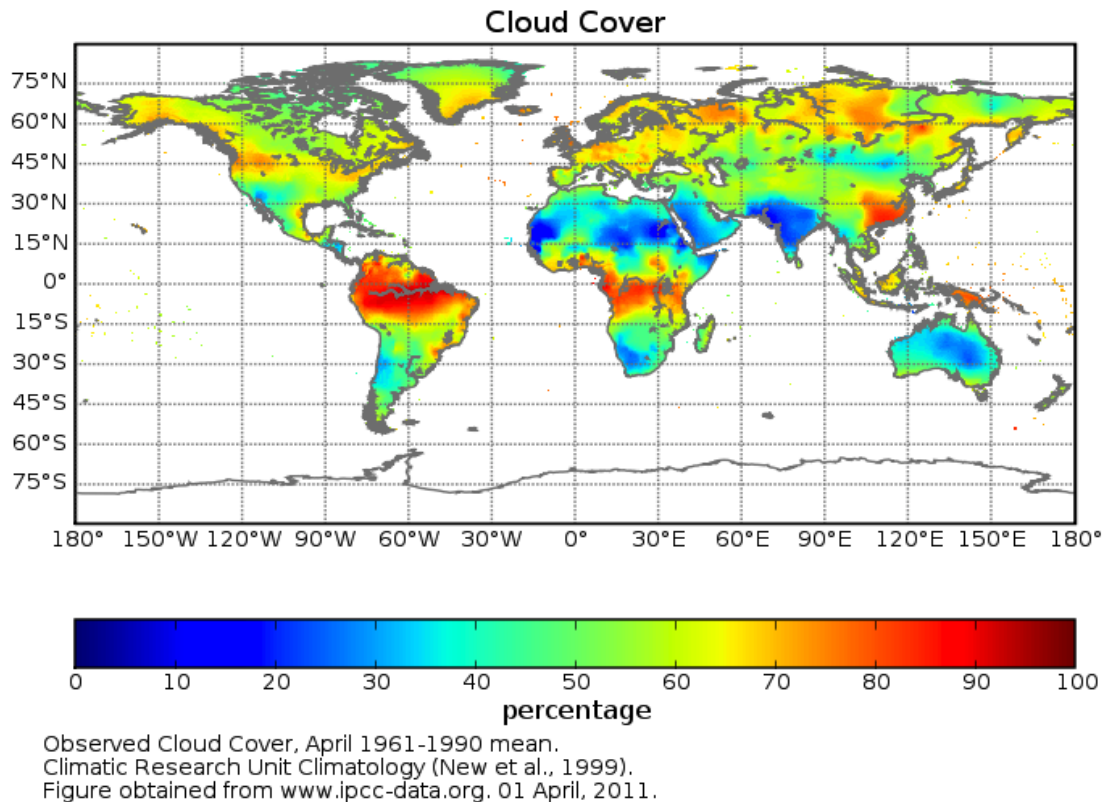
Figure 4.3.1 Fuzzy logic based classification of environments containing plant species

In **Fig. 4.3.1**, *CC* is cloud cover, *GFF* is ground frost frequency, *MaxTemp* is maximum temperature, *MinTemp* is minimum temperature, *P* is precipitation, *VP* is vapour pressure, *WDF* is wet day frequency, *A* is Altitude, FIS is Fuzzy Inference System, *E1*, ..., *E7* represent environments 1 to 7, analogous with the 7 plant strategies detailed in Chapter 3, subsection 3.6, pp. 44-45 and Chapter 2, subsection 2.3.1, pp. 15-19.

The basic fuzzy logic system of **Fig. 4.3.1** is a solution as to how we may model the occurrence of the 7 plant strategy environments. This model was formed using low-resolution variables, (exemplified in **Fig. 4.3.2**) on a global scale at 10-minute (approximately 20km depending on location) spatial resolution. It is postulated that we may form an approximation of environment type from the above 8 climatic variables (water and / or energy factors) and discretely ranged altitude data. However, the above model does not easily enable us to accurately specify the ranges, which may provide a unified statement or algorithmic base.

The basic model is described in terms of linguistic fuzzy predicates and gives a vague description of the conditions which best suit a strategy of plant growth. The proposed model effectively combines knowledge base and expert knowledge. This may be further refined using additional techniques. Development of the current model enables us to make use of the Water-Energy based patterning of variation (Hawkins et al 2003), used here to determine plant strategies (Grime et al., 1995). For this purpose rules are generated in the

FIS: for example IF P is high AND MeanTemp is high AND CC is medium AND A is Low to Medium THEN E1.

**Figure 4.3.2 a)****Figure 4.3.2 b)**

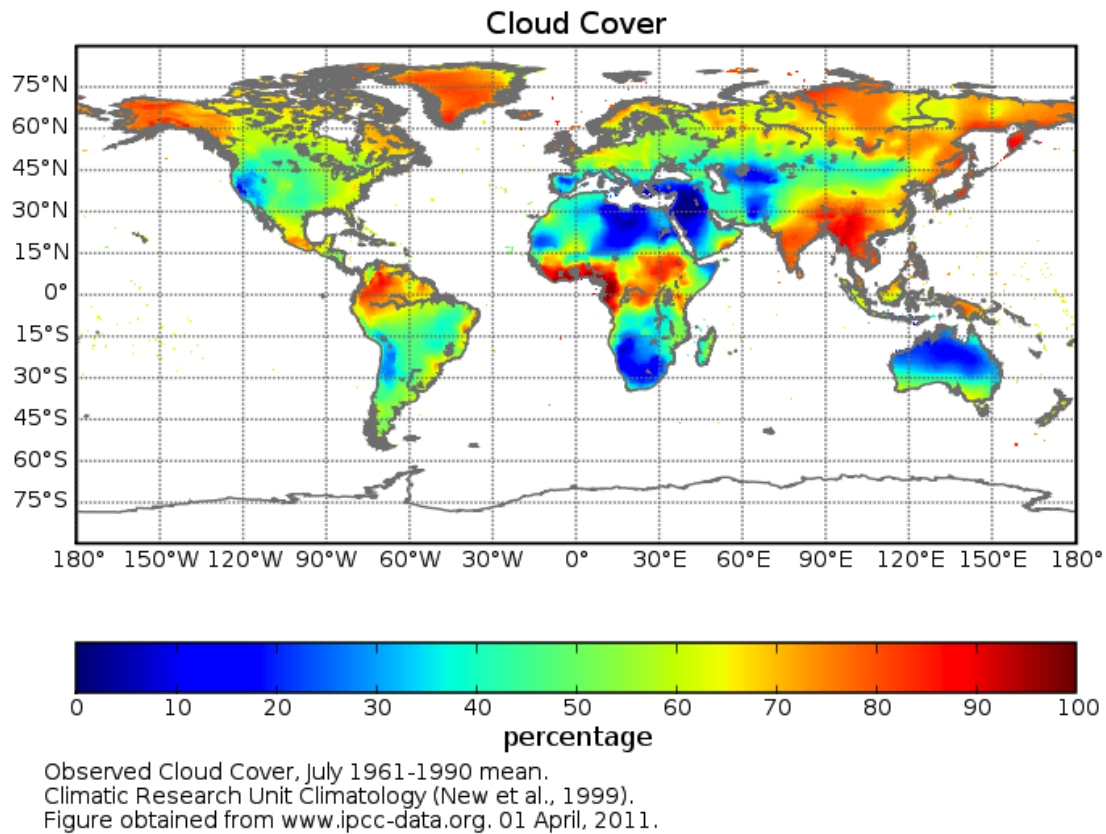
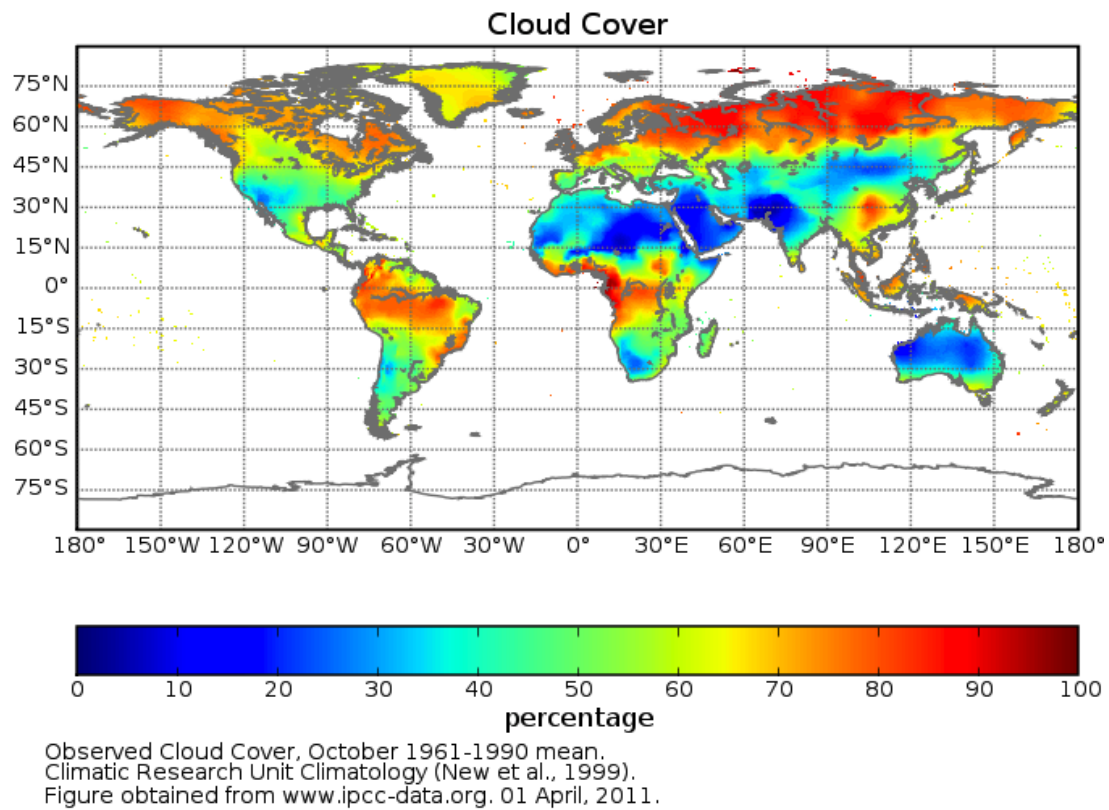
**Figure 4.3.2 c)****Figure 4.3.2 d)**

Figure 4.3.2 Quarterly measured global cloud cover percentage

Fig. 4.3.2 a) – d) shows quarterly values of one of the variables used in the current proposed model on a global scale.

Cloud cover is given as an example of a variable, which combines both elements of water and elements of energy. Data sets of other climatic variables are shown in Appendix 3. At this wide resolution, for example, Madagascar, within 30 and 60°East, 0-30°South shows 50-80% cloud cover in January, 40-60% in April, 40-50% in July and 50% in October.

The first step to increase the efficiency of the basic model shown in **Fig. 4.3.1** is to minimize the variables used to form the model. The variables are shown in terms of their water or energy type in **Tab. 3.2.2**.

Intuition for the distribution of plant strategies dictates that we need only use key elements of the water energy (W-E) dynamic, a variable to show disturbance within habitats and altitude, these are used due to the fact that both water and energy display different effects with altitude values (Bhatterai and Vetaas, 2003; Sommer et al., 2010).

Using these key elements we are able to make the following framework to form a rule based structure, removing co-variable, redundant variables:

$$If A1_{(n)} \prec A1_{(n)} \wedge A2_{(n)} \prec A2_{(n)} \wedge A3_{(n)} \prec A3_{(n)} \wedge A4_{(n)} \prec A4_{(n)} Then B_{(n)} = E1, ..., E7 \quad (4.3.1)$$

Where $A1$ is mean precipitation, $A2$ is mean temperature, $A3$ is mean ground frost frequency, $A4$ is altitude and $E1, ..., E7$ are environments 1 to 7. This is a concise statement linking the defined ranges of climatic and altitude antecedent variables to consequent plant strategy environments.

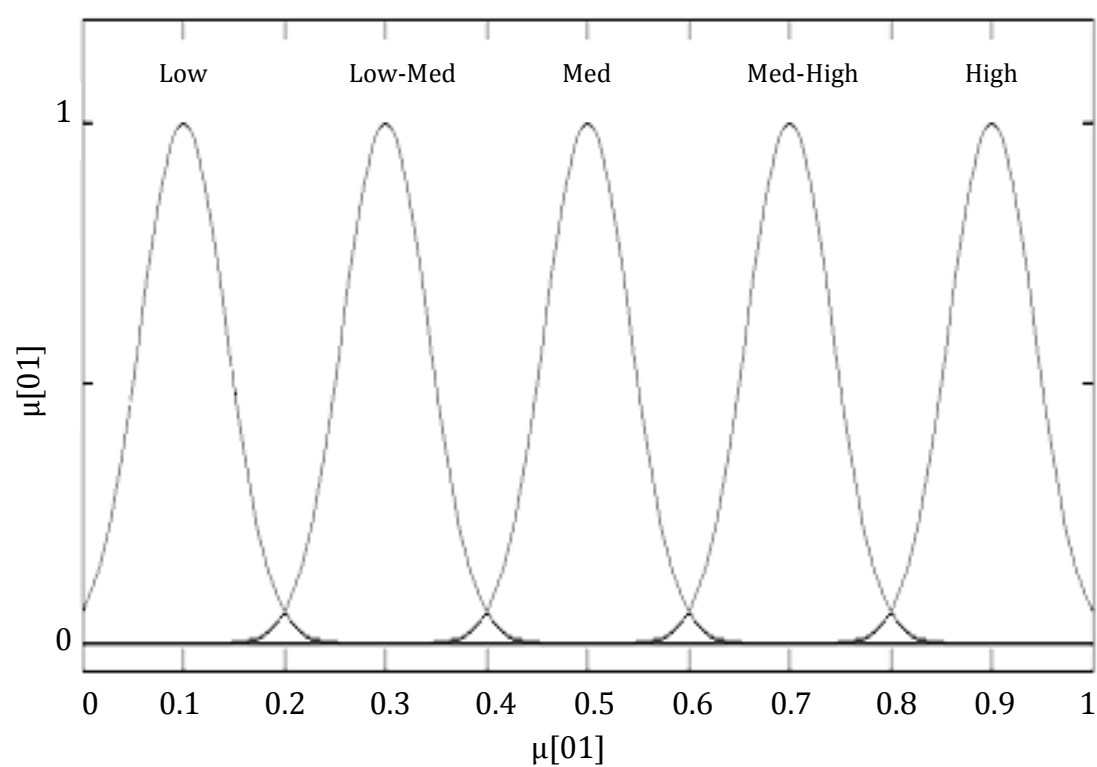
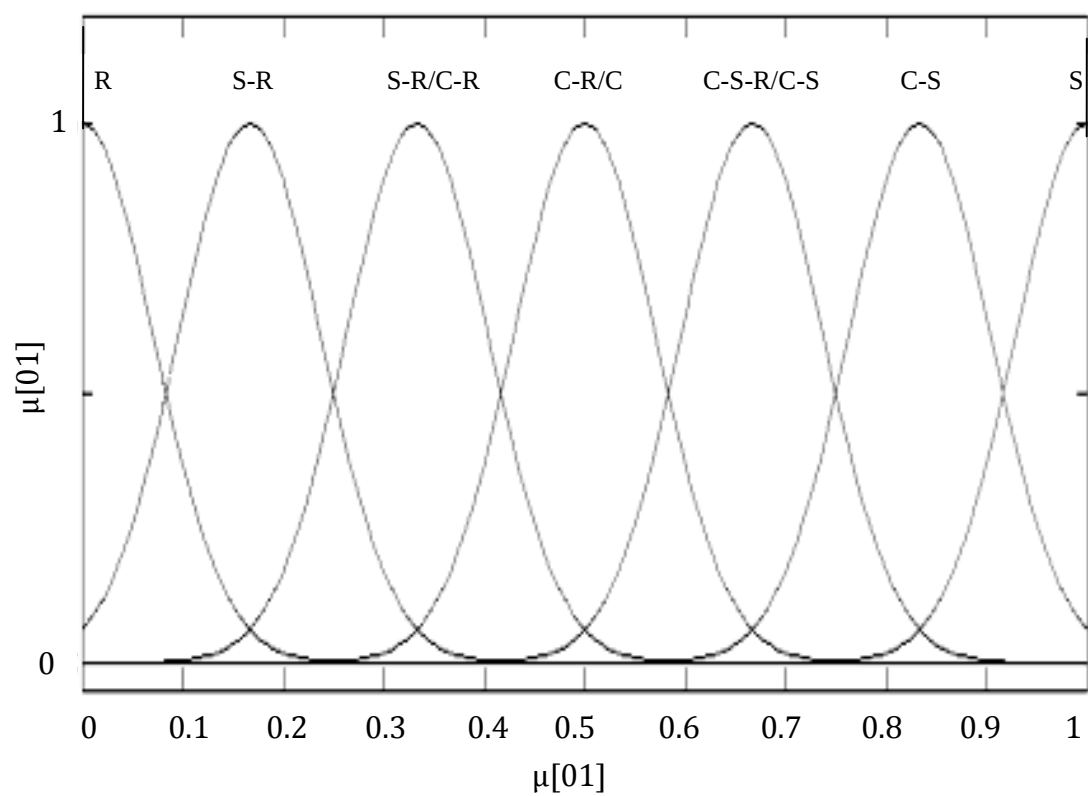
**Figure 4.3.3 a)****Figure 4.3.3 b)**

Figure 4.3.3 a) Gaussian membership functions as applied to the 5 divisions over the interval [0 1] of climatic and altitude input variables, b) Gaussian membership functions as applied to the 7 divisions (R is Ruderal, S-R is Stress tolerant to Ruderal, S-R/C-R is S-R or Competitive Ruderal, C-R/C is Competitive to Ruderal or Competitive, C-S-R/C-S is Competitive Stress tolerant Ruderal or Competitive Stress tolerant, C-S is Competitive Stress tolerant and S is Stress tolerant) over the interval [0 1] of plant strategies.

The FL system membership functions are for Mamdani type FL due to the fact that comparatively low resolution is used for the input variables and un-specified non-crisp numerical ranges are conceptualized for the plant strategy environments. The consequent surface area of the Mamdani FL system is shown in Appendix 2.2. The rules used to generate this system are shown in Appendix 2.3.

Fuzzy logic is an appropriate system by which one may develop the algorithms to characterise global plant distribution as the proposed characterisation is reliant on covariates of climatic factors which are used to calculate species presence based on probability of the species occurring (Furze et al., 2011).

In the next subsection higher resolution input variables and specific consequent ranges are employed for greater accuracy, enabling more specific statements to be made to further predict the occurrence of plant strategy types with regard to specific locations.

4.4 Validation of algorithms and global mapping of plant strategies

In this subsection the preliminary methodology for the construction and validation of the algorithmic framework shown in (4.3.1) is shown. The first step is to clarify the process required for formation of location specific fuzzy rules to ordinate the plant strategy environments. These are shown in **Fig. 4.4.1**.

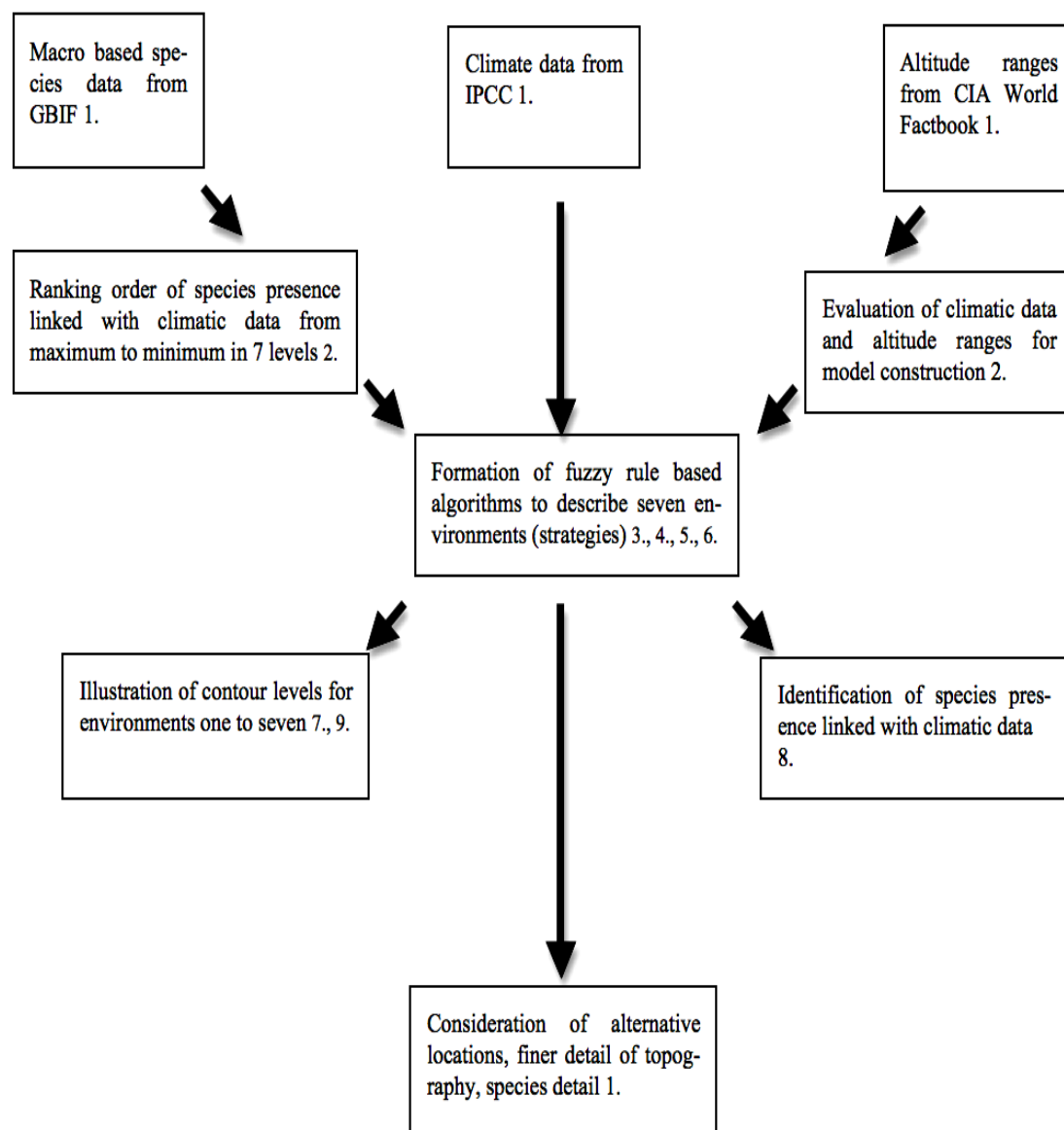


Figure 4.4.1 Stages of model construction

Fig. 4.4.1 shows the stages of methodology for the formation of fuzzy based algorithms to quantify plant life-history strategies on a global scale. Macro based species data from the GBIF, climate data from the IPCC and altitude ranges from the CIA world factbook are shown in Appendix 2.4.

Steps 1-9 of **Fig. 4.4.1**, are described as follows:

1. Data were selected to define model parameters.
2. The spread of numerical data and the total number of variables were defined.
3. The minimum number of key parameters required to build the model was determined.

4. Units and partitions within the data used were identified and concise linguistic description was made.
5. Linguistic description of the data to give fuzzy description of each variable and the fuzzy consequences for each plant strategy/environment (E1 to E7) were defined.
6. Seven estimates of the total number of individual plant occurrences found in each example geographic location, chosen at random from amongst areas containing more than 3000 plants per 10000 km², were made to infer E1 to E7.
7. Model parameters were numerically quantified.
8. Algorithmic instruction of E1 to E7 was constructed.
9. Plant life-history strategies with respect to the total number of individual plants were conceptually illustrated. The conceptual plot was undertaken using the contour plot element of Matlab (version R2010a ©), shown in **Fig. 4.4.2**.

The variables shown in **Tab. 4.3.1** were statistically reduced to four key variables in order to facilitate modelling of strategies. These variables were mean ground frost frequency (chosen for its effect in terms of disturbance), mean precipitation (chosen as the key water related variable), mean temperature (chosen as the key energy related variable) and altitude. Altitude was used as it is key in the water-energy modelling of plant species; at low latitudes (south of the equator) water is more important for high species numbers, whereas at higher latitudes energy is seen to be more important for species numbers (Hawkins et al., 2003; Vetaas, 2000).

Table 4.4.1 Unit percentage of key W-E modelling parameters.

Variable	Unit Percentage
Mean Ground Frost frequency	0.3 Days
Mean Precipitation	5 Kg m ⁻²
Mean Temperature	0.7 Celsius
Altitude	68.3 m

Unit percentage relates to 0.01 in the interval [0 1] of the variables, quantified according to the 5 linguistic terms (Alcala et al., 2007) (shown in **Fig. 4.3.3. a**)).

Overall seven estimates of the total number of individual plant occurrences were made, one for each of the example locations that were chosen randomly from areas containing more than 3000 plant species per 10000 km² - diversity zones (DZ) 8-10 (Barthlott et al., 2005). GBIF recorded number of individual occurrences in each location were summed. The resultant total numbers of individuals were ranked in decreasing order from one to seven. After modelling W-E and altitude variables to infer extremity of the environments' plant life history based strategy, each number was allocated to an environment (data and citations of collections for each example location used are shown in Appendix 1).

Contour Plot function was used to display conceptual levels of the seven plant strategies with respect to the total individual numbers of plants within each of the seven environments. Numbers of individuals were entered into Matlab (Version 2010b ©) to form a 2x2 matrix and a Z matrix of numbers was calculated (using a numerical space of 700 to reflect the seven plant strategies) from the 2x2 data. The magnitude of the number of individuals in each environment was used to reflect the difference between contour levels. The contours were plotted diagrammatically on a 700x700 axis.

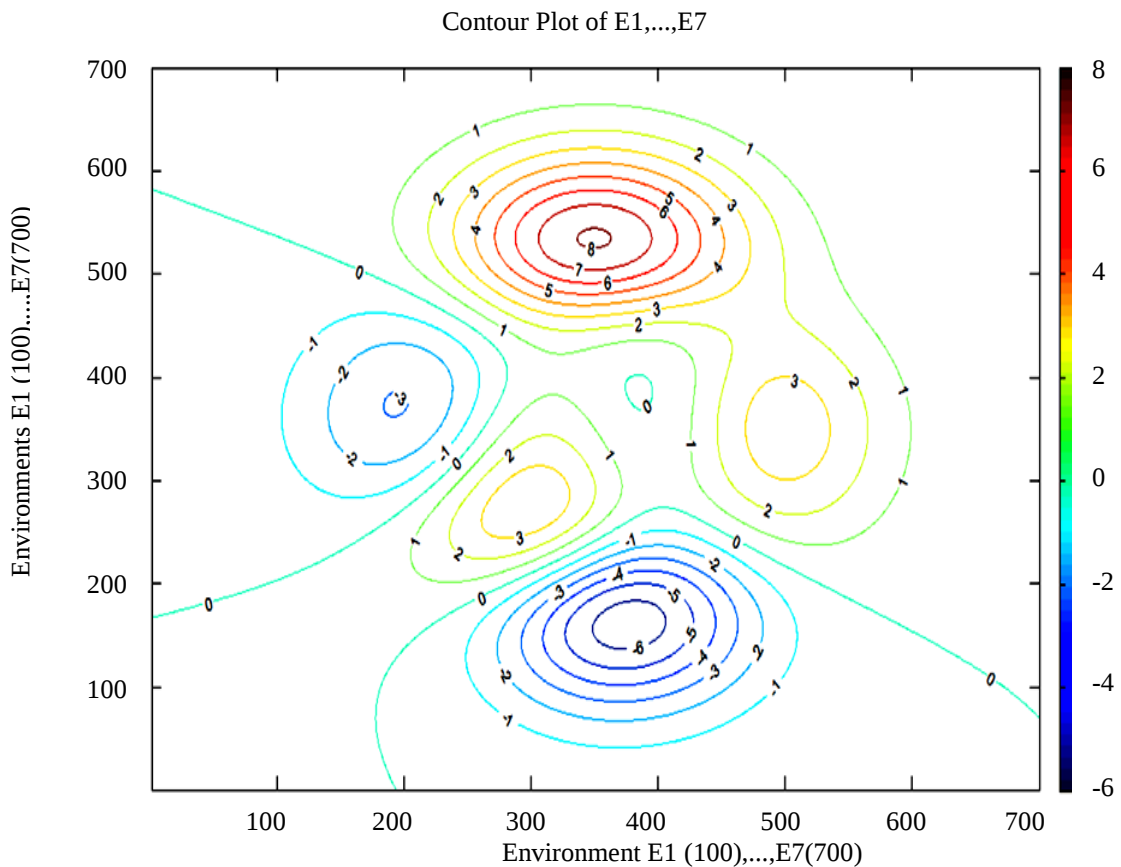


Figure 4.4.2 The concept of dimensionality (relative size) of environments one to seven (Furze et al., 2012a)

Zero and negative numbers were not present in reality, but were shown here representing the inverse of the positive levels. The legend defines the spectrum of colour used to display each contour level. Contour levels were as follows: 1 = numbers of individual plants for Mexico (Ruderal), 2 = numbers of individual plants for Guyana (Stress tolerant-Ruderal), 3 = numbers of individual plants for Cuba (Stress-tolerant to Ruderal, Competitive to Ruderal), 4 = numbers of individual plants for Democratic Republic of the Congo, Africa (Competitive to Ruderal, Competitive), 5 = numbers of individual plants for Georgia, Eastern Europe (Competitive to Stress-tolerant to Ruderal, Competitive to Stress-tolerant), 6 = numbers of individual plants for Guinea, Africa (Competitive to Stress-tolerant), 7 = numbers of individual plants for Macedonia, Southern Europe (Stress tolerant). **Fig. 4.4.2** means that as the environment number 1-7 increases, the number of species decreases. Limits are not stated precisely due to the fact that the contours are conceptualised from a Mamdani (imprecise) FLC. The power of this type of modelling is increased in Chapter 5, with use of greater precision in the modelling variables.

Data of climatic variables for locations were taken at 30 second, 18.5km resolution, to enhance accuracy of the algorithmic statements. These are shown below in **Fig. 4.4.3**

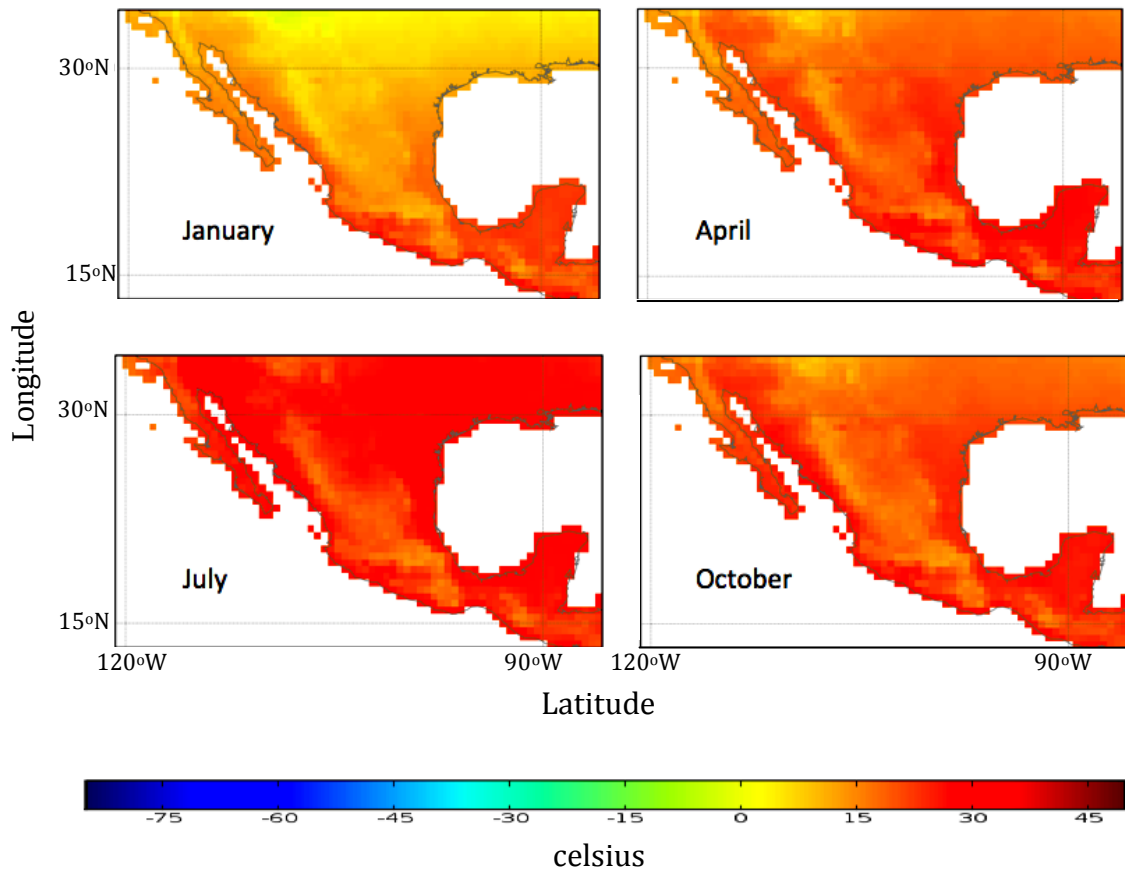
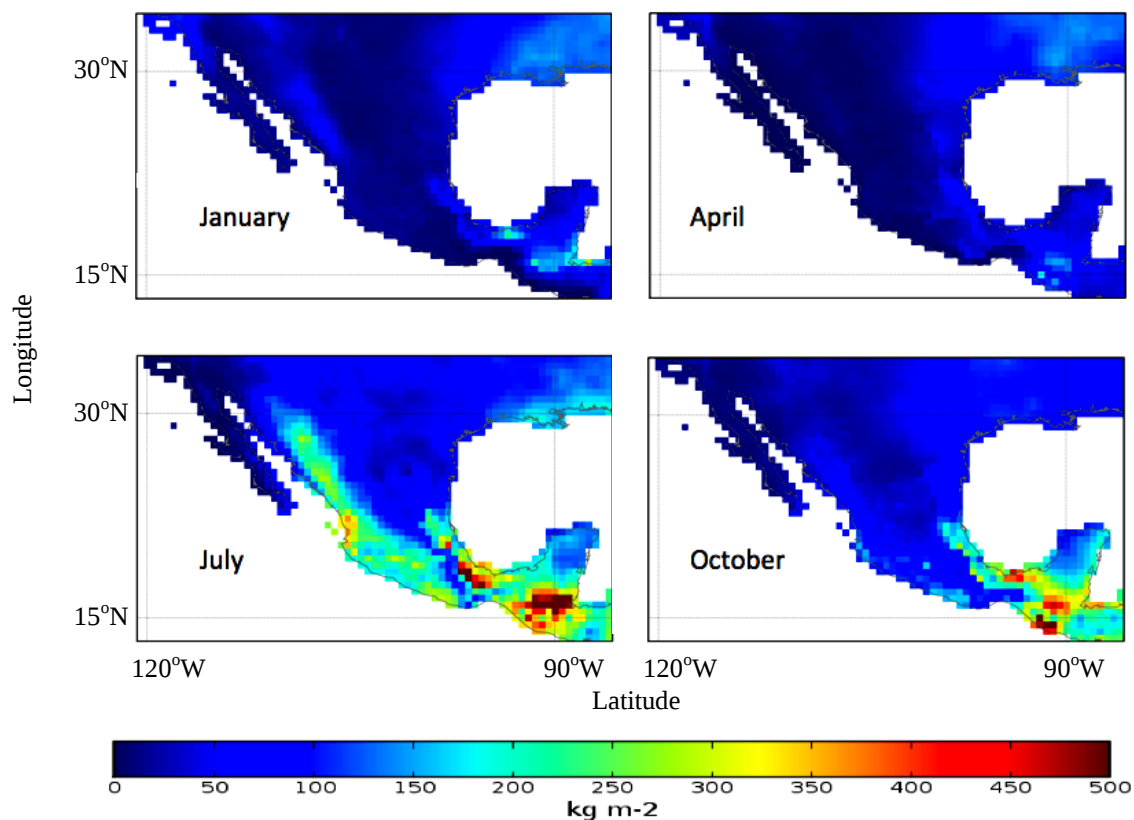


Figure 4.4.3 a)



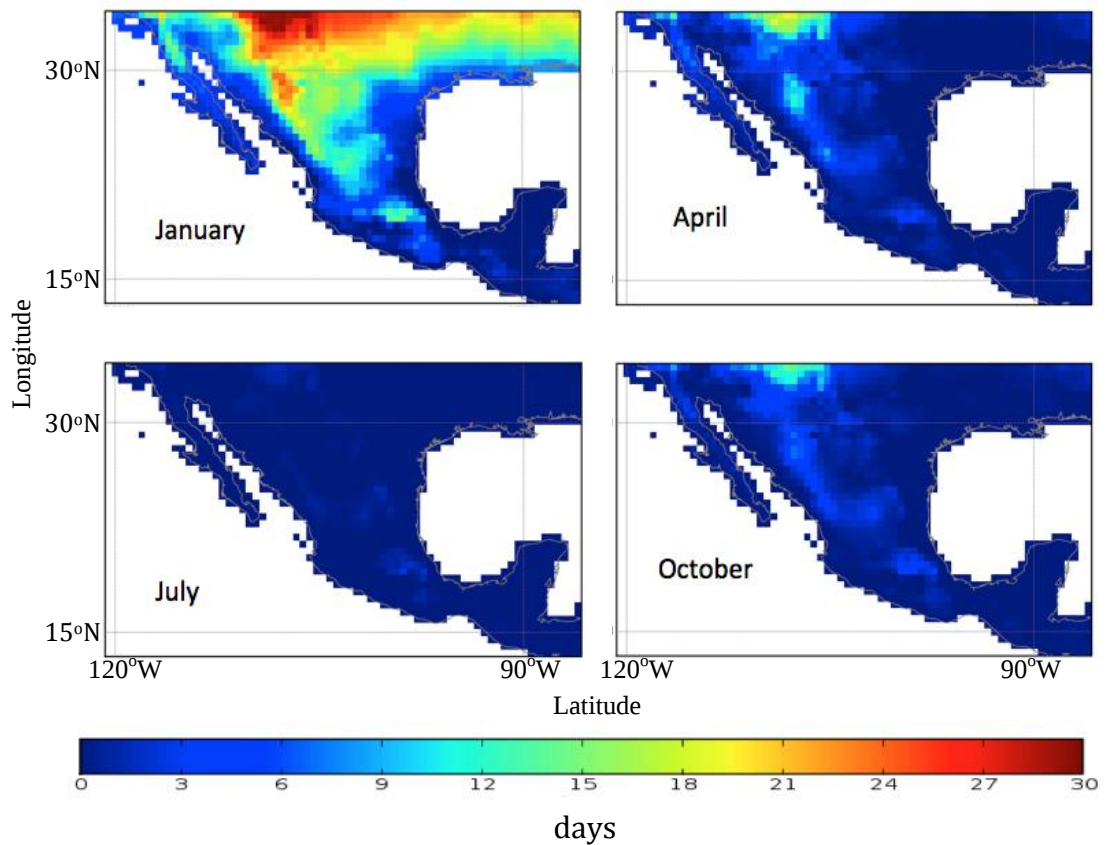


Figure 4.4.3 c)

Figure 4.4.3 Quarterly climatic data representing the water energy dynamic used for algorithmic statements for Mexico, a) Mexico quarterly mean temperature, 1961-90, b) Mexico quarterly mean precipitation, 1961-90, c) Mexico quarterly mean ground frost frequency, 1961-90.

Table 4.4.2 Partitions for climatic and topographic variables.

Variable	Low ⁽¹⁾	Low-Med ⁽²⁾	Med ⁽³⁾	Med-High ⁽⁴⁾	High ⁽⁵⁾
MT (%)	0-20	20-40	40-60	60-80	80-100
MP kg m ²	0-100	100-200	200-300	300-400	400-500
GFF (days)	0-6	6-12	12-18	18-24	24-30
Altitude (m)	-30 - 1366	1366 - 2732	2732 - 4098	4098 - 5464	5464 - 6830

In the above table MT is Mean Temperature, MP is Mean Precipitation, GFF is mean Ground Frost frequency and kg m² is Kilograms per metre squared.

The total range of each of the variables considered was quantified according to the partitions of **Tab. 4.4.2** and the unit percentage of **Tab. 4.4.1** to enable precise algorithmic statements to be formed according to (4.3.1). Latitude

Fig. 4.4.3 exemplifies the data of Mexico, which shows the following control algorithm for orientation of plant species:

$$\begin{aligned} & \text{If } A1_{(4)} \prec A1_{(5)} \wedge 0.5A2_{(1)} \prec A2_{(3)} \wedge 0.5A2_{(1)} \prec A2_{(5)} \wedge 0.25A3_{(1)} \\ & \prec A3_{(3)} \wedge 0.25A3_{(1)} \wedge A4_{(1)} \prec A4_{(5)} \text{ Then } B_{(65539)} = E1 \end{aligned} \quad (4.4.1)$$

In (4.4.1), antecedent expressions are mean temperature is $A1$, mean precipitation is $A2$, mean ground frost frequency is $A3$ and altitude is $A4$. Variable choice is made following climatic data available from the IPCC (Scenario T.S. 2.1) originally provided by New et al. (1999). Consequent expression B is the number of individual plant species occurrences and $E1$ is environment 1. The algorithm expands into the following conditions:

IF Variables (A) =

- Temperature = 40-60 % to 80-100 %
- Precipitation = 0.5 x 0-100 Kg m² to 200-300 Kg m², 0.5 x 0-100 Kg m² to 400-500 Kg m²
- Ground Frost frequency = 0.25 x 0-6 days to 24-30 days, 0.5 x 0-6 days to 12-18 days, 0.25 x 0-6 days to 0-6 days
- Altitude = -30-1366 m to 5464-6830 m

THEN Environment 1 (B) = >51847

Mexico is designated as $E1$ with ruderal species dominating due to the fact that the highest number was obtained for this location in comparison to the other example locations.

The algorithmic statements for the other example locations (Guyana, Cuba, Democratic Republic of the Congo (Africa), Georgia (Southern Europe), Guinea (Africa) and Macedonia (Southern Europe)) are shown in Appendix 2.

The patterning of plant species presence may be broken down into the seven life-history based strategies in the following way: The occurrence of a high water-energy dynamic results in the highest level of plant species diversity reflected in the greatest numbers of plant species presence. The highest plant species presence numbers reflects the dominant ruderal plant species. The effect of water and energy (mean rainfall and temperature) on plant distribution shown in this chapter may also be used to suggest application of more accurate modelling variables with the changing conditions within global warming / cooling cycles. Patterning of plant species in this chapter, provide first glimpses of the power of mathematical modelling. Decreasing numbers of individuals in each algorithmically described environment reflects the transition through competitive to stress tolerant species, which are present in the more extreme (hotter, dryer) environments.

Preserving the relationship between plant species presence and climatic and topographic variability requires the application of cooperatively controlled multi-agent systems. The use of a fuzzy-logic rule base is especially appropriate with respect to species presence as numbers of the latter involve mathematical dispersion based on the levels of water, energy and topographic dynamics. This chapter clearly shows the relevance of a mathematical approach with respect to water and energy dynamics and furthers the information rich patterning of plant species based on life-history strategy characterisation (Furze et al., 2011). The ecological relevance of the concept of plant strategies as derived from individual plant numbers is that the plant strategy patterns are shown in macro scale space. Topology has been simplified to discrete value ranges for the example locations given in this initial mathematical approach in order to show the validity of the modelling procedure in this thesis. Precise detail of the locations will be explored in later studies in order to enhance the accuracy of the algorithmic statement. Feeding location-specific data into the models will validate their application at finer spatial resolution and enhance regional interpretation of biodiversity patterns. Greater understanding may direct conservation management at local and national level, especially pertinent in future dynamic climatic scenarios.

The mathematical approach detailed is superior to other previously shown methods (Grime, 1979; Hodgson et al., 1999) as it enables simple quantification of many different elements and expression through specific algorithms. The methods used in this paper are not Boolean as resultant data may be distorted by uneven data sampling. The normal distribution of variables may be used to describe dynamic patterns with greater accuracy

(Zadeh, 1965). These methods are suggested above historical approaches as they result in minimal error (Furze et al., 2011). This is of great use in describing natural systems as the sensitivity of ecosystems with change can be eloquently stated. The application of the fuzzy rule base was shown along with the appropriate use of contour levels in order to reflect the numerical distribution of dimensionality between plant strategy groups. The information rich ordering of plant strategies shows the least severe environment (Mexico) to contain the highest numbers of individuals (ruderal plants) through competitive to the least number of individuals (stress tolerant plants) in the most severe environment (Macedonia) (Furze et al., 2012a). Although this is what we might expect from theory developed to the current time, the novel application of fuzzy logic in the subject of plant characterization at a global scale provides the basis for further analysis to be carried within temporal and spatial terms from which we can infer patterns in the categories of plant species, which are as yet undiscovered. Ultimately the novel application of fuzzy logic enables unification of previous theory by mathematic quantification, thereby substantiating concise statements for species community assemblages on a global scale.

4. 5. Summary

This chapter demonstrates the application of the SAR to real life data and shows that the relationship is not sufficient to describe the multi-objective approach that is required to ordinate plant species without distorted sample selection. This justifies the work of this chapter and that of remaining chapters.

An information-rich modelling approach (**Fig. 4.3.1**) is proposed for global characterisation of plant species and detailed at low-resolution. An initial model containing 9 antecedent variables by which plant species may be ordered into 7 consequent environments has been detailed. The initial model has been minimized to contain 4 antecedent input variables. After illustrating a conceptual model of the dimensionality of the seven environments, the seven combinations of plant strategies are allocated to environments. Examples of locations are given for the seven environments and reduced climatic data of an example location is shown at higher resolution for enhanced accuracy.

The control algorithm based on Gaussian distribution of the antecedents for the country of Mexico is given and expanded. Use of logic based mathematics given in this chapter is devoid of numerical error present in Boolean methods. Hence the only error in the control

algorithm approach shown is due to too wide a resolution of data, repetition of variable types and sampling errors present within initial knowledge bases.

In the following chapter we refine the fuzzy logic control method to show a concise description of plant strategy ordination.

CHAPTER 5

Using fuzzy logic control to ordinate plant strategies

5. 1 Introduction

In Chapter 4 it was shown that the basic species-area relationship shows an insignificant relationship when applied to actual numbers of species occurrence relative to the area of locations in more than twenty of the richest areas of diversity on the planet. Many studies have shown a relationship, but the methods employed in these studies are limited by the statistics used in them and represent examples of distorted Boolean methods, which have led to subsequent distortion of related patterns and possible misinterpretations in their conclusions. The major advantage of fuzzy logic control in description of species ordination is that fundamentally there is no error in the technique, due to the methods being centralised according to Gaussian distribution. Hence, significant or insignificant relationships or trends may be informatively explored with use of robust modelling frameworks. The novel application to plant characterisation enables inferences to be made throughout successive trophic levels of ecosystems. The application of fuzzy logic was proposed in order to differentiate the conditions in which plant species occur, using the water-energy dynamic. An expansive model allocated 7 environments based on the life-history strategy classification of plant species (Grime et al., 1995), successfully predicting plant species occurrences. The basic model used a Mamdani form of fuzzy logic, using broad ranges of input data and non-specific ranges of numerical consequence.

In this chapter a Takagi-Sugeno-Kang (T-S-K) fuzzy logic approach is taken to ordinate individual species occurrences in specific locations, using comparatively higher resolution of reduced numbers of climatic variables and specific ranges of altitude and consequent species occurrences. Species occurrence in Guyana, South America is exemplified, according to elements of the water-energy dynamic. Furthermore the analytical power of the original FLC proposed in the previous chapter is increased with use of defined ranges of consequential statements.

The remainder of this chapter is structured as follows: section 5.2 describes the methodology for construction of the T-S-K fuzzy logic control with respect to plant strategies. Section 5.3 exemplifies the location of Guyana, illustrating the fine scale resolution climatic and topographic data used in formation of the model and gives a developed algorithmic framework, with results of the control theory in other locations. Section 5.4 implements FLC to ordinate plant strategy in Guyana and finally section 5.5 summarises the chapter.

5. 2 Takagi-Sugeno-Kang modelling of plant strategies

In this section, the background of data types and sources is given along with their subsequent categorization to facilitate T-S-K modelling of plant strategies, with use of defined input data and specified ranges of consequence.

Biodiversity data, in the form of digitised data of individual plant occurrences identified to species level, were sourced from the Global Biodiversity Information Facility (GBIF, <http://www.gbif.org>). The total number of occurrences was then summed; this substantiates a component of the knowledge base used in T-S-K modelling. Seven locations were chosen at random from Barthlott's description of diversity zones 8–10. The zones were documented as containing more than 3,000 plant species/10,000 km² (Barthlott et al., 2005). The data have been validated (Yesson et al., 2007) and their quality proved sufficient to allow analysis using fuzzy techniques in classification (Zadeh, 1965).

Previous studies (Furze et al., 2011, 2012a) have made use of ranges of topographical measures based on publicly available broad-scale digital elevation model (DEM) data.

In this T-S-K modelling approach of plant life-history strategies, the sources of data for the modelling basis were as follows: topographical data (1 km resolution) was sourced from

the United States Geological Survey (USGS) DEM (USGS http://www.eros.usgs.gov/#/Find_Data/Products_and_Data_Available/GTOPO30), being 33 tiles with global coverage. The chosen areas were identified. Files were downloaded in compressed format. Data were extracted and processed using MATLAB (Version R2010a ©) and topographical maps were produced using the same platform. Data of climate variables (mean precipitation; mean temperature; mean ground frost frequency) at monthly intervals 1961-90 were sourced from the IPCC (<http://www.ipcc-data.org>). The geographical location (i.e. latitude, longitude) of the chosen area(s) was defined from the display of the DEM. Graphical images displaying quarterly data of 1961-90 (Mitchell and Jones, 2005; New et al., 1999) were obtained for the three required climate variables and altitude. The four images displaying the variables were processed in MATLAB in order to obtain the variables that express the image. The range of each variable was obtained from the data sources using the units of each source. These were then converted into percentage values and the percentages broken into five quintiles. The linguistic expressions, quantification and notations used to describe the data are shown in **Tab. 5.2.1**.

Table 5.2.1 Variable partitioning for T-S-K modelling of plant strategies

Ling exp	% Quant/Not'n	Range			
		MT ^{°C}	MP (kg m ²)	MGFF (days)	Alt (m)
Low	0-20/ ₁	-75 to -51	0-100	0-6	0-600
Low-medium	20-40/ ₂	-51 to -27	100-200	6-12	600-1200
Medium	40-60/ ₃	-27 to -3	200-300	12-18	1200-1800
Medium-high	60-80/ ₄	-3 to 21	300-400	18-24	1800-2400
High	80-100/ ₅	21 to 45	500-500	24-30	2400-3000

Ling exp = linguistic expression, Quant' = quantification, Not'n = notation, % = percentage, MT = mean temperature, °C = degrees Celsius, MP = mean precipitation, kg m² = kilogram per square metre, MGFF = mean ground frost frequency, Alt = altitude, m = metre.

The numerical data for each of the variables was considered in each of seven example environments. Using the maximum and minimum inference of each variable's linguistic definition ($A1_{(n), \dots, n_{(n)}}$), the fuzzy rule-based algorithms were constructed so that each

variable was expressed in terms of the number of species ($B1_{(n), \dots, n_{(n)}}$) of each geographic location ($E1_{(n, \dots, n)}, \dots, E7_{(n, \dots, n)}$). Mean temperature was noted as $A1_{(n, \dots, n)}$, precipitation was noted as $A2_{(n, \dots, n)}$, mean ground frost frequency was given as $A3_{(n, \dots, n)}$, altitude was noted as $A4_{(n, \dots, n)}$ and the number of species was noted as $B_{(n, \dots, n)}$. The numerical data substantiates the antecedent knowledge base, the ranges of values used in each case were extrapolated from the data sources. The linguistic connections 'IF', 'AND' and 'THEN' were used to construct the conditional fuzzy rule base.

The above elemental breakdown of the T-S-K modelling framework is in agreement with the basic rule structure (Takagi and Sugeno, 1985; Zadeh, 1965) as covered earlier in Chapter 3, subsection 3.7.2. In order to substantiate the elements of the T-S-K system, the following section details the enhanced resolution of data used for modelling of plant strategies.

5.3 Fine-scale resolution data enhances model formation

In this section, climatic data are shown which underlie the algorithmic structure employed based on set theory (Zadeh, 1965).

In order to build the fuzzy inference engine, which may predict the structure of seven strategy-based environments of plant species occurrence, the following steps were taken:

- 1 Define fuzzy inference system type (Sugeno for defined output type) variable names
- 2 Define membership functions of each variable
- 3 Define rules, weights in the interval [0 1] according to algorithm
- 4 Examine (adaptive neuro fuzzy/Sugeno) logic structure and modify as necessary
- 5 Test rules through data input
- 6 Display 2D and 3D views of resultant surface in order to graphically display the algorithm (efficiency).

The modelling system may also be used as a ‘stand alone’ engine. This was carried out by saving the fuzzy inference system in MATLAB workspace and to file. The code is given in Appendix 2.3. Enhanced resolution of antecedent variables, in turn, enables greater accuracy in consequential statements to be obtained. This is an imperative when using a T-S-K modelling basis as any errors received during the modelling process are borne out

during later stages of analysis (Sivanandam, 2007). Higher logic based mathematics used

in this study tolerate the presence of a degree of error due to the fact that they are based on continuous (Gaussian) process models. The error detected in the methods is minimized in this study by use of the triangular membership function, which gives a discrete perspective of continuous variation. An example of a method by which one can detect error present is the use of a random dispersal of objectively formed solutions (shown in Chapter 6).

Here (**Fig. 5.3.1**, **Fig. 5.3.2**) we show examples of the fine scale resolution data used to construct the algorithmic framework for ordination of (environment 2) stress

tolerant-ruderal species present in the location of Guyana, South America.

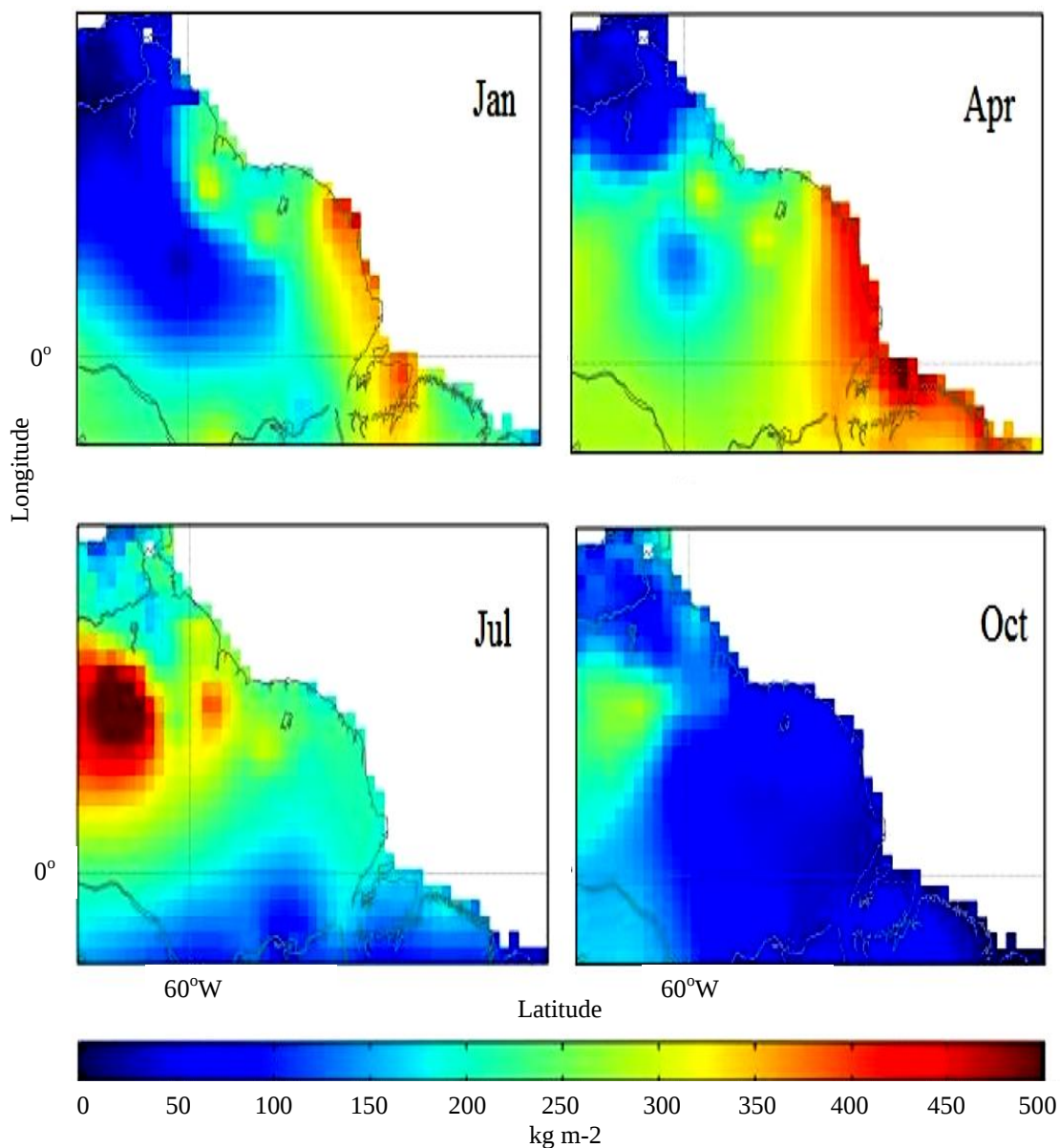


Figure 5.3.1 Guyana quarterly mean 1961-90 precipitation at 10 minute (18.5km) resolution

Fig. 5.3.1 shows example data, where Guyana mean precipitation is 0.75 (January, April, July) 0–100 kg m² to 200–300 kg m², and 0.25 (October) 0–100 kg m² to 300–400 kg m². The quantity of precipitation is shown in colours from low (dark blue) to high (dark red).

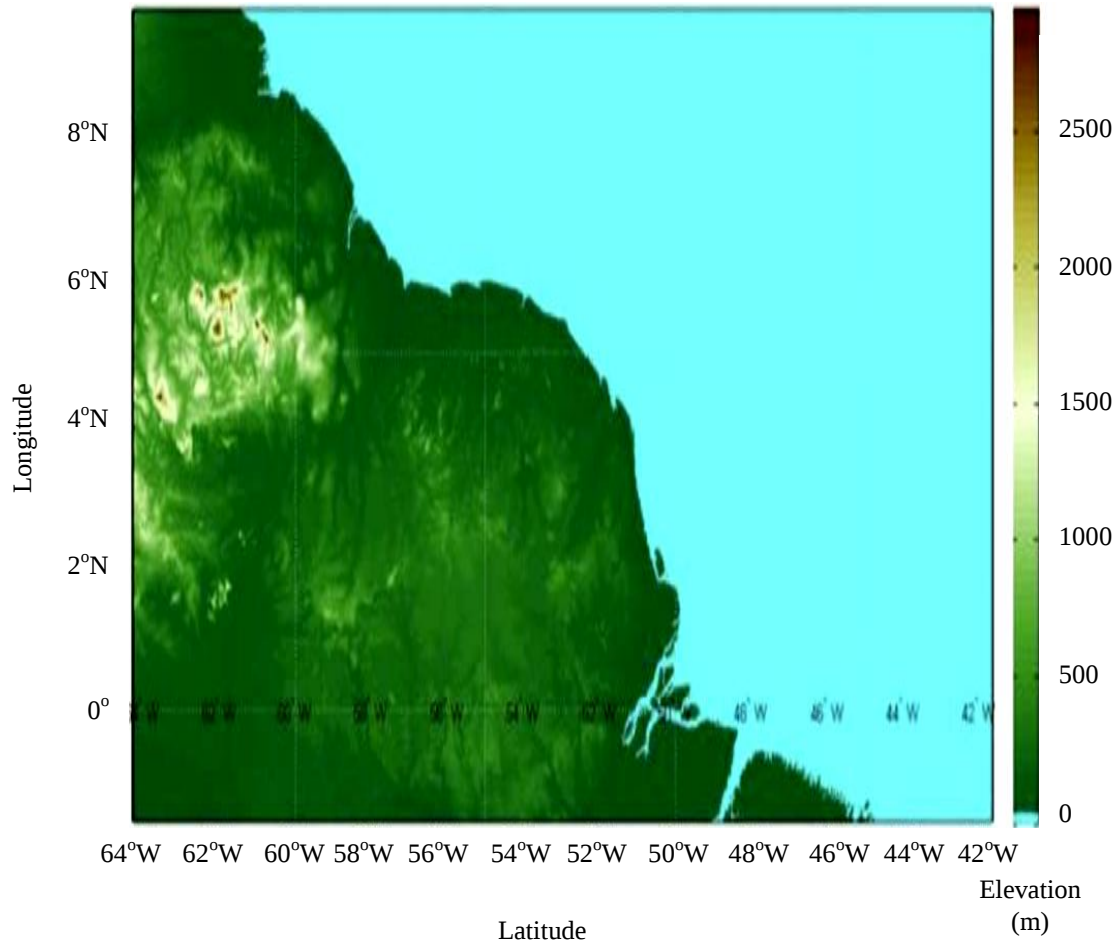


Figure 5.3.2 Guyana digital elevation model / topography at 30 second (1km) resolution

Fig. 5.3.2 is a digital elevation model (DEM) representation of Guyana, situated between latitude 60° – 55° West, longitude 0° – 7.5° North with an elevation from 0 – 1500 metres above sea level. Sea level is shown in blue, low elevation is in dark green and low-medium elevation in lighter green to white.

Climatic and DEM data for the other sourced environments, discussed in Subsection 5.4, are shown in Appendix 3.

The linguistically broken down T-S-K Fuzzy control algorithm integrating climate variables and DEM data for Guyana is written as follows:

$$\begin{aligned}
 & IFA1_{(5)} \prec A1_{(5)} \wedge 0.75A2_{(1)} \prec A2_{(3)} 0.25A2_{(1)} \prec A2_{(4)} \\
 & \wedge A3_{(1)} \prec A3_{(1)} \wedge A4_{(1)} \prec A4_{(2)} \text{ Then } B_{(51847)} = E2
 \end{aligned} \tag{5.3.1}$$

The control algorithm applicable for categorisation of the example location chosen at

random is (5.3.1). The dominant strategy of individuals in this location was stress tolerating-ruderal species ($E2$). Numerical quantification of the algorithm is as follows:

IF Variables A =

- Mean temperature = 80 – 100 % to 80 – 100 % ($A1_{(5)}$)
- Mean precipitation = 0.75 0 – 100 kg m² ($A2_{(1)}$) to 200 – 300 kg m² ($A2_{(3)}$), 0.25 0 – 100 kg m² ($A2_{(1)}$) to 300 – 400 kg m² ($A2_{(4)}$)
- Mean ground Frost frequency = 0 – 6 days to 0 – 6 days ($A3_{(1)}$)
- Altitude = –30 – 1366 m ($A4_{(1)}$) to 1366 – 1500 m ($A4_{(2)}$)

THEN $B_{(51847)} = E2$

Temperature and ground frost frequency are not shown in this chapter; however, they can be found in Appendix 3. Example locations of environments $E1$, $E2$, $E3$, $E5$, $E6$, and $E7$ were defined using the algorithmic control structure of (5.3.1), as shown in **Tab. 5.3.1**.

Table 5.3.1 Categorisation of environments and plant life-history strategies

Environment	Plant life-history strategy	Example location / number of individuals
1	R	Mexico / 51857 – 65535
2	S-R	Guyana / 50700 – 51847
3	S-R / C-R	Cuba / 33356 – 50700
4	C-R / C	Democratic Republic of the Congo / 11355 – 33366
5	C-S-R / C-S	Georgia / 8805 – 11355
6	C-S	Guinea / 2203 – 8805
7	S	Macedonia / 0 – 2203

In **Tab. 5.3.1** R is ruderal, S-R is stress tolerant-ruderal, S-R / C-R is stress tolerant ruderal / competitive ruderal, C-S-R / C-S is competitive stress tolerant ruderal / competitive stress tolerant, C-S is competitive stress tolerant, S is stress tolerant.

In the following section RQ1 and RQ2 are resolved with the implementation of the T-S-K

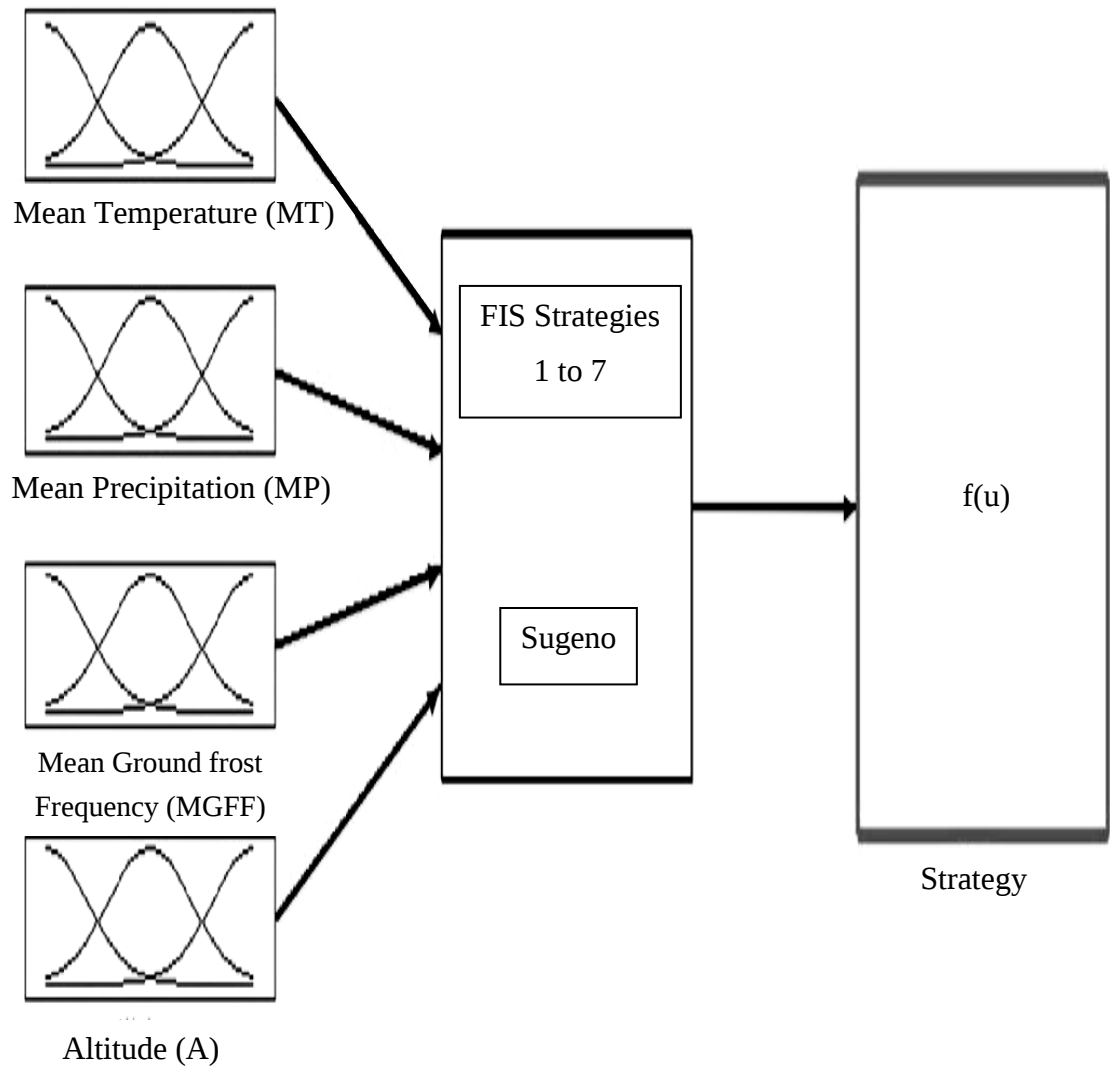
FLC applied for plant life-history environments.

5. 4 Implementation of fuzzy logic control to predict plant species occurrence

In this section, the choice of fuzzy inference system is presented by which plant life history strategies are mathematically ordinated according to the data presented in the previous section. The FL simulation engine was applied over the input vectors temperature, precipitation, ground frost frequency and altitude and the output vector of strategy is mapped according to the control algorithm of (5. 3. 1). Membership functions of each of the variables were defined as shown in **Fig. 5. 4. 2**, making use of triangular functions to discretely define each partition. Each inferential variable membership function was defined with use of the graphical user interface within the fuzzy toolbox of Matlab (Version 2010a ©), each vector function operated together under the instruction of the root algorithm of (5.3.1), which was expressed with the use of ten separately weighted rules as follows:

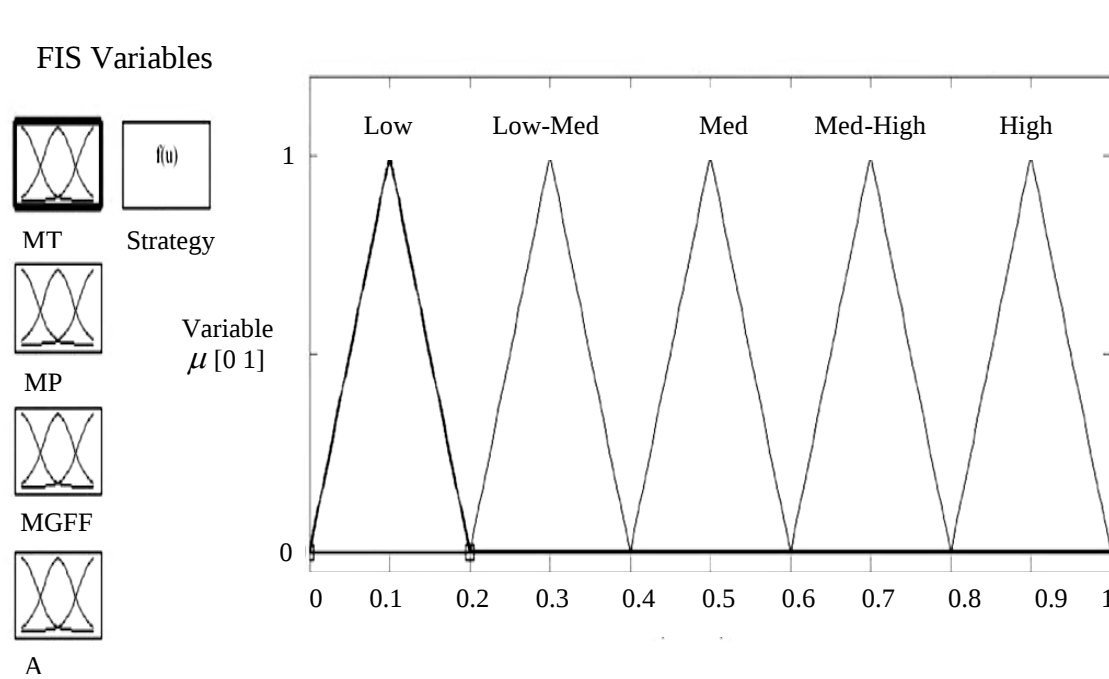
- 1 If (temperature is high) and (GFF is low) then (strategy is S-R) (1)
- 2 If (precipitation is low) then (strategy is S-R) (0.75)
- 3 If (precipitation is low-medium) then (strategy is S-R) (0.75)
- 4 If (precipitation is medium) then (strategy is S-R) (0.75)
- 5 If (precipitation is low) then (strategy is S-R) (0.25)
- 6 If (precipitation is low-medium) then (strategy is S-R) (0.25)
- 7 If (precipitation is medium) then (strategy is S-R) (0.25)
- 8 If (precipitation is medium-high) then (strategy is S-R) (0.25)
- 9 if (altitude is low) then (strategy is S-R) (1)
- 10 If (altitude is low-medium) then (strategy is S-R) (1).

The rules fed into the fuzzy inference system shown in **Fig. 5.4.1**.



FIS is fuzzy inference system, $f(u)$ is the fuzzy union consequent output.

Figure 5.4.1 Design of the fuzzy inference engine to differentiate plant strategies from the water-energy dynamic.



FIS is fuzzy inference system, $f(u)$ is fuzzy union consequent variable

Figure 5.4.2 Definition of triangular membership functions for ordination of plant strategies

As noted in Chapter 3, the discrete values for ‘feet’ and ‘peak’ of the triangular membership function are defined by (3.7.1.2).

Fig. 5.4.3 gives the nodal structure of the T-S-K FLC for the plant strategy environment of Guyana, South America.

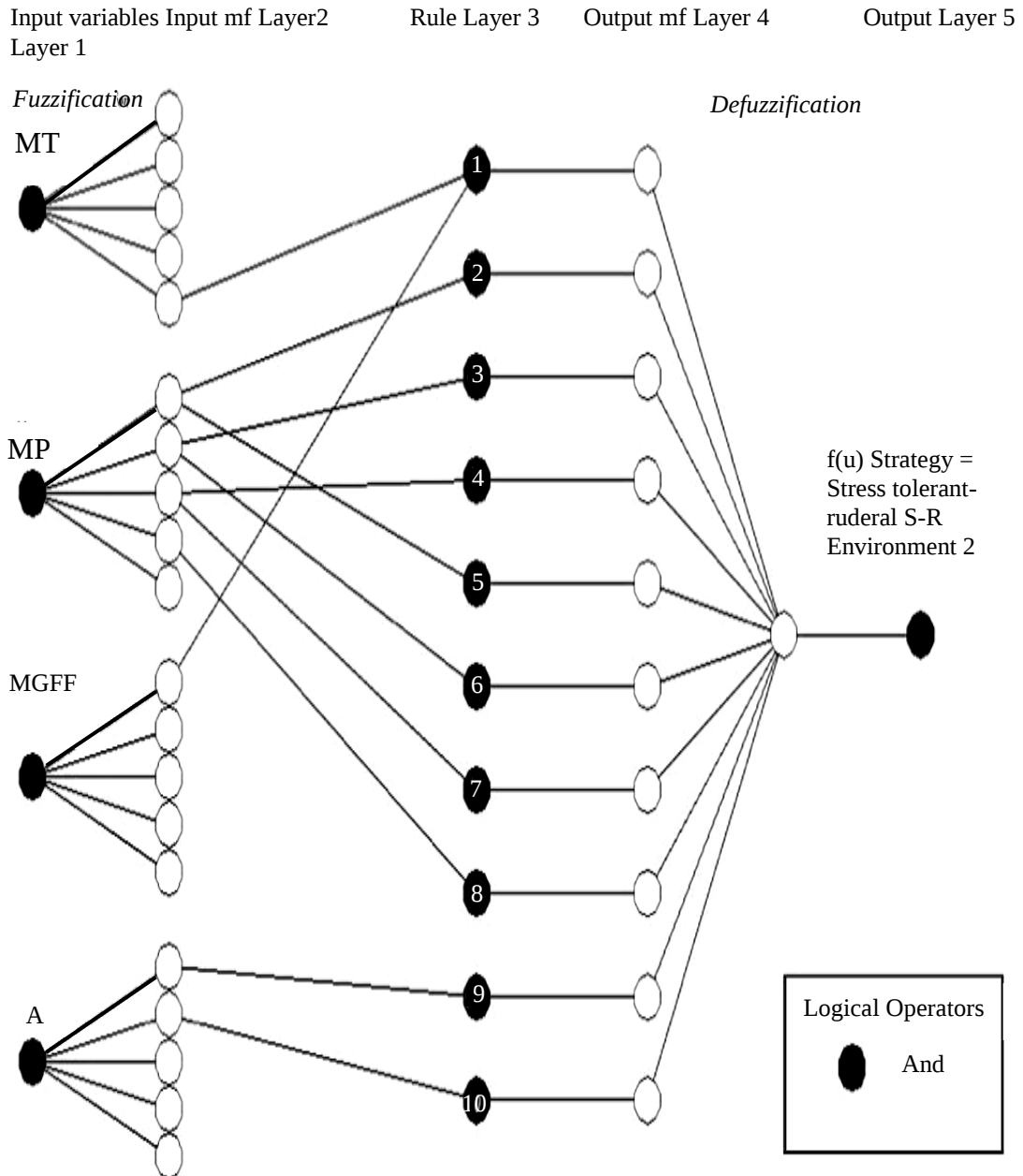


Figure 5.4.3 Node structure of the T-S-K model for differentiation of S-R plant strategy of Guyana, South America

During the fuzzification of Layer 1 the variables were split into their 5 membership values to result in Layer 2. The rules for S-R E2 (Furze et al., 2013b) listed above operated in Layer 3 of **Fig. 5.4.3**. Finally, Layer 4 resulted in the summation of membership values and were defuzzified to give the consequent Layer 5 union of strategy or environment.

The structure of the ANFIS for stress tolerant-ruderal plants is shown in **Fig. 5.4.3**. Within the node structure it follows that the input variables ($A1_{(n)}$, $A2_{(n)}$, ..., $An_{(n)}$) may be stated as vectors $\bar{x} = [x_1, x_2, \dots, x_n]$ and $\bar{y} = [y_1, y_2, \dots, y_n]$. The model was tested by inputting the data

within the ranges of (5.3.1) for Guyana into the engine shown in **Fig. 5.4.1**. The resultant strategy was S-R, the 2nd of the seven possible strategies. The result was also displayed in graphical form in two-dimensional (given in Appendix 4) and three-dimensional formats in order to monitor the efficiency of the algorithm when considering different variables of the dynamic. Combinations of single factors led to incorrect strategy (z) output. It was found that combining any two variables other than mean precipitation and mean temperature resulted in the correct measurement with the 1st (x) variable, but too high a measurement with the 2nd (y) variable. The three-dimensional diagram of **Fig. 5.4.4** shows the maximum efficiency obtained from the algorithm. It is imperative that at least two of the driving dynamic vectors are used – both altitude and ground frost frequency show static responses. This relationship is consistent with previous work (Omizegba and Monikang, 2009), which suggests that transference of variables into vector structure, such as that shown in (5.3.1), may provide a relational index (R) which may also be determined by expert knowledge or alternatively via genetic algorithm (Zadeh, 1973). Such a relational algorithm index may be used to quantify the cooperative link between the number of individuals and climatic modelling structure in different locations, since $\bar{y} = [\bar{x} * R]$.

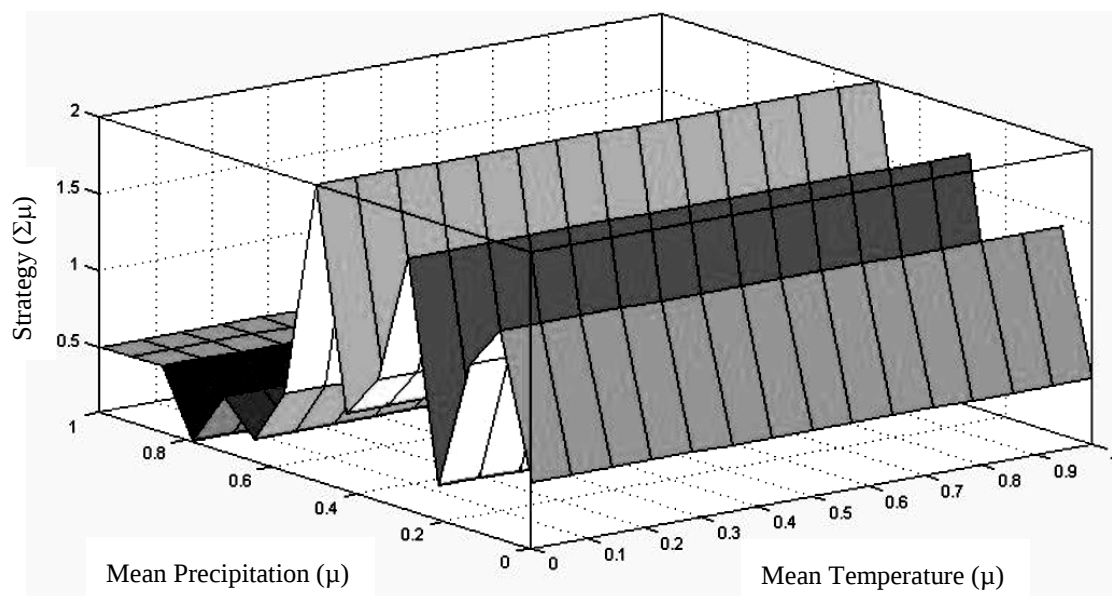


Figure 5.4.4 Three-dimensional surface view for differentiation of plant strategy environment 2

The patterning of plant species presence can be broken down into the seven life-history based strategies as follows: The occurrence of a high water-energy dynamic results in the

highest level of plant species diversity, reflected in the greatest numbers of plant species presence. The highest numbers of plant species presence reflect the dominance of ruderal plant species. Decreasing numbers of individuals in each of the algorithmically described environments reflect the transition through competitive to stress tolerant species, which are present in more extreme environments. The overall advantage of using the methods presented here are finer scale resolution in terms of the modelling framework and the resultant consequent strategy environment. Ultimately greater detail can be inferred within the strategies (environments) as follows:

Environment 1 contains ruderal plants, i.e., herbaceous and bryophyte species. These are of relatively small stature and have limited lateral spread; leaf forms show great variation and are often mesophytic; roots formed are short in length. The canopy layers of these species may be of various types. Each phase of the ruderal plants' life cycle is short. They have a short phase of production in periods of high potential productivity. In moderate to high temperature and water availability, flowers are formed early in the life history and there is a high frequency of flowering. The proportion of production devoted to seeds is high in annual growth cycles. In response to adverse conditions for growth (e.g. presence of competitive species, resource depletion), the plants show a rapid decrease in vegetative growth and re-allocation of resources into flowering. Consequently, photosynthesis may be described as opportunistic: it is optimal in environments with high temperatures and water availability. The total number of individuals present in environment 1 (e.g., Mexico) is very high as inferred in Chapter 4.

Environment 2 contains plants with elements of both ruderal and stress tolerant species, i.e. herbaceous, bryophyte, shrubs and trees. Species and conditions are intermediate between environments 1 and 7; hence, the number of species is consistently high. Temperature and water availability are also high, although there is a moderate intensity of disturbance/stress shown in the example produced by variable altitudes across the location and low ground frost frequency (e.g. Guyana).

Environments 3, 4, 5, and 6 contain either stress tolerant and/or ruderal species, as well as increasing numbers of competitive species (i.e. herbaceous, shrubs and trees). Competitive species have dense canopies of leaves around prominent, rapidly growing shoots. The leaf form is robust and exists in rapidly ascending monolayers. Phases of life cycle may be

short or long, flexibly matching the optimal growth for the environmental conditions. Leaf production is well defined and coincides with periods of maximum potential productivity. Root length is relatively short; flowers are most often produced after periods of maximum productivity on an annual basis. A small proportion of annual production is devoted to seeds. Resource depletion is addressed by great morphogenetic redistribution of leaf and root form. Optimal conditions of photosynthesis align with periods of vegetative growth during wetter and/or warmer seasonal months. Alternative strategies are out-competed given that the environmental conditions become more suited to competitors; only the stress tolerant species survive against competitors in the extreme environments, e.g., *Macedonia* (Kilinç et al., 2010; Furze et al., 2013a).

The novelty of the greater differentiation carried out using defined ranges of T-S-K consequent results enables further mathematic technique to be employed by which we can directly model characteristics of plant species in accordance with the identified pattern of climatic data. Use of the stochastic methods enable discrete patterns to be differentiated as indicated in Chapter 3, sections 3.5, 3.7.2 and 3.8. Such stochastic method enables novel characterisation of ecological systems within fixed time scales as appropriate for the current stage of progress identified within climatic science, thus linking it with the frontier research of biogeography (Lomolino and Heaney, 2004).

5.5 Summary

This chapter details the simulative use of fieldbased data to identify seven environments of plant species occurrence, via a T-S-K fuzzy logic modelling process. Although there are inevitably errors in data collection (e.g. for climatic data height of weather stations in collection of climatic data, inconsistent use of variables by collecting institutes (New et al., 1999); for species data human error in identification of species (Yesson et al., 2007)), the use of logic based mathematics shown in the chapter is devoid of semantic definition so any error or distortion in ordination of the individuals of species occurring in the locations investigated is extremely minimal. Fuzzy logic itself does not allow for distortion or arching of data patterns, which may be present in Boolean methods, hence the resultant union is an accurate portrayal of plant species occurrence with use of sophisticated control theory.

T-S-K modelling systems were first described in the context of plant strategies, after which the use of higher resolution data was detailed and some of the advantages of using these data were described. Fuzzy control algorithms were used to give structure to the spatial categorisation of plant species by integrating digital elevation model data at increased

resolution with data of selected climatic variables (mean precipitation, mean temperature, mean ground frost frequency and elevation). The latter is achieved via construction of rules using statements of variable ranges, and formation of a nodal structure by which control theory is used to resolve plant life-history strategy based environments. In previous studies covered in Chapter 4 the climatic variables have been obtained by minimising them to those essential for expression of the water-energy dynamic (the way in which water and energy are distributed in relation to diversity) in order to model life-history strategies of plant species. In this chapter, T-S-K fuzzy modelling has been applied to infer which factors of the water-energy dynamic may be used for efficient prediction of individual species occurrences in the groups of plant strategies. Additionally, this chapter balances the trade-off between accuracy and interpretability of individual plant species occurrences. In brief, more accurate modelling antecedent data enables more accurate consequent expression. In turn, the interpretability of the modelling process leads to the possibility of further mathematic differentiation of other characteristics of plant species and conditions in which they exist. Implications include modelling of plant life-forms and metabolic patterning, which are further discussed in Chapters 6 and 7.

CHAPTER 6

Stochastic processes to model uncertainty

6.1 Introduction

In Chapter 5 the presence of plant species within 7 nodes of plant life history were successfully predicted using a T-S-K modelling structure. In this chapter the higher mathematic theory of logic based systems progresses to illustrate how strategies may be further differentiated, split into their rudimentary elements, each specifically optimal in different water-energy conditions present in ecological systems. A combination of random dispersal methods are used in this chapter together with the stochastic evolutionary algorithm programming of strategies. This introduction serves as a brief discussion of stochastic methods in the context of biological systems.

The principal area of set theory with relation to development of evolutionary networks is species stochasticity. A stochastic process is generation of random variables, the key point being that evolution of variables is not uni-variate, but may potentially develop in many different shapes. Stochastic networks are well suited in the field of evolutionary algorithms and have extensive use in non-linear system modelling, computer technology and biological systems (Silvera et al., 2009). Just as any stochastic group, plants may be said to be functions of one or several deterministic arguments. To apply stochastic processes, the key point is that the variables determining the measured characteristics or vectors must share the same functional domain. In other words, they may be seen to show certain

probability distributions such as Poisson, Gaussian or other continuous or discrete pattern, as discussed in Chapter 3 and hence they often share complex statistical relationships (Zhang et al., 2012).

Plants show chaotic patterns of evolution in terms of their individual growth processes and plant species numbers (Cui et al., 2012; Furze et al., 2011; Su et al., 2009). Patterning of plant species may be determined by key factors of the water-energy dynamic (Hawkins et al., 2003; Kreft and Jetz, 2007; Sommer et al., 2010). Climatic variables such as rainfall and temperature often show discrete patterns across timescales, so are often made use of within fixed time ranges. As such, they can be said to be discrete stochastic patterns, which facilitate macro-level modelling of plants (Grime and Pierce, 2012; New et al., 1999; Silvert, 2000).

Gaussian functions, or normally deviated variables, are those that may be plotted with a bell-shaped curve or parabola; they are deeply rooted in probabilistic and statistical theory and have been made use of by many authors for patterning of natural or artificially created variables (Herrera, 2005). In particular, Lotfi Zadeh made use of Gaussian distribution functions in the inception of fuzzy techniques. The effect of the Gaussian pattern is that it can normalise a distorted statistical view of a variable or individual vector. The derivative of Gaussian or normal deviation is the central sample of the variation, which is proportional to the standard deviation or greatest incline of the population range of variables under consideration. Gaussian functions arise by applying the exponential function of variables to a general quadratic function; the key premise is hence the natural logarithm of the system's total. The latter makes Gaussian distribution a central function to minimise the error within Boolean statistics. As such, it may be seen to be a central premise of higher (logic based) mathematics of great application in the analysis of complex systems through inferential/differential equations (Angelov and Feliv, 2004; Zadeh, 1973; Zhao, 2012). Furthermore, the normal deviation can be estimated most effectively by an iterative estimation of differential weighted least squares until the tail ends of the bell curve are produced (Guo, 2011). Gaussian distribution is related to other functional spreads of variation such as Poisson distribution.

Evolutionary computing methods include genetic algorithms (Broekhoven et al., 2007; Zadeh, 1973). These algorithms make use of representation of the components of variation

as vectors (chromosomes) within strings (populations). The chromosomes recombine in an iterative process under specified conditions involving the same elements of genetic recombination as in natural systems. As in biological systems, operators are selection, crossover and mutation. In combination with fuzzy set theory, GAs are robust, stochastic evolutionary computational algorithms (Su et al., 2009). GAs are adaptive algorithms for finding the best (global) solution to optimisation problems.

A natural progression of GA and fuzzy systems is the estimation of a population's resultant distribution (introduced in Chapter 3, section 3.5); which forms an additional method of approximation: Estimation of Distribution Algorithms (EDAs) are algorithms with applications in computing, industry and natural systems. Copula theory is a concise, robust form of EDAs (Nelson, 2006). Copulas join multivariate distributions to one-dimensional marginal distribution functions, which make them ideal candidates for the analysis of patterns of genetic variation such as that which is produced by the multi objective optimisation (MOO) or Pareto front of a multi objective genetic algorithm (MOGA) process (Wang et al., 2012).

The objective of this chapter is to show proportions of plant strategies nested within fixed population sizes, which have been determined by a genetic (generational) algorithm fuzzy rule base. Subsequently the Pareto front for plant strategies is estimated, with further exploration into the utopia hyperplane (Erfani and Utyuzhnikov, 2011) of objective space for plant strategies via the construction of code (given in Appendix 5). The aim of this chapter is to show a minimised (efficient) modelling framework. Techniques employed within this chapter increase the accuracy of both data sources of plant variation employed and climatic modelling data using the novel method.

The remainder of this chapter is structured as follows: Section 6.2 gives a brief summary of the method used to form the T-S-K FL system of individual plant species occurrence of E6, the dynamic of Azerbaijan (identified as an E6 candidate) to form the algorithm. The robust nodal structure used to form the T-S-K model and algorithmic detail is given. Section 6.3 integrates genetic programming methods with fuzzy technique and the MOGA process is expanded to show examples of resulting data detected within elements of plant strategies. Section 6.4 gives further metabolic application of the MOGA process and expands linear and quadratic rules used in formation of elementally dispersed variables.

Finally, section 6.5 provides a summary of the information covered in the chapter.

6.2 Fuzzy-genetic programming of plant life-history strategies

Firstly, we must recall that the development of rule-based (eR) systems (FRBS) are hybrid systems which may provide knowledge guidance (Chapter 2, section 2.6). They use a summary of the information, in terms of the potential of a new data sample (such as accumulated spatial proximity information), to trigger the new rule base. Greater generality of the structural changes to the data can, therefore, be catered for (Angelov and Filev, 2004). In terms of plant strategy estimation, we may state the membership functions in terms of the seven environments (as demonstrated by Barreto (2008)) and obtain the data for the placement of species within each by considering optimisation limits, given that the initial number of species is known (Broekhoven et al., 2007). This method was covered in depth in Chapters 3, 4 and 5. Using the method, we may obtain stochastic matrices, effectively K_p or rK continuums. In combination with a generational algorithm approach, the rule base may be successfully used to generate the distribution present in natural systems (Elith et al., 2011). A ten-point summary for the combined methods is included here, representing a novel form of the scheme outlined in Fig. 3.8:

1. Determine structure via identification of complete sets of background/environmental data.
2. Derive minimal TSK IF–THEN rule base for strategical nodes of network organisation.
3. Identify elements of strategies (node structure).
4. Use expert-based intuition to rank ideal solutions of each element, represent within chromosome population structure.
5. Identify constraints (stopping conditions), key objectives and total number (maximum) individuals/generations for generation of ideal solution combination (MOO) to achieve the Z utopia hyperplane.
6. Generate random population of chromosomes and operate random selection of chromosomal values via MOGA.
7. Allow algorithm to run until sets of ideal solutions have spaced in objective (Z) space – this determines the Pareto Frontier.
8. Identify distribution of Ideal Pareto Frontier, fit linear, quadratic and polynomial functions as required to precisely describe relative distribution of chromosomes

within maximum number of individuals. The linear fit of the Pareto frontier is defined as the Utopia Line (The Utopia objective space is represented by the error of each individual Pareto solution; this may be determined by binary conversion of each Pareto (P) generating set, equal to 0-1).

9. Approximation of the distribution may determine real proportions of strategies within strategical nodes (stochastic distribution in this case).
10. Apply solutions to ranges identified through T-S-K FRBS, additionally structure may be checked via pattern identification (Angelov and Feliv, 2004; Juang and Hseih, 2012; Salah and Abdalla, 2011).

This chapter demonstrates the occurrence of the competitive-stress tolerant plant life-history strategy in the location of E6, Azerbaijan, Southern Europe, which falls between latitude 42.14 degrees north, 38.33 degrees south and longitude 50.53 degrees east, 44.78 degrees west. Coordinates of the location were obtained using United States Geological Survey (USGS) data and extracted using Matlab (Version 2010a©). Exemplified data is given in **Figs.** 6.2.1, 6.2.2 and 6.3.3.

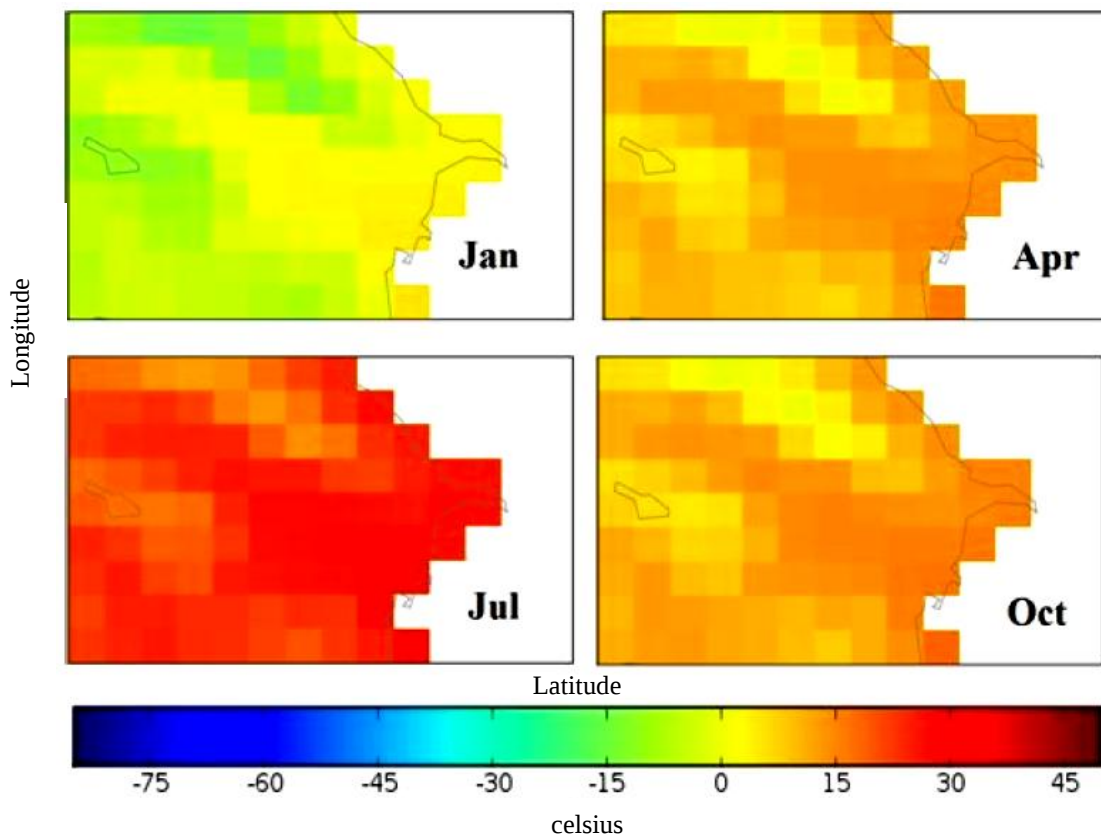


Figure 6.2.1 Azerbaijan energy (mean temperature 1961–1990) data, 18.5 km resolution

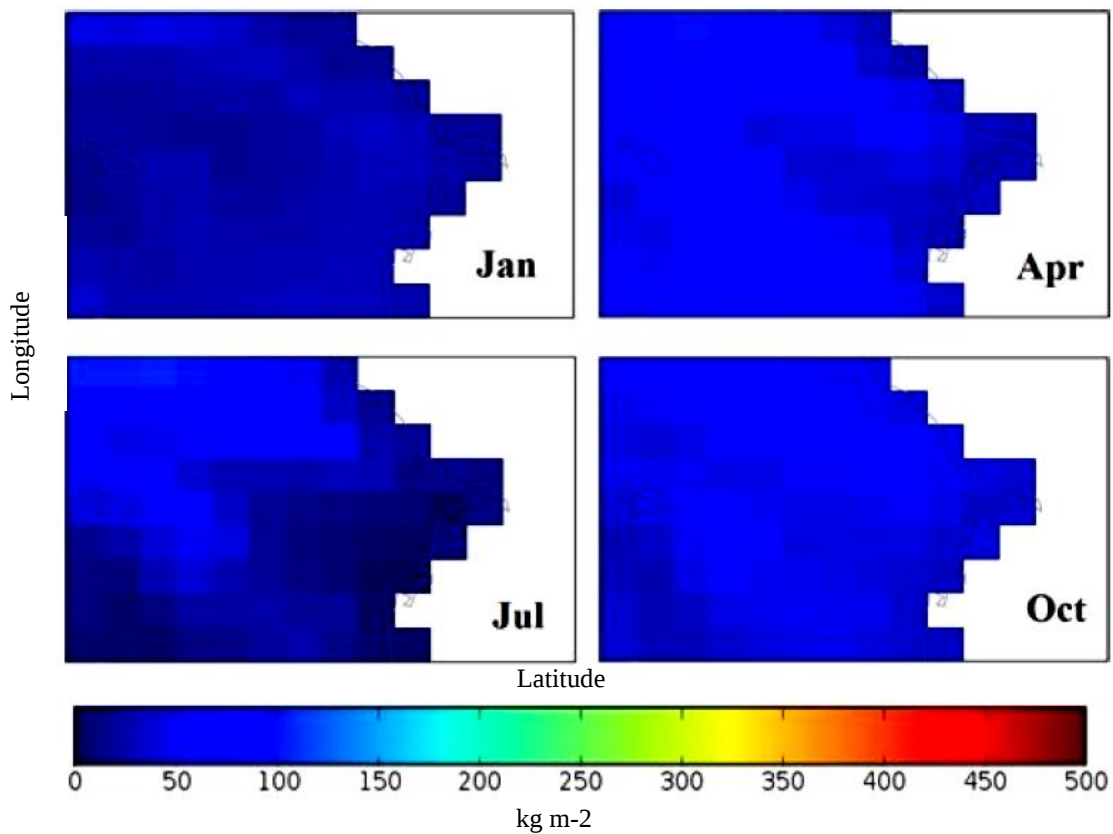


Figure 6.2.2 Azerbaijan water (mean precipitation 1961–1990) data, 18.5 km resolution

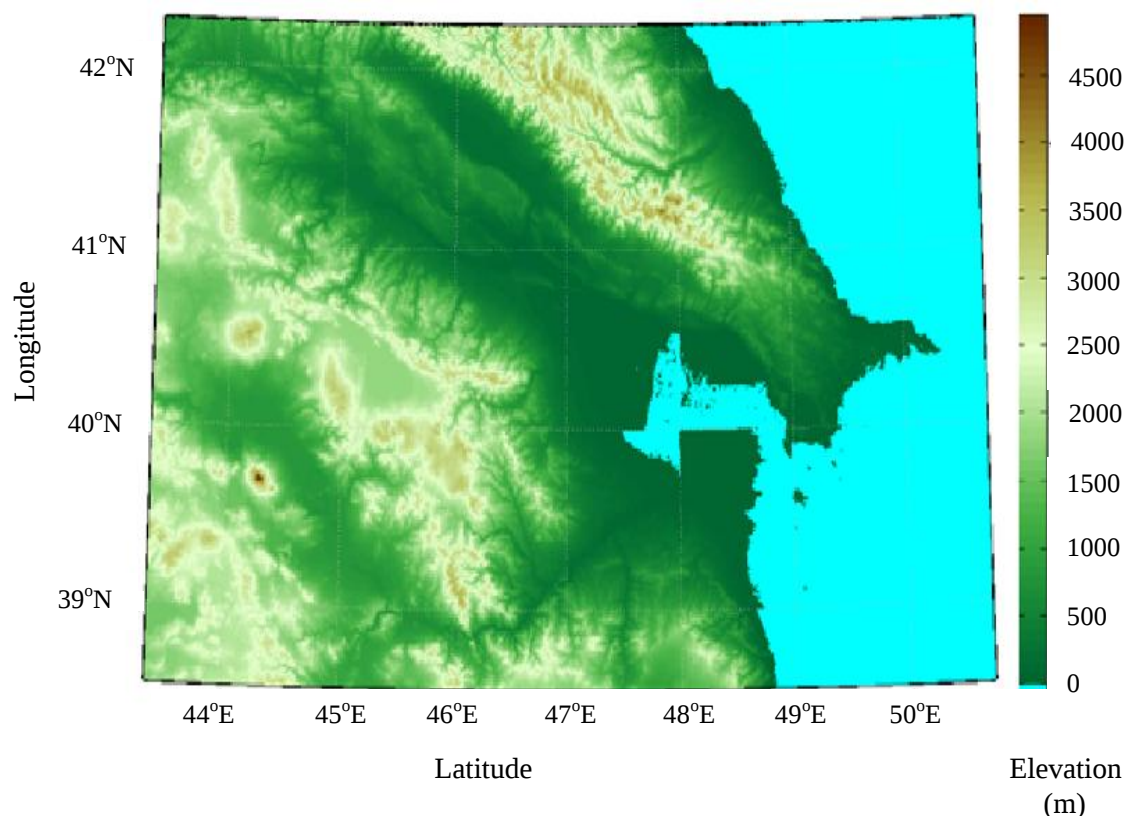


Figure 6.2.3 Area of Azerbaijan, Europe DEM data, 1km resolution

Azerbaijan quarterly mean temperature (1961–1990) at 18.5 km resolution is shown in **Fig. 6.2.1**, characteristic of the energy component of the water-energy dynamic used to form the algorithmic framework for modelling of strategies. Azerbaijan quarterly mean precipitation (1961–1990) is shown in **Fig. 6.2.2**. Data of **Figs 6.2.1** and **6.2.2** were originally recorded by New et al. (1999) and have been used by the Inter Governmental Panel on Climate Change (IPCC) for the use of climate modelling. They were made use of in this chapter after having been shown to be characteristic of the water-energy dynamic (Sommer et al., 2010).

The DEM data of GTOPO30, published through the United States Geological Survey (USGS), is shown graphically in **Fig. 6.2.3**, being at 30 s (1 km resolution). Data were processed using Matlab (Version R2010a ©). Code was constructed for mapping and image-processing sections of Matlab and is available in Appendices 2.1 and 2.1.1. The legends of the figures were used to obtain a unit percentage as detailed in Chapter 5 and subsequently the figures were summarised into ANFIS as shown in the following subsection.

6.2.1 Nodal structure and algorithmic detail

In layer 1 of the ANFIS procedure, input variables were estimated and quantified from the data sources (Intergovernmental Panel on Climate Change for climate variables, United States Geological Survey for altitude). The original data were partitioned into 5 linear ranges across 100% for the purposes of the algorithmic statement, seen to optimize model efficiency and accuracy whilst removing unnecessary data redundancy (Alcala et al., 2007). In the case of mean temperature the ranges were 0–20%, Low ($A1_{(1)}$); 20–40%, Low-Medium ($A1_{(2)}$); 40–60%, Medium ($A1_{(3)}$); 60–80%, Medium-High ($A1_{(4)}$) and 80–100%, High ($A1_{(5)}$). Mean precipitation, mean ground-frost frequency and altitude were all quantified in the same way. Weights were finally added to the variable ranges to make an accurate inference of the plant strategy present in Azerbaijan (E6). Layer 2 transferred the variable ranges through 5 equally partitioned triangular functions between 0 and 1, in the fuzzification process.

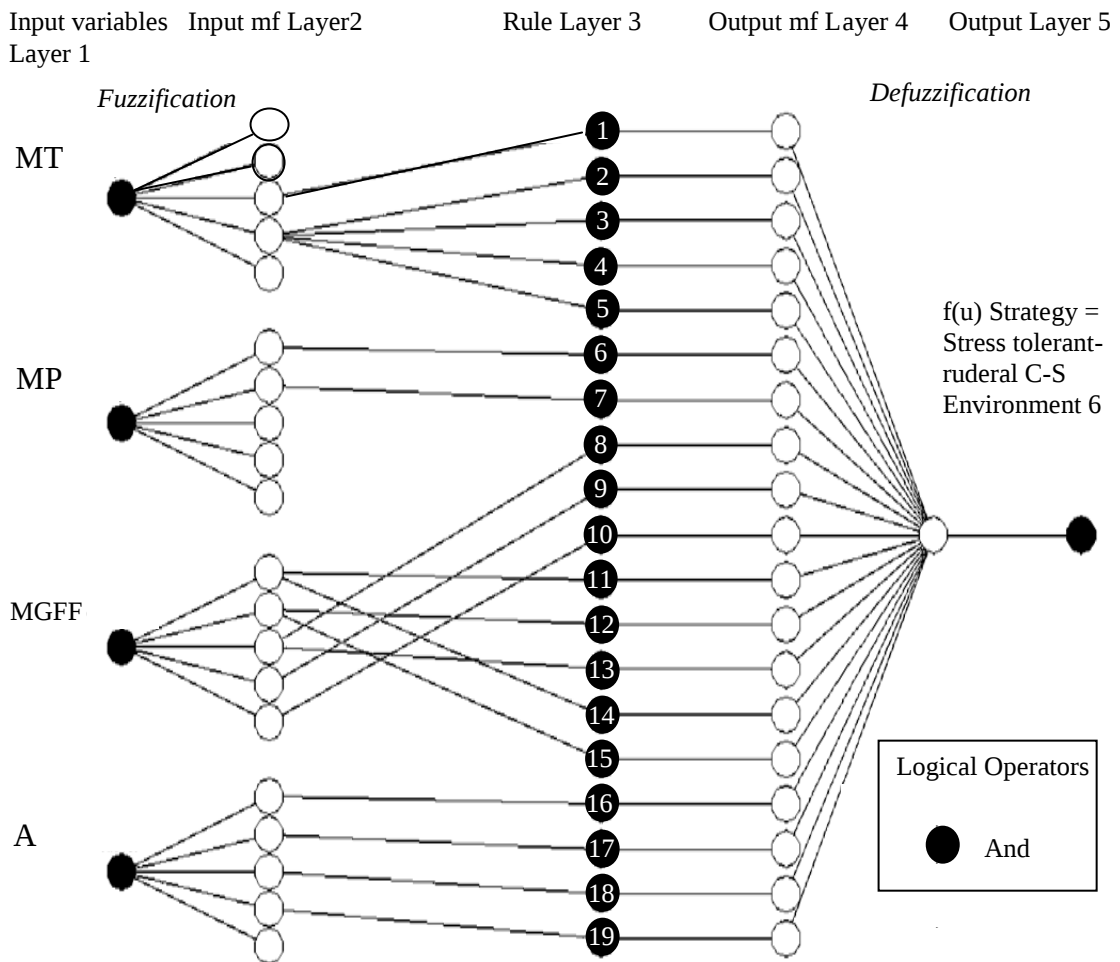


Figure 6.2.1.1 T-S-K Nodal structure of Competitive-Stress tolerant strategy of Azerbaijan, E6

The fuzzy ranges were separated in layer 3 according to the rules applied (19, listed in Appendix 5. 1, in the case of E6 shown here). Layer 4 returned fuzzy output depending on the combined distributed range as instructed. Finally, layer 5 gave the crisp output, sourced from the Global Biodiversity Information Facility.

The structure of **Fig. 6.2.1.1** is similar to that of **Fig. 5.4.3**, however the former contains additional rules and alternative quantification of the variables as shown in (6.2.1.1) below.

The T-S-K nodal structure of **Fig. 6.2.1.1** resulted in the following control algorithm for E6:

$$\begin{aligned} & \text{If } 0.25A1_{(3)} \prec A1_{(4)} 0.5A1_{(4)} 0.25A1_{(4)} \prec A1_{(5)} \wedge A2_{(1)} \prec A2_{(2)} \wedge 0.25A3_{(3)} \\ & \prec A3_{(5)} 0.5A3_{(1)} \prec A3_{(3)} 0.25A3_{(1)} \prec A3_{(2)} \wedge A4_{(1)} \prec A4_{(4)} \text{ Then } B_{(8805)} = E6 \end{aligned} \quad (6.2.1.1)$$

where the consequent number of individual plant species occurrences was 8805.

In fact, it was seen that the membership values of input variables mean temperature and mean precipitation were sufficient to instruct the occurrence of the life-history strategy (C-S) of E6. The maximum efficiency of the algorithm is displayed below.

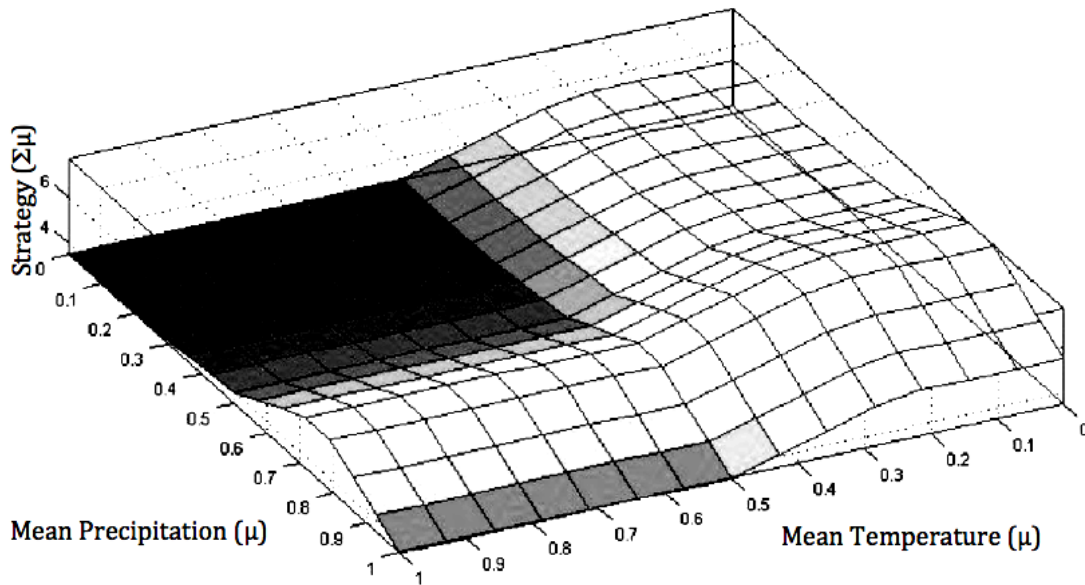


Figure 6.2.1.2 Three dimensional surface view of algorithmic control for differentiation of plant strategy environment 6

Areas of membership values of each input variable are seen at their most efficient values in the lighter (non-shaded) parts of the graph, particularly where Mean Temperature membership was <0.4 and Mean Precipitation was in the range >0.7 to 0.9 . One may estimate the distribution of plant species elements to be dispersed in a Gumbel copula within the water energy dynamic, which is in agreement with previous climatic studies (Schölzel and Friedrichs, 2007). This distribution is further investigated in the remaining sections of this chapter and in Chapter 7.

The following section proceeds to elaborate the hybridization process of fuzzy-genetic method, which was used to disperse elements of the algorithmically mapped characteristic / environment.

6.3 Fuzzy – genetic programming hybrid methodology

Global optimisation techniques are often used to calculate the dispersal of elements within objective space following application of the modelling technique (ANFIS). Genetic algorithms in particular calculate the optimal configuration of elements of characters given set objective values (see Chapter 2, sub-section 2.5.2 and Chapter 3, section 3.8 for further detail).

The process of multi-objective genetic algorithm involved the following main steps:

- 1 Define each vector for plant strategies.
- 2 Randomly generate an initial population of 20 solutions (chromosomes).
- 3 Evaluate each solution according to how well it fits into the desired environment (as defined in equation (6.2.1.1)).
- 4 Select chromosomes randomly (tournament selection). Keep those with the highest fitness function to improve the population and discard those with too low (value may be previously calculated) fitness.
- 5 Create new chromosomes by crossing selected solutions using crossover of proportions of the individual strings of solutions.
- 6 Mutate a previously determined proportion of the population's chromosomes.
- 7 Go back to step 3 until final population number is reached, then stop.

Code was constructed in Matlab for 4 sets of variables within defined parameters, variables

were mean temperature, mean precipitation, mean ground frost frequency and altitude). The variables were directed through two main objective types (water and energy) and expressed as a double vector population in Matlab. The algorithm was programmed to stop when the total number of individuals had reached that of C-S, E6. Code for the operation is available in Appendix 5.

The number of iterations required was recorded and the distribution of ideal solutions for each of the 20 chromosomes across objective space was noted. Linear and quadratic lines were fitted to the resulting Pareto in order that an approximation of the distribution be made. A plot was made and the position of the linear fit (representing the Utopia line) and the quadratic shape of the curve visualised in order that the evolution of the 20 characters could be estimated through future transgressions. This is given further discussion in Chapter 3, sub-section 3.8.1, section 6.4 and subsection 6.4.1.

6.3.1 Multi-objective genetic algorithm dispersal of plant strategies

In order to apply the multi objective genetic algorithm on the defined plant population, it was necessary to identify and quantify ideal solutions of the elements of strategies (Zadeh, 1973). In this section, elements of plant strategies (Grime et al., 1995) are detailed, defined and given quantifiable values.

In **Tab. 6.3.1**, the elements of plant strategies are: PT = plant type, sm = shoot morphology, lf = leaf form, c = canopy, loep = length of established phase, lor = lifetime of roots, lp = leaf phenology, rop = reproductive organ phenology, ff = flowering frequency, poaps = proportion of annual production for seeds, podup = perennating organs during unfavourable periods, rs = regenerative strategy, mpgr = mean potential growth rate, rrd = response to resource depletion, pumn = photosynthetic uptake of mineral nutrients, ac = acclimation capacity, sop = storage of photosynthates, lc = litter characteristic, psh = palatability to non-specific herbivores and nDNA = nuclear DNA amount. Ideal quantification is seen in brackets.

Table 6.3.1.1 Solutions and ranges for plant strategy chromosomes (Furze et al., 2012a)

Character/ Chromosome	Competitive (1,...,5)	Stress Tolerating (1,...,5)	Ruderal (1,...,5)
PT	Herbs, shrubs, trees (1,2,3)	Lichens, bryophytes, herbs, shrubs and trees (4,5,1,2,3)	Herbs, Bryophytes (1,5)
sm	Long with extensive above and below ground (3)	Long, short and intermediate (3,1 and 2)	Short stem, limited lateral spread (1)
lf	Robust, large often require high water (5)	Small (1), leatherly (1) or needle-like (2) low water requirement (1) (1,1 or 2,1)	Variable, often require high water (1,...,5)
c	Rapid upward growth of one layer (5,1)	Multi layered (5) if mono (1) layered, slow (1) upward growth (1,5,1)	Variable (1,...,5)
loep	Long or relatively short (4 or 2)	Long to very long ($4 \leq 5$)	Variable (1,...,5)
lor	Relatively short (2,3)	Long (4)	Very short (0,1)
lp	Well defined peaks of leaf production coincides with periods of maximum productivity ($5 \equiv 5$)	Short phase of production within period of high productivity (1,...,4)	Evergreen, with various patterns of leaf generation ($5 \equiv 1 \leq 5$)
rop	Flowers produced after periods of maximum productivity (1,...,5)	No relationship between productivity and flowering time (1,...,5 $\neq$ 1,...,5)	Flowers produced early in life-history (often before maximum growth) ($5 = 1$)
ff	Established plants flower every year (5)	Flowering intermittently over a long life-history (4)	High frequency of flowering (> 1 a year)
poaps	Small (1)	Small (1)	Large (4)
podup	Buds and seeds 1,...,5)	None (0)	Seeds (5)
rs	Vegetative (1), seasonal regeneration in gaps (1), wind dispersal of small seeds (1), persistent seed bank (5) (1,1,1,5)	Vegetative, wind dispersal of small seeds, persistent juvenile bank (1,1,5)	Seasonal regeneration in gaps (2), wind dispersal of small seeds (1), persistent juvenile bank (5)
mpgr	High (5)	Low (1)	High (5)
rrd	Rapid morphogenetic responses in form and distribution of leaves and roots (5,5)	Slow, small morphogenetic responses (1,1)	Rapid cessation of vegetative growth and reallocation of resources into flowering ($5 \equiv 1$)
rumn	Strongly seasonal coinciding with long continuous period of vegetative growth ($5 = 1$)	Opportunistic, uncoupled from vegetative growth ($1 \neq 1$)	Opportunistic, coinciding with vegetative growth ($3 = 1$)
ac	Weak (0,...,1)	Strong (5)	Weak (0,...,1)
sop	Rapid incorporation in vegetative structure and compartmentalized storage for growth in next season (1)	Storage in leaves, stems, or both (2,3 or 5)	Seeds (4)
lc	High volume (5), non-persistent (0)	Sparse (2), persistent (5)	Sparse (2), non-persistent (0)
psh	Various (0,...,5)	Low (1)	High (5)
nDNA	Small (1)	Small (1) and high (5)	Small (1) to very small ($0, \dots \leq 1$)

The elements identified in **Tab. 6.3.1.1** represented a chromosomal population and were used to form a MOGA, in which the chromosomes cycled (via roulette wheel selection)

through 1–5, resulting in a Pareto front. The process was carried out in order that estimation of the numerical distribution of each plant strategy in the previously algorithmically defined environments could take place. In the following section the resulting solutions are plotted, and the MOGA process is applied to plant metabolic (photosynthetic) types.

6.4 Multi-objective dispersal of plant strategies and metabolism

In this section the objective dispersal of the twenty elements detailed in **Tab. 6.3.1.1** is shown. Linear weighted least squares and quadratic expressions (with corresponding errors) summarise the distributions as shown in **Fig. 6.4.1**.

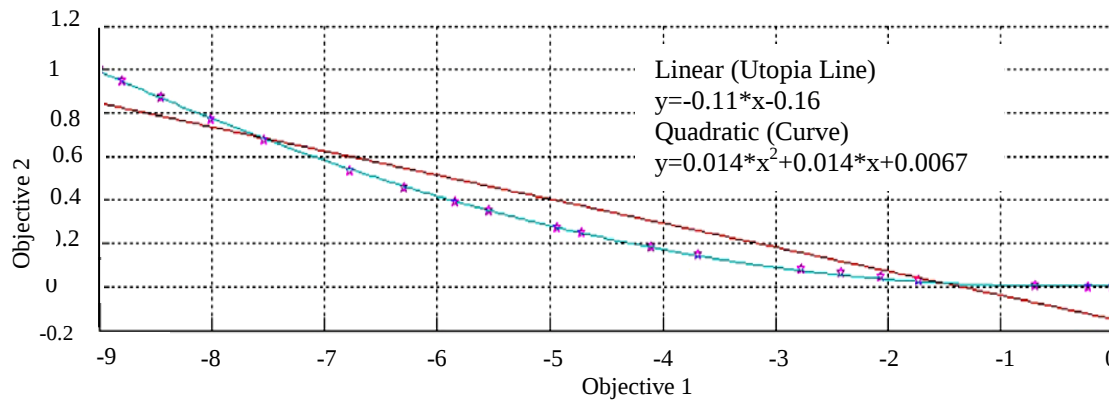


Figure 6.4.1 dispersal of twenty elements of plant strategies dispersed over the simulated water-energy dynamic of Azerbaijan

In **Fig. 6.4.1** the axis ‘objective 1’ (mean temperature, x) and ‘objective 2’ (mean precipitation, y) are n objective functions which may be expressed as Z in the following:

$$Z \subseteq R^n \quad (6.4.1)$$

Where vectors of Z (described in **Tab. 6.3.1.1**), are seen to be within the correlation (Relational) matrix multiplied by the number of objectives. **Fig. 6.4.1** represents the graphical dispersion / simulation via genetic programming of the T-S-K FIS summary expression of (6.2.1.1). Expansion of the rudiments of plant life history strategies across the numerical (Z) hyperplane is elucidated. This may be translated to geographical and

species characterisation via differential expressions. Expressions for the utopia line and curve are given in **Fig. 6.4.1**, and these are discussed in sub-section 6.4.1. From (6.4.1) it follows that there was a design variable (D) formed from the MOO (Erfani and Utyuzhnikov, 2011), being the numerical estimate of the utopian (Z) space. The following expression is used as a generic form of the optimisation:

$$\text{Min}F = \{F1(x), F2(x), \dots, Fn(x), \text{ subject to } x \in D \quad (6.4.2)$$

Where F represents different optimisation solutions (or functions in the case of $F1, F2, Fn$, with x representing set values of individuals). The residual error of the quadratic utopia curve and of D in (6.4.2) was seen to be 0.02717. The optimisation may be expressed across a given total number of individuals following a Poisson distribution with use of quadratic or varying degrees of polynomial expressions with greater differentiation to express greater numbers of categories within the optimisation (Zhao, 2012). We may approximate a similar distribution with a lesser number of grouped characters (such as strategies, or photosynthetic metabolism type) or a greater number of characters such as life-forms owing to expanding/contracting dimensional relations given through fuzzy and interval type 2 fuzzy inference systems (Raunkier, 1934; Zadeh, 1965; Zhao, 2012).

The beauty of the MOGA Strength Pareto shown in **Fig. 6.4.1** is that one may substitute any value of objective 1, temperature, and return a value for objective 2; these values being subject to the parameters entered (population size, number of characters to be dispersed). This novel method may, therefore, be used to identify accurate prediction of temperature or precipitation given a finite population number.

The efficient prediction of the presence of photosynthetic groups within the set population of individual plant species occurrences of E3, Cuba is carried out by a further reduced form of T-S-K FL (seen in Appendix 5.2). The reduced variables considered are the ‘driving variables’ (discussed in section 5.4) of Mean Temperature and Mean Precipitation. In this section, proportions of plant metabolism characteristics nested within the fixed population size, which has been determined by a hybrid genetic (generational) algorithm fuzzy rule base, were investigated.

Characteristics of plant species related to each type of photosynthesis are defined

according to Chapter 3, section 3.6. This section proceeds to qualify and quantify different characters shown in individuals of each photosynthetic type.

Table 6.4.2 Photosynthetic characters and quantification solutions for MOGA cycle

Character / Chromosome	C3 (1,...,5)	C4 (1,...,5)	CAM (1,...,5)
ws	(3,...,5)	(0,...,2.5)	(1,...,2)
fs	(1,...,2.5)	(2.5,...,5)	(2.5,...,5)
hs	(4,...,5)	(2.5,...,4)	(0,...,1)
fl	(0,...,2.5)	(2.5,...,5)	(2.5,...,5)
tl	(2.5,...,5)	(2.5,...,4.5)	(0,...,2.5)
nll	(0,...,2.5)	(0,...,1)	(0)
tr	(0,...,2)	(2,...,5)	(0,...,2.5)
crb	(0,...,3)	(0,...,3)	(0,...,2.5)
drb	(3.5,...,5)	(0,...,2.5)	(0,...,2.5)
gpspod	(5)	(5)	(0)
gpspon	(0)	(0)	(4,...,5)
pka	(0)	(5)	(1)
ppep	(0)	(5)	(0,...,2.5)
sspc	(0,...,2)	(2,...,5)	(2,...,5)
tspc	(0)	(0,...,2)	(5)
sac	(0,...,2.5)	(2.5,...,5)	(5)
soc	(5)	(3)	(0,...,2.5)
c	(0,...,2.5)	(3)	(3)
s	(0,...,2.5)	(2.5)	(5)
r	(5)	(2.5)	(2)

Plant characters in **Tab. 6.4.2** are: ws = woody stem, fs = fleshy stem, hs = hairy stem, fl = fleshy leaves, tl = thin leaves, nll = needle-like leaves, tr = tap root, crb = compact root ball, drb = dispersed root ball, gpspod = greater proportion of stomatal pores open during day, gpspon = greater proportion of stomatal pores open during night, pka = presence of Kranz anatomy, ppep = presence of phosphoenol pyruvate, sspc = spatial separation of

photosynthetic compounds, tspc = temporal separation of photosynthetic compounds, sac = storage of acidic compounds, soc = storage of carbohydrate, c = competitor, s = stress tolerant, r = ruderal. Ideal quantification is shown in brackets in the table.

In the following subsection the dispersal of photosynthetic elements identified in **Tab. 6.4.2** are shown and subsequently construction of combined objective (utopia) rules are detailed.

6.4.1 Linear and quadratic rules of multi-objective dispersed elements

In this penultimate part of the chapter an example of the dispersal of 20 elements within the water energy dynamic is shown, by exemplifying the plot of elements of photosynthesis for the individual plant occurrence data of Cuba (E3). The variance of the linear weighted least squares (utopia line) and quadratic (utopia curve) plots may be obtained as mentioned in the last subsection as shown in **Fig. 6.4.1.1**.

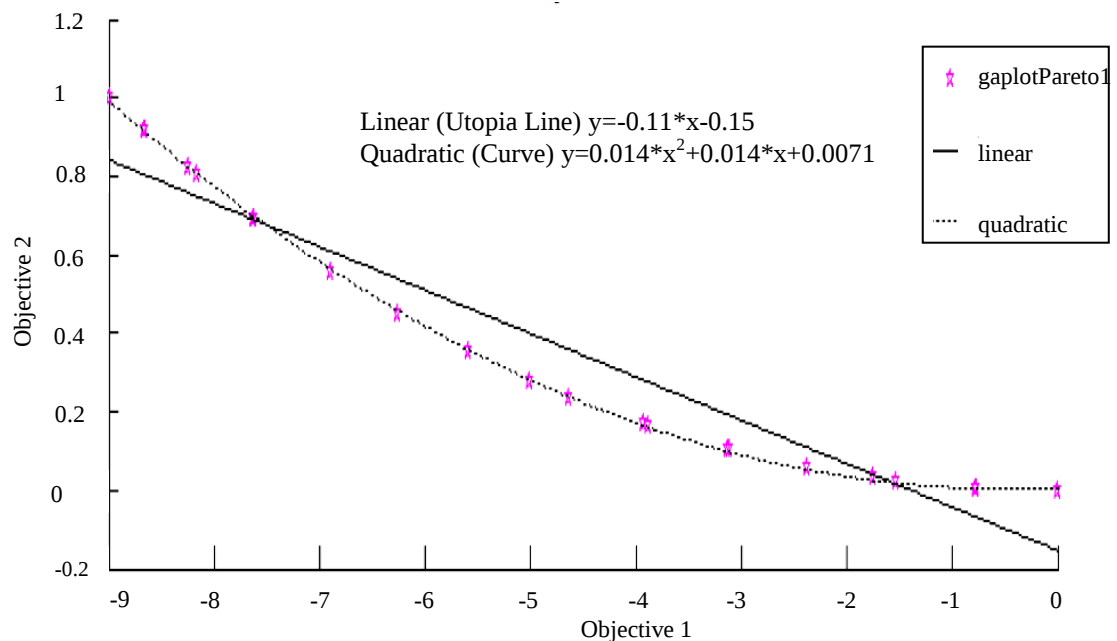


Figure 6.4.1.1 Plant photosynthetic evolutionary strength Pareto of Cuba

The evolutionary strength Pareto plot of elements of photosynthesis present in the species

of $E3$, Cuba, enabled estimation of the distribution of photosynthetic characteristics to be carried out, in accordance with the W-E dynamic as summarized in the utopia line and utopia curve expressions given in **Fig. 6.4.1**. As discussed in the previous section, these expressions are subject to the variance seen within the range of points plotted.

The residual variances of these expressions were also plotted as seen in Appendix 5.3 and in **Fig. 6.4.1.2**.

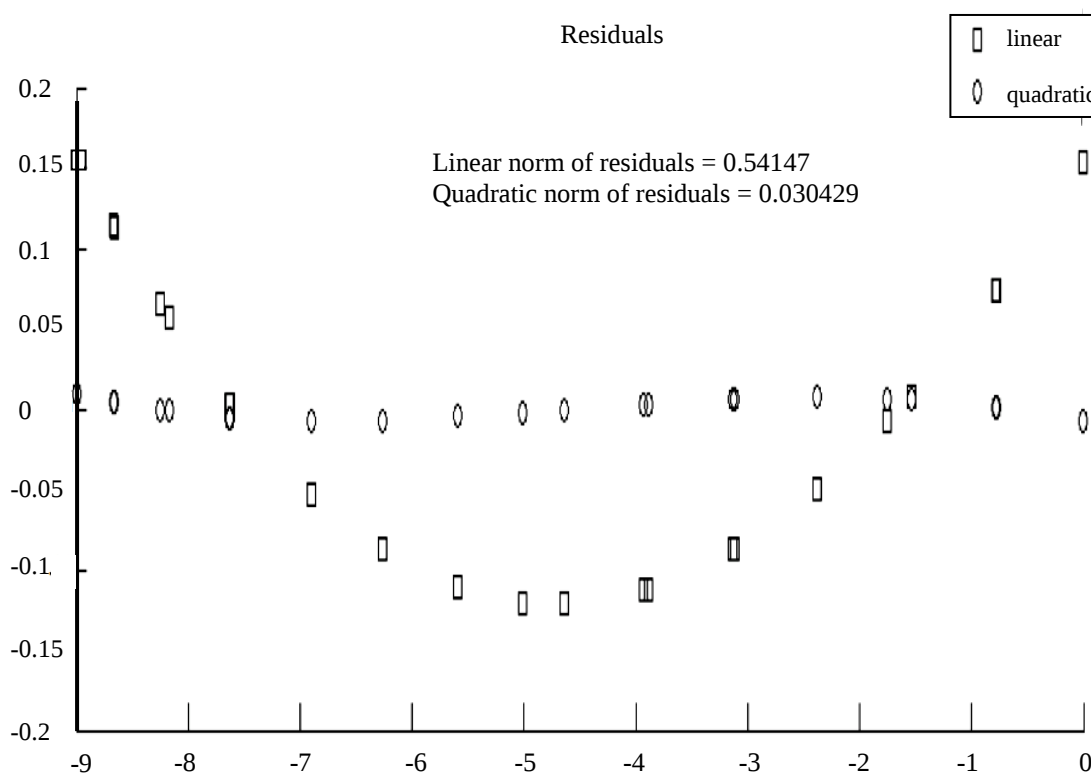


Figure 6.4.1.2 Residual error of the utopian space of plant photosynthetic characters

Plotting the residual error of utopian space enables the formation of rules for the distribution of photosynthetic elements to be formed as shown in table 6.4.1.1.

Table 6.4.1.1 Utopia rules of photosynthetic character rudimental dispersal

Rule	Variables (3 significant figures)			
	$\partial 1$	$\partial 2$	$\partial 3$	ε
1. $Z = \partial x + \partial 2 \pm \varepsilon$	-0.111	-0.154		0.541
2. $Z = \partial x^2 + \partial 2x + \partial 3 \pm \varepsilon$	0.014	0.014	0.007	0.03

In rule 1 of **Tab.** 6.4.1.1, y (objective 2, mean precipitation) is equal to an element of Z (given that x (objective 1, mean temperature) and y axis form the Z space), ∂_1 is -0.11061 and ∂_2 is -0.15387 following a weighted least squares structure subject to the error ($\varepsilon =$

0.54147). In rule 2, y (objective 2) is equal to an element of Z (given that x (objective 1, mean temperature) and y axis form the Z space), ∂_1 is 0.013653, ∂_2 is 0.013537, ∂_3 is 0.0071417, subject to the error ($\varepsilon = 0.030429$).

Construction of the above rules from linear and quadratic expressions enables estimation of each objective to be made given a fixed population number, as identified for each T-S-K algorithmically described location and character; hence the hybrid MOGA techniques enable us to predict climatic data and additionally species numbers subject to errors such as those detailed in **Tab.** 6.4.1.1. This point not only enables us to remove uncertainty in prediction of individual plant species occurrences, but also to enhance the knowledge bases on which the modelling system is founded (Furze et al., 2013a). Expression of the W-E dynamic in quintile terms together with quantification of plant characteristics in the same way allows the cyclical nature of the MOGA to produce optimal distribution of individual traits within the dynamic conditions. The benefit of carrying out the technique is that the likelihood of individual characteristics occurring in ecology / field conditions may be approximated within concise conditions. Knowledge of the characters being dispersed is very important, as it is this which supplies the root of intuition to infer which elements of a character are present given conditions of the dynamic in which they are dispersed. A major advantage of mathematics in this chapter is that it can be used to give highly accurate indications of both plant species presence / characteristics and given plant species occurrence, the conditions of the W-E dynamic are stated with greater accuracy. The novel application of these techniques provides time and financial savings for ecologists and biologists in otherwise purely field based studies.

Plots for dispersed elements of strategies, such as those detailed in **Tab.** 6.3.1, and plant life-forms (Chapters 3 and 7) are shown in Appendix 6 and Appendix 6.1 respectively. In the following section a summary of this chapter is given.

6.5 Summary

This chapter has presented the method and results of objective dispersal of plant strategies and primary metabolic (photosynthetic) patterns of individual plant species occurrences, citing example locations of Azerbaijan and Cuba. First introduction of stochastic distribution and mathematic methodological descriptive background was given. A brief

summary of T-S-K nodal structure employed in combined fuzzy genetic methods was shown furthering the background required for development of T-S-K rule based systems. The novel methodology required for fuzzy genetic differentiation of plant strategies was covered and subsequently elements of strategies are itemized for dispersal and dispersed using a process analogous to natural genetic recombination.

Genetic programming of strategies and metabolism enables the distribution of the characteristics of individual plant species occurrences to be numerically distributed in the combined objective space of the water-energy dynamic. The dispersal is in agreement with alternative methods used in climatic science which have resulted in similar shapes of data points (Schölzel and Friedrichs, 2007). The developed dispersal method of hybrid MOGA enables formation of expressions which allow summarization of the dispersed elements of strategies to be carried out in utopian space. Given that similar results are obtained for the Pareto fronts of 2 different environments (E6 and E3) used in the methods and results of this chapter it is reasonable to assess that the water-energy dynamic has the same effect regardless of location, though this point is further elaborated in Chapter 7.

This chapter gives the answer to research questions posed in Chapter 1, the sub groups of plant life-history strategies and photosynthetic types are differentiated. Informative value is given to inferential and antecedent variables used in the modelling of plant species characterisation. All research questions are revisited in Chapter 8. Subsequently, the mathematic distribution of elements has been shown and geographic coverage is given to 2 different areas (for others see Chapter 7 and Appendices 5.2 and 6).

Furthermore, linear and quadratic rules of utopian space of the water energy dynamic elucidate a method by which further characteristics of plant species may be investigated. Additional benefit is also found in enhancement of the climatic, species numbers and

characteristic knowledge bases used in modelling of plants.

The following chapter details the use of T-S-K FL and MOGA systems to inform geographic information systems. Further detail of mathematic exploration is also elaborated to substantiate statements of conservation and sustainability policies.

CHAPTER 7

Plant life-form distribution and instructive geographic information systems

7.1 Introduction

Chapter 3 identified plant life-forms as one of the potential categories by which plants may be grouped. Plant life-forms are a primary means by which to categorize forms of plant growth, together with life history strategies and metabolism. Distribution of life-forms is an effective way to show distribution of plant species on a macro scale with use of computational statements (Grime et al., 1995; Raunkier, 1934). Patterning of plant species may be determined by key factors of the water-energy dynamic (Hawkins et al., 2003; Kreft and Jetz, 2007; Sommer et al., 2010). Climatic variables such as rainfall and temperature often show discrete patterns across timescales, so are often made use of within fixed time ranges. As such, they can be said to be discrete stochastic patterns, which facilitate macro-level modelling of plants (Grime and Pierce, 2012; New et al., 1999; Silvert et al., 2000).

Utopian distribution refers to the informative combined objective Z matrices, which may be generated through the use of techniques including adaptive fuzzy neural inference

systems, genetic programming, and particle swarm optimization (Alcalá et al., 2007; Furze et al., 2013c; Omizegba and Monikang, 2009; Wang and Yang, 2010; Zadeh, 2011).

Functional relations may be explored within the products of evolutionary algorithms via the use of functional process models, which display continuous or discontinuous qualities (Büche et al., 2005; Schölzel and Friedrichs, 2007). Computational methods may be applied to break down biological systems such as reservoir capacity to support life and produce indices of essential ecological descriptors (Taheriyoun et al., 2010; Silvert et al., 2000; Wang et al., 2012).

There are five main groups of plant life-forms as documented in Chapter 3, section 3.6. The groups may be further differentiated to component sub-groups as detailed in section 7.3. Life-forms are often seen in differing proportions or spectra; plants show chaotic patterns of evolution in terms of their individual growth processes and numbers (Cui et al., 2012; Furze et al., 2011; Su et al., 2009). Life-form differences are often associated with variable gradients in topographical and climatic conditions (Bhatterai and Vetaas, 2003; Schmidt et al., 2005). In order to clarify the difference in life-form spectra this chapter considers two contrasting areas known to be rich in plant numbers: Ecuador, South America and Macedonia, Southern Europe (Barthlott et al., 2005; Bass et al., 2010; Furze et al., 2012b; Zlatković et al., 2011). The former of these has been well documented as being the most diverse location on the planet and the latter has previously been algorithmically defined as having the characteristics of a more extreme environment (elevated temperature, comparatively low rainfall).

In this chapter proportions of plant life-form characteristics are investigated within fixed population sizes, which have been determined from a combined genetic algorithm fuzzy rule base, furthered by field based studies (Bass et al., 2010; Zlatković et al., 2011).

The remainder of the chapter is structured as follows: section 7.2 illustrates DEM and climatic data of candidate locations for life-form categorisation, there is elaboration of the method for algorithmic breakdown of the ANFIS for plant life-forms and the resultant ANFIS is given. Section 7.3 explains the methodology for dispersal of life-form elements and gives the MOGA Pareto and selection process of the chromosomal population. The author explains the use of process models to distribute plant life-forms and expands the Z hyperplane with use of functional approximation. Section 7.4 gives a description of plant-strategy, plant photosynthetic and plant life-form ANFIS algorithms, and further life-forms

are used to elaborate the functional approximation algorithm approach for stable communities of plant species individual occurrence. Section 7.5 defines geographic information systems in the context of plant characterisation. The novel application of mathematical methods to substantiate national and international conservation and sustainability policy formation is briefly covered. Finally, section 7.6 summarises the chapter.

7.2 Using DEM and climatic data to substantiate T-S-K systems

Climatic variables of diversity zones 8-10 (Barthlott et al., 2005), of greater than 3000 species per 10000km², were investigated and given algorithmic definition (Furze et al., 2012a). Further, these areas are included within high resolution mapping tiles available from the Intergovernmental Panel on Climate Change (at 18.5km resolution) and the United States Geological survey (at 1km resolution). Two candidate areas were selected from the literature to show the breadth of difference in climatic, topographic and actual documented numbers of individual plant occurrences within the areas. The selected areas were Ecuador, including the reserve surrounding Tiputini Biodiversity Station (covering approximately 10000km²) (Bass et al., 2010; Bilsborrow et al., 2012) and the country of Macedonia (covering 25713km² (European Environment Agency, <http://www.eea.europa.eu/> accessed 30 06 13)). Coordinates for each area were obtained from the above sources and code was constructed in Matlab (available in Appendix 2.1.1) enabling display of the digital elevation model (DEM), precipitation, temperature and ground frost frequency data for each region. Variables were quantified to maximize computational efficiency and interpretability of the ANFIS. The DEM data of GTOPO30, published through the United States Geological Survey (USGS), are shown graphically in **Fig. 7.2.1** at 30 s (1 km resolution).

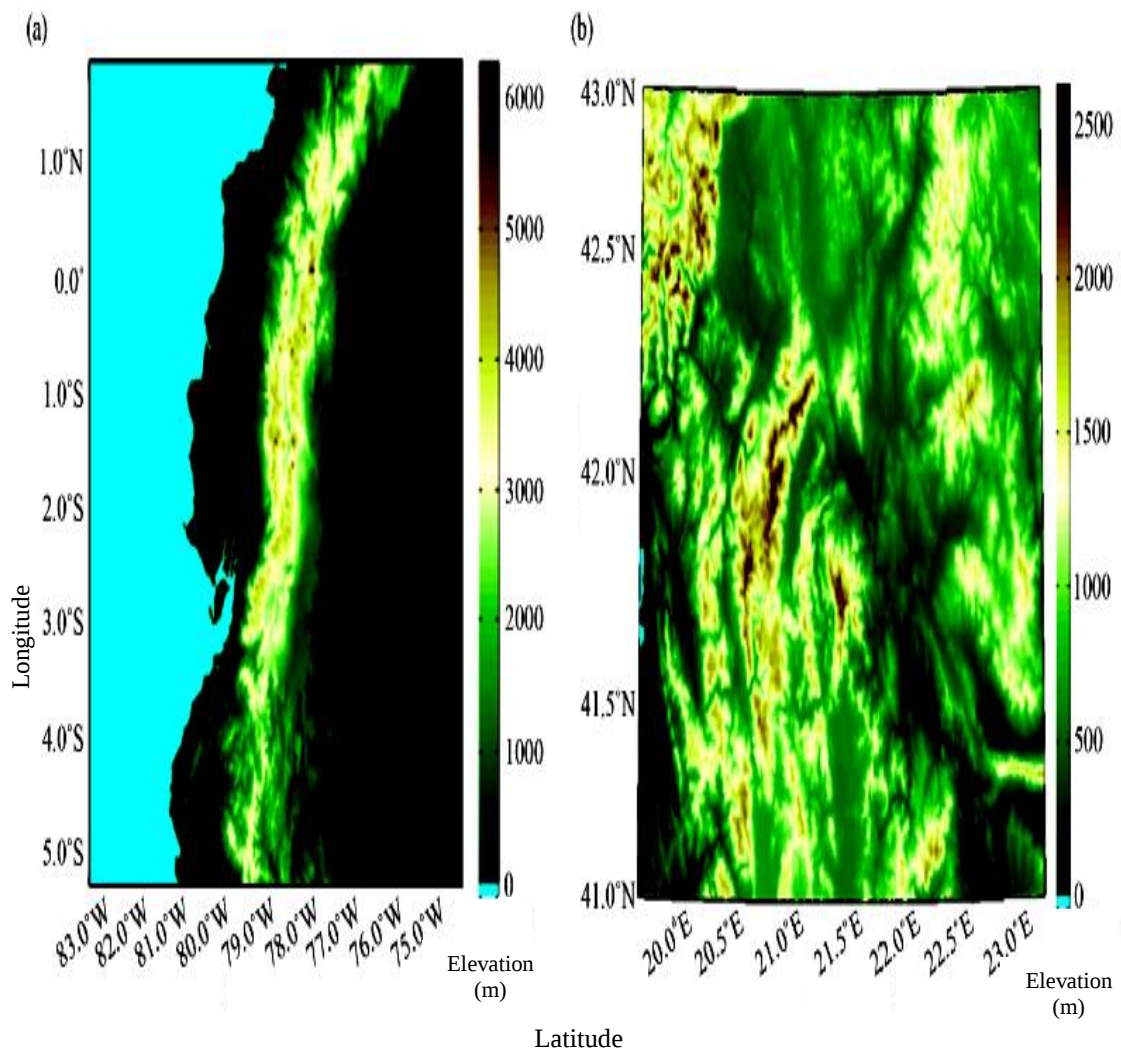


Fig. 7.2.1 Digital Elevation Maps for a) Ecuador, South America; b) Macedonia, Southern Europe (Furze et al., 2013d)

Ecuador shows an elevation range from 0m to 6300m above sea level, whereas Macedonia shows an elevation of 0m to just over 2520m above sea level. The elevation ranges of Ecuador and Macedonia were quantified according to a five-split partitioning of the range as shown in **Tab. 7.2.1**.

Tab. 7.2.1 Quantification of DEM data for Ecuador, South America and Macedonia, Europe

Elevation (m)	Notation
0 – 1260	$A4_{(1)}$
1260 – 2520	$A4_{(2)}$
2520 – 3780	$A4_{(3)}$
3780 – 5040	$A4_{(4)}$
5040 – 6300	$A4_{(5)}$

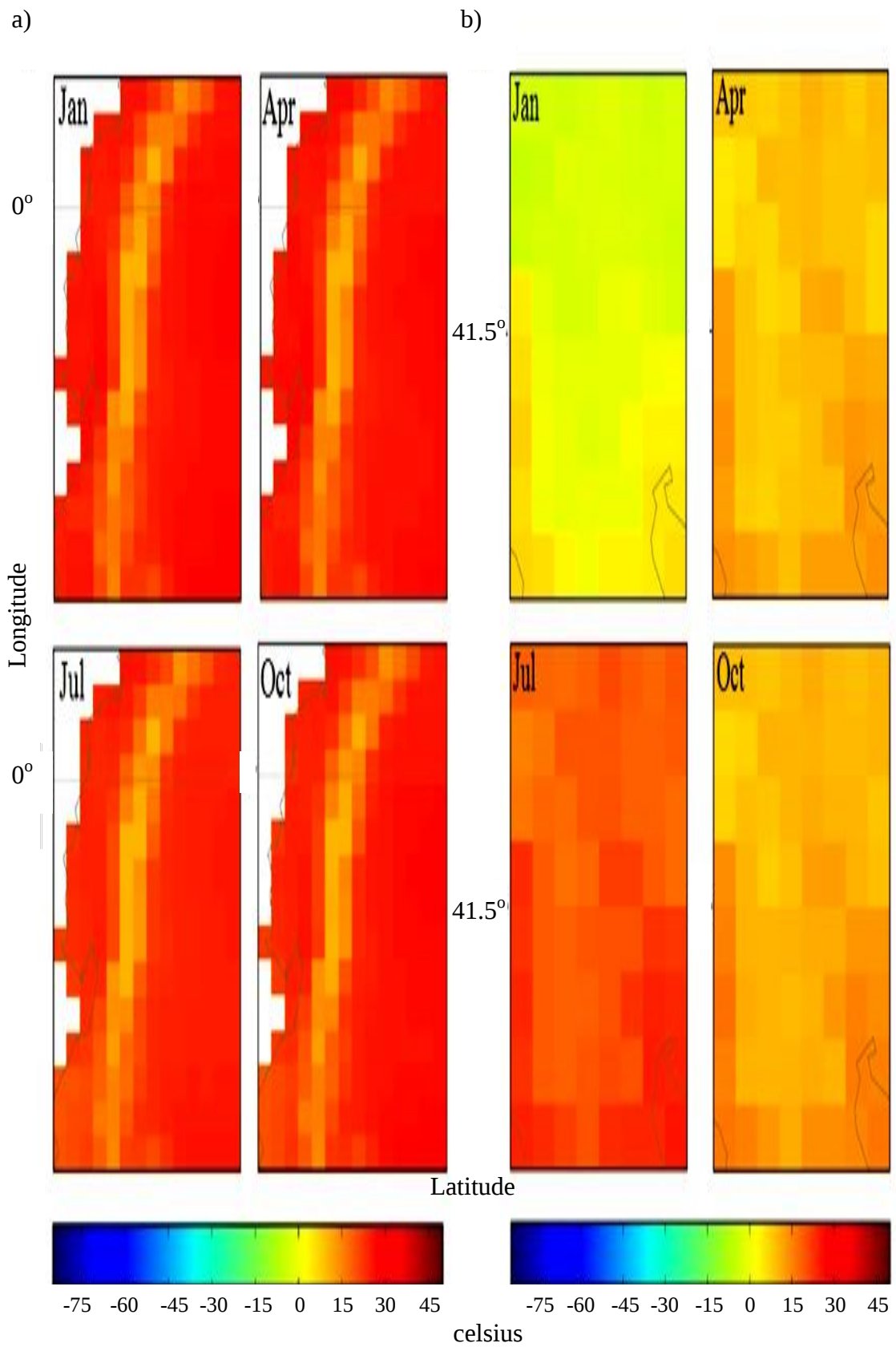


Figure 7.2.2 Quarterly mean temperature (1961-90) of a) Ecuador, South America and b) Macedonia, Southern Europe (Furze et al., 2013d)

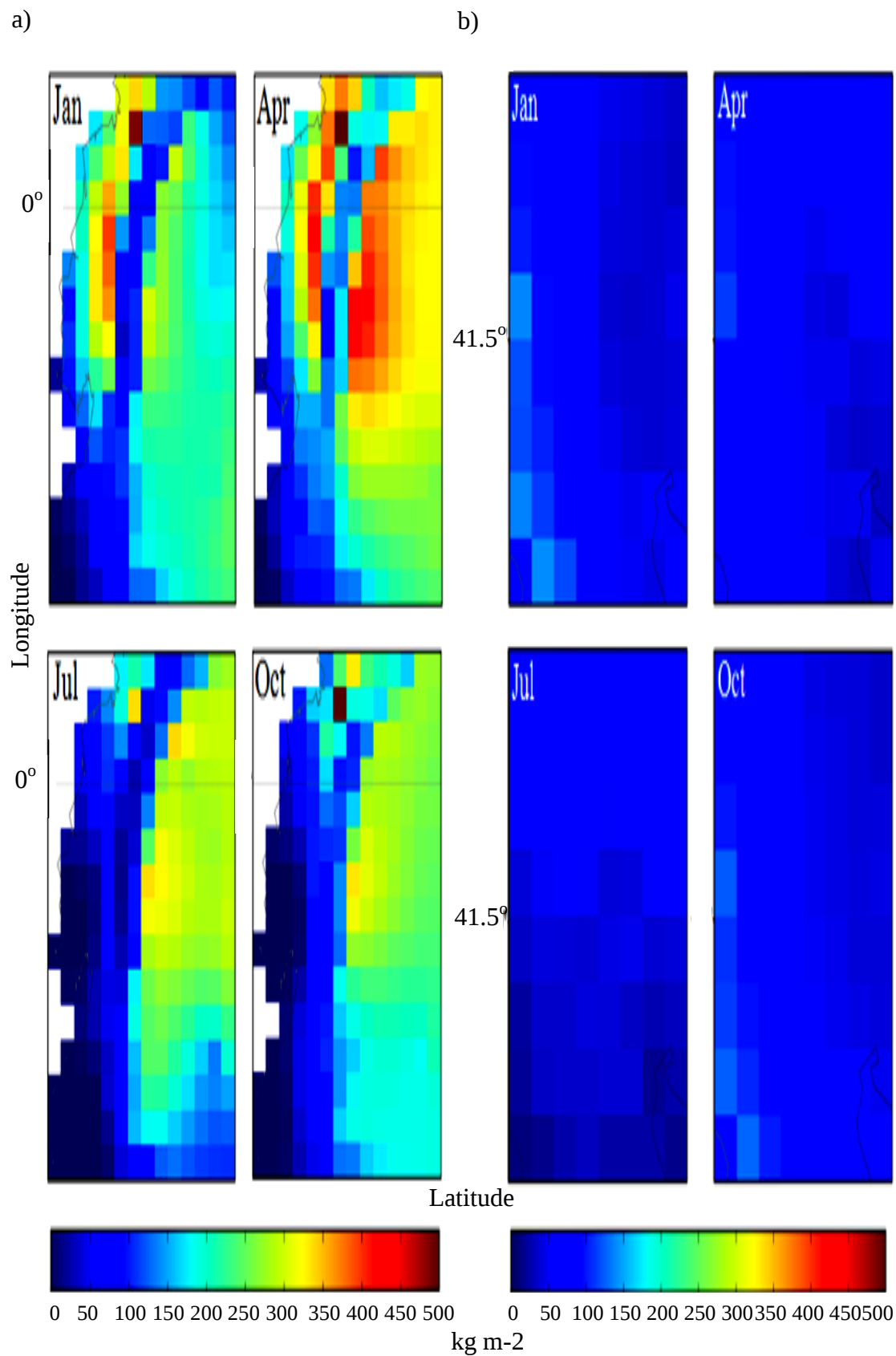


Figure 7.2.3 Quarterly mean precipitation (1961-90) of a) Ecuador, South America and b) Macedonia, Southern Europe (Furze et al., 2013d)

With reference to **Fig. 7.2.2** a) Ecuador shows mean temperatures of -3 to >21°Celsius in January, April, July and October; b) Macedonia shows temperatures of -27 to 21°Celsius in January, -3 to 21°Celsius in April, 21 to 45°Celsius in July, and -3 to 21°Celsius in October. Mean temperatures were quantified according to the method shown in (Furze et al., 2012a).

With reference to **Fig. 7.2.3** a) Ecuador shows precipitation of 0-500 kg m² in January, April and October and 0-400 kg m² in July; b) Macedonia shows precipitation of 0-200 kg m² in January, April and October and 0-100 kg m² in July. Data of ground frost frequency (New et al., 1999) is available on the IPCC web site. All climatic variables are displayed at 18.5 km resolution. The method for quantification of climatic variables is shown in Furze et al. (2012b). Mean temperature was designated as A1, mean precipitation as A2, mean ground frost frequency as A3 and elevation as A4 to enable the construction of the ANFIS engines, which is detailed in the following section.

7.2.1 ANFIS and efficiency of reduced model components

ANFISs are commonly built using Takagi-Sugeno-Kang (T-S-K) or Mamdani fuzzy logic. T-S-K fuzzy systems (Takagi and Sugeno, 1985; Zadeh, 1965) are more easily applied to multiple input and multiple (ranged) output as discussed in Chapter 3, section 3.7.

After quantifying the variables for the geographic locations of Ecuador and Macedonia, and sourcing the number of individual plant occurrences from the Global Biodiversity Information Facility (<http://gbif.org> , accessed December 2012), the following concise fuzzy singleton antecedent-consequent rule bases were applied in order to build the inference engines:

$$\begin{aligned} & \text{If } A1_{(3)} \prec A1_{(4)} \wedge 0.75A2_{(1)} \prec A2_{(5)} 0.25A2_{(1)} \prec A2_{(4)} \\ & \wedge A3_{(1)} \prec A3_{(2)} \wedge A4_{(1)} \prec A4_{(5)} \text{ Then } B_{(65535)} = E1 \end{aligned} \quad (7.2.1.1)$$

The above algorithm translated into the following conditions:

IF Variables (A) = Temperature = 60-80 °C; Precipitation = 0.75 x 0-100 Kg m² to 400-500 Kg m², 0.25 x 0-100 Kg m² to 200-300 Kg m²; Ground Frost Frequency = 0-6 days to 6-12 days; Altitude = 0-6300m THEN Environment 1 (Phanerophytes dominant ≤ Chamaephytes ≤ Hemicryptophytes ≤ Cryptophytes ≤ Therophytes) (B) = 65535 individual plant occurrences. (7.2.1.1) Creates 17 rules with variable weights as shown in **Tab. 7.2.1.1.**

Table 7.2.1.1 ANFIS rules for Ecuador, South America showing life-form dominance

Rule	Description	Weight [0, 1]
1	If Temp is Med-High the L-F is P	1
2	If Prec is Low then L-F is P	0.75
3	If Prec is Med-Low then L-F is P	0.75
4	If Prec is Med then L-F is P	0.75
5	If Prec is Med-High then L-F is P	0.75
6	If Prec is High then L-F is P	0.75
7	If Prec is Low then L-F is P	0.25
8	If Prec is Low-Med then L-F is P	0.25
9	If Prec is Med then L-F is P	0.25
10	If Prec is Med-High then L-F is P	0.25
11	If GFF is Low then L-F is P	1
12	If GFF is Low-Med then L-F is P	1
13	If Alt is Low then L-F is P	1
14	If Alt is Low-Med then L-F is P	1
15	If Alt is Med then L-F is P	1
16	If Alt is Med-High then L-F is P	1
17	If Alt is High then L-F is P	1

In **Tab. 7.2.1.1**, Temp = Mean temperature, Prec = Mean precipitation, GFF = Mean ground frost frequency, Alt = Altitude, L-F = Life-form, P = Phanerophyte dominated, Med = Medium. The terms Low, Low-Medium, Medium, Medium, Medium-High, High are quantified according to Furze et al. (2012b).

$$\begin{aligned}
 & \text{If } 0.25A1_{(4)} \prec A1_{(5)} 0.5A1_{(4)} 0.25A1_{(5)} \wedge 0.75A2_{(1)} \prec A2_{(2)} 0.25A2_{(1)} \wedge 0.25A3_{(1)} \\
 & \prec A3_{(5)} 0.5A3_{(1)} \prec A3_{(2)} \wedge 0.25A3_{(1)} \wedge A4_{(1)} \prec A4_{(3)} \text{ Then } B_{(2023)} = E5
 \end{aligned}
 \tag{7.2.1.2}$$

(7.2.1.2) is the algorithm for Macedonia, which translated into the following conditions:

IF Variables (A) = Temperature = 0.25 x 60%-80% to 80-100%, 0.5 x 60-80%, 0.25 x 80-100%; Precipitation = 0.75 x 0-100 Kg m² to 100-200 Kg m², 0.25 x 0-100 Kg m²; Ground Frost Frequency = 0.25 x 0-6 days to 24-30 days, 0.5 x 0-6 days to 6-12 days, 0.25 x 0-6 days; Altitude = 0-3780m THEN Environment 5 (Hemicryptophyte, Therophyte dominant ≤ Chamaephytes ≤ Hemicryptophytes ≤ Cryptophytes ≤ Phanerophytes) (B) = 2023 individual plant occurrences. (7.2.1.2) Creates 16 rules with variable weights (see Appendix 2.4).

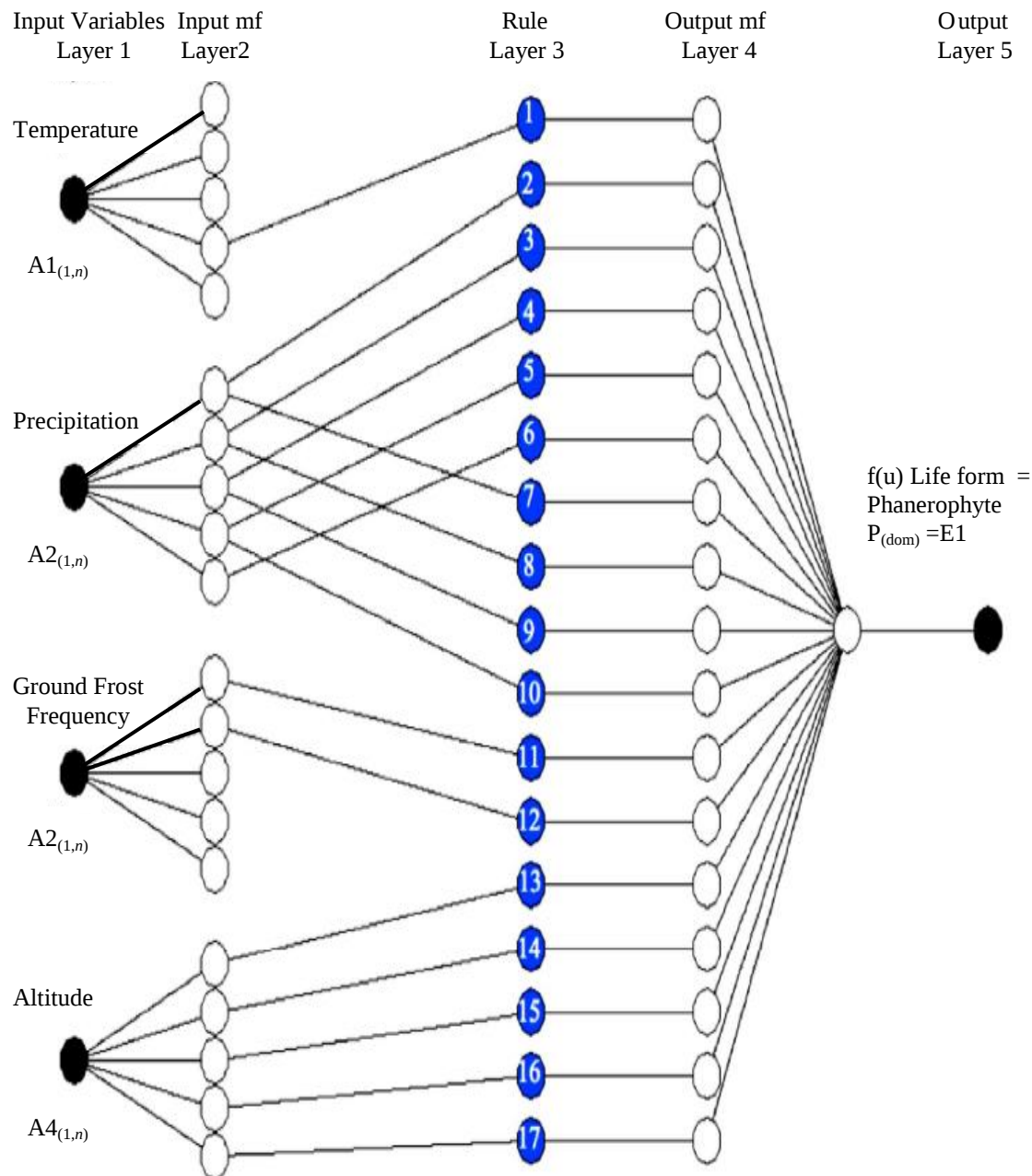


Figure 7.2.1.1 Adaptive neural fuzzy inference system quantifying the dominant plant life-form type of Ecuador (Furze et al., 2013d)

The first layer of the computational engine shown in **Fig. 7.2.1.1** accepted the crisp input variables, the second layer enabled conversion of the variables according to their membership functions values / terms, the third layer was where the rules of the engine operate (seen in **Tab. 7.2.1.1**), the fourth layer converted the values back through membership function partitioned terms and the fifth layer computed the specific (crisp) number applicable for the predominant life-form type. The estimated primary consequent nodal number was 65535 (individual plant occurrences) for Ecuador and 2203 (individual

plant occurrences) for Macedonia (GBIF (<http://gbif.org>, accessed December 2012) (Yesson et al., 2007)). The efficiency of each pair of variables was seen by viewing the

surface of the algorithm (**Fig. 7.2.1.2.2**). Clear definition is a good indication of accuracy achieved and helps to choose which variables are minimized in optimization techniques (Furze et al., 2013a).

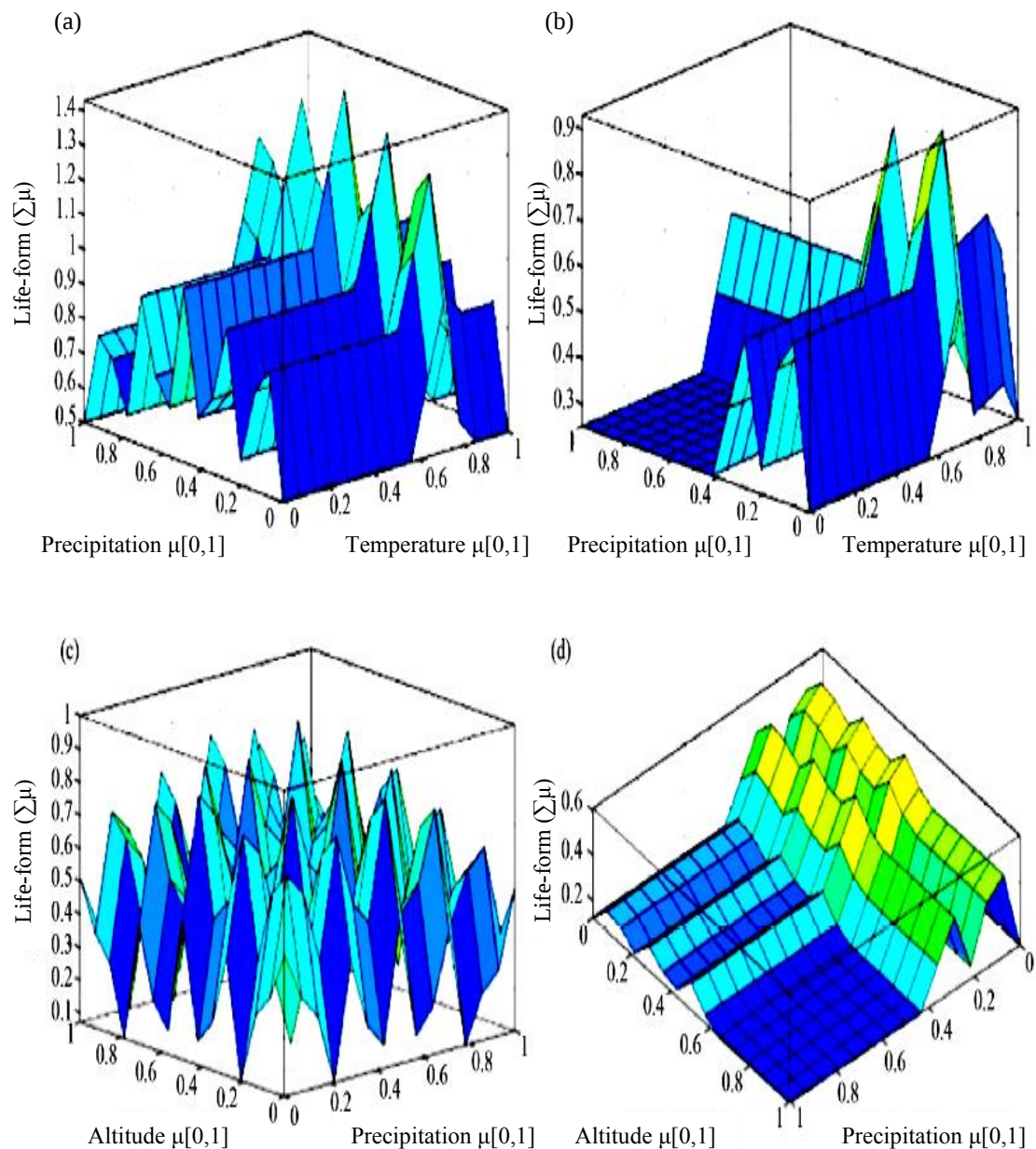


Figure 7.2.1.2.2 3-D Surface views of variables of the algorithm for plant life-forms. a) Ecuador, South America (precipitation versus temperature), b) Macedonia, Southern Europe (precipitation versus temperature, c) Ecuador, South America (precipitation versus altitude), d) Macedonia, Southern Europe (precipitation versus altitude) (Furze et al., 2013d)

Fig. 7.2.1.2.2 a) and b) show similar defined peaks of life-form differentiation, whereas c) and d) give no consistently defined life-form peaks. It is suggested therefore that the most effective variables for definition of the life-form categories are water (precipitation) and energy (temperature). Further dispersal of the life-forms is required in order that we may consider the distribution of the range of life-form sub categories present in the two candidate areas.

The following section proceeds to quantify the 18 life-form characterisation and shows the result of a multi-objective genetic programming allowing dispersal of the 18 life-form elements, employing the objectives temperature and precipitation to generate the utopian space via multi objective genetic algorithm (MOGA).

7.3 Multi-objective optimisation selection and dispersal of plant life-forms

In terms of plant life-forms we may state membership functions in terms of the five main groups of life-form within a final (known) population number, with an established distribution of the (objective) variables (Broekhoven et al., 2007). Integrating a mechanism to disperse elements (or sub-categories) of life-forms has proved to be in visualisation and estimation of the distribution of the associated life-form traits. A hybrid genetic algorithm approach is applied in order to achieve dispersal of elements and extrapolation of the Z hyperplane (Furze et al., 2013a). In this chapter the node structure of plant strategies is replaced with plant life-forms. The expanded number of life-forms (18) are given solutions and detailed in **Tab. 7.3.1**.

Table 7.3.1 Solutions and potential ranges for plant life-form chromosomes

Character / Chromosome	Phan (0,...,5)	Chamae (0,...,5)	Hemi-crypt (0,...,5)	Crypt (0,...,5)	Thero (0,...,5)
ws / 1	5	1,...,5	1,...,2.5	0,...,2.5	0,...,3
eg/d bs / 2	5	0	0	0	0
H>30m / 3	1.25,...,5	0	0	0	0,...,1.25
H 8-30m / 4	1.25,...,5	0	0	0	0,...,1.25
H 2-8m / 5	1.25,...,5	0	0,...,1.25	0,...,1.25	0,...,1.25
H<2m / 6	1.25,...,5	5	1.25,...,5	0,...,1.25	0,...,1.25
amgpud / 7	0,...,1.25	1.7,...,5	0,...,1.25	0	0
amgpuf / 8	0,...,1.25	1.7,...,5	0,...,1.25	0	0
sopag / 9	0,...,1.25	1.7,...,5	0,...,1.25	0	0
scpag / 10	0,...,1.25	1.7,...,5	0,...,1.25	0	0
agdoamgp / 11	0	0,...,1.25	1.25,...,5	0	0
lwdusspgb / 12	0,...,1.25	0,...,1.25	1.25,...,5	0	0
dlfbrlasgyal / 13	0	0,...,1.25	1.25,...,5	0	0
lrtbrlasaldf / 14	0	0,...,1.25	1.25,...,5	0	0
ufsoseigs / 15	0,...,1.7	0,...,1.7	0,...,1.7	1.7,...,5	0
gbaisbwsapaw / 16	0,...,1.7	0,...,1.7	0,...,1.7	1.7,...,5	0
gbuwagpsruw / 17	0,...,1.7	0,...,1.7	0,...,1.7	1.7,...,5	0
apcegcismawsas / 18	0,...,0.3	0,...,0.3	0,...,0.3	0,...,0.3	5

In **Tab. 7.3.1** Phan is phanerophyte, Chamae is chamaephyte, Hemi-crypt is hemicryptophyte, Crypt is cryptophyte, Thero is therophyte, ws is woody stem, eg/d bs is evergreen / deciduous bud scale, H is height, m is metres, amgpud is after main growth period upper shoots die, amgpuf is after main growth period upper shoots flat, sopag is shoots only produced along ground, scppag is shoots closely packed produced along ground, agdoamgp is all above-ground dies off after main growth period, lwdusspgb is leaves well developed up the sides of stems protecting growing buds, dlfbrlasgyal is developed leaves form basal rosette long aerial shoots grow year after leaf development for leaves and flowers, lrtbrlasaldf is leaves restricted to basal rosette long aerial shoot after leaf development for flowers only, ufsoseigs is underground food storage organs shoots emerge in growing season, gbaisbwsapaw is growing buds are in soil beneath water shoots

are produced above water, gbuwagpsruw is growing buds under water after growth period shoots remain under water, apcegcismawsas is annual plants complete entire growth cycle in summer months, after which survive as seeds.

Quantification values are seen in brackets in **Tab. 7.3.1**. The characters represented in the table represented a chromosomal population and were used to form a multi objective genetic algorithm (MOGA) in which the chromosomes cycled through 0-5, resulting in a Pareto front across the combined objective space. This substantiates one method by which the distribution of the life-form elements was estimated within the previous sections algorithmically defined statements, expanding the number of life-form nodes from 5 to 18.

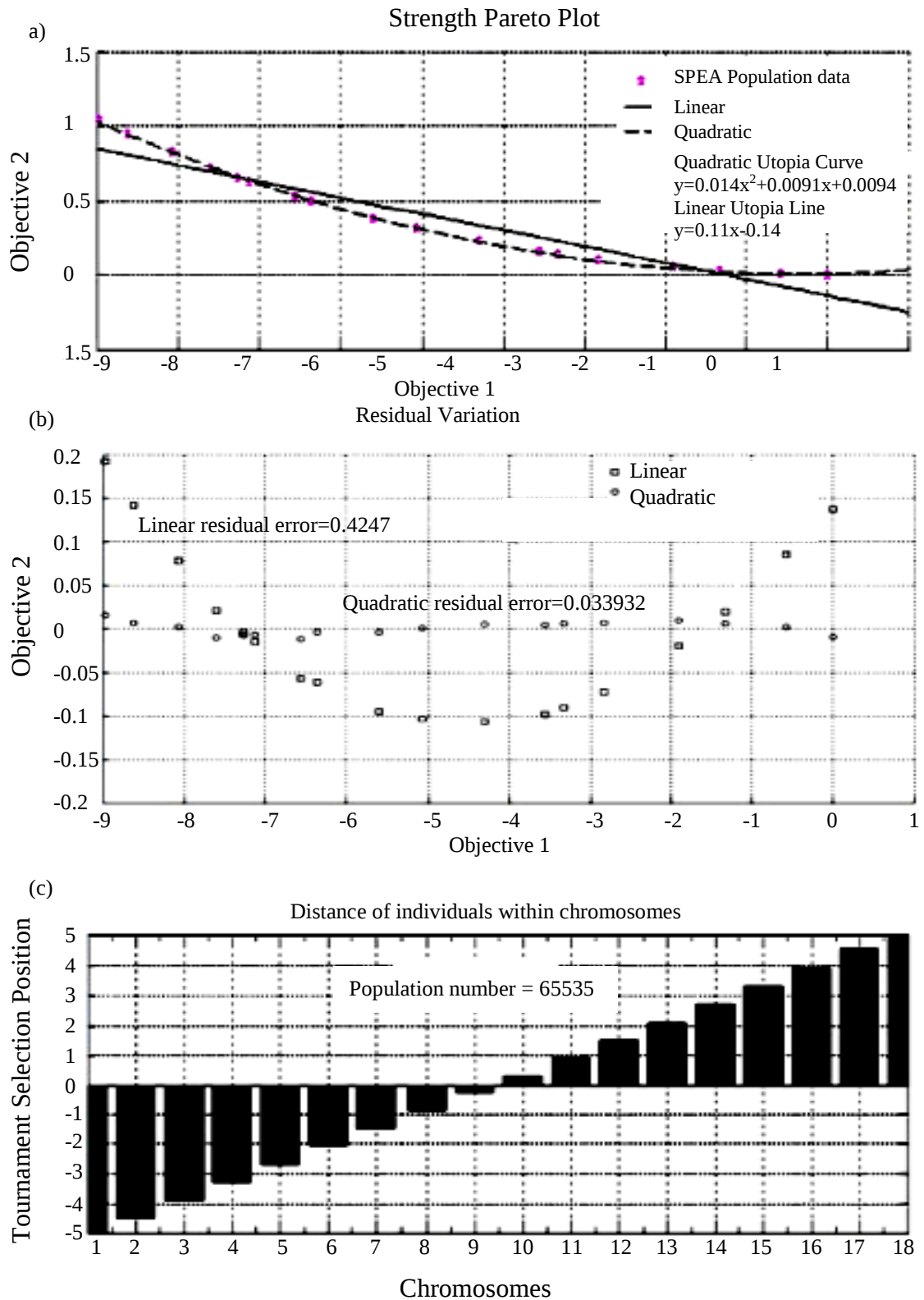


Figure 7.3.1 a) Objective dispersal of the Strength Pareto Evolutionary population obtained for Ecuador, b) Error of linear Utopia line and quadratic curve, c) Illustration of how selection within the chromosome populations took place (Furze et al., 2013d)

In **Fig. 7.3.1 a)** Life-form sub groups are distributed within the water-energy dynamic of Objective 1 (Mean temperature) and Objective 2 (Mean precipitation) of combined objective space. The number of individual plant occurrences is 65535, enabling the same number of generations over which the algorithm cycled. **Fig. 7.3.1 b)** illustrates the spread of hyperbolic residual variance obtained in the population of 18 chromosomes' utopia line and the much lower variance obtained given the utopia curve. This gave a similar result to that shown for plant strategies and photosynthetic elements, confirming the effect of the water-energy dynamic on individual plant occurrence characters from three alternative biological processes with implications as to the spread of characters in different environments and also to the spread of photosynthetic elements. **Fig. 7.3.1 c)** gives an illustration of the points at which the chromosomes were selected from within the MOGA process. The process of MOGA is covered in Chapter 6, (Furze et al., 2013a) with the essential difference being that 18 elements of plant life-forms as opposed to 20 elements of strategies are dispersed as covered in Chapter 6. Code for the MOGA is available in Appendix 7. The consequent rules for utopia line and curve are seen in Appendix 7.1.

The following sub-section develops an approximation of the distribution of the sub-groups of life-forms with use of a surrogate model of plant life-form distribution.

7.3.1 Functional approximation within utopia of plant life-forms

Approximation of the functional distribution of any character may be seen to be within the field of evolutionary algorithms and estimation of distribution algorithms as stated in Chapter 2, section 2.5.3 and Chapter 3, section 3.5. In this section a novel approach is formulated by making use of process models to summarise the distribution of plant life-forms seen within individuals of any set population. A further advantage of using a Gaussian process model in estimation of rudimentary distribution is the effective estimation of the life-forms' fitness function, which must be known when programming a MOGA such as that used in the previous sections. The fitness function is a numerical estimation of chromosomal 'fitness'. In previous application of MOGA the default fitness of 0.8 or 80% adaptation for selection has been made use of. The method of functional approximation algorithm (FAA) for plant variation consists of the following 4 steps:

- *Step 1.* Geographic and climatic study to establish the framework for modelling

species of plant life-forms in candidate areas.

- *Step 2.* Adaptive fuzzy neural inference system (ANFIS) using identified variables based on consequent primary nodal number.
- *Step 3.* Multi objective genetic algorithm dispersing expanded secondary nodal number.
- *Step 4.* Functional approximation of characters within secondary nodal number using a continuous/discrete surrogate process model.

Sub-section 7.2 completed step 1 and step 2 of the method given above. Section 7.3 numerically dispersed the sub-groups of plant life-forms, completing step 3. In this section a further method of generating the Z matrix is developed via the use of a surrogate function.

The assumption that individuals within the populations of plant species in each of the studied areas are normally distributed in their life-form characteristics is made, furthered by elaboration of the ANFIS in section 7.4. Additionally, standardization of the dispersed strength Pareto population to zero mean and unit variance allows the population to be expressed across a bell shaped (Gaussian) curve (Guo, 2011).

The choice of Gaussian models selected is Rastrigin's function, which served the dual purpose of expanding the dispersal of the 18 sub groups of life-forms simultaneously and verified the validity of the GA process described in section 7.3. This is elaborated below.

$$f(x) = 10 \cdot n \sum_{i=1}^n (x_i^2 - 10 \cdot \cos(2 \cdot \pi \cdot x_i)) \quad (7.3.1.1)$$

Where $i=1/n$, x_i is an element of the interval $[-5.12, 5.12]$. In order to make use of Rastrigin's function as a surrogate Gaussian process model in the current case, the number of life-form sub categories ($n=18$) occurs within the interval shown. **Fig. 7.3.1.1** is a visual representation of the function.

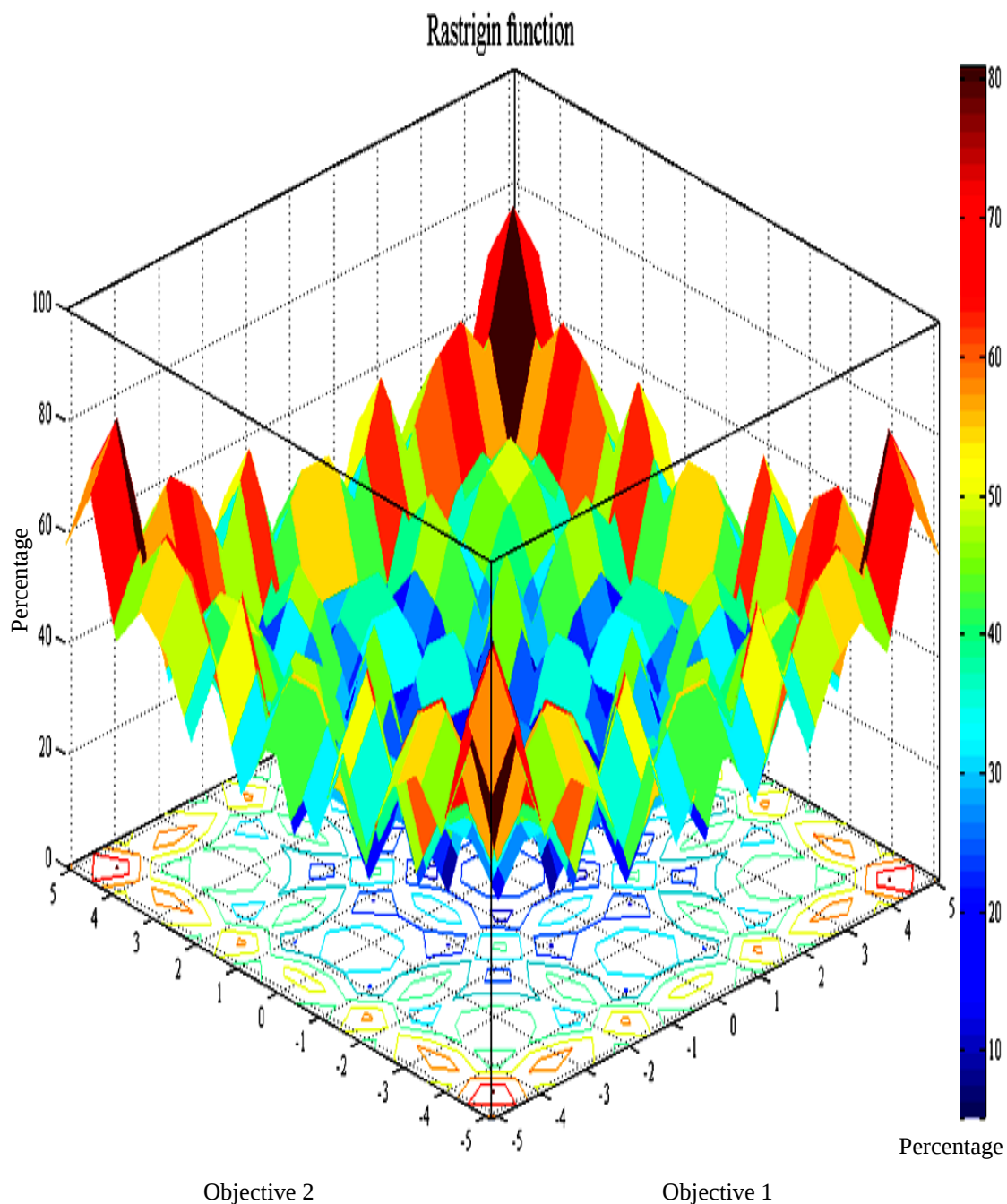


Figure 7.3.1.1 Rastrigin's function as applied to the standardised distribution $[0, \dots, 5]$ of 18 sub groups of plant life-forms (Furze et al., 2013d)

Objective 1 (standardised mean temperature) and Objective 2 (standardised mean precipitation) give the parameters in which individual plant species occurrence life form sub groups are distributed. The percentage of each of the 18 sub groups is shown in colours indicated in the legend. For clarity the valleys (minima) of the peaks can also be seen in contour form below the plot.

Plant life-form traits are distributed normally amongst the species occurrences of the candidate locations Ecuador (Bass et al., 2010) and Macedonia (Zlatković et al., 2011) as the strength Pareto shown in the previous section indicates. It is therefore possible to generate values of the Z plane with use of a surrogate Gaussian function model (Büche et al., 2005). Rastrigin's function is one such process function, which can be made use of in order to extrapolate values of a multimodal optimization. Code for the Rastrigin's function as applied to 18 life-form sub groups in Matlab (Version R2010a © is available in Appendix 7.2). Given that Rastrigin's function is particularly well suited for validating the MOGA approach shown in the previous section, we also visualised distributions of a lesser number of characters within a hyperplane (Appendix 7.3). Using this method, any value of either of the dimensions used in multimodal control models may be made use of to objectively extrapolate local set values of equations (3.7.2.3), (3.7.2.4), (3.7.2.5), shown earlier underpinning T-S-K logic FRBS, given the initial knowledge of the number of groups under investigation within the dynamic of the model.

The following section gives a summary of algorithmic approaches, which may be used to ordinate plant species and their characters.

7.4 ANFIS control strategy of plant strategies, plant metabolism and plant life-forms

This section gives examples of the T-S-K FL algorithms, formed in previous chapters, to predict the ordination of individual plant species occurrence in global locations. Expansions of the algorithms are seen in Appendices 2, 2.3, 2.4, 5.1, 5.2. Individual plant species occurrence is considered from the perspective of plant life-history strategies (Grime et al., 1995), metabolic (photosynthetic) pathway type (Chapter 3, section 3.6) and life-form (Chapter 3, section 3. 6, Chapter 7, sub-section 7.2.1). All ANFIS are based on the modelling framework provided by the water-energy dynamic together with high resolution altitude data, as antecedent data and individual plant species occurrence data providing consequent ranges, sourced from the Global Biodiversity Information Facility (GBIF) meta-database (Yesson et al., 2007). Citations of contributing institutions are included in Appendix 1.

Chapter 2, section 2.3.1 elaborates the numerical basis on which we ordinate strategies. This thesis proposes the sum of the strategy types always equates to 1 in order to classify an environment of plant species, there being 3 main nodes on which plant strategies are defined: Competitive, Stress tolerant, Ruderal or $n, n, n = 1$. Using this basis, the Ruderal Environment 1, consequent individual occurrence of plant species number (65535) = 0, 0, 1.

Table 7.4.1 Plant species C-S-R balance in numerical form.

Strategy / Environment	Competitive	Stress tolerant	Ruderal
R / E1	0	0	1
S-R / E2	0	0.5	0.5
C-R / E3	0.5	0	0.5
C / E4	1	0	0
C-S-R / E5	0.33	0.33	0.33
C-S / E6	0.5	0.5	0
S	0	1	0

In reality, the existence of a pure strategy without any elements of the others very rarely exists due to the polyploidy level of plant species. By extension to an ecosystem approach this means that in the plant-strategy environments, levels of competitive and of stress tolerant species do exist in ruderal environments, the limit after which, the strategy element is detected is equal to 1/3. The latter leads to the statement that each environment $E1, \dots, E7$ is dominated by the strategy elements shown in **Tab. 7.4.1**, following knowledge of the conditions in which each strategy element occurs (Grime et al., 1995). Proportionality (probability) of the 3 strategy elements may be investigated via the construction of a 3-dimensional linear mesh of the species elements as seen earlier (Chapter 3, **Fig. 3.6.2**).

T-S-K logic systems allow concise statements to be made due to the fact that the combined elements of plant strategies are seen to be membership values as stated in Chapters 3, 4, 5 and 6. Hence one may make an approximation of the proportionality of each strategy type within an estimated population number based on the applied T-S-K logic framework.

In fact, we may extend the above to interpolate the proportionality of plant photosynthetic types using the same principal, there being C3, C4 and CAM types of metabolism.

Dividing 1 by the number of groups under investigation leads us to consider that 1/18 probability equates to the probability of occurrence of each life-form category. There are 18 sub groups, as detailed in Chapter 3, section 3.6, in section 7.2 and 7.2.1, of the current chapter. The expanded probability table of life-form groups within each environment is seen in Appendix 7.4.

Algorithms for plant strategy environments ordinate individual plant species occurrence in terms of the probabilities indicated in **Tab. 7.4.1** with use of the following concise statements:

$$\begin{aligned} & \text{If } A1_{(3)} \prec A1_{(4)} \wedge 0.75A2_{(1)} \prec A2_{(5)} 0.25A2_{(1)} \prec \\ & A2_{(4)} \wedge A3_{(1)} \prec A3_{(2)} \wedge A4_{(1)} \prec A4_{(5)} \text{ Then } B_{(65535)} = E1 \end{aligned} \quad (7.4.1.1)$$

E.g. Ecuador, South America.

$$\begin{aligned} & \text{If } A1_{(5)} \wedge 0.75A2_{(1)} \prec A2_{(3)} 0.25A2_{(1)} \prec \\ & A2_{(4)} \wedge A3_{(1)} \wedge A4_{(1)} \prec A4_{(2)} \text{ Then } B_{(51847)} = E2 \end{aligned} \quad (7.4.1.2)$$

E.g. Guyana, South America.

$$\begin{aligned} & \text{If } 0.25A1_{(4)} \prec A1_{(5)} 0.75A1_{(5)} \wedge 0.5A2_{(1)} 0.25A2_{(1)} \prec \\ & A2_{(3)} 0.25A2_{(2)} \prec A2_{(3)} \wedge A3_{(1)} \wedge A4_{(1)} \prec A4_{(2)} \text{ Then } B_{(50709)} = E3 \end{aligned} \quad (7.4.1.3)$$

E.g. Cuba, America.

$$\begin{aligned} & \text{If } A1_{(4)} \prec A1_{(5)} \wedge 0.5A2_{(1)} \prec A2_{(4)} 0.5A2_{(1)} \prec A2_{(3)} \wedge \\ & 0.5A3_{(1)} \prec A3_{(2)} 0.5A3_{(1)} \wedge A4_{(1)} \prec A4_{(5)} \text{ Then } B_{(33359)} = E4 \end{aligned} \quad (7.4.1.4)$$

E.g. Democratic Republic of the Congo, Africa

$$\begin{aligned} & \text{If } 0.75A1_{(3)} \prec A1_{(4)} 0.25A1_{(4)} \prec A1_{(5)} \wedge 0.75A2_{(1)} \prec A2_{(2)} 0.25A2_{(1)} \prec A2_{(3)} \wedge \\ & 0.25A3_{(2)} \prec A3_{(5)} 0.5A3_{(1)} \prec A3_{(4)} 0.25A3_{(1)} \prec A3_{(2)} \wedge A4_{(1)} \prec A4_{(5)} \text{ Then } B_{(11359)} = E5 \end{aligned} \quad (7.4.1.5)$$

E.g. Georgia, Southern Europe.

$$\begin{aligned} & \text{If } 0.25A1_{(3)} \prec A1_{(4)} 0.5A1_{(4)} 0.25A1_{(4)} \prec A1_{(5)} \wedge A2_{(1)} \prec A2_{(2)} \wedge 0.25A3_{(3)} \\ & \prec A3_{(5)} 0.5A3_{(1)} A3_{(3)} 0.25A3_{(1)} \prec A3_{(2)} \wedge A4_{(1)} \prec A4_{(4)} \text{ Then } B_{(8809)} = E6 \end{aligned} \quad (7.4.1.6)$$

E.g. Azerbaijan, Southern Europe.

$$\begin{aligned}
& \text{If } 0.25A1_{(4)} \prec A1_{(5)} 0.5A1_{(4)} 0.25A1_{(5)} \wedge 0.75A2_{(1)} \prec \\
& A2_{(2)} 0.25A2_{(1)} \wedge 0.25A3_{(1)} \prec A3_{(5)} 0.5A3_{(1)} \prec A3_{(2)} \\
& \wedge 0.25A3_{(1)} \wedge A4_{(1)} \prec A4_{(3)} \text{ Then } B_{(2023)} = E7
\end{aligned} \tag{7.4.1.7}$$

E.g. Macedonia, Southern Europe.

Algorithms for plant photosynthetic type environments share the same basis as the above, there being 3 main types as detailed in Chapter 3, section 3.6. These provide the following concise statements:

$$C3 \geq C4 \geq CAM \forall E1; E2; E3 \tag{7.4.1.8}$$

$$C4 \geq C3 \geq CAM \forall E4; E5 \tag{7.4.1.9}$$

$$CAM \geq C4 \geq C3 \forall E6; E7 \tag{7.4.1.10}$$

Algorithms for plant life-form dominance must be expressed in different terms, the primary-nodal number of life-form groups being 5. The secondary-nodal number of (18) combined probabilities, indicated in Appendix 7.4, may be obtained using the dispersive optimisation method detailed in section 7.2 and the functional approximation method of section 2.1 above. Although the spectra of life-forms are dispersed in a non linear fashion the two ends of a continuum may be obtained with use of the following:

$$PhanerophyteDom \forall E1 \tag{7.4.1.11}$$

$$Therophyte; HemicryptophyteDom \forall E5 \lim E7_{(max)} \tag{7.4.1.12}$$

Use of genetic methods covered in section 7.2 enable expansion to the secondary nodal number of plant life-forms. Further, a C-S-R based model was split into 6 further categories as 18/3 returns 6. However, unless the investigators have field-based knowledge of the life-forms within each location it is difficult to combine the probabilities formed in Appendix 7.4. It is suggested that the localised use of the Gaussian process model formed in sub-section 7.3.1 is implemented in combination with field data. The standardised mean of individual populations life-form categories will enable the statements of the water energy dynamic and elevational conditions within which each life-form category occurs to be stated. From the latter statements of objective dispersal we may proceed to make statements linked with further character dispersal.

The T-S-K based algorithmic structure may be employed to implement geographic information systems as detailed in section 7. 5.

7.4.1 Functional approximation algorithm for plant communities

Gaussian process models may be used to simulate the distribution of species within individual environments and equally of elements of life-form characteristics of individual species. Complex dynamical relations at both the individual plant level due to the number of sets of chromosomes and the population level in different ecosystems may be easily diagnosed and certainties may be formed using T-S-K FL based method. Further, genetic computation methods elaborated in section 7.2 enable the dispersal of expanded orders (nodal numbers) of plant characters in set objectives such as that which is provided by the water-energy dynamic and altitude.

Using Rastrigin's function to approximate the distribution of a localised population provides the population with an estimate of zero mean and unit variance. Given a normally distributed set of individuals within the plant life-forms, imposing variable water or energy related elements will result in transgressing the population range proportional to the pressure on the population. Positive or negative transgressions of the distribution maintain a stable population dependent on pressures imposed. Increasing altitude leads to positive skewing of the life form characterisation, increasing individual occurrences and overall richness (Bhatterai and Vetaas, 2003; Gentry and Dodson, 1987; Küper et al., 2007). Decreasing altitude leads to negative skewing of the population.

There are thought to be more than two distinct niche processes operating in convergence and divergence of plant strategies and life-forms (Furze et al., 2013b; Kraft et al., 2008). It has been proven that strategy differentiation in plant species contributes to the maintenance of diversity in highly diverse locations (Jenkins et al., 2013). The logical progression is that life-forms provide the potential of a diverse range of plant strategies so are of equal importance. Although it is complex to dissect the life-form spectra, further genetic approaches are advised in combination with the logic based mathematics of this thesis (Kearsey and Pooni, 1995).

The following section implements geographic information systems in the context of plant characterisation to demonstrate the application of the mathematic methods formed in this thesis, by which teams of mathematicians and biogeographers can make substantive statements for national policy concerning ecosystems and human populations who are dependent on them.

7.5 Geographic information systems – applying plant characterisation

Geographic information systems (GIS) are integrated computer systems that can be employed to visualise geographic and species data in order to display a combination of statistical and raw data. As such, they offer a great tool to geographers, mathematicians and those whose work is in various disciplines which may relate to species characterisation and indeed to the wider cataloguing of information science. As such there are many uses of GIS, for example displaying species demographic information, land use change software, industrial development and so forth. GIS can be used to integrate characteristics of databases, spread-sheets and other soft computing methods (Trauth, 2006 [ch. 1]).

An example of a large integrated GIS is the data portal of the Global Biodiversity Information Facility (GBIF), which acts as a gateway to information catalogued by multiple interfaces (users) in many institutes on a global scale. The GBIF in particular has been praised for its capacity to be made use of in species characterisation (Yesson et al., 2007).

The use of species data is often standardised to a given unit area in order that it may be compared and given further analysis. Section 2.2 of Chapter 2 gives detail of the justification of ecological species data being displayed at the scale of 10000km^2 .

In this section, simple geographic information systems are developed employing the output of the mathematical models developed in this thesis to provide a working tool (for example **Fig. 7.5.1**) for national organisations involved in policy formation. The framework was provided by the mapping graphical user interface of Matlab (Version R2010a ©). Code developed in order to make use of the GIS framework of **Fig. 7.5.1** is shown in Appendix 8. The algorithms developed in section 7.4 are implemented in terms of plant life-history

strategies and plant photosynthetic type. As such the GIS shown makes use of techniques of evolutionary computation and global optimisation, thereby providing substantial mathematic foundation to take knowledge and decision support further, providing recommendations of concentrated conservation and research efforts with mathematic reasoning.

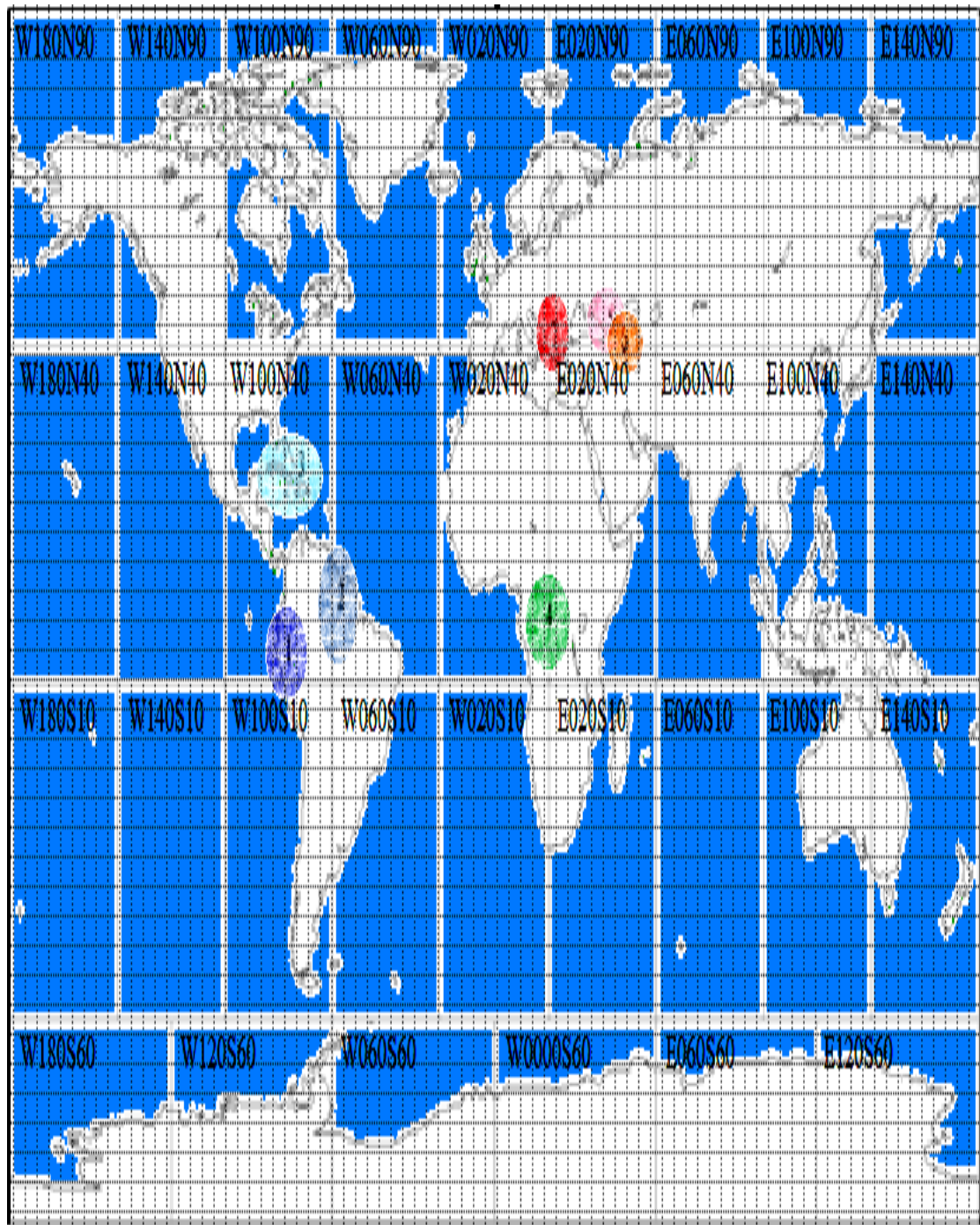


Figure 7.5.1 Outline of simplistic GIS based on plant strategy based environments (Furze et al., 2013g)

Environments *E1-E7* are indicated on the above GIS map, representing dominant strategies recorded in the locations to form the environments. Each environment is highlighted with a different colour to reflect the combination of water and energy that prevail in the location and which dictate life-form. Colours are ordered from blue (*E1*, e.g. Ecuador), grey blue (*E2*, e.g. Guyana), light blue (*E3*, e.g. Cuba), green (*E4*, e.g. Democratic Republic of the Congo), rose (*E5*, e.g. Georgia), orange (*E6*, e.g. Azerbaijan) and red (*E7*, e.g. Macedonia). The location of each environment was found by selecting the tile code (e.g. W100N40 for Ecuador), number of columns from left to right and row number from top to bottom. Each area above may be mapped with latitude, longitude with use of Matlab code, as seen in Appendix 8.

The three types of photosynthesis are represented across the 7 environments as specified in (7.4.1.8), (7.4.1.9) and (7.4.1.10). Furthermore, life forms are ordinated according to the expressions given in (7.4.1.11) and (7.4.1.12). This modelled map has been finessed and could not otherwise be produced without labour intensive fieldwork, as detailed by Yesson et al. (2007). The map could be tested and compared to smaller areas where we do have such data, for example for herbarium data, Ponder et al. (2001), for field data, Bass et al., (2010). The importance of such maps which can be produced for any location is the predictive function they provide for biogeographers, plant scientists and field workers. Dissection of plant' processes provide inferences of complex conditions in which they exist within set time scale scenarios. Photosynthesis is the primary metabolic process of plants, hence the simplistic statements outlined give the method by which subsequent metabolic processes may be ordinated within populations on a global scale, with great use in human and ecological terms due to the uses of plant products (e.g. food, construction and medicinal compounds).

The algorithmic basis of GIS shows great potential in simulation development as the algorithms developed in this thesis are constructed from initial knowledge bases, which are publicly available with use of the internet. Linking of different time scenario climatic data, topographic data, image processing, map processing, species occurrence records and finally digital mapping are techniques by which new software may be designed for design of control systems with both mathematical and geographical global inferences. This represents novel simulation development using driving dynamic factors of the W-E dynamic and continuously monitored numbers of individual plant species occurrence. The

implications of such a GIS allow the unification of multiple subject areas to substantiate policy formation and enable subject development in the component areas (e.g. geography, biochemistry, physiology and mathematics). Generally, GIS offer cost and time savings involved in data collection and interpretation, may lead to better decision making and improved communication between contributors, geographers, mathematicians and policy makers. The following sub-section discusses the recommendations for policy formation using mathematical method.

7.5.1 Recommendations for policy formation using mathematical method

Substantiation of the methods of this thesis, and in particular this chapter, are useful not only to explore mathematical relations (niches and functional traits) but also to reinforce the requirement for enhanced protection of the areas covered by this study. The implementation of logic-based mathematics adds strength to related interdisciplinary fields of plant characterisation. Additionally, modelling of climatic variables and the characters of plants modeled therein is enhanced in terms of accuracy and pattern distribution. The unification of different approaches (geographic, physiological and mathematic) allows plant characteristics to be expressed in a common language. Mathematic methods provide the unified approach in this thesis. Authoritative descriptions of plant strategies, life-forms and photosynthetic type characterized plant species. Mathematic substantiation is therefore added to ecology, biology and biochemical characterization in this thesis. Examples of the potential uses of this work include the finer scale structuring of phylogenetic trees, the patterning of prey-taxis relations (Grunewald et al., 2013; Ma et al., 2013; Huson et al., 2004) and measurement of quantitative trait loci such as those involved in biochemical pathways (Kearsey and Pooni, [ch. 8] 1996; Kraft and Ackerly, 2010).

There are many potential areas of research which are fundamental to protective policies. The accessibility of higher mathematics to related subject areas and therefore policy makers is important to emphasize. Policy formation in vulnerable locations and affecting indigenous populations in locations such as Ecuador, which is under threat of development (Pappalardo et al., 2013), benefit from the systematic approaches taken in the thesis. Identification and expression of life within priority conservation areas under threat of destructive human activity is of increasing importance, given the nature of the activities

and the immediate effect on the concentrated biodiversity. The GIS illustrated in the previous section also assists in further policy formation for planting of predicted plant strategies in conservation efforts, in the areas of *E1-E7* for example, ensuring high levels of the predominant plant strategy of the environment enables balanced populations of life forms and plant photosynthetic types to form in naturally ranked continual distributes present in ecology. Continual distribution becomes a key premise as it has been well documented that plant traits themselves are expressed from quantitative patterns of genetic loci (Kearsey and Pooni, [ch. 8] 1996).

There are thought to be more than two distinct niche processes operating in convergence and divergence of plant strategies (Furze et al., 2013b; Kraft et al., 2008). Kraft et al. (2008) states that strategy differentiation in plant species contributes to the maintenance of diversity in highly diverse locations. The work of this thesis may assist with national policy formation in justification of direction of resources towards increased conservation and protection of vulnerable locations (Jenkins et al., 2013). A starting place for implementation of conservation policy could manifest through local or national government partnerships with increased numbers of research based organizations, in order that the mathematical substantiation provided in this chapter could be further investigated.

The implemented geographic information system (GIS), shown in the previous section, provides yet another tool which may be viewed by policy makers in local and national governments. The GIS also provides a gateway through which intergovernmental bodies such as the United Nations Environment Programme (UNEP) can view informative patterns in summary in order to enhance future sustainability of the communities within the highlighted areas. Ideally, surrounding areas may also be supported by local and national governmental policies formed within these areas as the beneficial nature both of research and enhanced protection will take effect at various levels, from local community understanding of diversity and sustainability through to in-country industries with emphasis on natural products and their potential in producing novel opportunities for trade and further research. Beneficial effects are also seen at many levels within trophic systems, ensuring the maintenance of predator-prey relations as higher trophic levels are reliant on the productivity of natural systems of a primary level (Ma et al., 2013).

7.6 Summary

In this chapter we have presented a T-S-K logic based structure for the ordering of plant life-forms, carried out genetic programming of the life-forms and developed a method by which one can form functional approximation algorithms to elucidate distribution of elements within the Z utopia hyperplane. The crucial difference between the use of Boolean mathematics to describe systems and the higher mathematic logic based methods employed are that the latter are devoid of semantic definition, establishing certainty in a previously distorted view (Jongman et al., 1995 [ch. 5], Kent and Coker, 1992 [ch. 4], Sivanandam et al., 2007 [ch.1]). In order to visualise the distribution of elements within a stochastic population this thesis used a Gaussian process model, from which enhanced detail of the Z matrix was extrapolated. An efficient minimized algorithmic approach was implemented, using key elements of the water-energy dynamic for 2 candidate areas. A spectrum of life-forms were distributed, within given environments ideal for plant growth and comparatively more extreme conditions (Hawkins et al., 2003; Furze et al., 2013b). Creation of the closed loop system for the areas covered allowed bounds of life-form spectrum distributions to be perceived as a continuum and primary nodes of life-forms identified, amongst which the location of Ecuador was the most diverse and Macedonia was less diverse. Further use of optimization methods enabled dispersal of the primary nodal number to give the secondary nodal number of life-forms. The distribution of the subcategories of life-forms was of a binomial, Poisson nature in agreement with previous climatic studies (Schölzel and Friedrichs, 2007). The distribution was summarised using linear and quadratic rules.

Using the multi-modal Rastrigin's function showed that each dispersed life-form was expressed continuously for each individual life-form element, with zero mean and 0-1 variance (of the Z plane) within alternate environments. This method may be used to estimate proportions of individual species occurrences according to life-form (Furze et al., 2013a). The additional use of alternate functions (e.g. Sphere function, Schwefel's function, Rosenbrock's function) is proposed in order to indicate the functional approximation of all characters of plant species individual occurrence, either involving a simulative base, field data or an integration of the two (as in this chapter). Using this method further unveils the dimensions of multi-objective orientated characters such as those of plant metabolites. This chapter adds a more structured approach to strategy

differentiation in plant species, which contributes to the maintenance of diversity in highly diverse locations (Kraft et al., 2008).

Plant life-history strategies, plant photosynthetic pathways and plant life-form dominance within example locations have also been summarised in order to make further recommendation towards which areas should be allocated limited resources and national policy focus detailed above, similar to the priority conservation zonation carried out by Soosairaj et al. (2007). Community linked knowledge of the issues with regard to diversity patterns should be dispersed through as many channels as possible, in order to increase awareness of the pressure imposed on natural habitats. Trophic webs and individual ecosystems themselves are fragile systems which may be put out of sink by well meaning, but in the longer term, ineffective conservation efforts, due to the lack of quantified knowledge of the range of strategies, metabolism and life forms. Taking the elements covered in this chapter into consideration should help to implement more effective conservation policies over the longer term. Section 7.4 and 7.4.1 are included to make simplistic representation of the mathematic substantiation included in earlier chapters. These sections are by no means meant to be considered as the final word in the power of modelled plant characterisation but hopefully will provide the readers with some inspiration in future characterisation of complex (linear or non-linear) systems. This chapter partially answers the research questions posed in Chapter 1. Mathematic and geographic distribution of individual plant occurrence has been shown and discussed. Predictions of climatic conditions are made within utopia after having calculated linear and quadratic relational curves. A brief discussion of conservation and sustainability policy contribution in the light of this thesis has been given. The required structure of groups of plant species has been shown to be most tolerant to change given that the number of strategies and life-forms are maximised within each location. The functional fitness of the groups (individuals) can be plotted using the Gaussian Process model of Rastrigin's function.

The following chapter summarises the fundings of this thesis, concludes the work and makes a brief discussion of the future potential of plant characterisation.

CHAPTER 8

Future research work and conclusions

8.3 Thesis summary

The research work in this thesis has been divided into two main sections. Firstly theoretical studies were covered and consequently simulations based on mathematic technique were developed.

In the first part of the thesis, the background of the research was outlined and fundamental knowledge revisited, the related concepts and definitions were clarified, adaptive fuzzy neural algorithms were developed both for the strategical nodes applicable to 300 000 plant species and the variable dynamics in which they exist. T-S-K fuzzy systems were introduced into optimization techniques to enhance the knowledge bases employed and for predictive value of the modelling framework.

In the second part of the thesis, seven case studies were explored with different fuzzy neural strategies. A GIS was designed with plant diversity at its core in order to provide a

platform on which the dispersal of characters can be founded. The distribution of plant strategies was addressed in Chapter 5. Uncertainty of plant strategies and of their basic metabolism was simulated in Chapter 6 to verify the effectiveness of hybrid genetic nonlinear systems with no strict constraints and prior knowledge. In Chapter 7, the dynamics of plant life-forms were characterised using an additional algorithmic basis and the differentiated life-history strategies, life forms and metabolic patterning of plants combined for the integrated GIS to answer research questions concerning climatic prediction and contributions of plants to biotic and abiotic systems. The GIS simulation generalised the application of the developed control algorithms and verified their effectiveness following available species data of the GBIF (Yesson et al., 2007), giving a promising opportunity to further investigate both the theoretical aspects and practical applications of plant processes.

8.2 Answering the research questions

The research questions posed in the first chapter have been answered throughout the thesis. In this final chapter the answers to the research questions are brought together and summarised in order that conclusions can be drawn and further research is stimulated.

Answering RQ.1, prediction of individual occurrence of plant species according to the species-area relationship of Arrhenius was carried out in Chapter 4. However, the relationship identified between species and area (with inclusion of non-differentiated environmental and taxon related variables) was insignificant (at $p=0.05$). The species-area relationship has been significantly identified in other studies, e.g. island systems of MacArthur and Wilson (1967). Further the relationship has been made use of by more recent authors (Kreft et al., 2007) and it offered the base by which biotic and abiotic factors were introduced to successfully document species categorisation in different ecological settings. Such justification of the use of the species-area relationship was enabled by the continued use of Boolean statistical method within the sampling method, which allowed distortion of data. However, when non-standardised, field based individual plant species occurrence data were used in this thesis, prediction was seen to be affected

using fuzzy logic based systems. The use of both Mamdani and T-S-K FLC in description of individual occurrence data enabled a continual monitoring of both species occurrence and climatic conditions upon which the individual occurrences were based. Fuzzy logic based categorisation was therefore recognised as being the most efficient form of prediction.

Answering RQ.2, regarding the balance of the trade-off between accuracy and interpretability of models used, T-S-K FLC systems accurately predicted individual plant species occurrences with use of the modelling framework provided by the water-energy dynamic. It was recognised that additional variables which affect distribution of plant species were included in the modelling framework. Use of FLC enabled minimisation of variables to show key elements of the W-E dynamic which increased the efficiency of modelling. Grime's plant strategy (Grime et al., 1995) theory proved to be a highly effective grouping system for plant species, with inferences within ecological systems. Plant strategy based models gave inferences for ecological systems which ensure the maintenance of high levels of diversity throughout trophic levels of ecology. Additionally, increased resolution of the modelling antecedents gave highly efficient predictive patterns. Subsequently GA techniques were made use of to further relate species prediction in feedback to climatic systems due to the relationship between the individual occurrences and objectives (climatic variables). Linear weighted least squares and quadratic polynomial systems identified trends between individual occurrences and climatic data in Chapter 5 and 6. The interpretation made was limited by the rate of current progress in identification of climatic variability and species data available.

Answering RQ.3, within the wide groupings of plant strategies, plant photosynthetic type and plant life forms, rudimentary sub-groups were identified which expanded the informative value of plant characterisation modelling carried out. After illustrating the use of GA based technique and making use of novel MOGA, combined objective distributions were shown, which enabled further interpretation and informative value to be gained. Functional approximation algorithms were implemented via the use of process models which enabled further extrapolation of Z hyperplanes. Component subject areas (e.g. biogeography, mathematics, and plant science) benefited from the functional expansion, as standardising mean and variance within the hyperplane allowed approximations of modelling variables within Gaussian frameworks. The explorative informative value of genetic algorithm stochastic techniques were such that patterns were identified in utopian

distributes to encourage further research on subsequent categories of plant characteristics which may as yet be undiscovered.

Answering RQ.4, it was shown that the groups of plant species were distributed according to their functional types within ecological systems. These functional groups related to the form expressed within each group of plant strategy, life-form and metabolic type. The groups were formed as a result of their combinatorial parts, hence expanding distribution from the number of groups to secondary and tertiary nodal numbers of rudimentary parts. Within each group the individual characteristics of the group were distributed with use of Gaussian process models, i.e. by mathematical function (Rastrigin's function). Mathematic expansion of the combined objectives within which the primary group of life-forms was expressed, for example, results in expansion of the Z hyperplane of plant life-forms, from which indirect inferences may be drawn regarding the expression of each objective on which the modelling of life-forms was carried out. The numerical distribution of plant species primary groups within the W-E dynamic was seen to be stochastic, Poisson distribution. Latitudinal and longitudinal ordering of the groups of plant species was carried out to an extent with the use of minimised partial differential equations of FLC. Greater numbers of plant species, in ruderal based environments were found in equatorial regions (e.g. Mexico, Ecuador), where the levels of water related variables are high, whereas competitive (e.g. Democratic Republic of the Congo) and stress tolerant based plant environments were found in locations where energy related variables are expressed with greater weight acting on the distribution (e.g. Macedonia), such areas are shown by example in Chapters 5, 6 and 7. The areas of competition and stress tolerance were principally ordinated above the equator, e.g. Macedonia. The finite distribution of plant species requires further research in areas of physical and biochemical characterisation. It was suggested that the Gaussian patterns identified may be further extrapolated into discrete distributes, e.g. using discrete functions to enable characterisation of plant species in terms of different metabolic groups. It is postulated that further functional approximations of plant species will be useful in identification of biotic and abiotic conditions within ecological systems as well as the conditions in which they are found.

Answering RQ.5, it is possible to make predictions of climatic systems using individual occurrence data of plant species. The basic process makes use of the relationship between plant species occurrence and the prevailing conditions in which they occur. After forming simulated distribution via genetic computation of final population numbers it was seen to

be possible to predict the main objectives (climatic conditions) of these distributions using weighted least squares and polynomial regression differential expressions. Relationships between set variables of x , y and z allow estimation of each of the variables dependent on the other parts of the linear or non-linear expressions formed. Expansion of combined objective numerical planes (Z) via Rastrigin's function allows fitting of objectives with differential Z values. Further research in this expansive functional area will make use of more functions and characteristics to give concise statements of characterisation and distribution in rational mathematic (Zhu et al., 2013) and separable real terms (Furze et al., 2013d). In the latter study key integration of in-country and international research institutes is proposed as a result of the mathematic foundation discussed above.

Answering RQ.6, GIS of plant characters can be developed with use of FLC modelling systems, as these systems are based on Gaussian process models. Conclusions may be drawn for conservation and sustainability policy formation within the limits of error of the distribution of the modelling frameworks employed. The mathematic substantiation provides incentive for national organisations to recommend protection of these areas and safeguarding from developmental pressures which may exist on the respective areas. Protection of the land from industrial development is recommended due to the sensitive nature of the relationship between the numbers of species occurrences and climatic conditions. Monitoring and planting programmes are advised to maintain the predominance of the strategy types in each location respectively. Indigenous populations of the areas are preserved by such policies, given that their cultures and life-styles are extremely vulnerable to developmental change. Maintaining diversity in plant strategies is linked with diversity in plant species and indeed through trophic levels in ecology. The parabolic distribution of life-form types enables balanced ecological systems of the vulnerable areas. Peaks in life-form types exist, indicating the dominant type and giving stochasticity to plant populations related to characteristics of life-history strategy and photosynthetic type. Thus maintenance of the dominant life form type in natural ecosystems ensures high diversity and distribution of successive trophic levels. The dominant life form type within the population of individual plants is analogous to a major quantitative trait locus of an individual plant (Kearsey and Pooni, 1998). Accordingly quantification of the environments in which plants exist enables the measurement of each life-forms impact on the environment and trophic system, giving a further avenue for research leading on from this thesis. The photosynthetic types predominantly found in each

location are strongly linked with the climatic conditions of the areas themselves. Photosynthesis is the primary metabolic process of plants and the dominance of each of the respective types identified in this thesis should therefore allow secondary metabolite development through different plant products to be formed with maximum efficiency. The greatest recommendation that the latter point can make is for further integration of the protected areas with research based institutes, stimulating research in the range of plant products present. Research in plant products is likely to be linked with additional functional distribution, which will provide further incentive for limited financial resources of conservation to protect the areas. Completing functional approximation of levels of plant products in each area provides an avenue for generating revenue for the locations from use of the plant products. The latter point enables sustainability of diversity of plant populations in the areas and of subsequent trophic levels (including humans) dependent upon them.

Answering RQ.7, this thesis has identified that groups of plant species require a Gaussian structure within dynamic climatic systems and in Chapter 7 Rastrigin's function was proposed as being suitable. Ordering dominance of plant strategies, plant life-forms and photosynthetic type within climatic systems helps to quantify and partition non-linearity which is exhibited in distribution of plant characterisation. In this thesis Boolean mathematics, novel FLC, MOGA and Functional Approximation have been made use of to generate the combined objective numerical plane within which plant characters are distributed. A framework for an original GIS has also been presented. It is suggested that future field studies and biochemical characterisation will allow use of discrete functions to further characterise plant species within individual ecosystems.

8.3 Future research directions

- 1) The research work of this project develops a control strategy, which furthers recommendation of patterns of plant numbers to be maintained in order to sustain stability of other trophic levels of living organisms. However, it remains that species dynamics of all trophic levels vary with non-linearity. Therefore, it is necessary to

develop control systems for complex systems with space allowed for the stochastic pressures of trophic levels which feed-back onto plant species.

- 2) The integrated knowledge guided control strategies in this research work suffer from one main defect, that is, their direct interpretability, which is one of the trade-offs with accuracy of the developed model. It is practically feasible at the current stage, though in comparison to previous alternatives there is still room to improve the interpretability to reach an increased audience.
- 3) Neural network based fuzzy logic control algorithms and genetic methods are developed in this thesis. Databases of plant information, such as that developed here, should use fuzzy sliding mode control combined with alternative optimization techniques to further amplify the significance of species loss through ecosystems.
- 4) Applying hybrid optimization techniques to climatic and biotic systems has resulted in the fundamental properties of plants being characterised. Much work remains in classes of plant metabolites and morphology on both a national and more local basis.
- 5) It is a global imperative that the technique developed in this thesis is continually progressed in order that species characterisation is not lost by increasing pressures on the environment. Hence this study calls for more publication and integral research in the domain in order that policies may be formed to safeguard and conserve ecosystems. Additional implementation of integrated localised research, eg. of plant diversity and production in urban settings provides more potential for application of research of this thesis.
- 6) The algorithms and optimization developed in this research work are analysed mathematically and verified by simulations and literature across a wide range of disciplines. In order to further verify and demonstrate validity, field based and laboratory work should follow, but for the problems in finance and time limitations, this is only a desire of the author.

8.4 Conclusions

A novel adaptive neural fuzzy control algorithm has been developed to tackle the problems of numerical distribution of plants within a GIS. Practical simulation with integrated species occurrence data and detail of case study locations were made to validate the developed control strategies.

Table 8.3.1 Summarising the main research findings, research questions and answers

RQ	Chapter(s) / Summary	Notes
1	4,5 / pp.152-153	It is possible to predict plant species occurrence, the most efficient method is with use of Mamdani and T-S-K FL
2	5 / p.153	Accuracy and interpretability of predictive models is balanced by use of FL and expansion of the methods seen
3	6 / pp.153-154	Sub-groups of the main groupings of plants were identified. The novel hybrid MOGA by which this is carried out is detailed
4	6, 7 / p.154	Functional groups of plants are distributed. FL, determining stochastic distribution within the W-E dynamic, hybrid MOGA and the Gaussian Process (Rastrigin's function). The Z plane of plant characteristics gives informative value to plant characteristics and Geographic distribution alike. The latter is shown by example on a global scale
5	7 / pp.154-155	Predictions of climatic conditions may be made by making use of the relationship between occurrence of plant species and conditions in which they occur. This is carried out by the use of differential equations
6	7 / pp.155-156	Planting programmes can be recommended and policy formation structured with long term developmental importance
7	7 / p.156	Gaussian structure is required within dynamic conditions of ecosystems, use of novel FLC, MOGA and Functional Approximation are key techniques of related subject development in plant characterisation, all of which can be mathematically stated

The main research findings of this thesis are summarised in **Tab. 8.3.1**. The research work has benefited the family of hybrid computational intelligence and conventional control algorithms. It has also improved the integral value of GISs. The work has provided contributions towards national policy development, and offers an efficient alternative for areas of biogeography and plant science to consider, with conclusive publications made across the time of the research (Furze et al., 2011; Furze et al., 2012a; Furze et al., 2012b; Furze et al., 2013a; Furze et al., 2013b; Furze et al., 2013c; Furze et al., 2013d; Furze et al., 2013e, Furze et al., 2013f, Furze et al., 2013g).

References

- R. Alcalá, M. J. Gacto and F. Herrera, (2007), A multi-objective genetic algorithm for tuning and rule selection to obtain accurate and compact linguistic fuzzy rule-based systems, *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, Vol. 15, No. 5, pp. 539-557.
- P. P. Angelov and D. P. Filev, (2004), An approach to online identification of Takagi-Sugeno fuzzy models, *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 43, No. 1, pp. 484-498.
- O. Arrhenius, (1921), Species and Area, *Journal of Ecology*, Vol. 9, pp. 95-99.
- L. S. Barreto, (2008), The reconciliation of the r-K, and C-S-R-models for life history strategies, *Silva Lusitana*, Vol. 16, No. 1, pp. 97-103.
- W. Barthlott, A. Hostert, G. Kier, W. Küper, H. Kreft, J. Mutke, M. D. Rafiqpoor and J. H. Sommer, (2007), Geographic patterns of vascular plant diversity at continental to global scales, *Erdkunde*, Vol. 61, pp. 305-315.
- W. Barthlott, W. Lauer and A. Placke, (1996), Global distribution of species diversity in vascular plants: towards a world map of phytodiversity, *Erdkunde*, Vol. 50, pp. 317-328.
- W. Barthlott, J. Mutke, D. Rafiqpoor, G. Kier and H. Kreft, (2005), Global Centers of Vascular Plant Diversity, *Nova Acta Leopoldina NF*, Vol. 92, No. 342, pp. 61–83.
- M. S. Bass, M. Finer, C. N. Jenkins, H. Kreft, D. F. Cisneros-Heredia, S. F. McCracken, N. A. Pitman, P. H. English, K. Swing, G. Villa, A. D. Fiore, C. C. Voigt and T. H. Kunz, (2010), Global conservation significance of Ecuador's Yasuní National Park, *PLoS One*, Vol. 5, No. 1, pp. 1-22.

- K. R. Bhattarai and O. R. Vetaas, (2003), Variation in plant species richness of different life forms along a subtropical elevation gradient in the Himalayas, east Nepal, *Global Ecology and Biogeography*, Vol. 12, pp. 327-340.
- R. E. Bilsborrow, S.J. Walsh, and B. Frizzelle, (2012), LBA-ECO LC-01 National, Provincial, and Park Boundaries, Ecuador. Data set, Oak Ridge National Laboratory Distributed Active Archive Center, Oak Ridge, Tennessee, U.S.A, URL: <http://daac.ornl.gov/> , 2013.
- S. Bornhofen, S. Barot and C. Lattaud, (2011), The evolution of CSR life-history strategies in a plant model with explicit physiology and architecture, *Ecological Modelling*, Vol. 222, pp. 1-10.
- S. Bornhofen and C. Lattaud, (2008), Evolving CSR Strategies in Virtual Plant Communities, In Bullock, S. *et al.* (Eds.): *Artificial Life XI*, Proceedings of the Eleventh International Conference on the Simulation and Synthesis of Living Systems, pp. 64-71.
- E. V. Broekhoven, V. Adriaenssens and B. D. Baets, (2007), Interpretability-preserving genetic optimisation of linguistic terms in fuzzy models for fuzzy ordered classification- an ecological case study, *International Journal of Approximate Reasoning*, Vol. 44, pp. 65-90.
- B. Brzeziecki and F. Kienast, (1994), Classifying the life-history strategies of trees on the basis of the Grimean model, *Forest Ecology and Management*, Vol. 69, pp. 167-187.
- D. Büche, N. N. Schraudolph, and P. Koumoutsakos, (2005), Accelerating Evolutionary Algorithms with Gaussian Process Function Models, *IEEE Transactions of Systems, Man and Cybernetics – Part C: Applications and Reviews*, Vol. 35, pp. 183-194.
- Y. J. Cao, Q. H. Wu, (1999), Teaching genetic algorithm using matlab. *International Journal of Electrical Engineering Education*, Vol. 36, pp. 139-153.
- B. Cleveland and A. Meystel, (1990), Dynamic predictive planning + fuzzy logic control = Intelligent control, *IEEE Symposium on Intelligent Control*, pp.1217-1224.

-
- O. Cordón, F. Gomide, F. Herrera, F. Hoffman and L. Magdalena, (2004). Ten years of genetic fuzzy systems: current framework and new trends, *Fuzzy Sets and Systems*, Vol. 141, pp. 5-31.
 - J. M. Craine, (2005), Reconciling plant strategy theories of Grime and Tilman, *Journal of Ecology*, Vol. 93, pp. 1041-1052.
 - J. M. Craine, (2007), Plant strategy theories: replies to Grime and Tilman, *Journal of Ecology*, Vol. 95, pp. 235-240.
 - Z. Cui, X. Cai and J. Zeng, (2012), A new stochastic algorithm to direct orbits of chaotic systems, *International Journal of Computer Applications in Technology*, vol. 43, pp. 366-371.
 - B. Damer, K. Marcelo and F. Revi, (1998), Nerve garden: a public terrarium in cyberspace, *Lecture Notes In Computer Science*, Vol. 1434, pp. 177-185.
 - M. Ebner, (2006), Coevolution and the red queen effect shape virtual plants. *Genetic Programming and Evolvable Machines*, Vol. 7, No. 1, pp. 103-123.
 - J. Elith, S. J. Phillips, T. Hastie, M. Dudik, T. E. Chee, C. J. Yates, (2011), A statistical explanation of MaxEnt for ecologists, *Diversity and Distributions*, Vol. 17, pp. 43-57.
 - H. Ellenberg, H. E. Weber, R. Düll, V. Wirth, W. Werner, and D. Paulissen, (1991), *Zeigerwerte von pflanzen in mitteleuropa*. *Scripta Geobotanica*, Vol. 18, pp. 1–248.
 - T. Erfani and S. V. Utyuzhnikov, (2011), Directed search domain: a method for even generation of the Pareto frontier in multiobjective optimization, *Engineering Optimization*, Vol. 43, No. 5, pp.467–484.
 - K. L. Evans, J. J. D. Greenwood and K. J. Gaston, (2005), Dissecting the species-energy relationship, *Proceedings of the Royal Society B*, Vol. 272, No., pp. 2155-2163.
 - J. Furze, Q. M. Zhu, F. Qiao and J. Hill (2011), Species area relations and information rich modelling of plant species variation, *Proceedings of the 17th International Conference on Automation and Computing (ICAC)*, 10 September, 2011, pp. 63–68.

- J. Furze, J. Hill, Q. M. Zhu and F. Qiao, (2012a), Algorithms for the Characterisation of Plant Strategy Patterns on a Global Scale, *American Journal of Geographic information System*, Vol. 1, No. 3, pp. 72-99.
- J. N. Furze, Q. M. Zhu, F. Qiao, and J. Hill, (2012b), Facilitating description of fuzzy control algorithms to ordinate plant species by linking online models, *Proceedings of the Fourth International Conference on Modelling, Identification and Control (ICMIC2012)*, pp. 933-938.
- J. N. Furze, Q. Zhu, F. Qiao and J. Hill, (2013a), Implementing stochastic distribution within the utopia plane of primary producers using a hybrid genetic algorithm, *International Journal of Computer Applications in Technology*, vol. 47, pp. 68-77.
- J. N. Furze, Q. M. Zhu, F. Qiao and J. Hill, (2013b), Linking and implementation of fuzzy logic control to ordinate plant strategies, *International Journal of Modelling, Identification and Control*, Vol. 19, No. 4, pp.333–342.
- J. N. Furze, Q. Zhu, F. Qiao and J. Hill, (2013c), Mathematical methods to quantify and characterise the primary elements of trophic systems, *International Journal of Computer Applications in Technology*, Vol. 47, No. 4, pp. 314-325.
- J. N. Furze, Q. Zhu, F. Qiao and J. Hill, (2013d), Functional enrichment of utopian distribution of plant life-forms, *American Journal of Plant Sciences*, Vol. 4, No. 12A, pp. 37-48.
- J. N. Furze, Q. Zhu and J. Hill, (2013e), The use of plant characterisation modelling studies to substantiate national conservation and sustainability policies, *Global Education Magazine*, *International Day of Democracy*, 15 Sept., Issue No. 5, pp. 68-75.
- J. N. Furze, Q. Zhu, F. Qiao and J. Hill, (2013f), Utopian exploration of global patterns of plant metabolism, *Proceedings of the Fifth International Conference on Modelling, Identification & Control (ICMIC 2013)*, Cairo, Egypt, Aug. 31st-Sept. 3rd, pp. 47-52.
- J. Furze, Q. Zhu and F. Qiao, (2013g), Collaborative proposal development towards sustainable communities, *Global Education Magazine*, *Human Rights Day*, 10 Dec., Issue No. 6, pp. 28-32.

- M. Gaspar, J. Suñer and J. Torrens, (2007), Sklar's theorem in finite settings, *IEEE Transactions on Fuzzy Systems*, Vol. 15, No. 3, pp. 410-416.
- A. H. Gentry and C. Dodson, (1987), Diversity and biogeography of neotropical vascular epiphytes, *Annals of the Missouri Botanical Garden*, Vol. 74, pp. 205–233.
- D. B. Gesch and K. S. Larson, (1996), Techniques for development of global 1-kilometer digital elevation models, In: *Pecora Thirteen, Human Interactions with the Environment - Perspectives from Space*, Sioux Falls, South Dakota, United States of America.
- S. Godefroid and E. D. Dana, (2007), Can Ellenberg's indicator values for mediterranean plants be used outside their range of definition?, *Journal of Biogeography*, Vol. 34, pp. 62-68.
- D. E. Goldberg, (1989). *Genetic Algorithms for Search, Optimisation, and Machine Learning*. MA: Addison-Wesley.
- E. J. Gumbel, (1960), Bivariate exponential distributions. *Journal of the American Statistical Association*, Vol. 55, pp. 698–707.
- H. Guo, (2011), A simple algorithm for fitting a Gaussian function [DSP Tips and Tricks], *Signal Processing Magazine, IEEE*, Vol. 28, No. 5, pp.134–137.
- J. P. Grime, (1979). *Plant Strategies and Vegetation Processes*. John Wiley and Sons.
- J. P. Grime, J. G. Hodgson and R. Hunt, (1995). *The abridged comparative plant ecology*. Chapman & Hall.
- J. P. Grime and S. Pierce, (2012). *The Evolutionary Strategies that Shape Ecosystems*. Wiley-Blackwell, Chichester, UK.
- S. Grunewald, A. Spillner, S. Bastkowski, A. Bögershausen and V. Moulton, (2013), SuperQ: Computing Supernetworks from Quartets, *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, Vol. 10, pp. 151-160.
- C. H. Haswell, (1920). *Mechanics' and engineers' pocket book of tables, rules and formulas*. Harper and Brothers, New York, United States of America.

- B. A. Hawkins, R. Field, H. V. Cornell, D. J. Currie, J. F. Guégan, D. M. Kaufman, J. T. Kerr, G. G. Mittelbach, T. Oberdorff, E. M. O'Brien, E. E. Porter and J. R. G. Turner, (2003), Energy, water, and broad-scale geographic patterns of species richness, *Ecology*, Vol. 84 No. 12, pp. 3105-3117.
- F. Herrera, (2005), Genetic fuzzy systems- status, critical considerations and future directions, *International Journal of Computational Intelligence Research*, Vol. 1, No. 1, pp. 59-67.
- J. L. Hill, P. J. Curran and G. M. Foody, (1994), The effect of sampling on the species-area curve, *Global Ecology and Biogeography Letters*, Vol. 4, pp. 97-106.
- M. O. Hill, D. B. Roy, J. O. Mountford, R. G. H. Bunce, (2000), Extending Ellenberg's indicator values to a new area: an algorithmic approach, *Journal of Applied Ecology*, Vol. 37, 3–15.
- J. G. Hodgson, P. J. Wilson, R. Hunt, J. P. Grime and K. Thompson, (1999), Allocating C-S-R plant functional types: a soft approach to a hard problem, *Oikos*, Vol. 85, pp. 282–294.
- J. H. Holland, and J. S. Reitman, (1978), Cognitive systems based on adaptive algorithms, in: D.A. Waterman and F. Hayes-Roth (Eds.), *Pattern-Directed Inference Systems* (Academic Press, New York, 1978), pp. 313-329.
- G. Houle, (1990), Species-area relationship during primary succession in granite outcrop plant communities, *American Journal of Botany*, Vol. 77, pp. 1433-1439.
- A. V. Humboldt, (1806). *Ideen zu einer physiognomik der gewächse*. Tübingen.
- A. V. Humboldt, (1808). *Ansichten der natur mit wissenschaftlichen erläuterungen*. Tübingen.
- C. C. Hung and L. H. Chen, (2009), A fuzzy TOPSIS decision making model with entropy weight under intuitionistic fuzzy environment, *Proceedings of the International MultiConference of Engineers and Computer Scientists*, Vol. 1, pp. 13-16.
- R. Hunt and R. L. Colasanti, (2007), Self-assembling plants and integration across ecological scales, *Annals of Botany*, Vol. 99, pp. 1023-1034.

-
- D. Huson, T. DeZulian, T. Klöpper and M. Steel, (2004), Phylogenetic Super-Networks from Partial Trees, IEEE/ACM Transactions on Computational Biology and Bioinformatics, Vol. 1, No. 4, pp. 151-158.
 - C. N. Jenkins, S. L. Pimm and L. N. Joppa, (2013), Global patterns of terrestrial vertebrate diversity and conservation, Proceedings of the National Academy of Science Early Edition, Vol. 110, 1-9.
 - W. Jetz, H. Kreft, G. Ceballos and J. Mutke, (2009), Global associations between terrestrial producer and vertebrate consumer diversity, Proceedings of the Royal Society B, Vol. 276, pp. 269-278.
 - R. H. G. Jongman, C. J. F. Ter Braak and O. F. R. Van Tongeren, (Editors) (1995). Data analysis in community and landscape ecology. Cambridge University Press.
 - C. F. Juang and C. D. Hsieh, (2012), A fuzzy system constructed by rule generation and iterative linear SVR for antecedent and consequent parameter optimization, IEEE Transactions on Fuzzy Systems, Vol. 20, No. 2, pp.372–384.
 - M. J. Kearsey, H. S. Pooni, (1998). The genetical analysis of quantitative traits. Chapman & Hall.
 - J. E. Keeley and P. W. Rundel, (2003), Evolution of CAM and C₄ carbon concentrating mechanisms, International Journal of Plant Science, Vol. 164, S. 55-S.77.
 - M. Kent and P. Coker, (1992). Vegetation description and analysis a practical approach. Belhaven Press, Pinter Publishers.
 - M. Kiliç, N. Karavın and H. G. Kutbay, (2010), Classification of some plant species according to Grime's strategies in a *Quercus cerris* L. var. *cerris* woodland in Samsun, northern Turkey, Turkish Journal of Botany, Vol. 34, pp. 521-529.
 - G. Kier, H. Kreft, T. M. Lee, W. Jetz, P. L. Ibishch, C. Nowicki and J. Mutke, (2009), A global assessment of endemism and species richness across island and mainland regions, Proceedings of the National Academy of Sciences, Vol. 106, pp. 9322-9327.
 - G. Kier, J. Mutke, E. Dinerstein, T. H. Ricketts, W. Küper, H. Kreft and W. Barthlott, (2005), Global patterns of plant diversity and floristic knowledge, Journal of Biogeography, Vol. 32, pp. 1-10.

- N. J. B. Kraft and D. D. Ackerly, (2010), Functional trait and phylogenetic tests of community assembly across spatial scales in an Amazonian Forest, *Ecological Monographs*, Vol. 80, pp. 401-422.
- N. J. B. Kraft, R. Valencia and D. D. Ackerly, (2008), Functional Traits and Niche-Based Tree Community Assembly in an Amazonian Forest, *Science*, Vol. 322, pp. 580-582.
- H. Kreft, and W. Jetz, (2007), Global patterns and determinants of vascular plant diversity, *Proceedings of the National Academy of Sciences*, Vol. 104, No. 14, pp. 5925-5930.
- H. Kreft, W. Jetz, J. Mutke, G. Kier and W. Barthlott, (2008), Global diversity of island floras from a macroecological perspective, *Ecology Letters*, Vol. 11, pp. 116-127.
- H. Kreft, N. Köster, W. Küper, N. Jürgen and W. Barthlott, (2004), Diversity and biogeography of vascular epiphytes in Western Amazonia (Yasuní, Ecuador), *Journal of Biogeography*, Vol. 31, pp. 1463-1476.
- H. Kreft, J. H. Sommer and W. Barthlott, (2006), The significance of geographic range size for spatial diversity patterns in neotropical palms, *Ecography*, Vol. 29, pp. 21-30.
- W. Küper, H. Kreft, J. Nieder, N. Köster and W. Barthlott, (2004), Large scale diversity patterns of vascular epiphytes in neotropical montane rain forests, *Journal of Biogeography*, Vol. 31, pp. 1477-1487.
- Y. H. Lee, S. K. Park, D. E. Chang, J. C. Ha and H. S. Lee, (2009). Parameter estimation using the genetic algorithm in terms of quantitative precipitation forecast, *Annales Geophysicae*, Vol. 24, No. 12, pp. 3185-3189.
- M. V. Lomolino, (2001), The species-area relationship: new challenges for an old pattern, *Progress in Physical Geography*, Vol. 25, No. 1, pp. 1-21.
- M. V. Lomolino and L. R. Heaney, (Editors) (2004). *Frontiers of biogeography: New directions in the geography of nature*. Sinauer Associates Incorporated.
- A. Lindenmayer, (1968), Mathematical models for cellular interaction in development, Parts I and II, *Journal of Theoretical Biology*, Vol. 18, pp. 280-315.

-
- U. Lüttge, (2003), Ecophysiology of Crassulacean Acid Metabolism, *Annals of Botany*, Vol. 93, No. 6, pp. 629-652.
 - M. Ma, Z. Han, J. Tao and D. Wu, (2013), A food chain model with preytaxis and chemotaxis, *International Journal of Modelling, Identification and Control*, Vol. 19, No. 4, pp. 235-247.
 - R. H. MacArthur and E. O. Wilson, (1967). *The theory of island biogeography*. Princeton University Press.
 - L. I. Malyshev, (1975), The quantitative analysis of flora: spatial diversity, level of specific richness, and representivity of sampling areas, *Botanicheskiy Zhurnal*, Vol. 60, pp. 1537–1550.
 - E. H. Mamdani, (1974), Applications of fuzzy algorithms for simple dynamic plants, *Proceedings of IEE*, Vol. 121, pp 1585-1588.
 - W. Massant, S. Godefroid and N. Koedam, (2009), Clustering of plant life strategies on meso-scale, *Plant Ecology*, Vol. 205, pp. 47-56.
 - T. D. Mitchell and P. D. Jones, (2005), An improved method of constructing a database of monthly climate observations and associated high resolution grids, *International Journal of Climatology*, Vol. 25, pp. 693-712.
 - J. M. Mustard, D. B. Standing, M. J. Aitkenhead, S. Robinson and J. S. McDonald, (2003), The emergence of primary strategies in evolving virtual-plant populations, *Evolutionary Ecology Research*, Vol. 5, pp. 1067-1081.
 - J. Mutke and W. Barthlott, (2005), Patterns of vascular plant diversity at continental to global scales, *Biologische Skrifter*, Vol. 55, pp. 521-531.
 - J. Mutke, H. Kreft, G. Kier and W. Barthlott, (2010), European plant diversity in the global context, In: J. Settele, L. Penev, T. Georgiev, R. Grabaum, V. Grobelnik, V. Hammen, S. Klotz and I. Kühn (Eds.). *Atlas of Biodiversity Risk*. Pensoft Publishers, pp. 4-5.
 - N. Myers, R. A. Mittermeier, C. G. Mittermeier, G. A. B da Fonseca and J. Kent, (2000), Biodiversity hotspots for conservation priorities, *Nature*, Vol. 403, pp. 853-858.

- E. Nasibov and S. Peker, (2011), Exponential membership function evaluation based on frequency, *Asian Journal of Mathematics and Statistics*, Vol. 4, pp. 8-20.
- R. B. Nelson, (2006). *An introduction to copulas*. Springer, New York, United States of America.
- M. New, M. Hulme and P. Jones, (1999), Representing twentieth century space-time climate variability. Part I- Development of a 1961–90 mean monthly terrestrial climatology, *Journal of Climate*, Vol. 12, pp. 829–856.
- S. Niu, Z. Yuan, Y. Zhang, W. Liu, L. Zhang, J. Huang and S. Wan, (2005), Photosynthetic responses of C3 and C4 species to Water availability and competition, *Journal of Experimental Botany*, Vol. 56, No. 421, pp. 2867-2876.
- E. E. Omizegba and F. A. Monikang, (2009), Fuzzy logic controller parameter tuning using particle swarm algorithm, *International Journal of Modelling, Identification and Control*, Vol. 6, No. 1, pp.26–31.
- S. E. Pappalardo M. D. Marchi, and F. Ferrarese, (2013), Uncontacted Waorani in the Yasuní Biosphere Reserve: Geographical validation of the Zona Intangible Tagaeri Taromenane (ZITT), *Plos One*, Vol. 8, pp. 1-15.
- S. J. Phillips, R. P. Anderson and R. E. Schapire, (2006), Maximum entropy modelling of species geographic distributions, *Ecological Modelling*, Vol. 190, pp. 231-259.
- E. Pianka, (1970), On r and k selection, *American Naturalist*, Vol. 104, pp. 592-597.
- W. F. Ponder, G. A. Carter, P. Flemons, R. R. Chapman, (2001), Evaluation of museum collection data for use in biodiversity assessment, *Conservation Biology*, Vol. 15, No. 3, pp. 648-657.
- P. Prusinkiewicz and A. Lindenmayer, (1990). *The algorithmic beauty of plants*. Springer-Verlag.
- C. Raunkier, (1934). *The Life Forms of Plants and Statistical Plant Geography*, Oxford University Press.
- M. L. Rosenzweig, (1995). *Species diversity in space and time*. Cambridge University Press.

- F. Salah and H. Abdalla, (2011), A Knowledge based system for enhancing conceptual design, *International Journal of Computer Applications in Technology*, Vol. 40, Nos. 1–2, pp. 23–36.
- F. B. Salisbury and C. W. Ross, (1992). *Plant Physiology Fourth Edition*, Wadsworth Publishing Company, California, USA.
- M. Schmidt, H. Kreft, A. Thiombiano, and G. Zizka, (2005), Herbarium collections and field data-based plant diversity maps for Burkina Faso, *Diversity and Distributions*, Vol. 11, pp. 509-516.
- A. K. Shackelford and C. H. Davis, (2003), A hierarchical fuzzy classification approach for high-resolution multispectral data over urban areas, *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 41, No. 9, pp. 1920-1932.
- C. Schölzel and P. Friedrichs, (2008), Multivariate non normally distributed random variables in climate science – Introduction to the copular approach, *Non linear Processes in Geophysics*, Vol. 15, 2008, pp. 761-772.
- K. Silvera, L. S. Santiago, J. C. Cushman and K. Winter, (2009), Crassulacean acid metabolism and epiphytism linked to adaptive radiations in the orchidaceae, *Plant Physiology*, Vol. 149, pp. 1838-1847.
- W. Silvert, (2000), Fuzzy indices of environmental conditions, *Ecological Modelling*, Vol. 130, pp. 111-119.
- D. S. Simberloff, (1974), Equilibrium theory of island biogeography and ecology, *Annual Review of Ecology, Evolution and Systematics*, Vol. 5, pp. 161-182.
- S. N. Sivanandam, S. Sumathi and S. N. Deepa, (2007). *Introduction to fuzzy logic using matlab*. Springer-Verlag.
- A. Sklar, (1959), Fonctions de repartition á n dimensions et leurs marges, *Publications de L’Institut de Statistique de l’Université de Paris*, Vol. 8, pp. 229-231.
- S. Smith, (1980). *A learning system based on genetic algorithms*. Doctoral dissertation, University of Pittsburgh, United States of America.

- J. H. Sommer, H. Kreft, G. Kier, W. Jetz, J. Mutke and W. Barthlott, (2010), Projected impacts of climate change on regional capacities for global plant species richness, *Proceedings of the Royal Society B*, Vol. 277, pp. 2271-80.
- S. Soosairaj, S. J. Britto, B. Balaguru, N. Nagamurugan and D. Natarajan, (2007), Zonation of conservation priority sites for effective management of tropical forests in India, *Applied Ecology and Environmental Research*, Vol. 5, No. 2, pp. 37-48.
- Y. Su, G. Zhu, Z. Miao, Q. Feng and Z. Chang, (2009). Estimation of parameters of a biochemically based model of photosynthesis using a genetic algorithm. *Plant, Cell and Environ.* 32, 12, 1710-1723.
- M. Taheriyoun, M. Karamouz and A. Baghvand, (2010), Development of an entropy-based fuzzy eutrophication index for reservoir water quality evaluation, *Iranian Journal of Environmental Health Science and Engineering*, Vol. 7, No. 1, pp. 1-14.
- T. Takagi and M. Sugeno, (1985), Fuzzy identification of systems and its applications to modeling and control, *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 15, No. 1, pp. 116–132.
- M. Trauth, (2006). *MATLAB® Recipes for Earth Sciences*. Springer.
- N. I. Vavilov, (1926), Studies on the Origin of Cultivated Plants, *Bull. of Applied Botany*, Vol. 16, No. 2, pp. 1-245.
- O. R. Vetaas, (2000), Comparing species temperature response curves: population density versus second-hand data, *Journal of Vegetation Science*, 11, 659–666.
- A. R. Wallace, (1878). *Tropical nature and other essays*. Macmillan.
- C. Wang, L. Guo, Y. Li and Z. Wang, (2012), Systematic Comparison of C3 and C4 Plants Based on Metabolic Network Analysis, *BMC Systems Biology*, Vol. 6, S.2, pp. 1-14.
- N. Wang and Y. P. Yang, (2010), An iterative fuzzy identification method hybridising modified objective cluster analysis with genetic algorithm, *International Journal of Modelling, Identification and Control*, Vol. 10, No. 1/2, pp. 44-49.

- K. Wendt, A. Cortés and T. Margalef, (2010), Knowledge-guided genetic algorithm for input parameter optimisation in environmental modelling, *Procedia Computer Science*, Vol. 1, No. 1, pp. 1367-1375.
- D. H. Wright, (1983), Species-energy theory: an extension of species energy theory, *Oikos*, Vol. 41, No. 3, pp. 496-506.
- E. W. Wulff, (1935), Versuch einer einteilung der vegetation der Erde in pflanzengeographische gebiete auf grund der artenzahl, *Repertorium Specierum Novarum Regni Vegetabilis*, Vol. 12, pp. 57-83.
- C. Yesson, P. W. Brewer, T. Sutton, N. Caithness, J. S. Pahwa, M. Burgess, W. A. Gray, R. J. White, A. C. Jones, F. A. Bisby and A. Culham, (2007), How global is the global biodiversity information facility?, *Plos One*, Vol. 11, pp. 1-10.
- L. A. Zadeh, (1965), Fuzzy sets, *Information and Control*, Vol. 8, pp. 338-353.
- L. A. Zadeh, (1973), Outline of a new approach to the analysis of complex systems and decision processes, *IEEE Transactions on Systems, Man and Cybernetics*, Vol. SMC-3, No. 1, pp.28–44.
- L. Zhao, (2012), Application of interval type-2 fuzzy neural networks to predict short term traffic flow, *International Journal of Computer Applications in Technology*, vol. 43, pp. 67-75.
- X. W. Zhang, F. Gong and L. Z. Xi, (2012), Inspection of surface defects in copper strip using multivariate statistical approach and SVM, *International Journal of Computer Applications in Technology*, Vol. 43, No. 1, pp.44–50.
- Q. Zhu, Y. Wang, D. Zhao, S. Li and S. A. Billings, (2013), Review of rational (total) nonlinear dynamic system modelling, identification and control, *International Journal of Systems Science*, pp. 1-12, DOI: 10.1080/00207721,2013.849774
- B. L. Zimmerman and R. O. Bierregaard, (1986), Relevance of the equilibrium theory of island biogeography and species-area relations to conservation with a case from Amazonia, *Journal of Biogeography*, Vol. 13, pp. 133-143.

- B. Zlatković, L. Nikolić, V. Randelović, N. Randelović, and V. Stevanović, (2011), Comparative analysis of the vascular flora of the Pčinja river gorges in Serbia and Macedonia, Archives of Biological Sciences, Vol. 63, pp.1157-1166.